An Empiricist Solution to the Radical Concept Nativism Puzzle

Khashayar AriaAmjad

Major Research Paper

in

The Department

of

Philosophy,

presented in Partial Fulfillment of the Requirements

for the Degree of Master of Arts (Philosophy) at

Concordia University,

Montreal, Quebec, Canada

May 2021

Abstract

Jerry Fodor (1975) argued for a view according to which virtually all ***concepts*** are **innate**, that is no concept can be learned. This so-called "Radical Concept Nativism Puzzle" has engendered a lot of controversies. A number of solutions have been proposed to it. A significant one belongs to *Stephen Laurence and Eric Margolis*. They try to solve the puzzle from a nativist perspective, although a non-Fodorian non-radical one. Another proposal is *Jesse Prinz*'s *Proxytype Theory of Concepts* which is wholeheartedly empiricist. The two views share several features with a general inclination towards empiricism. I shall argue that an empiricist solution is plausibly available to *Fodor*'s puzzle. Moreover, it is agreed by *Fodor* and others that any proposal for a theory of concepts needs to account for, inter alia, the *Publicity* of concepts, viz., their capacity of being shared by different individuals, and by one individual at different times. I will show the extent to which each of the two solutions would successfully satisfy this desideratum. With regard to the concept of color, I shall discuss how they should account for each of these two extreme cases: the congenitally blind people and the tetrachromat (the one who enjoys super color vision.) A contrast between these cases will show how an altered perceptual capacity, i.e., sight, can affect one's conceptual system. Finally, I will use the evidence from the studies made on "Feral children", the ones who have grown up in isolation, to argue against nativism.

1. Introduction

In his (1975) book, **The Language of Thought** (hereafter, *LOT*)**,** *Jerry Fodor* argued for a view that continues to seem controversial among scholars. He claimed that all ***lexical concepts*** are **innate**. The argument had two main steps. First, he begins with the broadly-accepted[1] view that to ***learn*** a concept is to learn its *meaning* or *content*. The second step of his argument, however,

is what departs from any previous accounts. He maintains that if one is to ***learn*** a concept, then she will need to **form** a **hypothesis** (hereafter, HF for Hypothesis Forming) to confirm whether her next encounter in the world is an instantiation of that concept or not. The result is that complex concepts are not problematic since their structure allows for such hypotheses. One can confirm, for instance, that some human being is a BACHELOR by considering his being an UNMARRIED ADULT MALE. That is to say, she has ***learned*** the meaning of the concept 'bachelor'. But on these occasions, one must already be in possession of some *lexical* concepts like male, or unmarried. The problem for such an account arises with respect to how one learns these *lexical* concepts. There would seem to be **no** hypothesis, whatsoever, to confirm something's being a *lexical* concept. He concludes, that lexical concepts ***cannot*** be learned and, consequently, are all **innate**. This leaves his account with a commitment to some tremendous stock of innate concepts including CAR, INTERNET, ZOMBIE, and so on. It is puzzling to see how such concepts might be innate. This is the reason why *Fodor*'s account is sometimes called '*Radical Concept Nativism.'*

The view has few defenders today, while *Fodor*'s dialectic is considered highly influential for any theory of concepts. A number of solutions have been proposed. One significant attempt belongs to *Laurence and Margolis* (hereafter *ML*). They offer a model by appeal to the notion that they call '*Sustaining Mechanisms.'* Another solution to the puzzle is offered by *Jesse Prinz* within his *Proxytype Theory* of concepts. These two works have their foundation in the **empirical** study of the mind and share several features. Hence, they are good rivals to each other, and of course, as their authors persistently claim, to Fodor's proposal in *LOT*. I shall examine each proposal along with their strengths and weaknesses, and see how they would apply to some real-life cases. In particular, I shall consider why *ML*'s model, despite a great number of advantages, may still face difficulty *as a theory* of concepts.

Let us start with a reconstruction of Fodor's concept acquisition puzzle with regards to his radical concept nativism:

> "1. Apart from miracles or futuristic super-science all concepts are either learned or innate.
>
> 2. If they're learned, they are acquired by hypothesis testing.
>
> 3. If they're acquired by (non-trivial) hypothesis testing, they're structured.
>
> 4. Lexical concepts aren't structured.
>
> 5. So lexical concepts aren't acquired by hypothesis testing.
>
> 6. So lexical concepts aren't learned.
>
> 7. Therefore, lexical concepts are innate." (*ML* 2002)

A decisive part of this formulation with respect to our concern is:

2) If concepts are "learned", they are acquired by Hypothesis Testing (or, as Fodor puts it, by HF, viz., Hypothesis Formation.)

To illustrate the force of Fodor's argument, consider this example of a typical "concept learning experiment." The experimenter has a certain concept in mind that is labeled by a novel predicate, e.g., "flurg." So every object around should either be a flurg or not. The subject is presented with various stimuli and in each case guesses if it is a flurg. If it was a flurg then the experimenter would give an affirmative answer to the subject and if it was not, then the experimenter would let the subject that was is not. In other words, the subject would only rely on the feedback she receives from the experimenter after each trial. If the FLURG is, e.g., GREEN^2

then when the subject is presented with a card with a green circle on it and says it is a flurg, the experimenter will say that she is right, but if the subject is presented with a card with a red circle on it and says it is flurg, the experimenter will disconfirm. After a while of doing such a trial and error activity, the subject becomes able to distinguish flurgs from other (non-flurg) things. (*ML* 2002)

Fodor thinks that the subject will, eventually, reach a criterion that enables her to tell flurg from non-flurg. Only at this point, Fodor states, the subject has **learned *the concept flurg***. The odd thing about accounts of this kind is that in order to frame and test hypotheses about whether an encountered object is a tokening of FLURG (or any other concept) one would need to possess the concept in the first place. In the above example, one would need to possess the concept GREEN before a sufficient number of trials and errors could lead her to learn that "FLURG = GREEN." In the case of complex concepts with some internal structure such hypotheses could safely be formed by appeal to their constituent concepts. However, when it comes to the most basic concepts (also called lexical or primitive concepts), the ones without an internal structure (as Fodor takes them to be) **learning** would seem impossible. In such cases, Fodor argues, no learning is going on. Accordingly, he thinks such concepts must come from another source, namely that they are **innate**.

There is controversy about the potential *double entendre* of "**learning**" in the above claim. Most famously, Margolis (1998; Laurence 2011), Carey (2009, 2014) also Beck (2017) in the same vein, and Shea (2016) have argued against that interpretation of "learning"[3].

In Fodor's own explication, a concept is "**learned**" if it is "***acquired***" through a "***rational***" process, rather than just a "***brute causal***" process, e.g., swallowing a pill or hitting one's head against a hard surface (Fodor 2008, 135). Prinz (2011) notes that in Fodor's positive view concepts are words in Mentalese, rather than being ***prototypes*** or **stored impressions** and those words, like

English words, are arbitrarily related to their referents. So Fodor, unlike the empiricist, owes us an explanation of why we seem to attain concepts by means of encounters with their instances. Here, says Prinz, things get a little weird:

> "In Concepts (1998), Fodor offered a bizarre metaphysical solution. Concepts. He said, refer (at least initially) to mind-dependent properties, and, in particular, C refers to the property of being the kind of thing typical instances of which cause attainment of C." (Ibid.)

Prinz calls this an "idealist" theory of reference, without making an argument. I have the impression that he did not find such an argument even required, as the suggestion resonates with famous idealist views (George Berkeley's, for instance). Prinz adds:

> "This idealist theory of reference does not ring true even for concepts that are widely held to refer to response-dependent properties. FUNNY does not refer to things that lead us acquire FUNNY; it refers to things that make us laugh (these are modally distinct properties.)" (Ibid.)

Prinz, of course, continues to say that in *LOT2*, Fodor re-asserts the mind-dependence proposal, despite objections in the literature (e.g., in Prinz 2004) and his account gets even more radical by introducing the "sea metaphor"[4]. An idealist account of reference does not fit, however, with semantic externalism, the view maintained by Kripke/ Putnam that Fodor (1998) sees as his preferred account. Putnam (1975), in particular, famously argues that "psychological state does not determine extension" and emphasizes two other factors rather than "individual's observations", namely "other people" and "the world", to be encompassed by a better philosophy and a better science of language. Another remark in the same vein by Putnam is concerned with

"un/conceptualized reality." He thinks that it is "nonsense" to say that all we have access to is "conceptualized reality." Nobody has, he notes, encountered an instance of the concept of DEATH as a personal experience. Putnam quotes Reichenbach, with regards to this latter example, that if there is no reality as such, then "why should I buy a life insurance policy?"[5].

