Essays in Philosophy

Volume 5
Issue 1 *The Philosophy of Perception*

Article 9

1-2004

Hunting Fat Gnu: How to Identify a Proxytype

David DeMoss
Pacific University

Essays in Philosophy

A Biannual Journal

Vol. 5 No. 1, January 2004

Hunting Fat Gnu: How to Identify a Proxytype

Think about a gnu, a fat one. Ready? You have just tokened a proxytype. Or at least that's how Jesse Prinz would have it in his recent book, *Furnishing the Mind: Concepts and Their Perceptual Basis* (MIT, 2002). According to Prinz, "concepts are proxytypes, where proxytypes are perceptually derived representations [in long-term memory networks] that can be recruited by working memory to represent a category,"[1] like the gnu category. Your GNU proxytype stands in as a proxy for what it represents, the gnu category. Your GNU proxytype may be active or inactive. It is inactive when you have the GNU concept but are not currently using it in a thought. Inactive proxytypes are stored in long-term memory networks along with your standing knowledge; for example, your knowledge that gnu are four-legged creatures. But when you think the thought that gnu have four legs or that some gnu are fat, you activate your GNU concept and store it briefly in working memory; that is, you token your GNU proxytype. On this view, 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."[2] So when you thought about a fat gnu a few moments ago, you presumably tokened your GNU proxytype and perhaps extruded its belly shape.[3]

Prinz argues that his proxytype theory will apply to all of our concepts. Hence he defends Concept Empiricism: "All (human) concepts are copies or combinations of copies of perceptual representations."[4] It is the goal of my essay to first sort out the relationship, as Prinz would have it, between concepts and perceptual representations, and then proceed to examine the relationship between concepts and proxytypes. As I tease out some further details of Prinz' theory , I will compare his account of representational content (informational semantics) with a rival account (inferential role semantics), and also discuss the compatibility of his Concept Empiricism with Connectionism. I am led to conclude that a concept, for Prinz, is really a certain kind of causal-role player. Finally, I will suggest that the distinction he makes between concepts and proxytypes late in his book is an ad hoc move to shore up his faith in "realism."

To understand in what sense concepts are copies of perceptual representations, we must have an account of perceptual representations. What makes a perceptual representation a perceptual one? Prinz considers and rejects various proposals to distinguish perceptual representations by their syntactic properties.[5] For example, perhaps (unlike conceptual representations) perceptual ones lack "generality" in the sense that they cannot be recombined with other features in a new context. Prinz dismisses this move, in part because cognitive scientists often do assume that recombination of perceptual representations is possible, but also because denying recombination would immediately undermine his Concept Empiricism: if concepts are combinable but perceptual representations are not, then it seems unlikely that concepts could *be* copies of perceptual representations. Prinz also

rejects the purely semantic proposal to identify perceptual representations as mental states that represent perceivable properties, because even though they do represent perceivable properties, they do so only because a perceivable property is one that can be detected using perceptual machinery. Thus Prinz is led to a theory of perception that defines perception in terms of the machinery of detection, so that the faculty of the senses, as opposed to the faculty of the intellect, is distinguished by its detection mechanisms.

Prinz, then, defines perceptual representations as "representations that have their origins in the senses"[6] and "senses [he says] are dedicated input systems."[7] Within such systems, distinct collections of cooperative neural populations receive inputs from outside the brain and respond to a proprietary input class using different kinds of representational coding. "For example, vision responds to wavelengths of light, audition to frequency of molecular motion, and smell responds to molecular shapes."[8] Two clarifications are in order.[9] First, the senses are receptive, but not passive. We do transform, reconstruct, and interpret signals at the perceptual level, and we can form mental images by willfully reactivating our input systems. Second, although the senses are modular in the sense that each sense is domain specific and uses a dedicated system with its own representational coding, the senses are not modular in the sense of being informationally encapsulated; processing in perceptual systems can be influenced by information contained in other systems.

