

In Search of the Concept CONCEPT

Joel Parthemore

PAICS Research Group
Centre for Research in Cognitive Science (COGS)
Department of Informatics
University of Sussex
Falmer, Brighton, UK

1 Introduction

We use concepts, and we reflect on our use of concepts, all the time. But what actually *is* a concept? One needn't mean the question in any narrow realist sense; we don't expect to find our concepts defined in physical reality, any more than we look to find infinity (though many researchers do look for and hope to find neural correlates of consciousness, and many of the same people might hope to find neural correlates of concepts). Rather, one need only be asking something like: what do we mean by "concept" for practical purposes? What role does it play in e.g. our Wittgensteinian-style language games?

In answering these questions, not much can be taken for granted. As Stephen Laurence and Eric Margolis point out, "...It's even controversial whether concepts are objects, as opposed to cognitive or behavioral abilities of some sort." **[9, p. 3]**

Why are these questions important? I find myself roughly in agreement with Jerry Fodor on this point: "For reasons I'll try to make clear, the heart of a cognitive science is its theory of concepts. And I think that the theory of concepts that cognitive science has classically assumed is in a certain way seriously mistaken." **[4, p. vii]**

In finding a suitable definition of concept, we may find that we need to build a representation or model of a concept. We may even consider a concept to be a kind of representation, in the sense of providing, at some level of abstraction, a kind of *likeness*. But while a concept of concept is, definitionally, itself a concept, that a representation of a concept could itself be a concept is not, perhaps, so obvious. Indeed, as some have pointed out, there appears to be a real danger in conflating representations with representeds, which may have

very little in common.¹ After all, a representation of a waterfall, in the form of a depictive painting or a photograph (or a fountain?), is not, itself, a waterfall. A concept of a bird is clearly not a bird. The danger, in conceptualizing about concepts, is that we may no longer be dealing with what we think we are.

This paper will consider various standard and less standard answers to the questions: what is a concept, and what is the relationship of concept to knowledge? Many will overlap. Some owe much to philosophy, some to psychology; some owe much to symbolic AI, others to autopoiesis; some to abstract theory, some to practical engineering.

The CYC project, an attempt at practical engineering, offers one answer. What is useful about the CYC approach, and what perhaps is not? The purpose to which one puts one's theory of concepts is going to have a lot to say about how one answers these questions. If the goal is a "true" AI, we might be inclined toward one sort of answer. If the goal is tools for helping people build external models of their conceptual domains, we might be inclined toward another.

This leads on to thoughts about "building a better concept". Are concepts in the main structured or unstructured (atomistic) entities – or is it possible for them to be both? How should one account for the way concepts can appear both as discrete entities (my concept of DOG, my concept of WOLF) and as part of some continuous, web-like structure, where there is no clear line between the end of one concept and the start of another?

Concepts are always concepts *of*. The referents can be categorized along broad lines: referents that are spatially (the sun) or temporally (the rising of the sun) extended; referents that seem to exist "in the world" (the examples above) or referents (the experience [quale?] of happiness) that seem to exist "only in the mind"; referents that are abstract (triangle) or specific (*this* triangle). This leads us back to a reconsideration of the relationship between representation and represented and occasions when the distinction breaks down.

The final section of the paper tries to bring as many as possible of the answers to "what is a concept?" together, with a metaphor of concepts as interlocking building blocks. What, if any, insights does the metaphor offer? Benefits and limitations are considered.

¹"Contrast an image of a cat (such as a photograph) with a real cat. The real cat is a mammal, furry, alive, eighteen inches long (say), and composed of flesh and blood, while the image is not a mammal, not furry, not alive, five inches long (say), and composed of paper and Kodak chemicals". [6, p. 74]

2 What is a Concept?

2.1 Concept as (potential) component of propositions

One standard way of defining concepts is as sub-propositional components of thought: i.e., the building blocks of propositions. Certainly for lexical concepts, this works well. But, though Fodor has toyed with the idea that they do, not all non-complex concepts² map neatly to lexical concepts. Indeed, there may be concept-like structures that play no direct role in propositions, whether in a language of thought or in public discourse. *Knowing-that* knowledge is typically approached propositionally; *knowing-how* knowledge is not. It is at least conceivable that both forms of knowledge might be explained in terms of common underlying structures. Why should *knowing-how* knowledge be fundamentally different in nature?

Furthermore, in defining concepts as sub-propositional entities, are we unnecessarily excluding the possibility that propositions themselves may be conceptual entities, albeit of a special sort (e.g., having or implying truth values)? If concepts *can* be propositions, then we need a new definition.

Note that when Fodor defines concepts as “constituents of mental states”³, he’s not saying anything different, because for Fodor and his Language of Thought hypothesis, mental states *are* by their nature propositional.⁴

2.2 Concept as mental representation

In the widest sense, concepts *are* representations, in that they stand (in) for their referents. There are three stronger claims one might be making when one takes a representational approach to concepts.