So we may think of, at least, four objections against the contention behind the **flurg** example and the consequent explications by Fodor (especially in his 1998 and 2008.) First, a number of authors have argued that HF should not be taken as the only path to "**learning**". As we saw above, HF is crucial for that example to make sense. Fodor does not allow for stages that would lead to "***learning***" and considers all such stages as mere "***acquisition***". Of course, *ML* (2002) provide a reformulation of Fodor's argument that does not appeal to HF. Their revised version supposes that "concepts are either constructed from primitives or innate" (premise 1) and concludes that "lexical concepts are innate." That "lexical concepts are innate" is the conclusion of Fodor's argument rather than one of its premises. So *ML*'s reformulation is not valid on pain of begging the question for there being innate concepts. Fodor makes his argument only by an appeal to the distinction between lexical and complex concepts. The conclusion of the argument, namely that lexical concepts cannot be learned, along with the fact that we, in fact, do (and *a fortiori* can) possess lexical concepts leads to the final claim that lexical concepts are innate. Nonetheless, if we accept such a revised version of the argument, then it would be a problem for *ML*'s proposal as well to solve the issue that Fodor's argument turns on of "*conceptual structure*". *ML* admit, Fodor's argument "isn't affected by skepticism about the significance of hypothesis testing," but it is also worth noting that their own model (as we shall see in the following sections) appeals to experimental results, hence is susceptible to the same objection. A conceptual puzzle could not be

solved by appeal to experimental information[6]. Fodor (1998) thinks there are no grounds to associate the concept acquisition story to cognitive neurology (as opposed to neurology *tout court*.) In other terms, Fodor would outsource the job of stipulating detection mechanism(s) to the scientists who study the brain and the neural system. His methodology appeals to an a priori argument and his approach to cognitive science does not require a specification from inside of philosophy unless philosophers are asked to do the job of neuroscientists[7].

Second, Fodor's proposal (both in flurg example and in later modifications such as his 1998 book) does not work well for response-dependent properties, e.g., FUNNY[8]. Third, as Putnam (1975) argues, psychological state does not determine extension (although there is, definitely, a contribution.) Finally, an idealist account of reality is not the preferred view for a good many of authors, especially due to serious objections famously made by Russell and Moore[9].

2. Two Empirical Solutions

We saw that Fodor's modifications to his initial proposal in *LOT* (1975), in effect, raised new questions while leaving the critics unconvinced. Now let us see two other solutions, one nativist (though not Fodorian, as the authors note) by *ML* (2002, 2011, 2012) and the other, empiricist by Prinz (2002).

2.1. Sustaining Mechanisms and Kind Syndromes

ML (2002) summarize a variety of failed attempts to solve the radical concept nativism puzzle. Subsequently, they try to develop a pathway through which they hope the puzzle may get solved. They introduce a main idea which they refer to as "*Sustaining Mechanisms*". These are detection mechanisms for categories that have the capacity of establishing *mind-world* dependency relations for a concept. Such relations can constitute content with respect to a "causal theory of

content", mainly the one proposed by Fodor. In their next step, they provide an illustration of a sustaining mechanism by using evidence from psychology. They suggest a ***syndrome-based*** sustaining mechanism and, specifically, a ***kind syndrome***, i.e., one that accounts for *categorizing* **natural kinds**. They think of such a syndrome as

> "a *collection* of *highly indicative properties* of a [natural] kind, accessible through **perception**, that allow for an encountered thing to be categorized under a certain concept. Examples of such properties may include things like the typical shape, motions, markings, sounds, colors, etc., associated with a kind. The significance of this type of sustaining mechanism, *ML* emphasize, is that "it readily translates into a **learning** model[10]" (*Ibid*., emphasis added)

ML admit, in a modest way, that their claim is not meant to be the exact and exclusive way in which the mind works. They, rather, see their general strategy as a "***collection of models***" that is open to further development. I shall argue that their strategy can be pushed further by emphasizing the role of **perception** via an *empiricist account of concept learning.*

In ***The Language of Thought 2***, *Fodor* (2008) argues against the model suggested by *ML* (2002), claiming that their model falls short of accounting for "*genuine*" cases of learning due to a lack of *rationality* in the process of concept acquisition that they suggest. During this revision of *LOT*, which led to his *LOT2* theory, he shows sympathy with his critics in there being "*something*" wrong with the argument in *LOT1*, i.e., his 1975 book. This time, however, he makes the stronger claim according to which ***no*** concept is capable of being ***learned***, be it either *complex* or *lexical*. Instead he explains *concept acquisition* in terms of biological processes.

ML ("**Learning Matters: The Role of Learning in Concept Acquisition**", 2011), provided answers to *Fodor*'s criticism against their model. Furthermore, they argued that *Fodor*'s new suggestion would fail to satisfy his initial method, which is a commitment to *HF*, as the only legitimate way of learning. *ML* note that "learning" (or acquiring) a concept's **stereotype** would necessitate, as *Fodor* suggests in *LOT2*, putting forward hypotheses about the stereotype's individuating conditions and testing these against relevant observations. But then, putting forward the correct hypothesis would require having the stereotype prior to learning it (*LOT1)*, which would entail that the stereotype can't be learned after all. One remaining way for *Fodor* (in the context of *LOT2*) might be to allow other ways of learning for stereotypes, rather than reducing all learning to *HF*, but then this qualification would undermine the circularity against concept learning, which was his initial motivation for introducing "the puzzle" in *LOT1*. In this latter work, *ML* claimed that there are a number of **other** legitimate ways for *learning* to take place (compare this with Prinz's proposal about acquiring default proxytypes in Sec. 2.2.2. below)

Contra *Fodor*, they claim that *learning* is a "**progressive process**," rather than a single occurrence at a certain moment. Accordingly, it can take place during several steps since, unlike the HF test, it is not a single moment of *decision* (to determine whether or not the HF test is confirmed.) They use a parallel example made by Davidson (1969) to show that *learning* need not wait until its **final *stage*** to have taken place. The **process** of *learning*, they maintain, happens within a number of ***stages*** over the course of time, before the concept in question is ***fully*** *learned*, hence a sort of hierarchical model for learning. Again, there are similarities to the empiricist account defended by Prinz (see Sec. 2.2.1 below.) Since these *stages* are arrived at in a *rational* way, this whole process of *concept acquisition* counts as a genuine case of *learning*. They spend less time, nonetheless, discussing the ways, if any, in which their alleged cases are *rational*, so that

what they call "*Fodor's dogmatic* position with regard to rationality" may be explained away. In his objection against their model, *Fodor* reduces the status of a *sustaining mechanism* to a *theory*, whence the common criticism against the *theory*-theory of concepts. They show how *Fodor*'s biological solution to his own puzzle would leave, inter alia, the ***Publicity*** desideratum of a theory of concepts unsatisfied. In other words, his biological account will fall short of providing an explanation of the fact that different individuals can share concepts. They provide clear examples that confirm such a failure. With a focus on *Publicity* of concepts, I shall argue that *Prinz*'s suggestion is less demanding than postulation of a *kind syndrome*. I will, also, investigate how *Publicity* is to be approached in some real-life cases wherein it is seemingly not satisfied. *ML* (2011) defense against *Fodor*'s objections in *LOT2* is highly empirical, in a number of important ways. One obvious way, among others, is that they emphasize the **non-nativist** part of their model, that is, their *sustaining mechanisms* strategy. I shall discuss that that part has much to share with *Prinz* (2002) suggestion, where he introduces concepts as *detection mechanisms.*

There is another twist in *ML*'s model, though, beyond any detection mechanism. They think that such mechanisms are motivated by an ***innate*** ability or *syndrome*. All the empirical evidence that *ML* (2002) use as the basis of their suggestion seems to come from experiments on young children of an age when some *meaningful* verbal ability is already developed, rather than newly-born infants, or, at least, children of an age when the verbal ability is still undeveloped. In other words, those young children were, ironically, too old for the conclusion that *ML* may favor, since the development of verbal capacity leaves room for many other interpretations than a serious need for positing an innate element to explain the results (see Soja, Carey and Spelke 1991, and Landau 1994.)[11]

2.2. Proxytype Theory of Concepts

A crucial aspect of the empiricist theory proposed by *Prinz* (2002) is concerned with the principle of ***compositionality***. A non-negotiable desideratum for any theory of concepts would, admittedly, be that simpler concepts can combine to form more *complex* ones. Otherwise, there should be no explanation for the fact that we humans as finite beings are able to form a virtually infinite system of concepts, despite having a finite mind. For instance, a compositional system allows us to combine AARDVARK concept with NOCTURNAL concept whence forming the thought that aardvarks are nocturnal. Similar combination rules can be used to push that thought further, e.g., that aardvarks consume insects and bats. Different combinations can emerge from the same stock of primitive concepts by using the same rules (*Prinz* 2002, 12.) *Fodor* rejects the proposal that concepts are *prototypes* most importantly on the grounds that prototypes do not compose. He, nonetheless, does not deny that prototypes exist and, in fact, this plays an irreplaceable role in his own theory. *Prinz* provides an answer to this objection of Fodor's, in the first place, and then goes on to propose a *modified* notion of prototypes, the one that he eventually calls **proxytypes**. So *Prinz*'s proxytypes are, to a first approximation, a *modified* notion of prototypes.