Prinz suggests that such influence does not require a common code for representation. He rejects "common-code rationalism" (the view that "perceptual modalities all use the same kinds of symbols as each other and as the more central systems associated with high-level cognition"[10]) as well as "central-code rationalism" (the view that thought is couched in a special code of its own not shared by any perceptual modality). Prinz' view is that perception and cognition use the same representational codes, but that these codes are fundamentally sensory and vary from sense to sense. Thus if concepts are indeed copies of perceptual representations, they will be "couched in representational codes that are specific to our perceptual systems."[11]

Yes, but why believe that there *is* such a direct link between perceptual representations and concepts? Well, suppose one is committed to informational semantics, the view that concepts obtain their identity by carrying information about aspects of the environment[12] and thus refer in virtue of standing in causal nomological relations to their referents.[13] The idea here is that a concept like GNU is the concept that it is, has the content that it has, because it is reliably caused or activated by gnu. An informational atomist, like Jerry Fodor, treats concepts as atomic (unstructured) indicators; for Fodor, the GNU concept gets switched on when perceptual systems detect the presence of gnu, or gnu sounds, or gnu pictures, or gnu-related words, etc. Prinz buys Fodor-like informational semantics (and in early chapters of his book argues for it over rivals), but rejects the atomism. Prinz drops the atomic indicators, and instead identifies concepts directly with the perceptual detection mechanisms which can be internally structured, such that "[t]heir parts detect parts of the things that cause them to be engaged."[14] Thus Prinz forges the link between perceptual representations and concepts. "Perceptual representations must be associated with concepts if concepts, [as required by informational semantics], attain content by reliable detection. The link between concepts and perceptual representations is forced on us by the best theory of how concepts get their intentional contents."[15]

That best theory is informational semantics from Prinz' point of view. But not everyone agrees. Without attempting to canvas all the leading theories of conceptual content, I would like to contrast briefly informational semantics with what Fodor regards as his arch enemy: inferential role semantics (or IRS, for short). According to informational semantics, a mental representation can carry information about *x* because the mental representation has a reliable causal relation with *x*. But, notoriously, this seems to leave no criterion for distinguishing, for example, the concept WATER from the concept H2O. IRS turns from the referent to the referee to fix the content; oversimplified, the idea is that your WATER associations will differ from your H2O associations even if your concepts of WATER and H2O have the same extension. Fodor describes IRS this way: "*what concepts one has is determined*, at least in part, *by what inferences one is prepared to draw* or to accept."[16] But Fodor abhors IRS: "I don't want content to be constituted, even in part, by inferential relations."[17] Why? Mostly because IRS has intolerable holistic implications; if the associations and inferences ramble on too far, then conceptual content cannot be precise. (My own reaction to such holism objections is to just accept the fact that content can be and is vague.) Prinz too would be resistant to IRS, though he does not use that term. He objects to theory theory (in which "concepts are construed as mini theories of the categories they represent"[18]) on the grounds that "they typically fail to explicitly represent essential features."[19] In other words, Prinz wants to ensure that concepts hook up to the *real* world, and the inferences one is prepared to make about *x*'s might fail to so hook. This objection is acute only for realists, like Prinz and Fodor.

Informational semantics also has its critics. Fodor's version of informational semantics is what Andy Clark calls a "symbol system" semantics. On this view, an intelligent cognitive system is a physical symbol system, "a physical device that contains a set of interpretable and combinable items (symbols) and a set of processes that can operate on the items (copying, conjoining, creating, and destroying them according to instructions)"[20]; the brain is seen as a symbol manipulator in which symbols are vehicles for psychological contents (concepts, propositions, and attitudes like believing or wanting) and their manipulation constitutes cognition. Clark argues that the symbol system account of mental contents "is false because there *is* no pervasive text-like inner code, and it is false because (*a fortiori*) the folk solids [concepts, beliefs, desires] do not have inner vehicles in the form of context-free syntactic items in such a code."[21]

Clark's rejection of a "text-like inner code" is akin to Prinz' rejection of "common-code rationalism." Prinz dumps the need for an inner common code by advocating a concept empiricism that allows cognition to borrow and piggyback on the varied representational codes used in perception. Likewise, Clark advocates a version of connectionism in which the line between perceptual representations and cognitive representations is not sharply drawn. Like Clark, I think that connectionism is on the right track, so I am intrigued by Prinz' account which retains an informational semantics without the need for a common or central inner code. A brief summary of connectionism will help here.

According to connectionism a cognitive system is a networked system of interacting units from which higher-level cognitive effects are emergent. On this view, the brain is a connectionist system, a maze of interconnections trained to recognize and respond to patterns of stimulation. Properly trained connectionist systems manage to categorize patterned inputs in the activation patterns of their hidden layers. Such categorizations are representations in a connectionist system. Thus,

conceptual representations emerge as stable patterns of activation within adjusted networks. Further, the training that results in categorization of an input pattern also disposes the system to respond with an output pattern appropriate to the given categorization. Thus the stable activation patterns that constitute connectionist representations are goal-directed; to categorize an input pattern is at once to recognize a goal and to become disposed to pursue it by certain means. Finally, it is important to note that connectionist representations are distributed over a large number of networked units. There is no coded unit or serial string of units whose activation means dog. Rather DOG shows up as an activation pattern in the higher dimensional spaces defined by the hidden units of the network. In fact, there is no single repeatable pattern that means dog, but a family of related hidden unit activation patterns that when carefully analyzed using various statistical methods can be interpreted as meaning dog.