First is that concepts are *mental* representations (or components of mental representations), which might be taken to imply *internal* representations. I think this is consistent with Fodor’s Representational Theory of Mind (RTM). Concepts map internal states to (mainly) external referents. If one takes an autopoietic view that distinctions of “internal” and “external” are ultimately artificial and from an operational perspective disappear, then one probably won’t be happy with RTM. Second is that concept possession and use are part of a *computational* process based on content-preserving operations on strings or structures of symbols.⁵ This is standardly taken to require a strict separation

²...Complex concepts being phrasal constructs like “brown cow” or “old shoe”.

³“...Concepts are constituents of mental states. Thus, for example, believing that *cats are animals* is a paradigmatic mental state, and the concept ANIMAL is a constituent of the belief that *cats are animals*”. [4, p. 6]

⁴“It may strike you that mental representation is a lot like language, according to my version of RTM [Representational Theory of Mind]. Quite so; how could language express thought if that were not the case?” [4, p. 25]

⁵“In a nutshell: token mental representations are symbols”. [4, p. 10]

between syntax and semantics. Fodor takes it that way (it is, indeed, fundamental to his version of informational semantics); many anti-symbolists would as well. Third is that concepts are in some sense and in some cases likenesses of their referents. Note that one must be careful with what one means by “likeness”, not to limit oneself to *visual* likeness. Fodor writes: “The Idea idea is historically intertwined with the idea that Ideas are images, and I don’t want to take on that commitment. To a first approximation, then, the idea that there are mental representations is the idea that there are Ideas *minus* the idea that Ideas are images.” [4, p. 8] Note that, while Fodor is agnostic whether *some* mental representations are likenesses, he is committed to (lexical) concepts *not* being so, for in his account they are unstructured atoms.

2.3 Concept as likeness

Representation is commonly taken to imply *some kind* of likeness: generally not a *physical* resemblance but rather, at some level of abstraction, a structural isomorphism: parts of the concept correspond one-to-one with parts of the referent. One needs to be careful when one talks about representation and resemblance: there are well-known logical problems with the idea that resemblance yields representation.⁶ Rather it may be the case, part of the time at least, that representation yields resemblance.

Jesse Prinz’s proxytypes theory might be taken to fit this account. His proxytypes perform as they do in mental simulations precisely because of certain, albeit very superficial, structural similarities they bear to their referents, allowing them to behave in mental simulations as their referents would in the “real” world. As a bird has wings, so my BIRD concept has WINGS. My BIRD, of course, is not really a bird; at some point sooner than later, the structural isomorphism breaks down and what remains is as unrelated as the painting is to the waterfall. Still, if I THROW a STONE at a BIRD and HIT it, the result in my mental simulation should be as if I threw a stone at a “real” bird and hit it.

Note that the likeness – if we should call it that – is not, for Prinz, what makes this process possible. Rather, it is the process that creates the likeness, and then only to the external observer, not to the agent engaged in the simulation.

The problem with concepts as likeness is that not all concepts seem to work this way. It is easier, for example, to imagine a likeness of sorts between my concept BIRD and a bird than between my concept LOVE and love. But the problem may be only the visual metaphor once again intruding. After all, one can imagine love having certain components (e.g., affection) and certain intrinsic (required) and extrinsic (typical but optional) properties that are reflected in the structure of LOVE.

⁶Still the standard reference on this is [7].

2.4 Concept as product of pattern recognition

The account, at least according to a concept empiricist such as Prinz, might go like this: regularities in the perceptual stream are recognized by the organism and form the basis for low-level concepts: concepts generally, perhaps, far below the level of lexical concepts. Regularities in the regularities and regularities in *those* regularities are likewise recognized and form the basis for higher-level concepts. The resulting concepts can then be matched in a top-down fashion against their non-conceptual analogs in present perception⁷, layer by layer, expectations based on past experience guiding and simplifying our interaction with our environment: we see, and hear, and feel, what we expect to, until the matching-up breaks down, and we are forced to take a closer look. In a way this is repeating what was said earlier: our encounter with the world is necessarily, to greater or lesser extent, a conceptual one. Concepts are the expectations that drive experience.

It need not, of course, be the case that only *conceptual* expectations drive experience. Non-conceptual expectations may drive experience as well, as Ron Chrisley and I have been exploring in modeling the visual experience of an agent embodied in similar fashion to an AIBO robot.[**2**] Rather, what makes concepts special is their reusability, grounded in pattern recognition.

2.5 Concept as model

Can something be a representation of itself? – most of the time arguably not. A waterfall is not a representation of a waterfall, and a painting of a waterfall, though it is a representation of a waterfall, is not a representation of a painting. When we are talking of concepts as representations, then a concept of concept is surely a special thing.

On the other hand, philosophers often *do* talk about things being models of themselves: so for example, there is Rodney Brooks’ famous dictum that “the world is its own best model” [1, p. 5]. It is a not infrequent observation that any object models itself “perfectly”.⁸ What, if any, subtle distinction is being made here, or should “model” be taken as roughly synonymous with “representation”? Is concept-as-model the same as or different from concept-as-representation?