Here is how *Prinz* answers to the compositionality objection. The fact that prototypes "do not **always** compose", *Prinz* admits, "is readily explained." It does not follow, he adds, "that prototypes cannot figure into our explanation of how people generate novel thoughts from finite means." But he also adds one more twist to the compositionality desideratum that he later claims is "wholly consistent" with the crucial point: We modify the resulting compound prototypes using the power of reasoning". Still such modifications are consistent with the point that "prototypes *can* combine compositionally. […] In cases where I have reason to think the compound will not be like its components, I can simply amend my prototype." This is, of course, not applicable to an

atomistic theory with "a pair of structureless atoms", but Prinz's theory is not atomistic. (*Prinz* 2011, original emphasis) The RCA model provides a detailed explanation about how this might be possible. (See Sec. 2.2.2. below) A good example would be "carnivorous plant" that fits neither with prototypical features of plants, nor with carnivores (e.g., "eats insects.") We do not rely on compositional mechanisms when acquiring prototypical features of such a concept due to our encounters with carnivorous plants. But now think of a truly unfamiliar concept like "carnivorous chair" which could be the subject of a new horror movie. Upon hearing or reading such an English phrase, we may form a mental representation of it by compositionally combining prototypical features of its components, namely, that "carnivorous chairs have mouths, stalk their victims, have four legs, back support, and can be sat on, etc." In a similar manner, *Prinz* maintains, "we can generate a prototype for any arbitrary compound." We implement our "power of reasoning" to rationalize the results when generating these compounds. For instance, we may consider that a "tinfoil chair" could not be able to bear a human's weight and hence could not be sat on. *Prinz* concludes that:

> "prototypes *can* be combined compositionally, and that is *sufficient* for explaining how they could be used to represent arbitrary combinations." (*Prinz* 2011, my emphasis)

It is worth-noting that before a new concept is acquired, it has almost the same status of an arbitrary one. Only after being provided the evidence from the *actual world* one can decide whether it is a real-world or a fictional concept. But that decision is not even always so straightforward. A SWAN prototype was, for example, changed after the first Europeans encounter with a BLACK SWAN in Western Australia on 1697. There are undiscovered facts that can partially (or, even, sometimes completely) change our prototypes (e.g., the *Einsteinian* concept of gravitation

compared to *Newtonian* concept of it.) These game-changing facts have always been lurking in nature and hence it is not logically implausible to think that the same story can aptly go during the process of concept formation. What's more the prototypes need not always combine, as we are not always (i.e., in each and every moment of our lives) in the actual business of forming new concepts. When direct encounters happen, the corresponding concept would form without a need for compounding prior ones. It is, some of the time, possible to make such a compounding story by telling that, e.g., SODA is (flavored or unflavored) sparkling (/fizzy) water to someone who had no prior encounters with sodas. Still upon her first real encounter with the drink, such a story would become unnecessary.

2.2.1 An Updated Empiricist View

When we allow for compositionality of prototypes, the door is wide open to new horizons for a theory of concepts. Empiricists have traditionally proposed that concepts are *images* or pictures in the mind. *Prinz* holds that the proposal needs some revisions in light of more recent findings in cognitive science:

> "To bring concept empiricism up to date, one must abandon the view that concepts are conscious pictures. Contemporary cognitive science helps in this endeavor by identifying a rich variety of highly structured, unconscious perceptual representations. […] I argue that such representations can be used to form concepts." (*Prinz* 2002, 139)

However, elaborating various aspects of such a theory to a nicety would go far beyond the limitations of this paper. Accordingly, I shall briefly recount his story about those "unconscious representations" and how "structured" they happen to be, and then confine myself to only those facets of Prinz's theory of concepts that directly relate to my purposes here. One should,

nevertheless, bear in mind that his theory is capable of providing answers to a vast number of other questions that are plausibly expected to be addressed by a theory of concepts, or so he claims.[12]

A significant point in this theory is concerned with the notion of a "*long-term memory network*." Perceptual representations *link* together to form such a network. The link could be formed via a "*link principle*" and through particular ways that can be called "*link types*." *Prinz* suggests a variety of link types including what he calls "hierarchy links", "transformation links", "binding links", "situational links" and "predicative links", each based on the nature of the concept in question. In other words, he implies that different concepts need different tools to settle in the mind. A stored group of linked perceptual representations, he suggests, can be called a "*long-term memory network*." Long-term memory networks, he adds, can come to store various kinds of information about commonly encountered categories. (*Prinz* 2002, 144-6)

An initial proposal is that concepts are *mechanisms of detection*. Note how similar this proposal is to the *sustaining mechanism* suggested by *ML* (see Sec. 2.1. above.) But what information in our memory networks can contribute to *detection*? *Prinz* think that the obvious answer is "All of it." (Ibid, 148) If we are to identify concepts with the totality of category knowledge then one problem would be that it is not always straightforward to decide where knowledge of one category begins and that of another ends. But this, argues *Prinz*, should not be a serious concern: "There is no reason to insist that boundaries between concepts must be sharp. Different concepts can encode some of the same knowledge." (Ibid)

For example, my knowledge that "dogs make good pets" need not ***exclusively*** belong to either my DOG concept or my PET concept. It is, also, important to "distinguish between a *thought* and a piece of *standing knowledge*", that is, the "knowledge that I possess when I am not thinking about it." There is a division of labor between long-term memory, which stores standing

knowledge, and short-term memory, which stores thought for their brief duration. *Prinz* suggestion is that concepts are proxytypes, "where proxytypes are perceptually derived representations" [stored in long-term memory] that can be recruited by working memory to represent a category." A proxytype can, especially, be

> "a detailed *multimodal* representation, a single *visual* model, or even a mental representation of a *word* (e.g., an auditory image of the word "dog"). Every long-term-memory network of perceptual representations contains many overlapping proxytypes. Just about any concise subset from such a network can do the job. **Context** determines what proxytype is used in working memory on any given occasion." (Ibid, 149, my emphasis)

An example to illustrate the idea of a proxytype would be like this. We may see a gorilla as standing still, or as beating its chest. Because the same object was observed in both states, the perceptual representations that match these two states of it get grouped together. I may also store a record of the sound I hear when the gorilla beats its chest along with the visual representation of arm movements because the two are coinstantiated. Finally, there are cases where two objects are not physically bounded together but the two are observed in co-occuring scenes, props or other things. We might see a gorilla eating a banana and make a link between that situation and our representation of the gorilla. All these informations will contribute to formation of a gorilla proxytype.

As we can see, a significant part of this proposal has to do with the "**context**". For our present purposes, this should account for a fundamental difference between empiricist accounts of concepts and nativism. What's more, since proxytypes are stored representations of perceptual

experiences the fact that a concept, e.g., RED is acquired by having a red experience comes as no surprise.

2.2.2. Publicity of concepts: Default Proxytypes

Concepts, *Prinz* suggests, often put constraints on proxytypes. This may help different individuals **share** their proxytypes. For example, when one is in a restaurant her fish-proxytype would be more like a *fish filleted on a plate* rather than the *swimming creature in the ocean*. But why is it that people in different contextual situations share concepts and hence communicate? The filleted fish on my plate was, after all, the fisherman's catch some time ago. This would lead us to another important notion within the present proposal, namely, the "**Default Proxytypes**." There is reason, holds *Prinz*, to believe that we have default proxytypes:

> "A default proxytype is the *representation* that one would token if one were asked to consider a category ***without*** being given a ***context***. When no context is specified, we use representations of the kind that figure **most frequently** in our interactions with category members. I speculate that default proxytypes are relatively stable, widely shared, and frequently responsible for guiding ordinary category-directed behavior. If I am right, then default proxytypes may play an important role in explaining publicity." (Ibid, 154, emphasis added)

One may ask which features are ultimately included in a default proxytype. According to *Prinz*, a number of factors would contribute:

> "What ultimately determines whether a feature is included in a default proxytype is how frequently one represents a category as having that feature. [...] The ***frequency*** with which

a feature is used can be affected by its *cue validity*, *category validity*, *perceptual salience*, or *conformity to theories*. Even if none of these factors can perfectly predict which features get included in our default proxytypes, they can all exert some influence." (Ibid, 155, my emphasis)

Cue-validity is the subjective probability that a given object, belongs to a particular category given that it has a particular feature. Category-validity is the subjective probability that an object has a particular feature given that it belongs to a particular category. Theoretical beliefs, as another factor, may sometimes cause us to neglect features that are both ***highly perceivable*** and cue-valid. Note that this can readily be made as an objection against *ML*'s model (see Sec. 2.1. above.)

As I noted early in this section, default proxytypes are akin to prototypes. An important difference is that prototypes (in their traditional sense) do not compose on pain of their failure to represent **relations** between features that compose them. Also, they do not contain **linguistic information**. For instance, prototypes do not include representations of *category names* like the word "dog". Default proxytypes, on the other side, are immune to these issues because they typically encode a considerable amount of **relational *information*** between features, rather than just simple spatial relations (see Sec. 4.1.), e.g., the complex explanatory relation with regards to the fact that *having wings* in birds is responsible for *being capable of flight*. (Ibid, 156)

The claim in here is that proxytypes, unlike prototypes (in the traditional sense), ***do*** compose. Fodor uses the concept PET FISH as an example against compositionality of prototypes. The prototype for PET FISH, Fodor argues, is neither similar to the prototype for PET (e.g., has fur, moves around the house, sleeps with closed eyes, etc.) nor similar to the prototype for FISH (e.g., lives in *mare liberum*, has dark colors, is eaten by humans etc.) In other terms, our best

mechanism for detecting pet fishes is not formed from our best mechanisms for detecting pets and fishes. Since the nature of proxytypes share a lot with prototypes, this objection needs to be addressed before we can proceed.