If Prinz' use of informational semantics in his Concept Empiricism does not involve moveable symbol tokens,[22] then a connectionist like me ought to pay attention, because Prinz' account resonates in some ways with the connectionist account of conceptual content sketched above—specifically, the perceptual representations are the precursors to conceptual representations and they make use of the same coding. However, although Prinz believes that his own proxytype theory and connectionism may be compatible,[23] he would have trouble with a connectionist account of content that depends on statistical analyses to pick out "prototypes" (sometimes called "state-space semantics"). Prinz argues that prototype theories pick out superficial appearance features that cannot establish reference; (like theory theories) they do not identify the essential features of the alleged referent.[24] Of course, this is a problem only if one insists on a real appearance/reality distinction. I don't; Prinz does—but more on Prinz' realism later.

Setting the realism issue aside, perhaps a connectionist like me should buy into some version of Concept Empiricism. However, if Prinz' endorsement of informational semantics includes the Fodorian refusal to make any use of IRS to fix content, then I cannot buy it. Here's why. The informational content of a representation is always relative to an interpreter who must assume that the system using the information is goal-driven and hence acting for reasons. Thus the informational content posited for a representation must be constrained, not only by its reliable causal links with what it is supposed to be detecting, but also by the reasons for doing attributed to the system. The need to attribute reasons to the user of information means a necessary role for IRS in order to fix content, because reason-giving is an inferential activity. (I am confident that the dog has a TREAT concept not only because he reliably detects the presence of treats, but also because of what he is prepared to do (or "infer") to get them—namely, rollover on command. And if I am wrong about the powers of doggie-reasoning here, then so be it. The point remains that the positing of representation presumes the attribution to the system of a reason for using the represented information.)

If I am correct that a theory of concepts cannot do without some version of IRS, then what becomes of Concept Empiricism? Can Concept Empiricism have its informational semantics and its IRS too? Fodor will obviously say no. But what about Prinz? Let's return to his Concept Empiricism.

Recall that concepts on Prinz' view are internally structured perceptual detection mechanisms. He claims that the many facets of proxytype theory "can be unified under a single overarching idea: concepts are mechanisms that allow us to enter into perceptually mediated, intentionality-

conferring, causal relations with categories in the world."[25]

But now here is a puzzle. Since Concept Empiricism says that concepts are copies of perceptual representations, it would seem that the perceptual detection mechanisms that are concepts are in some sense copies of perceptual representations. So can a perceptual detection mechanism be a copy of a perceptual representation? Well, not exactly. The dual characterization of concepts as both detection mechanisms and copies of perceptual representations makes some sense if we recall the active/inactive distinction used in the opening paragraph of this essay, a distinction clearly used by Prinz though not explicitly formulated. As inactively stored in a long-term memory network the concept is a detection mechanism; but when the mechanism is engaged the concept is activated or tokened as a "copy" of the original perceptual representation which itself was a causal factor in creating the detection mechanism. This interpretation fits the (possible) account of copying found in chapter five:

> [R]epresentations in perceptual systems leave behind records in other systems that allow those representations to be regenerated in their original perceptual systems on subsequent occasions. Imagine that a stimulus causes a state in the visual system, and then some other system stores a record that can cause the visual system to generate a state of the same kind when the stimulus is no longer there. Stored records are not themselves copies, on this proposal; rather they are instructions for producing copies. . . . [A]n active token of a concept qualifies as a copy of a perceptual representation.[26]

Given such an account, Prinz can say that "concepts are mental representations of categories that are or *can be* activated in working memory."[27] When they are activated in thinking they are concept tokens, copies of perceptual representations. When they *can be* activated they are stored records, activatable detectors.