A model can be defined as a copy or a miniature representation that is able to stand in for its referent (e.g., in a children’s game of cowboys and Indians) or serve as a pattern for building the actual thing (e.g., in a sculptor’s or painter’s studio). Prinz describes his proxytypes as being like “scale models” of their referents. In his proxytypes theory, one of the primary roles of concepts is to *stand in for* their referents in mental simulations.

⁷... Non-conceptual in particular because they do not meet the Generality Constraint[3], existing only in the moment.

⁸Simon McGregor, personal communication.

If concepts are proxytypes, thinking is a simulation process.... Tokening a proxytype is generally tantamount to entering a perceptual state of the kind one would be in if one were to experience the thing it represents. One can simulate the manipulation of real objects by manipulating proxytypes of them in their absence. The term 'proxytype' conveys the idea that perceptually derived representations function as proxies in such simulations. They are like the scale models that stand in for objects during courtroom reenactments. They allow us to reexperience past events or anticipate future events [13, p. 150].

Not all thinking is simulation. As Brooks would have it, much or most of thinking is not. It is, rather, the result of direct engagement with the world.

But that engagement is, as we've said already, inevitably conceptualized, to greater or less extent; and it's not clear where, in Brook's view, concepts fit in.[14] And here, too, Prinz's proxytypes-as-models may have their value, confirming our expectations of the world around us: concepts projected over top of non-conceptualized experience and all but obscuring it, so that we might even think sometimes that the non-conceptualized experience does not exist.

2.6 Concept as perturbation

From the operational standpoint, according to autopoiesis⁹, the sharp distinction between representation and represented (or between model and modeled) disappears. What appeared before as representation and represented exist instead, for the operationally closed system, along a continuum from initial perturbation to traces of that perturbation – echoes if you will – reaching levels of conscious awareness and full self-awareness.¹⁰

Representations, in the usual sense of the word, exist only on the boundary between the organism and its environment. Because they are not internal to the organism, it may be tempting to call them *external* representations, as some philosophers do¹¹; but it is possibly more accurate to say that they are neither internal nor external, for neither are they in the environment. Note that whether the representational relationship holds is something that can only be determined by an external observer, “not by the operationally closed system.

⁹For a good introduction to autopoiesis, consult [8].

¹⁰Marek McGann, in a recent seminar presentation, presented just this argument in relation to Walter Freeman's work with rabbits and smells: “Whatever it is that's happening between the smell hitting the nose and the brain responding to the smell is not simply the brain being pushed into one representational state or another by the chemicals hitting the nose. Freeman... argues through his work that the olfactory bulb is a closed system...”

¹¹Ezequiel Di Paolo takes this view (personal communication).

The system cannot distinguish between externally and internally generated operational activity.”¹²

The argument from autopoiesis is not that humans (and potentially many other animals) do not represent things, nor that representational talk may not be unavoidable, nor even that representations do not exist. It is rather that representational talk is potentially very misleading and confusing.

So if concepts are representations (or models), then they exist not internally to the organism but on the boundary between the organism and its environment. The alternative is that concepts are perturbations – or the results of perturbations – caused by interaction with the environment. Think of a hand pressed into wet sand: the surface of the sand is the interface between sand and environment, the imprint of the hand its “representational” content. The compression of the hand beneath the hand print is the perturbation. There is an obvious mapping between the initial stimulus and the perturbation, but it is not a one-to-one mapping and not in any customary sense of the word representational.¹³

2.7 Concept as concept

However we try to define “concept”, there is something curiously self-referential going on. When we ask, “what is a concept”, what we’re really saying is: what is our concept *of* concept, or simply, what is the concept CONCEPT?

There is an obvious problem in using concepts to define what a concept is: it seems to invite a vicious circularity. Better to define “concept” in terms of non-conceptual or at least proto-conceptual things. It is a general rule in definitions never to use the term (or a derivative of the term) in its definition. On the other hand, the only tools we have for defining what a concept is are conceptual tools. Experience, as said before, is necessarily, to greater or lesser extent, conceptualized. But the problems don’t stop there.

Defining “concept” with concepts raises all the usual self-referential paradox problems. So there is the concept of all possible concepts, which surely must be a concept, but then it must also contain itself, and contain itself containing itself, and so on in an infinite regression.

Or consider this: one can imagine dividing all concepts into self-referential concepts (such as the concept CONCEPT) and non-self-referential concepts (such as the concept DOG). This invites the concept of all self-referential concepts and

¹²Tom Froese, email (personal communication).

¹³“It is the structure of the living system and its previous history of perturbations that determines what reactions the new perturbation will induce.” [8, p. 54]

the concept of all non-self-referential concepts. But is the concept of all non-self-referential concepts itself a self-referential concept or a non-self-referential concept? This is, of course, a variation on Grelling's Paradox¹⁴, itself a variation on Russell's Paradox. Unless one simply banishes universal quantification or creates some artificial separation between concepts and meta-concepts (say, decreeing that the concept of all non-self-referential concepts is *not* a concept, as Russell would doubtless have preferred), there seems to be an inescapable paradox in understanding concepts conceptually.