Prinz's initial response is that such examples may have to do with our interpretation of compositionality desideratum.

> "The compositionality desideratum should be interpreted as saying that we *can* generate phrasal concepts and thoughts compositionally, not that we always do. There is no need to demand that the contents of phrasal concepts always be inherited from their constituents. Such a demand would be imprudent." (2002, 291 original emphasis)

For example, we may come across a place with a huge number of red plants and they might all have been poisonous. Being poisonous is neither a prototypical feature of red things nor of plants. But we, in fact, include such a feature in our RED PLANT prototype and hence avoid hazards. **Exemplar knowledge** in such cases leads to survival!

One proposal could be formulated as "we combine concepts compositionally when, and only when, we are unfamiliar with the exemplars falling under them." Fodor and Lepore (1996) object that there are cases in which a lack of exemplar knowledge fails to predict compositional combination. We have never seen exemplars of the concept PET FISH LIVING IN ARMENIA THAT HAVE RECENTLY SWALLOWED THEIR OWNERS. Even so, this concept has emergent features, e.g., being large and voracious. So non-compositional combination cannot be explained by the possession of exemplar knowledge.

Prinz responds to this objection by noting that non-compositional combination also occurs when we have **background knowledge** (Prinz 2002, 292). In the case of PET FISH and RED

PLANT we have exemplar knowledge, and in the case of KILLER FISH we have background knowledge for our corresponding prototypes. Hence, Prinz considers the proposal that predicts that "compositional mechanisms will be used only when we lack *both* exemplar knowledge *and* relevant background knowledge. Compositionality is a fallback system, not a mandatory mode of operation." (Ibid., 293)

One might still object that this proposal is untestable. Are there examples in which we lack relevant background knowledge? Ironically, Fodor and Lepore (1996) give us a good one. They provide the example of COWS BELONGING TO PEOPLE WHOSE NAMES BEGIN WITH "w". On the present proposal, one can hardly think of any background knowledge that can be brought to bear on this compound, whence a lack of emergent features. Intuition lines up perfectly with this prediction (Prinz 2002, 293). Another objection that could be made by Fodor and Lepore is that because prototypes lack logical forms, prototype theory cannot ensure that, e.g., combining PURPLE and APPLE will produce a PURPLE APPLE concept. On Fodor's account the same combination could be generated in virtue of the fact that the reference of compound concepts is determined by the reference of their constituents together with logical forms (this is discussed under what is called ***systematicity desideratuem***, namely that compound concepts could be generated from their constituents in a systematic way.) Prinz thinks that the objection is not compelling.

> "The prototype theorist can stipulate that when concepts are combined by selective modification, they refer to the *intersection* of what their constituents refer to. On this view, **combination rules** function as a kind of ***ersatz*** logical form. The compound created by combining purple and apple refers to purple apples, not in virtue of its form, but in virtue of the fact that it is generated by a certain rule." (Ibid., 299, my emphasis)

It may, also, be objected that prototype theory is inadequate because it exploits mechanism of combination that are not *fully* systematic. Prinz argues that the objection backfires. Take the example OSCAR ATE AN APPLE. Fodor's language of though model has it that thoughts are strings of unstructured symbols. Accordingly, there would be no reason to predict any asymmetries between thoughts of the form *a*R*b* and thoughts of the form *b*R*a*, e.g., AN APPLE ATE OSCAR. But such asymmetries, Prinz holds, seem to be pervasive. Prototype theory selectively modifies the schematic representations of objects when they are combined with verbal concepts and consequently excludes thoughts of the form AN APPLE ATE OSCAR.

> "It is easier to conceive of a person eating an apple than an apple eating a person. Prototype theory predicts this asymmetry, and Fodor's theory does not. The limitations predicted by prototype theory coincide with the limitations of our cognitive systems. The use of context-sensitive rules does not prevent prototype theory from explaining systematicity. It merely imposes psychologically realistic constraints on systematicity." (Ibid. 300)

Now that we established that prototypes (and *a fortiori* proxytypes) do compose in an unproblematic way let us take a moment considering how so. Based on numerous suggestions by psychologists (especially Hampton 1991, Murphy 1988, 1991 and 1992, Smith et al. 1988, Wisniewski 1996 and 1997) Prinz proposes a three-stage model that he calls RCA.

> "Concept combination begins with the ***retrieval*** stage. When two concepts are brought together, one first attempts to look for an appropriate representation in memory. This can be done by locating a *stored compound concept* or by locating *exemplar representations* that can be **cross-listed** under both of the concepts being combined. If this stage is unsuccessful, it is followed by the ***composition*** stage. Here, a pair of concepts is compositionally combined by *aligntegration* [= *alignment and integration*, that is the

process of identifying an attribute dimension in one concept and then replacing its default value with another concept] or *feature pooling* [that is the process during which features with very low **weights** are dropped from the representation]. Once a compound is generated, the collection of features that constitute it must be **analyzed**. In the analysis stage, apparent *conflicts* or *gaps* are resolved by eliminating, modifying, and introducing features. I call this the RCA model (for retrieval, composition, and analysis)." (Prinz 2002, 301-8, emphasis added)

The only remaining issue before the present theory can provide an explanation for the publicity of concepts is to see why default proxytypes are shared. An initial response is that they are comparatively ***resilient to change***. Although the features that are included in proxytypes are determined by *subjective measures* (namely by cue-validity, category-validity and salience), they are **grounded** in *objective fact*. An important concern would be about theoretical knowledge that differs from one person to another and can affect their default proxytypes. This concern, *Prinz* thinks, can be met by relaxing the publicity requirement:

"Rather than demanding strict identity between default proxytypes, we can settle for **similarity**. If you and I agree about the most conspicuous walrus features, then we understand each other when we use the word "walrus," and we engage in similar walrus-directed behaviors. If the publicity desideratum is intended to explain such examples of coordination, a theory that predicts considerable conceptual similarity will suffice." (Ibid, 158)

Fodor and Lepore (1992) argue against an appeal to similarity rather than strict identity between concepts. The question arising from such an appeal, they hold, is what it is for two concepts to be similar. If ***similarity*** is understood in terms of ***partial identity*****,**

then we would have a problem with the above proposal, namely that we should define identity in the first place before being able to legitimately talk about partial identity, viz. similarity. If we simply replace identity with similarity (that is, partial identity) and say nothing about the former, then the latter would seem problematically implausible. On this account, one must show that distinct concepts can have identical features, that is a proper subset of their features are identical, if the one is to show that similarity is possible. A feature is individuated by its role in a network of variables, so if the network differs even slightly then the same feature cannot play the same role in two distinct networks, and this makes similarity unmeasurable.

Fodor and Lepore hold that theories of concepts defended by, e.g., connectionists should be rejected on these grounds. Those theories suggest holistic criteria for feature individuation, that is an appeal to a feature's functional role in a network of variables. A significant problem is that functional roles are too abstract. Proxytype theory explains this problem away by giving a nonholistic proposal for feature individuation that appeals to primitives. A set of primitive concepts that are not identified by their functional roles can help functional roles be identified with networks of these primitives rather than networks of variables. In other words, some concepts, viz. primitives, must be individuated by something other than functional roles. This is, of course, a problem for the functional role theorist, since she will need to either replace or supplement her theory with one capable of individuating primitives without an appeal to functional roles.

The RCA (retrieval, composition, and analysis) proposal that we saw above suggests, inter alia, that proxytype features are ***weighted***. Two proxytypes are type identical if they are formed by comparably ***weighted***, ***perceptually*** derived feature representations

of the same type. But what makes, Prinz asks, two features count as the same? One answer is that two features can be taken as the same if they share more primitive features of the same type. This "decomposition", however, must stop at some point. Put another way, we must arrive at "a set of primitives that are not type-identified by further features." In the lack of a way to say when primitive features are identical, we will have no way to say when two proxytypes are identical or to quantify **similarity** between proxytypes. Prinz, following neurophysiologists, suggests that we should individuate perceptual primitives by "distal properties" that they (qua proxytypes) detect. Detection is driven by appearances, but only "perceptual primitives" can be individuated solely by appearances. In other words, "perceptual primitives" can be individuated solely on the basis of what they detect, and what they detect are classes of things that appear alike. Primitives can be individuated by their nomological causes. Proxytype theorist should still face a problem with feature ***weights***, in the same way that functional role theorist would. Namely, two mental representations composed of the same primitives can qualify as distinct proxytypes if those primitives are weighted differently. Those two proxytypes would have different detection tendencies. To treat this problem, Prinz proposes that proxytypes be individuated by "sets of sets of properties" rather than "mere sets of properties." (Prinz 2002, 273-5)

A couple of examples may help us understand what is intended in here.