But does this really help? Can the GNU concept be both the inner state poised to play a causal role in producing a gnu thought *and* an inner state that (in part) constitutes the gnu thought? Can the GNU concept be both part of the cause and its effect? Prinz wants to identify concepts in terms of their causes, but he is also tending to identify them in terms of their effects when he characterizes detection mechanisms in terms of what they produce, namely, the alleged active "copies." If concepts are detection mechanisms that represent in virtue of being reliably caused by one thing to activate another thing, then those mechanisms are just as much transmitters as detectors. "A transmitter is something that represents in virtue of reliably causing something."[28] (As an aside, Prinz at one point characterizes states within our motor systems as transmitters; they represent what they reliably cause, movements.) If he insists that concepts are detectors, then he should also say they are transmitters. And if he insists that the representational content of the GNU concept is a function of its activating causes, then he should also say that the content is a function of the activated mechanism's effects, thoughts about gnu. Given Prinz' oft-stated claim that concepts are detection mechanisms, I think his concepts turn out to be inner states that are reliably caused by certain inputs *and* reliably cause certain outputs. Concepts are mechanisms identified by their input-output function.

The last couple of paragraphs have nudged Concept Empiricism a bit closer to a functionalist dispositional account of concept content, which is amenable to my brand of connectionism in which

to categorize an input pattern is at once to recognize a goal and to become disposed to pursue it by certain means. Still, informational semantics has its role to play; for example, the GNU concept-as-detection-mechanism serves as a conduit of information as it reliably detects gnu and thereby causes gnu thoughts (and behaviors). However, it seems to me, that this nudged version of Concept Empiricism is in obvious need of a little IRS, precisely because its conceptual representations are goal-oriented. Andy Clark calls these "action-oriented representations": "representations that describe the world by depicting it in terms of possible actions [and/or possible thoughts, I would add]."[29] Once again, assuming as interpreters that the information is being used by the system to achieve goals, the contents of action-oriented representations will be in part constrained by the inferential role they play in the reason-giving economy attributable to the system. So I am arguing that Prinz' Concept Empiricism (as I have been sketching it) could use a dose of IRS. However, as we shall soon see, Prinz' realism will throw it right back up.

For most of his book, Prinz uses the terms "concept" and "proxytype" interchangeably. However, in chapter ten, he separates the terms, which allows him to draw a distinction between what he calls intentional content and cognitive content. This is the same sort of distinction that Frege makes between reference and sense, which is very roughly the difference between what a term actually picks out in the world and what a term, as understood by its user, picks out in the world. Prinz' intentional/cognitive distinction applies to the content of concepts, (and roughly parallels what other philosophers call wide and narrow content). A standard example is that concepts expressed by the phrases "the morning star" and "the evening star" have the same intentional content (Venus) but not the same cognitive content; likewise, two different uses of the concept expressed by the term "water" could have the same cognitive content (liquid stuff in oceans and streams that we use for drinking and fire-extinguishing) but different intentional contents (H2O on Earth and XYZ on Twin Earth).

One more shift before we get down to business. In chapter ten (see especially section 10.3), Prinz substitutes the terms "real content" and "nominal content" for "intentional content" and "cognitive content," respectively. He does this because he is piggybacking on Locke's distinction between real and nominal essences. A real essence is out there in the world; it is whatever it is that gives a natural kind thing its observable properties. (H_2O gives water its fire-extinguishing properties.) Real contents are simply real essences. A nominal essence is in the mind; it is a complex representation of a natural kind thing's observable properties. Nominal contents are not simply nominal essences; rather, nominal contents are the properties out there in the world that are represented by the representations in the mind that constitute nominal essences—or more succinctly: nominal contents are the properties in the world that are represented as nominal essences in the mind. Prinz sometimes refers to these properties as "appearance properties" or just "appearances" for short. Thus the nominal contents of concepts refer to the appearances of natural kinds, while their real contents refer to natural kinds.

Prinz uses this distinction among contents to differentiate concepts and proxytypes (see section 10.3.2). According to Prinz, proxytypes (although they have real contents) must be individuated by their nominal contents, because that is how the perceptual representations from which proxytypes are derived are individuated (see section 10.2.2). More technically, Prinz claims that "we can identify a proxytype [by] the set containing the sets of [appearance] properties sufficient for causing the proxytype to exceed its critical detection threshold."[30] Concepts, however, are not simply

individuated by their nominal contents. Concept tokens can be grouped into types according to the contents they share. Member tokens of real concept types share real contents and member tokens of nominal concept types share nominal contents. Now, as nominal types, concepts just are proxytypes. "But [says Prinz] this type identity fails when concept tokens are individuated by their real contents."[31] This is because distinct real contents can correspond to the same proxytypes (as in the H_2O/XYZ case), and distinct proxytypes can correspond to the same real contents (as in the morning star/evening star case). It is a version of the latter sort of case that best illustrates the difference between concepts and proxytypes. Consider the following pair of proxytypes individuated by their nominal contents: (1) the representation of whiskey at t_1 as a golden translucent liquid and (2) the representation of whiskey at t_2 as a golden translucent liquid that causes inebriation. Both proxytypes have the same real content: whiskey. And we could add a third, fourth, and fifth proxytype in a series indicating a progressively more accurate understanding of whiskey as a natural kind. The conclusion drawn by Prinz is that although a concept *as a nominal type* is a proxytype, a concept *as a real type* is a set of distinct proxytypes that share the same real content.