3 Posing a Replacement Question¹⁵

The question, "what is a concept?" could be taken to imply some objective and empirically decidable answer. In its place, one might be inclined to ask instead, "how is it useful to think about concepts, with respect to certain goals?" For example, if our goal is capturing common-sense reasoning in a format suitable for data mining, or creating a "true" AI, or providing tools for users to build externalized models of their conceptual domains, our approach to understanding concepts may be different in each case.

3.1 The CYC model

The CYC Project¹⁶ takes one, often very controversial, approach to the nature of concepts and knowledge. It might be summarized like this:

1. Knowledge, or at least "common sense" knowledge, is mainly conceptual.
2. All conceptual knowledge can be propositionally expressed.
3. All propositional knowledge can be captured in a form of higher-order predicate logic.
4. Disembodied knowledge, or knowledge without an agent, is still knowledge.

Non-conceptualists, enactivists and many others would have a problem with (1)¹⁷. (2) might be less controversial, but it still seems conceivable that there might be components of *non*-propositional knowledge that meet the Generality Constraint and have a similar abstract structure to sub-propositional units and so, on those measures at least, qualify as concepts. (3) will raise some concerns, both because higher-order logics are generally considered less "well-behaved",

¹⁴Divide all adjectives into two classes: self-descriptive ones (e.g., polysyllabic) we'll call "autological"; non-self-descriptive ones (e.g., green) we'll call "heterological". Question: Is "heterological" itself heterological?

¹⁵The title and something of the spirit of this section I owe to Blay Whitby.

¹⁶<http://www.cyc.com>.

¹⁷There is another assumption they will balk at: that knowledge is programmable. Consider this quote from Marvin Minsky, defending the CYC Project: "People have silly reasons why computers don't really think. The answer is we haven't programmed them right; they just don't have much common sense." (quoted on the CYC website)

and because Grelling’s Paradox is considered difficult if not impossible to avoid (which may, of course, be precisely what makes it attractive for describing concepts). (4) raises an interesting philosophical question: is a concept a concept without an agent to possess it, one that, furthermore, should be embodied and richly embedded in its environment?

A concept, to the CYC project, is a discrete and explicitly represented sub-propositional structure, of which there are currently 300,000 in the CYC database. Concepts are, by and large, richly structured entities that can be recursively defined. Concepts have definitions of a sort, but they are flexible, being constantly updated in light of present circumstances. Furthermore concepts are represented in a way that is said to permit global inconsistency while retaining local consistency.¹⁸

3.2 “True” AI?

Defining concepts this way is fine if your goal, as CYC’s is at least part of the time stated to be, along the lines of improved data mining, better search engines and data consistency checking. CYC’s concept CONCEPT allows it, in certain contexts, the ability to give a good semblance of common sense propositional reasoning. Looked at this way, the CYC project isn’t concerned with what a concept *really* is but what it can do with the concept CONCEPT it has.

Part of the time though, CYC’s goal appears to be something far more ambitious. Consider this quote from John De Oliveira, the president of the CYC Foundation: “In the Cyclify project. . . we ask you to imagine a world in which every single person is given free access to *programs that reason with* the sum of all human knowledge.”¹⁹ The CYC project is meant not just to process data but to understand, in a rich *human*-like way, the data that it processes: to achieve agentiality. The conviction is that once a critical threshold of concepts and propositions has been reached, the CYC database will start reasoning with the full depth of human common-sense reasoning. The CYC Project has been “priming the pump” for over 20 years, and Doug Lenat, the founder of the project, has made public statements that the threshold is close and the database is, in fact, very close to reasoning on its own.²⁰ When the focus shifts from what the CYC system *can* do to what it *will* do, from one that mines data to one that is aware and that reasons, then the choice of the concept CONCEPT may bear much closer scrutiny.

¹⁸This is not an uncommon approach in representing “common sense” knowledge and is, indeed, the one taken by Alan Rector in his Medical Informatics Group at the University of Manchester (personal communication).

¹⁹From the website.

²⁰A video tape of a recent presentation Lenat made at Google is available at <http://video.google.com/videoplay?docid=-7704388615049492068>.

It's unclear whether the CYC team understands this common-sense reasoning as requiring any kind of self-awareness; but it's also unclear how they can avoid it. So one requires a concept of self, and probably not just one but several: self as autonomous agent in the world ("third-person" self), self as myself ("first-person" self), and so on, all of which are intimately related to each other. [11] The more recognizably human-like the reasoning is meant to be, the more similar not just some of the concepts may be expected to be to human concepts but the concept CONCEPT itself: at least, there should be functional if not structural equivalence.

3.3 Externalizing the conceptual domain

What if our goal is not to re-create human conceptual abilities but to complement them? What will our concept CONCEPT look like then? What if, for example, our goal is to allow users of a computer system to build external models of portions of their conceptual domains, the better to make areas of implicit knowledge explicit and to check knowledge for consistency?