> "If a DOG proxytype includes the features FURRY, BARKS, and FETCHES and causing any two of these to be tokened is sufficient for tokening DOG, then that proxytype can be individuated as a set containing four members: the set of appearances detected by FURRY and BARKS, the set of appearances detected by FURRY and FETCHES, the set of appearances detected by FETCHES and

> BARKS, and the set of appearances detected by all three of these representations. This gives us a way to compare proxytypes. Two proxytypes can be ***identified*** if they detect the same appearance sets, and two proxytypes are ***similar*** to the extent that the appearance sets they detect *overlap*. Two proxytypes can also be said to be ***similar*** if they detect *similar appearances*. The appearance of red, for example, is more like the appearance of orange than the appearance of blue. That similarity is **objective** in the sense that the physical features in virtue of which we coclassify things as red (e.g., certain reflectance properties) may be closer to the features in virtue of which we classify things as orange than to the features in virtue of which we classify things as blue. Such objective similarities in the magnitudes by which we perceive the world provide a basis for measuring similarities between internal states. Thus, similarities between proxytypes can be measured by ***overlapping*** *sets of appearances* or by independently measurable similarities between *members* of those sets. This contrasts with functional roles, which are too abstract to support similarity assessments." (Ibid. 275-6, emphasis added)

Accordingly, if two individuals happen to have different default proxytypes under unusual circumstances, then we may unproblematically expect that communication and psychological generalizations collapse. "My failure to understand why you do not flinch when I say there is a snake by your foot", *Prinz* notes, "can be regarded as a very localized communication failure." (Ibid, 159)

In other words, *Prinz* suggests that publicity of concepts is apt to come in **degrees** and this would not harm the possibility of communication except in anomalous circumstances (see sec. 5

below for an example of such circumstances) It is worth-noting that nativist theorists do not seem to allow for such considerations and therefore will have almost no explanation for a situation which is far from normal. Finally, one should note that proxytype theory is an informational theory but, contra Fodor, not atomistic (Ibid. 164).

2.2.3. Abstract Concepts

With regards to abstract concepts, *Prinz* suggests they may be identified with **operations** on proxytypes and those operations (e.g., verbal skills, attention, transfer operations, scanning operations etc.) do not require representations outside of our perceptual systems. (Ibid, 187-8) A final point that is of importance for the purposes of this paper is concerned with the proposal made by the famous psychologist, *Antonio Damasio*, called the "**convergence zones**" theory (Damasio 1989.) One might point out that more anterior regions of the brain's cortex, associated with higher cognitive functions, also use cells that are not unique to individual modalities, and hence could deserve to be called "*amodal*" cells. *Prinz* maintains that such cells that appear to be amodal might serve as convergence zones:

> "[…] convergence zones are cell populations that record simultaneous activity in sensory areas and serve to reactivate those areas during cognition. Damasio's theory predicts, and is partially based on, the existence of cells that cannot be associated with a single modality. At the same time, it is consistent with the empiricist view because occurrent cognitive activity and conceptual performance rely on activity within the modalities. Convergence zones may qualify as amodal, but they contain sensory records, and they are not the actual

vehicles of thought. Convergence zones merely serve to initiate and orchestrate cognitive activity in modality-specific areas." (*Prinz* 2002, 137)

So the proposal is that "if an amodal code exists, it works ***on credit*** rather than serving as the ***primary currency*** of thought." (Ibid. my emphasis) It remains to be shown that such a theory is able to gain credit on its own. Fortunately, a good number of experiments have confirmed the existence of "convergence zones" over the years after the initial proposal. Here is a short summary of a long story[13].

Damasio had access to the records assembled by the group of doctors in the department of neurology at the *University of Iowa Hospital*. Over the course of 15 years, the group had collected the records of about 1,500 brain-injured patients in a single registry. Indicating to that collection, the philosopher *Patricia Churchland* once said "It's incredible, you can say you need people with such and such conditions, and *Tony* will say, 'Right, let's call up Mrs. *B*. in *Des Moines*.' There's never been anything quite like it." In the late 1970's the spark of the idea was provided when one day a 65-year old woman came to *Damasio*. The old lady was unable to recognize faces, even her own in the mirror, while her memory was working perfectly. *Damasio* thought that it was not just about faces, she was also unable to distinguish between a *Buick* and a *Chevy*. He concluded that there should be areas in the brain in charge of combining sensory data and making sense of the world. He called those areas "**convergence zones**", because the sensory neurons intersect at those areas. It is worth-noting that he was able to see the place of injuries in the brain by using X-rays, and tell whether a "convergence zone" is injured. *Stephen Kosslyn*, a neuroscientist at *Harvard University*, once said "If someone says it's not a theory…well, it's testable, so it is a theory." In the 1980's, when MRI was available, *Patricia Churchland* volunteered to participate in an

experiment that could show whether the theory of "convergence zones" is true. She was supposed to perform a mental task and they were able to see her brain's activity. The experiment was repeated on many other volunteers and each time the results confirmed the theory.

3. Two Objections against Kind Syndrome Model

Prinz's suggestion enjoys an obvious advantage over a *kind syndrome* model. Namely, that it is not committed to a *certain* account of natural kinds. To say that something is a *kind* is different than saying that a particular *kind* is a natural kind (Hawley and Bird, 2011). Since concepts are supposed to do a categorization job, a model should expectedly be able to account for **all** *kinds*, rather than just *some* of them, viz. *natural* ones. It is not clear how the model is to be generalized from *natural kinds* into other, i.e. *non-natural* kinds. Another adjacent issue is that the model will need to presume the genuine existence of *natural kinds*, which is another subject of debate among philosophers. Those debates are, normally, supposed to establish that under appropriate circumstances, the *kinds* that appear in scientific theories refer to the groupings that are part of the *structure of the world* (in Fodor's terms, the way "the world seems to God (1998, 152). This contrasts with the view that ***human*** *interests* or *actions*[14] *determine such groupings*, an anti-realist view on science[15].

We can still make this objection against *ML*'s final model that they do not specify their preferred account of natural kinds before they posit the *kind syndrome*. In other words, one may criticize, so to speak, *ML*'s account "as a theory", rather than "an approach", because even the *paradigmatic* cases of natural kinds are subject to controversy among philosophers (Ref.) I would argue, then, that *ML* will need to choose between different proposed criteria of determining natural kinds, since they may each include/exclude some objects as natural kinds, before they can

legitimately use the notion in their model and this is due to the differences among various proposed accounts. Accordingly, it appears that there is a methodological problem with *ML*'s **model** as such.

Another objection could be made on the grounds that, according to *ML* (2002), a *kind syndrome* depends on picking the ***properties*** of a certain kind which are ***more*** *significant* such that the kind in question may qualify as a token of the respective concept. For *ML*, properties related to a kind syndrome are the ones that are "highly indicative" while still "perceptually detectable" (*ML*, 2002). They do not point out what ***measures***, if any, are to make some perceptually accessible properties "highly indicative".

In their later work, *ML* (2011) state that "different people will often acquire different concepts on exposure to the same physical environment", when they argue against *Fodor*'s biological account of concept acquisition. They spend less time, notwithstanding, explaining how their *own* final model is to survive the same issue. They discuss a number of situations in arguing against Fodor's *LOT2*. For instance, they consider, among other examples, the case of the basic color concepts, on the normal assumption that people all over the world are equipped with the same fundamental sensory systems (Davidoff et al. 1999). We shall return to this very issue in the next section.

4. Some real-life cases

In the following I will show how this very assumption undermines the kind syndrome model, and how an empiricist view can remain safe from similar criticisms. With respect to some of the most recent empirical studies, I will discuss two cases of congenitally blind people's concepts and tetrachromacy. On the assumption that all people have the same *kind syndromes*, these cases may hardly satisfy the "desideratum on Publicity of concepts."

4.1. Congenital Blindness

An early study on "word meaning among congenitally blind children" compared a number of blind children with sighted ones of matching age, sex, and socioeconomic status. Their study:

> "failed to demonstrate that the underlying concept of the objects used – i.e., their mental representation as revealed through descriptive attribution – was substantially different for blind and sighted children." (Anderson and Olson, 1981)

Visual properties are, admittedly, among "*highly indicative*" properties of objects[16]. So it is plausible to think that sighted children may enjoy a *kind syndrome* (or a set of them) that is less accessible to congenitally blind children, or so *ML*'s model would predict. The empirical study above shows the opposite results. The *Proxytype* theory, on the other hand, does not rely on a *kind syndrome*, and consequently allow for the same concept in both blind and sighted children. Anderson and Olson concluded that:

> "In sum, **communication** between blind and sighted children about the objects used in our study does not appear to be seriously disrupted." (Ibid., emphasis added)

A *kind syndrome* is supposed to be, *ex hypothesi*, different in blind children than in sighted ones. In *ML*'s terms, we may say that a *snow syndrome*, for instance, would be composed of "highly indicative properties", e.g., whiteness, coldness, softness, cotton-like shape etc. Some of those properties are accessible to both sighted and blind individuals (like coldness or softness), while others are not accessible (at least directly, as we shall consider shortly) to the blind. Hence,

the *kind syndrome* for the two should be different. Why is it that they can still communicate reliably? What would be the source of publicity for their concept of *SNOW*? Surely, two *kind syndromes* cannot account for a single concept, or *ML*'s model would be committed to a disjunctive way of representing things (a *snow syndrome* is either so and so if the individual is sighted, or such and such if the individual is blind) which is not acceptable.