In fact, he takes this as an illustration of how human conceptual schemes can make progress. For as real types, concepts can be a series of evolving proxytypes that progressively track the world more and more accurately. This move, late in the book, to distinguish concepts from proxytypes is a clever one. It does offer a solution to some standard problems about sense and reference, or narrow and wide content. However, it does so via an ad hoc insistence that our concepts must hook up to the really real world, even though we observers of appearance properties might not (do not? can not?) know this. Prinz himself admits, "This interplay between nominal and real stems from an underlying faith in the reality of natural kinds."[32] That was also Locke's leap of faith.

Is there anything wrong with this realist leap? And is Concept Empiricism obligated to take it? My answers are yes and no (I don't think so), respectively. I see Prinz as moving away from an items-in-the-head view of concepts (and percepts), and that is certainly the part of Locke that empiricism should leave behind. Prinz is right to resist symbol system semantics like Fodor's, because the moveable symbol tokens in such a system are once again items in the head that need to be paired off with some *really* special items out there that in turn need to be *accurately* (or not) represented in the head. Once one drops the items-in-the-head picture, then the circle of problems about accuracy of representation dissolve, and one needs no special explanation about how we track essences. But if Prinz drops his realism, then doesn't he lose that clever solution to those narrow/wide content problems? Yes, he loses *that* realist solution, but there is no reason (beyond realist fears about losing touch with reality) that IRS cannot be used to explain how the concept WATER differs from the concept H2O in cognitive content, while informational semantics explains how the concept H2O differs from the concept XYZ in intentional content. Without the realism, Concept Empiricism seems compatible with allowing IRS to help fix content right alongside a non-symbol-system version of informational semantics.

David DeMoss
Pacific University

Notes

1. Jesse Prinz, *Furnishing the Mind: Concepts and Their Perceptual Basis* (Cambridge, MA: MIT

Press, 2002), 149.

2. Prinz, 150.

3. See Prinz, 151.

4. Prinz, 108.

5. See Prinz, 110-113.

6. Prinz, 113.

7. Prinz, 115.

8. Prinz, 117.

9. See Prinz, 113-120.

10. Prinz, 117

11. Prinz, 119.

12. See Prinz, 89.

13. See Prinz, 123.

14. Prinz, 124.

15. Prinz, 127.

16. Jerry Fodor, *Concepts: Where Cognitive Science Went Wrong* (New York: Oxford University Press, 1998), 13.

17. Fodor, 13.

18. Prinz, 76.

19. Prinz, 86.

20. Andy Clark, *Mindware: An Introduction to the Philosophy of Cognitive Science* (New York: Oxford University Press, 2001), 28.

21. Andy Clark, *Associative Engines: Connectionism, Concepts, and Representational Change* (Cambridge, MA: MIT Press, 1993), 13.

22. And apparently it does not, for he clearly rejects the classical symbolic approach: “The picture that I have been defending differs significantly from orthodox theories of thinking in cognitive science. According to the orthodoxy, inspired by classical computing, thinking occurs in a symbolic medium, whose representations have subject-predicate structure and are manipulated by logical

rules. Thoughts are more like verbal descriptions [according to the orthodox symbolic approaches] than reenactments [of past or future perceptual events, as Prinz would have it]" (Prinz, 151).

23. See Prinz, 321n3.

24. See Prinz, 60.

25. Prinz, 164.

26. Prinz, 108-109.

27. Prinz, 149.

28. Prinz, 258.

29. Clark, *Mindware*, 95.

30. Prinz, 275. He actually says (emphasis added): "we can identify a proxytype *as* [not *by*] the set containing the sets . . . ." But he must mean "by." He uses "by" in the previous sentence. Also, this sentence clearly does not mean that proxytypes *are* sets of sets of appearance properties, since the last half of the sentence obviously presumes that a proxytype is a detection mechanism.

31. Prinz, 279.

32. Prinz, 281.