That was the goal of the Pharos Project at the University of Manchester in the early 1990s, which was seeking tools to support the social survey design process.²¹ The sociologists who design these surveys typically spend an hour or more with the wording of each question and still discover problems in the completed survey. It was felt that, if the designers could be assisted through creating an external model of their research domain, that this could provide the basis for an "intelligent" writing environment that would help them spot problems in the wording and the overall construction of the survey much more quickly.²²

The goal being not autonomous agency but collaborative support, not human-like reasoning but natural-feeling conceptual structures, the requirements on the concept CONCEPT could arguably be relaxed. Instead of functional equivalence, functional similarity might be enough. Structural similarity might or might not be useful. Rather than being bound to psychologically or physiologically grounded theories as starting points, or any other evidence for what a concept "really" is, researchers may find it sufficient merely not to violate any of our understandings of how humans represent knowledge.

A concept, to the Pharos Project, was a sub-propositional entity that *could* be viewed discretely; but it could equally be viewed as part of a larger, highly connected network. It could be viewed as an indivisible atom or as itself encompassing an area of the larger network. It drew on the work in linguistics into

²¹Described in [12].

²²"Equipped with a model of the research domain and an extensive model of syntax, the writing environment will, we hope, provide meaningful advice and feedback on the structure and content of questions written by the designer." [12, p. 90]

context-sensitive phrase structure grammars (CSPSGs). It was an attempt to provide a common structural basis to all concepts one might want to represent: be they concepts of physical objects or events or abstract qualities or mental states. It was based on propositional rather than predicate calculus, with the intention that the tools of predicate calculus would need to be derived from more primitive elements in the propositional calculus. Otherwise the goals were not unlike those of the CYC Project, albeit on a much more modest scale: a rich sense of context, local consistency with global inconsistency, “definitions” that are always provisional, always subject to revision. Still, arguably it suffered from many of the same over-ambitions!

4 Building a Better Concept

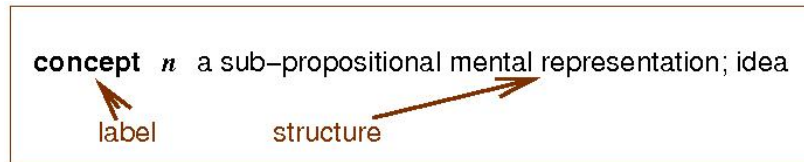
In deciding what a concept is – in determining our concept CONCEPT – we have many possible answers to guide us, most of which emphasize concepts as objects, most of *those* portraying concepts as sub-propositional representations, but some (as, perhaps, the enactivist approach does) preferring to see concepts instead as abilities or “sensorimotor contingencies”. We need to decide what our motivations are: are they theoretical or practical, or some mixture of both? Finally, we need to take into account those properties that concepts seem to exhibit.

4.1 Structured vs. unstructured

I take it as non-controversial that non-complex concepts *can* be viewed as unstructured atoms. Take GREEN for example, which might be understood as simply a pointer to all things green, or a label. Take the pointer (the analog of the word “green”) apart, and while you may get components of the pointer (the analog of the letters “g”, “r”, “e”, “e” and “n”), you will not get any components of the concept: that is to say, whatever sub-components you find, there will be nothing green or green-like or related to green about them.

I take it as likewise non-controversial that all concepts can be viewed as structured entities, *roughly* along the lines of classical definitionism: the structure provides the “definition”, or conversely, the “definition” determines the structure. So a bird has certain required components, including (normally) two wings; and a BIRD concept can be understood to reflect this. It has certain intrinsic properties, including not giving birth to live young, and extrinsic but typical properties, including an ability to fly; and the BIRD concept can be understood to reflect that as well. All of these – components, intrinsic properties, and extrinsic properties of the bird – can be *understood* as within the structure of the BIRD concept. At least, we often talk about concepts this way. Definitionism prevailed as a theory of concepts for a very long time, and it has never fully gone away.

This is like the distinction between the words in bold face and the definitions after them in a dictionary. The words in bold face and the definitions *together* constitute the dictionary entries.



The difficulty comes when we try to decide what concepts *actually* are, since however one understands concepts, one has to be able to account for this appearance both as structured and unstructured entities. For Fodor, all lexical concepts (which he would seem to equate most of the time to all non-complex concepts) are unstructured atoms. They are like the words in bold face in the dictionary. They may have structure associated with them, true enough – like the definitions that follow the words in bold face, and not unlike the association of birds with flying. But the structure is not *part of* the concepts, any more than flying is part of the essence of being a bird.²³ Question: Fodor’s concept CONCEPT seems like it surely should be a lexical concept itself and hence an atom. All the rich explanation he brings to his concept CONCEPT is only associated structure. What makes it a concept is that it is satisfied by and only by concepts. Is this possible, or is it viciously circular?