A recent empirical study by Sterim-Amit et al. (2018) using fMRI investigated how congenitally blind people process concepts whose referents are imperceptible to them ("rainbow", "red") due to their visual deficiency. The study made a comparison between brain activity for these concepts and the ones whose sensorially-perceptible referents, e.g., "rain", classical abstract ones, e.g., "justice", and concrete ones, e.g., "cup." *ML*'s model would predict that people born blind will lack concepts like RED, due to the absence of the related *kind syndrome*. In other words, the property "redness", not only falls short of being "highly indicative", but also is inaccessible to those individuals due to the visual nature of it and their lack of visual abilities. So those people should not be able to possess or communicate such concepts according to the *kind syndrome* model. The empirical results, however, show the opposite. With regard to the fMRI results, Sterim-Amit et al. (2018) think that congenitally blind people have color concepts in an "**abstract**" form, that is, similar to the way in which we (sighted people) acquire concepts like "*justice*", e.g., by reading or hearing about them. The study finds that

> "Anterior Temporal Lobe (ATL) responses track concept perceptibility and objecthood: preference for imperceptible object concepts was found in dorsal ATL, for abstract (non-object, non-referential) concepts in lateral ATL, and for perceptible concepts in medial ATL. These findings point to a new division-of-labor among aspects of ATL in representing conceptual properties that are abstract in different ways." (Ibid.)

This division-of-labor is not a fixed one. It is rather affected by the presence of sensory stimuli. In blind people the activity of one part of ATL may be more, while the other part's is less in facing different experiences. The important fact, however, is the study found that similar connectivity patterns between those parts are shared by both blind and sighted people. This connectivity pattern can explain how the congenitally blind people can communicate, in the exact expected way, with the sighted people about concepts like *color*. The *kind syndrome* model, I assume, will have less to say about such results. The Proxytype theory, on the other side, can easily explain this by appealing to *default proxytypes* and *convergence zones*, as we saw earlier.

One may still ask "do blind people understand what the sighted one means by "red"?" or even, in a broader sense, "do we understand a concept when we possess it?" This is a big question in its own right and there is still-ongoing research with respect to it. For instance, a recent study at MIT suggests that

> "it is possible to couple first-person subjective phenomenological data with objective third-person brain fMRI measures. Consequently, one can think of monitoring the activity within this neural network and give feedback to the participants in order to create neuro-modulatory interventions to either augment or diminish brain activity and the associated subjective sensations." (Bauer et al. 2014)

In the philosophical context, these issues are usually discussed with regards to the notion of "Qualia." There are plenty of arguments for and against that notion, so perhaps such a question should be addressed in a wholly independent paper. For my part, though, I think I can confine

myself to the notion of concept acquisition/possession by taking it as the ability of *using* a concept in an *appropriate* ***communicative*** way. This should be read as something more like a "Turing Test" (Turing, 1950). So for the purpose of our discussion, we may think of talking to an unknown person on the phone (or other communicating device) without knowing whether or not they are congenitally blind. The talk could be about any concept involving a visual feature such as color. Now the suggestion is that you won't be able to tell whether the other person on the phone is congenitally blind or not. For instance, you might try uttering a sentence like "I cut my hand accidentally yesterday and there were "green" drops of blood on the table." The unknown person on the phone would respond "How could it be? Isn't your blood "red"? Are you sure you are not a Martian?" You would, accordingly, consider that the person on the phone has the concept "red" and "green", by looking at the fact that she was surprised by your unusual claim about the color of your blood. They knew that the "blood is red" and, that in a fictional scenario, "a Martian's blood might be green". You conclude that she possesses both concepts of "red" and "green", and perhaps by going ahead and asking similar questions all involving color concepts, you would come to the conclusion that the other person can talk about colors and, indeed, in the most imaginable appropriate way.

On the face of it, you might be surprised when you come to realize that the one you had a decent conversation with on the phone was a congenitally blind person. However, you will easily overcome that feeling of astonishment by simply looking at the fact that it just happened. You may further justify yourself by considering that the blind one is still in a verbal community where people talk about colors and they might have learned the appropriate way of talking about "red" and "green" by hearing others talking about it. Finally, you may ask "ok, but does she *understand* a concept like *redness*?" This, however, should beg the question that "do we (sighted people)

understand such concepts when we use them?" and "if yes, in what sense?" Now our best bet is Turing's answer, namely that one should be taken as understanding a concept if she is able to use it appropriately in the conversation. So there remains no other way (unless we have more advanced accounts of Qualia) than accepting that the congenitally blind person possesses color concepts, perhaps not exactly in the way sighted people do, but we are not so sure even about our own (sighted) way of it.

The experimental evidence that I referred to in my paper draws an initial distinction among three sorts of concepts with respect to their referents and the corresponding brain activity when the one entertains those concepts in her mind. There were sensorially-perceptible referents like "rain", classical abstract referents such as "justice" or "democracy", and concrete objects like "cup". According to this division of labor, the corresponding concepts invoke activity in three different areas of the brain. For instance, when a sighted person entertains an abstract concept like "democracy" to which one can refer only by hearing and/or reading about (but, definitely, not by touching, smelling etc.) the fMRI shows neural activity in the "lateral" ATL (Anterior Temporal Lobe) of the brain. On the other hand, when the same (sighted) person entertains a perceptible concept such as "red" the brain activity is in "medial" ATL.

The interesting fact is that the congenitally blind people, despite the sighted ones, show brain activity in the "lateral" ATL when entertaining a concept like "red" whose referents are sensorially-imperceptible to them due to their visual deficiency. This is the exact part of the brain used by sighted people when entertaining concepts whose referents are abstract, like "democracy" or "justice". In other words, we may say that congenitally blind people have rewired their brain activity (although during an interaction with their environment, rather than at or before their birth) in a way that enables them to communicate with sighted people in an appropriate way. This is

confirmed by the fact that a shift in brain activity (with respect to the areas I mentioned above) happens in people who are not born blind but at a later stage became blind. A kind syndrome can provide no explanation for this. Default proxytypes can easily account for such a case by appealing to the operations on proxytypes like verbal skills (see Sec. 2.2.3. above), while an innate *kind syndrome* can barely do this.

4.2. The Tetrachromat

Tetrachromacy is the condition of having extra vision capacity, often read as having a wider spectrum of colors, ostensibly as a result of there being an extra cone cell in a species' eyes. The one who has such a condition is called a 'tetrachromat'. The condition has been found both in animals and human, although different studies make varying approximations about the population of human tetrachromats. A recent study that has examined various aspects of earlier inquiry about it suggests that:

> "Tetrachromacy, a phenotype arising from X-linked genotypic variants, is an exciting prospect and it now seems likely that there are individuals who fit the expected profile. However, the relationship between the existence of a fourth class of retinal cone and the dimensionality of color vision is more complex than previously thought. A number of factors may determine whether functional tetrachromacy arises, such as the spectral distance between the peak sensitivities of the cones, the distribution and relative number of cone types across the retina, and the relative strength or weighting of chromatic signals that are sent from the retina to the cortex." (Jordan and Mollon, 2019)

The inquiry with regard to the biological origins of the phenomena is still going on. Nonetheless, the mere existence of such a condition will suffice for our purposes in the present discussion, and there is, fortunately, enough consensus about it. So the tetrachromat is appeared via more colors. We might wonder whether a tetrachromat can be a more successful artist, due to her exclusive natural gift? Jordan and Mollon respond,

> "If she uses her color palette veridically to render her sensations, then it is not clear how she can communicate the added richness of her private gamut to those who live in an impoverished perceptual world." (Ibid.)

This time we have a majority of people in the "*impoverished perceptual world*" compared to the last case of the congenitally blind minority. In other words, we (non-tetrachromat people) are blind, to some extent, compared to tetrachromats. It is plausible to think that we are unable to enjoy a more colorful world, the way they do, just like the way a blind person cannot enjoy a painting. Even so, the tetrachromat can talk about her visible ultraviolet-range colors and we may acquire, as it is, an abstract concept of them, just the way the congenital blind individual acquires a concept of, e.g., "redness". In a similar way to the last case, the *kind syndrome* model has less to say about a private or *minority-possessed* syndrome with respect to the publicity desideratum of a theory of concepts. Within the Proxytype theory, however, the tetrachromat enjoys a richer set (viz. hyper-set) of default proxytypes that shares a certain range with normal people but has some extra proxytypes.

The same considerations that I mentioned in last section could be reiterated. In this case, I may add that seeing "*more colors*" is not, in principle, anything inconceivable for normal people,

hence the communication can, in principle, take place. We understand that if a painting has the color 'red' in addition to black and white, then it has "*more colors*" than the one painted only in black and white. We may communicate with the tetrachromat in the same way that congenitally blind people communicate with us sighted ones. Perhaps the trickiest part here is the adjective "normal" in its mere statistical sense. Surely, when tetrachromats consist of no more than a few percent of the human population, they may not have as many *words* as we have when they want to refer to their extra colors, while in principle they might. We may think of all those colors as just extra colors, unless we gather a number of tetrachromats with shared real (extra) visual abilities and ask them to make conversations with each other involving those extra colors after naming them. At that point, we may gradually develop an understanding of those extra color names, and I expect we (as normal people) will eventually be able to take part in a conversation with a tetrachromat that involves a number of those extra colors.

This is similar to what happens when congenitally blind people take part in conversations with us (sighted people) involving the color concepts. We, as normal people (in a statistical sense), are ***partially*-blind**, if you like, compared to a tetrachromat person. I think this should never be problematic. The same could be said with regards to a concept like "romantic affections" for a homosexual person from the view-point of the heterosexual, or vice versa. There is neither a lack of communication and, let us all hope, nor a lack of understanding.