Prinz is not, by any means, a definitionist, in the sense of posing his concepts as static, context-independent definitions. But his proxytypes are similar to the definitions that follow the words in bold face, in that, although they may have pointers or labels associated with them, it’s the *structured entities* that constitute the concepts. In place of static, propositional definitions, his proxytypes are “perceptually derived representations that can be recruited by working memory to represent a category. A proxytype can 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’).” [13, p. 149] Note that Prinz is seeking to maintain Fodor’s informational semantics without the atomism [13, p. 164]: though a BIRD concept is a concept because of a rich structure that, in one way or another, captures prototypical properties of birds (like having wings), nevertheless a BIRD concept is a concept of a bird *because it reliably tracks (what other people likewise identify as) birds*; and a BIRD concept has the structural features that it does precisely because they are chosen as the most reliable detectors. [13, p. 156]

²³“... Though the *satisfaction* conditions of a concept are patently among its essential properties, it does not follow that the *confirmation* conditions of a concept are among its essential properties. Confirmation is an epistemic relation, not a semantic relation...” [4, p. 25]

As Prinz points out, atomism buys elegance at a high price: at the least a difficulty with explaining categorization²⁴, and at worst a tendency toward radical nativism, which Fodor has in the past flirted with, though he more recently (and wisely?) rejects. Prinz writes, “if concepts are structurally uniform (or uniformly unstructured), a uniform theory of concepts is easier to achieve.” [13, p. 94] Prinz’s proxytypes theory rejects atomism at the cost of losing that uniformity.

Are an atomistic account of concepts and a non-atomistic account in any way compatible with each other? Can we accommodate the way we seem to treat the same concepts sometimes as unstructured pointers, sometimes as richly structured composites of other concepts? I suspect we toggle between the two perspectives all the time. Is there any way to preserve that dictionary metaphor of the entries being both the word in bold face (the atom or pointer or label) and the structured description? I think we can, so long as we do not allow concepts to be both structured and unstructured at the same time, in the same context.²⁵ Following on from this, structured concepts should take only unstructured (with respect to the current context) sub-components.

There is also a relationship worth exploring between structured/unstructured concepts and levels of consciousness. Concepts at the level of self-awareness often appear unstructured (or only minimally structured). Much or all of the structure of “unstructured” concepts may lie below the level of consciousness – a point that Prinz makes himself. [13, p. 143] Both of these points require further development.

4.2 Discrete vs. continuous

When we introspect about concepts, they *seem* like discrete, individuable entities, and yet it is difficult, perhaps impossible, ultimately to account for them as such. Where, exactly, does one concept end and another begin? Does my DOG concept, perhaps, shift gradually into my WOLF concept, as my BALD concept shifts gradually into HIRSUTE?²⁶ The situation may be analogous to the way we treat small portions of the Earth’s surface as flat, even though we know the Earth is (essentially) round; so we make use of Euclidean geometry, in which parallel lines can exist, even though the Earth’s surface defines an elliptical plane, in which they cannot. It may also be not so dissimilar to the situation with structured vs. unstructured concepts: we can treat concepts as

²⁴... To which Fodor responds that that is not the job of a theory of concepts.

²⁵Needless to say, we should ditch any lingering tendency to say that concepts “just are” definitions, in any customary sense of the word.

²⁶“... It is difficult to determine where knowledge of one category begins and that of another category ends. Does my knowledge that dogs make good pets belong to my DOG concept or my PET concept?” [13, p. 148] The same problem extends, as Prinz realizes, to concepts themselves. He continues: “This worry may be surmountable. There is no reason to insist that boundaries between concepts must be sharp.”

being discrete and individuable, or continuous and blending into one another, but probably not at the same time in the same context. That would invite local inconsistency, which is to say, contradiction. Likewise we treat the same concepts now as classes, then as instances: e.g., DOG, which can be an instance of PET or MAMMAL or can be the class of all the different types of dog breeds or all the different individual dogs. Of course at least in most accounts, there will be some (primitive) concepts that can *only* be unstructured, just as there will be some concepts that can only be instances (“MY DOG FELLA”), and, perhaps, some concepts that can only be continuous. ²⁷

5 Concept of *What*?

To talk about conceptual referents is, necessarily, to use conceptual language: this should not be new, by now. Therefore categories of referents are necessarily categories of concepts. That much should be untendentious. But one could be seen to be making a stronger claim: much of the time we may afford ourselves the illusion that we are dealing in our conscious deliberations with the actual things-in-themselves, when what we surely must be dealing with are our conceptualized understandings of them (if, indeed, there is any difference). Concepts are representations; referents are represented. But, as autopoiesis reminds us, there are occasions where the distinction between representation and represented breaks down. Does the distinction between concept and referent likewise break down?

For a conceptualist like McDowell, concepts extend “all the way out”, into the world. That surely must mean that the referents of concepts are – in some sense – other concepts. What else could it mean?²⁸ For a non-conceptualist, there are *degrees* to which we conceptualize our experience – but if experience is never fully non-conceptual then neither, for practical purposes, is the experienced.

5.1 Physical vs. mental

There may be only one underlying substance – that the physical monists and neutral monists can agree – but there are, or there appear to be (as the conceptual dualists like David Papineau will concede, by way of defending physicalism), conceptually two, very different substances. There are things in the world, and there are things in the mind. As Descartes pointed out, things in the world tend to have properties like weight, mass and color; things in the mind seem to have very different properties like degree of pleasure or degree of pain. A memory of a summer afternoon is no less real – or valuable – for weighing nothing and consuming no space.

²⁷The last is an idea that Prinz considers: [13, p. 151].

²⁸This need not mean that McDowell is taking any anti-realist or skeptical stance. Quite the contrary: one can take in “how things are” precisely because they are already conceptual. [10, p. 25] “There is no gap between thought as such and the world.” [10, p. 27]

5.2 Spatially vs. temporally extended

But physical objects are not the only things “in the world”. Of things in the world, there are, or appear to be, again conceptually two very different kinds. There are those that are spatially extended: physical objects, with length, width and breadth; and there are those that are temporally extended: actions or events, with duration, with alternative ways they might have transpired, and with alternative perspectives from which they might be (or have been) viewed.

Note that things “in the mind” may inhabit a kind of space as well, and concepts themselves may inhabit what Salvatore Gaglio, drawing on the the work of Peter Gärdenfors, calls a “conceptual space”: “A conceptual space is a space similar to our tridimensional space. . . .” [5, p. 109]²⁹ Those dimensions might include, say, one of abstraction (from more specific to more abstract), one of similarity (from more similar to less similar, along some measure of “similar”), and one of typicality (from more typical instances to less typical, again along some measure of “typical”).

5.3 Homogeneous vs. heterogeneous

Some things are homogeneously structured: take water, which can be divided into smaller amounts of water. Some things are not: my bicycle, which can not be divided into smaller bicycles but is instead an assembly of specialized parts: the wheels, the chain, the pedals, and so on. This doesn’t apply only to physical things: some actions or events are homogeneously structured as well, others not. An act of walking can be divided into smaller acts of walking. A complex set of movements to remove myself from the path of an oncoming car cannot.

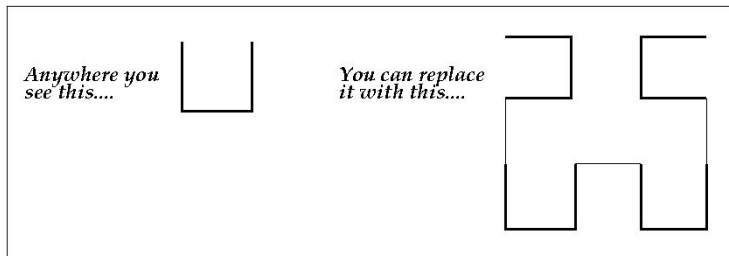
Of course, for those things that are homogeneously structured, at some point the homogeneous structure always “bottoms out”. Chocolate chip cookie dough can only be divided into smaller quantities of chocolate chip cookie dough so far: at some point, you’re faced with the underlying ingredients, the chocolate chips, the butter and so on. The “primitive element” for water is a water molecule. Divide a water molecule, and what you get will not be water. The “primitive element” for walking is a single step. Question: if concepts are homogeneously structured – as I think both Fodor and Prinz believe they are (concepts, to the extent that they are built up of anything, are built up of concepts) – what are the primitive elements for concepts? Are they perceptual primitives, as Prinz would have it? . . . Or are all non-complex concepts (non-perceptual) primitives, as Fodor would have it?

²⁹He continues (same page): “Each point of the space is characterized by the values of the coordinates in the qualitative dimensions and can be considered as an elementary concept, we could call it a *knoxel*, like a pixel is an elementary point of an image. . . .”

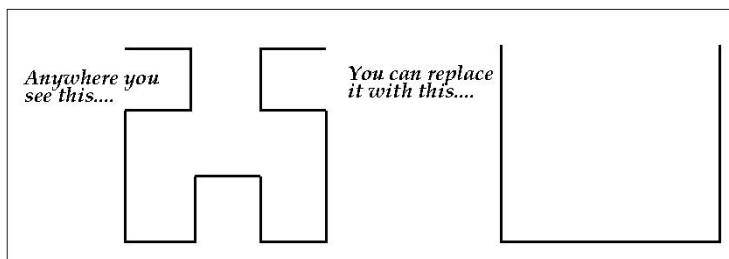
6 Concepts as Building Blocks

Both Fodor and Prinz are inclined not just toward concepts being built up from concepts but, at least ideally, those concepts being *uniformly* structured. This can best be achieved by keeping the number of primitive elements – the basic building blocks – small, as well as the rules for combining them. Chris Thornton: “I think if we could find one single building block, that would be great!”³⁰ The smaller the number of building blocks, and the simpler they are, conceptually and structurally, the wider the range of surface structures that can be built with them. Anywhere one examines the conceptual network, the building blocks and the rules should be the same. If concepts are similarly structured at *any* level, then one has *scale invariance*.

Par our earlier discussion, a concept should be viewable either as an unstructured atom or as structured entity itself composed of concepts. The atom should be replaceable with the structured entity (a process which could be repeated endlessly). The structured entity should likewise, going the other direction, be replaceable with the unstructured atom. An illustration using a two-dimensional Hilbert curve should make this clearer:



From unstructured atom to structured entity.



From structured entity to unstructured atom.