5. *Solvitur Ambulando*: The case of Feral Children

A nativist theory of concepts, be it radical Fodorian or milder versions like *ML*'s *kind syndrome* model, would predict that even if a single human is isolated from all others at the

moment of birth, he should still be able to enjoy a conceptual system, at least, to a significant extent. Otherwise, the whole story of nativism is pointless. A dog, for instance, innately starts to bark before a certain age, even if isolated from all other dogs. There need not be any additional conditions for a dog under which it will start too bark at some age after its birth, while lacking which it will fail to start barking. One may, of course, argue that concepts are more complicated than being compared to the innate barking of a dog. I do agree, but is it an argument for or against concept nativism?

We may think of some empirical cases like the ones I discussed above. Such an experiment on an innocent human child, however, could never be morally done for obvious reasons. The child may receive unfixable mental damage and psychological harm. Nevertheless, there are approved cases of unfortunate children who are called "feral children" (or wild children), who happened to be in a similar situation to our imagined unethical experiment, due to some sort of abuse or other causes. Fodor's account or *ML*'s model, would predict a Mowgli as depicted in *The Jungle Book* with, more or less, the same (openness to) social abilities as normal people. Sadly, those children were not able to show even the most primitive conceptual abilities. They could not speak (or even learn to speak), were not responding to others (due to a *functional*, rather than *physical* deafness), had trouble walking upright rather than on fours, refused to eat cooked food, and in one case was killed in a fire. (so, probably not aware of the *danger* qua an "innate ***concept***^{IN}") These cases were subject to attention by, among others, linguists:

> "In the last 300 years, there have been several reports of children found in the wild, who apparently survived on their own or were brought up by wild animals (Singh & Zingg 1939). These feral children were regarded as critical evidence for nature –

what is built into the genes – or nurture – what is learnt from socialization – in determining human behavior." (Clark 2016, 365)[18]

Clark indicates two arguments with regard to linguistic abilities of these children:

> "The argument from nature runs as follows: If the capacity to learn language is the characteristic that distinguishes humans from other animals, then feral children removed from the wild should be able to learn language. From nurture, the position is a little different: It is unsurprising that feral children have no language because they have had no exposure to human society and so no opportunity to learn. Once in society, they would of course learn language."[19] (Ibid.)

ML endorse Chomsky's considerations with regard to innateness of the language:

> "Noam Chomsky has argued that, contrary to empiricist doctrine, the real difficulty in accounting for cognitive capacities such as language is one of postulating a *sufficiently rich innate* mental endowment. Were we to limit ourselves to the methodological constraints of empiricism, we simply wouldn't be able to explain how children rapidly develop these capacities in such a *uniform* manner across **widely varying and impoverished** environments. While hardly uncontroversial, Chomsky's forceful case for nativist approaches to language has had a liberating effect." (*ML* 2002, my emphasis)

In the case of feral children, as mentioned above, it seems that "rapid development of cognitive capacities such as language" is not the way *ML* imagine, not in an "impoverished" environment. Perhaps the type of environment they call (cognitively) "impoverished" is, after all, a far richer one. I believe by (cognitively) "impoverished" they refer to what we usually call "normal" human environment[20]. In comparison, one might ask if our normal environment is cognitively "impoverished", then what should be the adjective for the environments where those helpless feral children were grown up in? Here is E. V. Clark's idea as a linguist:

> "Learners can be conservative or bold, or somewhere in between. When children learn language, they could go step by step, one form at a time, waiting for evidence from adult speech and rarely going beyond it – go, run, fall, fell, cat, cats, feet. They could generalize from a few forms to new instances – from jump/ jumped to run/runned, from cat/cats to man/mans. They could go item by item then make some limited generalizations, with different children following different paths. Or they could generalize broadly, acting as if all of language is orderly and rule governed (it isn't), and so regularize many irregular forms (e.g., bringed, sitted, goed, foots, sheeps, mouses)." (Clark 2016, 8)

The case in here is handsomely relevant to *ML*'s model and the sort of nativism that they defend. They use the results from "isolation experiments" on animals to draw evidence for the argument from "poverty of stimuli", the one proposed by Chomsky that we saw above.

> "Squirrels raised in isolation from other squirrels, and without any solid objects to handle, spontaneously engage in the stereotypical squirrel digging and burying behavior when eventually given nuts. […] The stereotype of the movement becomes particularly obvious

in captivity, where inexperienced animals will try to dig a hole in the solid floor of a room, where no hole can be dug. They perform […] covering and patting the nut, even though there is no earth available." (*ML* 2012)

As we discussed so far, there seems to be something deeply incomparable between squirrel instincts and the human conceptual system, especially when we regard concepts as word-like entities in the mind as *ML* would do. In the same paper they make another crucial proposal after reserving a list of concepts that they regard as innate (see note 11).

"[…] Nativists should embrace a substantial range of concepts that are acquired through learning. No doubt, **some** concepts are learned via *empiricist learning systems*, but at the same time […] there is no incompatibility between nativism and learning." (Ibid. emphasis added)

That seems, at last, a charitable approach towards empiricism. It is yet to be said, nonetheless, that what are the criteria (if any) to see which concepts are learned and which are learned via empiricist learning systems. *ML* claim their reserved list of innate concepts apparently by relying on "either ample empirical evidence, or persuasive theoretical considerations" (Ibid. footnote 22) From early years after Chomsky made his famous initial proposal, there appeared responses that brought persuasiveness of those theoretical considerations seriously into question[21]. Empirical evidence for such claims of innateness, on the other hand, seem to be open to other interpretations. During a live stream debate between Jesse Prinz and Susan Carey in 2016 at NYU Center for Mind, Brain and Consciousness[22], Prinz explained that how small manipulations in the

alleged experiments would lead to radically different results. Hence, he argued, the results from those experiments could not legitimately be considered as evidence for innateness of the concepts in question.

6. Conclusion

Fodor's "Radical Concept Nativism Puzzle" has engendered a lot of controversy. It has led to new horizons and brought many authors' attention into a good number of subtle issues that one may face when trying to develop a theory of concepts. The nature of the original puzzle seems to be *logical* and *a priori* as we saw in the discussion surrounding the HF (Clarke 2011). *Fodor* provided a metaphysical solution to the original problem in his later works, especially in *LOT2*. Some authors, nonetheless, were not fully convinced by that modification and criticized the view ever since on a number of grounds. *ML* tried to propose a model that could line up with *Fodor*'s original contention while remaining not ***radically*** nativist. Over the course of years, they showed a general tendency towards the empiricist side of their solution (particularly in *ML* 2011 and 2013). As we discussed here, their initial model, viz. *kind syndrome*, suffers at least from two theoretical flaws. Firstly, they did not clarify which account of natural kinds relates best to their posited notion of a *kind syndrome*. Secondly, they did not introduce clear measures that may help identify what their model would take for a property to be "highly indicative" of something while "accessible through perception." Even if such measures were introduced, the model could hardly account for many concepts, e.g., concepts that are highly theory-laden. That was the reason why Prinz proposed further measures than, say, just cue-validity in his own theory. Finally, we saw that *ML*'s

appeal to results from experiments on animals' instincts (taken as innate traits) may hardly amount to such an extremely complicated structure as the human conceptual system.

As an updated version of the long-standing tradition of concept empiricism, Prinz's theory is able to explain phenomena which would come as a mystery on *ML*'s account. Proxytype theory is also capable of making predictions that, regardless of how radical the situation might happen to be (like the case of feral children), will perfectly reflect real-life observations. It is held by nearly all authors of the field that any theory of concepts should account for what we called the "publicity" desideratum. *ML*'s *kind syndrome* model, as we explored, may provide almost no explanation for the cases of "congenitally blind people" who share concepts with sighted ones, as well as the case of "tetrachromat." On the contrary, the proxytype theorist handles such cases easily.

By and large, it seems that ***concept nativism*** could scarcely get deradicalized in the sense that *ML* envisioned, or at least their proposal would not easily pave the way for it. Put another way, to the contrary of what they suggest, *ML* (2011) could not plausibly claim to be "concept nativists, just not Fodorian radical concept nativists." Experimental evidence for innateness claims are hard to combine because the results are almost always open to alternative interpretations. In a similar way, one may argue that experimental evidence will fall short of rejecting a nativist account, and this has been maintained by *Fodor* himself throughout his career.

In addition to a long-standing tradition, an empiricist view benefits from a number of highly-advanced technologies, such as fMRI, that are used by neurologists and neuroscientists as a means of studying the brain and the neural system. It is possible that such an empiricist view may still lack the logical and theoretical firmness of Fodor's "informational atomism". In particular, Prinz's theory is less clear about "how a structured theory of concepts can be coordinated with informational semantics," (Prinz 2002, 237) the latter being the main source of

theoretical firmness for such a theory[23]. Nevertheless, when we decide to allow for an empiricist view, the explanatory resources of Prinz's proxytype theory can hardly be understated.

Notes

[1] See Rives, Bradley (2010). Jerry Fodor. *Internet Encyclopedia of Philosophy*.
[2] This is *ML*'s reconstruction of the example. In the original example something, e.g., a card, is "flurg" iff it is either green or square. See Fodor (1981, 266).
[3] See *ML* 2019.
[4] See Fodor 2008.
[5] INTELECOM, "Unconceptualized Reality," YouTube Video, 2:55, April 18, 2018, https://www.youtube.com/watch?v=nglKepVqW0o
[6] One may hold that the same objection applies to the empiricist account proposed by Prinz while he provides no direct response to HF version of the radical concept nativism puzzle or the non-HF version of it. We should, however, bear in mind that Prinz does not allow for innate concepts or innate (that is domain-specific) detection mechanisms. All he finds plausible are domain-general detection mechanisms. So the radical concept nativism puzzle, on Prinz's account, does not even arise and hence requires no treatment. For the full discussion see Prinz 2002, 189-235. Also compare with *ML* 2013 for a discussion about domain-specific/domain-general detection mechanisms.
[7] See Clarke 2011 for a fuller discussion.
[8] One can also think of the case where the experimenter has something like "FLURG = GAME" in mind. Chances are the subject of such an experiment never reaches a "criterion" for telling flurgs from non-flurgs, or so Wittgenstein would argue.
[9] In brief, we normally take "real" as synonymous with "mind-independent". Fodor's argument can be reformulated as "our minds are real, therefore doorknobs (as the things that strike our minds as such) are real". But this, simply, begs the initial question of "is there a reality independent of our minds?" To put it another way, one might ask "are our minds real?" is a question that needs to be addressed, before we can go on to Fodor's argument. But that question should have its answer at a higher level of reality, viz. the mind-independent reality, or we will simply face Putnam's "brain-in-vat" problem. Putnam's solution to that problem assumes a mind-independent reality.
[10] For sustaining mechanisms as a "general strategy" they count other types of such mechanisms like "possession of a complete theory of the kind in the extension of the concept" or "the use of experts" etc. However, they do not go further to explain those suggestions in detail. Their focus is rather on one type of sustaining mechanisms, namely the *kind syndrome* which they consider as "the most important type of sustaining mechanism." See ML (2002.)

[11] *ML* could argue that innate mechanisms come online during later years in the developmental plan. Some allegedly innate abilities are not present at birth but come on board at an early age. There are two problems with this response. First, we need independent bases for such a claim because the mere fact that "some" cognitive mechanisms emerge at a later age and they are

(allegedly) innate, does not establish grounds for the claim that the same holds for "conceptual system development", although makes us think of such a possibility. Second, the case of "feral children" provides significant evidence against that possibility (see Sec. 5).

[12] For the full discussion see *Prinz*, 2002 especially, Ch. 6.

[13] The full story was published in the National edition of The New York Times Magazine on Oct. 18, 1992, Section 6, Page 44 with the headline: Mapping the Mind. An online archived version is available at:

https://www.nytimes.com/1992/10/18/magazine/mapping-the-mind.html

[14] This is why Fodor strongly rejects a natural kind view, because he thinks concepts are completely mind-dependent.

[15] See for example Bird and Tobin 2018

[16] In Fodor's "flurg" thought experiment, in effect, the visual aspect of the object is considered all and everything it might possess as a property. See Fodor (1998, 59)

[17] ML (2011) in a last footnote point out that, despite their openly empirical defense against Fodor's criticism in *LOT2*, they still consider a list of concepts as innate. The concept "danger" is in their list. They rehearsed the same list in a later paper, this time, in the main body. See ML (2012)

[18] Eve V. Clark, *First Language Acquisition* (Cambridge University Press, 2016).

[19] For a fuller discussion see Jesse Prinz, "Beyond Human Nature," 2012, 1-14.

[20] This should be after "poverty of stimuli argument" by Chomsky (1980).

[21] See Putnam (1967), also see Cowie (2008) for a later analysis.

[22] NYU Center for Mind, Brain and Consciousness, "Debate: "Are There Innate Concepts?" (Susan Carey and Jesse Prinz)," YouTube Video, 2:06:20, March 26, 2016, https://www.youtube.com/watch?v=cVnbyB6gtes

[23] See Woodfield (2004) for a critical discussion.

References

- Anderson, David W., and Myrna R. Olson. "Word Meaning Among Congenitally Blind Children". *Journal of Visual Impairment & Blindness* 75, no. 4 (1981): 165-168. doi:10.1177/0145482x8107500404.
- Bird, Alexander and Emma Tobin, "Natural Kinds", *The Stanford Encyclopedia of Philosophy* (Spring 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2018/entries/natural-kinds/>.

- Bauer, C., Díaz J.L., Pasaye, E., Barrios F. (2014). “Subjective sensations in the brain: an fMRI approach to neurophenomenology,” Mind and Life Summer Research Institute (MLSRI) Conference, Volume: 5.
- Brunswik, E. (1956). *Perception and the Representative Design of Psychological Experiments*. Second edition. Berkeley: University of California Press.
- Clark, E. V. (2009). *First Language Acquisition*. Cambridge University Press.
- Clarke, M. (2011), “The Radical Concept Nativism Debate,” Proceedings of the 9th International Hawaii Conference on the Arts and Humanities.
- Cowie, Fiona, "Innateness and Language", *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2017/entries/innateness-language/>.
- Damasio, A. (1989). 3. Concepts in the Brain. Mind & Language, 4(1–2), 24–28. https://doi.org/10.1111/j.1468-0017.1989.tb00236.x
- Fodor, J. A. (1975). The Language of Thought (The Language & Thought Series, Volume 1) (1st ed.). Crowell.
    - — (1981). The Current Status of the Innateness Controversy. In J. A. Fodor, *Representations*. Cambridge: MIT Press.
    - — (1990). A Theory of Content and Other Essays by J.A. Fodor(1990-08-22). MIT Press.
    - — (1992) and E. Lepore, *Holism: A Shopper’s Guide*. Oxford: Basil Blackwell.
    - — (1996) and E. Lepore, The Red Herring and the Pet Fish: Why Concepts Still Can’t Be Prototypes. *Cognition* 58: 253–270.

— (1998). Concepts: *Where Cognitive Science Went Wrong* (*Oxford Cognitive Science Series*) (1st ed.). Oxford University Press.

— (2008). LOT 2: The Language of Thought Revisited (1st ed.). Oxford University Press.

- Gasparri, Luca and Diego Marconi, "Word Meaning", *The Stanford Encyclopedia of Philosophy* (Fall 2019 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2019/entries/word-meaning/>.
- Jordan, G., & Mollon, J. (2019). Tetrachromacy: the mysterious case of extra-ordinary color vision. *Current Opinion in Behavioral Sciences*, 30, 130–134. https://doi.org/10.1016/j.cobeha.2019.08.002
- Margolis, E. (1998). How to Acquire a Concept. *Mind and Language*, 13(3), 347–369. https://doi.org/10.1111/1468-0017.00081
- Margolis, E., Laurence, S. (1999). *Concepts: Core Readings*. MIT Press.

— (2002). Radical concept nativism. Cognition, 86(1), 25–55. https://doi.org/10.1016/s0010-0277(02)00127-0

— (2003). Concepts and Conceptual Analysis. *Philosophy and Phenomenological Research*, 67(2), 253–282. https://doi.org/10.1111/j.1933-1592.2003.tb00290.x

— (2007). *Creations of the Mind: Theories of Artifacts and Their Representation* (Illustrated ed.). Clarendon Press.

— (2007b). The Ontology of Concepts—Abstract Objects or Mental Representations? *Nous*, 41(4), 561–593. https://doi.org/10.1111/j.1468-0068.2007.00663.x

— (2011). Learning Matters: The Role of Learning in Concept Acquisition. *Mind & Language*, 26(5), 507–539. https://doi.org/10.1111/j.1468-0017.2011.01429.x

— (2013). In defense of nativism. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, Vol. 165, No. 2 (September 2013), 693-718.

— (eds) (2015). *The Conceptual Mind: New Directions in the Study of Concepts* (The MIT Press). The MIT Press.

— (2019). "Concepts", *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2019/entries/concepts/>.

- Prinz, J. (2002). Furnishing the Mind: Concepts and Their Perceptual Basis (Representation and Mind). A Bradford Book.

— (2011). Has Mentalese Earned Its Keep? On Jerry Fodor's LOT 2. *Mind*, 120(478), 485–501. https://doi.org/10.1093/mind/fzr025

— (2017). *The return of concept empiricism.* In H. Cohen & C. Lefebvre (Eds.), *Handbook of categorization in cognitive science* (p. 931–950). Elsevier Academic Press.

- Putnam, Hilary. "The 'Innateness Hypothesis' and Explanatory Models in Linguistics." *Synthese* 17, no. 1 (1967): 12-22.

- Rescorla, Michael, "The Language of Thought Hypothesis", *The Stanford Encyclopedia of Philosophy* (Summer 2019 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2019/entries/language-thought/>.
- Striem-Amit, Ella, Xiaoying Wang, Yanchao Bi, and Alfonso Caramazza. "Neural Representation of Visual Concepts In People Born Blind". *Nature Communications* 9, no. 1 (2018). doi:10.1038/s41467-018-07574-3.
- Turing, A., 1950, “Computing Machinery and Intelligence,” *Mind*, 59 (236): 433–60.