³⁰Personal communication.

If one replaces the two-dimensional Hilbert Curve with a three-dimensional Hilbert Curve and the piece of paper (or computer screen) with something like Gaglio's or Gärdenfors' conceptual space, one should get something of the idea I'm trying to convey. To wit: if one *could* find a "single building block", then the "building block" could be described with an iterated fractal equation. Things are more complicated when there's more than one basic type of building block, but the principle is the same: in this case, iterated simultaneous equations producing the fractal pattern collectively.

7 Conclusions

The goal of this paper was to attempt to answer the question, "what is a concept?" Having tried on various possibilities – sub-propositional entity, (mental) representation, resemblance, stand-in model, product of a pattern recognition process, autopoietic perturbation – it's not clear that we're any closer to an answer. But some things emerged:

- The answer is not nearly as simple as it might at first appear. Any answer we choose will be problematic. In particular:
- All attempts to understand concepts are conceptual. But any conceptual understanding of concepts is problematic. It invites not just vicious circularity but the usual self-referential paradoxes.
- How one answers the question depends a lot on one's motivations for asking it. I considered the CYC Project as an exercise in practical engineering; I considered the aspirations of strong AI; and I considered an application that may fall between the two.
- In practice, concepts seem sometimes like (unstructured) pointers or labels, sometimes like richly structured composites of other concepts. It would be nice, if possible, to have it both ways.
- Likewise, concepts seem sometimes like discrete entities, sometimes as part one holistic conceptual network; sometimes as classes, sometimes as instances. It would be nice to hold onto these insights as well and not be forced to choose one side or the other.
- Concepts and referents seem from one perspective like completely unrelated things (compare the painting of a waterfall and an actual waterfall), from another so close as to be hard to distinguish. In any case, the categorization of referents informs the categorization of concepts. Some things seem to exist "in the mind", some "in the world"; some are spatially extended, some temporally extended; some are homogeneously structured, some heterogeneously structured. However we choose to understand the concept CONCEPT, we need to be able to encompass all these things.

- Homogeneously structured things must, at some point, “bottom out”. Their “primitive elements” may themselves be structured, but not of the same material. If concepts are homogeneously structured they, too, will have primitive elements – which raises the question: what *are* those conceptual primitives?

The principle we kept finding ourselves returning to is that experience is necessarily, to greater or lesser extent, conceptual. Conceptual experience overlays and, most of the time, all but obscures non-conceptual experience.

Some brief thoughts are given on how the various definitions and insights might be brought together using an iterated pattern that would provide the structural uniformity both Fodor and Prinz consider, all other things equal, to be a good thing.³¹ The metaphor is given of a Hilbert curve filling a Gaglio- or Gärdenfors-style conceptual space.

References

- [1] Rodney A. Brooks. Elephants don’t play chess. *Robotics and Autonomous Systems*, 6:3–15, 1990.
- [2] Ron Chrisley and Joel Parthemore. Synthetic phenomenology: Exploiting embodiment to specify the non-conceptual content of visual experience. *Journal of Consciousness Studies*, 2007 (forthcoming).
- [3] Gareth Evans. *Varieties of Reference*. Clarendon Press, 1982.
- [4] Jerry Fodor. *Concepts: Where Cognitive Science Went Wrong*. Clarendon Press, Oxford, 1998.
- [5] Salvatore Gaglio. Intelligent artificial systems. In Antonio Chella and Riccardo Manzotti, editors, *Artificial Consciousness*, chapter 6, pages 97–115. Imprint Academic, 2007.
- [6] Sanford Goldberg and Andrew Pessin. *Gray Matters: An Introduction to the Philosophy of Mind*. M.E. Sharpe, Armonk, New York, 1977.
- [7] Nelson Goodman. *Languages of Art: An Approach to a Theory of Symbols*. Hackett Publishing Company, Cambridge, 1976.
- [8] Pier Luigi Luisi. Autopoiesis: A review and an appraisal. *Naturwissenschaften*, 90(2):49–59, February 2003.
- [9] Eric Margolis and Stephen Laurence, editors. *Concepts: Core Readings*. MIT Press, Cambridge, Massachusetts, 1999.

³¹Fodor gets it, essentially, for free, by having most concepts be uniformly unstructured!

- [10] John McDowell. *Mind and World*. Harvard University Press, Cambridge, Massachusetts, 1996.
- [11] Joel Parthemore. The representation of self in the representation of knowledge. poster presentation, Association for the Scientific Study of Consciousness 2006, Oxford, UK, June 2006.
- [12] Joel Parthemore and Jon Taylor. Charley: A linguistic formalism applied to writing environment design. *Intelligent Tutoring Media*, 3(2/3):85–92, May/August 1992.
- [13] Jesse Prinz. *Furnishing the Mind: Concepts and Their Perceptual Basis*. MIT Press, 2004.
- [14] Sharon Wood. When being reactive just won't do. In *Integrated Planning Applications: Papers from the 1995 AAAI Spring Symposium*, page 102ff, Menlo Park, California, 1995. AAAI Press.