Prinz's Theory of Conceptual Content

Marc Artiga (Logos; Universitat de Girona)

May 1, 2012

Abstract

In many of his arguments, Prinz's has heavily relied on a naturalistic account of conceptual content, which he has put forward and defended in several works (Prinz, 2000, 2002, 2006). In this essay, I outline his account of conceptual content and raise certain objections that suggest that this account should be abandoned.

1 Introduction

In this paper, I would like to discuss Prinz's naturalistic account of conceptual content. This is an aspect of his theory that has not been much discussed in the literature, if even some of his main arguments heavily rely on it. For instance, when Prinz (2006) argues that we can perceive abstract entities, he supports his argument with a particular view of how conceptual content is determined. In this essay, I would like to show that his own theory of content determination falls prey to important difficulties.

More precisely, here I will focus on Prinz's account of *referential* content (that is, truth-conditions), which Prinz distinguishes from something he calls 'Nominal Content' (Prinz, 2000) or 'Cognitive Content' (2002). The are two main reasons for that preference: first of all, Prinz's theory of referential content is the one he uses in most arguments in which a theory of content is playing an important role. Secondly, an account of Nominal Content (which, in any case, Prinz has not developed in much detail- see Prinz, 2000) will probably ride piggyback on a theory of referential content, so I think some of the problems of the former will probably carry over to any theory of Nominal Content.

The main goal of Prinz's theory of conceptual content is to explain in virtue of what process conceptual states acquire their content. In other words, Prinz wants to describe the process by means of which certain mental states come to have certain meanings. Why does my concept DOG mean dog rather than *cat* or *Obama*? This is a deep problem in philosophy that has generated an extense philosophical literature. Here I would like to outline Prinz's contribution to this important topic.

2 Prinz's Account

As he admits, Prinz's (2000, 2002, 2006) account is intended to be a combination of Fodor's (1990) Asymmetric Dependence Theory and Dretske's (1981, 1986) Informational Theory. According to him, for a concept C to have X as its content (that is, for C to mean X) two conditions need to be met: (1) there has to be a *nomological covariance* between C and X and (2) X must be C's *incipient cause*. Let us define both notions in some detail.

First of all, Prinz appeals to the notion of causal covariance between the concept and its referent. The intuition that the reference relation is determined by some notion of covariance is a common claim that has lead different proposals (e.g. Dretske, 1981, 1986; Rupert, 2008). However, Prinz *'s* concept of *nomological covariance* differs from other proposals in not being based on a covariance within the actual world, but across possible worlds. That is, C does not covary with X in virtue of the fact that the presence of C increases the probability of X's occurrence, as it is usually assumed. Nomological covariance has to do with covariance in proximate worlds. According to Prinz (2002, p. 241):

> COVARIATION Xs *nomologically covary* with concept C when, ceteris paribus, Xs cause tokens of C in all proximate possible worlds where one possesses that concept.

That is, John's concept DOG means *dog* partially because in all proximate possible worlds where John has DOG, tokens of this concept have been caused by dogs.

By appealing to causal relations that would hold in counterfactual situations, Prinz intends to to solve the 'Swampman problem'. The 'Swampman problem' is an objection based on a thought experiment, that was originally raised against certain historical theories of mental content, such as Millikan's (1984) and Papineau's (1984). Suppose that a lightening bolt strikes a swamp and a creature is produced (a 'Swampman') that happens to be microphysically identical to a normal human. Now, many people have the intuition that Swampman has representational states; since he is microphysically identical to a normal human, it seems he would behave and even talk in the same way as we do. However, any theory of content that requires that in order for a state C to represent X, there must be a causal relation between X and C is committed to the denying that Swampman has representational states, because nothing has caused his brain states. That is an unwelcome result for causal and historical theories of mental content.

But notice that, while Swampman lacks causal history, it seems his brain states support the same counterfactuals as we do*,* since *ex hypothesi,* swampman is microphysically identical to normal humans and the truth of many counterfactuals seem to be grounded on internal properties of human beings. So Prinz's notion of covariation seems to be in position to attribute representational states (and concepts) to swampbeings.

Nevertheless, Prinz is well aware that COVARIATION alone is too weak a relation for grounding semantic relations because there are many things men-

tal states nomologically covary with. First, my concept WATER nomologically covaries with water (H_2O), but it also nomologically covaries with XYZ, if in proximate worlds the transparent and colorless liquid that fills oceans and ponds is XYZ. In other words, anything that sufficiently resembles WATER would be included in he content of John's concept WATER (in the actual world). That seems to make concepts highly disjunctive. It seems we need to narrow down the set of possible candidate for content.

For this reason, (2000) Prinz adds a second condition: C means X only if X has caused the origin of the concept, that is, only if X is what Prinz calls the 'incipient cause' of C. In that respect, Prinz was inspired by Dretske's appeal to a learning period (1981). In a similar fashion, Prinz claims that a concept's reference should be identified with the cause that originated the concept.

In short, Prinz's view (Prinz, 2002, p.251) is the following:

INCIPIENT X is the intentional content of C if:

1. Xs nomologically covary with tokens of C and, in accordance with COVARIATION
2. An X was the incipient cause of C.

Let me now argue why I think this account is unlikely to be satisfactory.

3 Discussion

First of all, notice that there is some tension between 1 and 2. While 1 was designed to attribute representational states to Swampman, 2 precludes this attribution. Since nothing has caused Swampman's thoughts, there is no incipient cause of their mental states, and hence they are not about anything. In other words, by including incipient causes within the definition we are undermining the main motivation for endorsing preferring COVARIATION. Of course, there is still the intuition that concepts somehow covary with their referents, but it is not clear that the kind of covariation that has intuitibe support is the one put forward by Prinz. Furthermore, by adding 2, not only fails one of the main motivations for the thory: it shows that Prinz's account falls prey to the Swampman problem.

Secondly, Prinz does not provide any theoretical or empirical motivation for 2: why should we think the incipient cause plays such an important role? Why should we think the first cause of a mental concept plays a crucial role in fixing content? It is not obvious that this claim has intuitive support (though I admit that my intuitions my be biased at that point). Thus, as a first approximation, it seems INCIPIENT is not sufficiently motivated.

Indeed, I will argue that, even if independent reasons for motivating INCIPIENT were put forward, I think it suffers from serious difficulties. In particular, let me present 4 objections to Prinz's view. The first two arguments involve condition 1, the third argument involves condition 2 and the final remark is a general worry about this approach.

3.1 Indeterminacy

First of all, even if INCIPIENT can avoid including entities that exist in other possible worlds and ressemble very much the entities in the actual world (such as H_2O and XYZ), there are still many sources of indeterminacy that he does not properly address. For instance, John's MONARCH concept nomologically covaries with monarchs, but also with butterflies, and also certain retinal imagines (in particular, the retina image that is produced when seeing a monarch) because all of these states also cause John's MONARCH concept in all proximate worlds where monarchs cause them.[1] This is what most people call the 'Indeterminacy Problem' (which Prinz also calls the 'qua and chain problem'). Prinz is well aware if this difficulty, but he thinks condition 1 can deal with it:

> The first clause solves the qua and chain problems and can be embellished with further detail about the nature of the nomological relations involved to solve the semantic-marker problem.(...). For example, nomological covariance determines that my MONARCH concept refers to monarchs and monarch mimics but not to butterflies or retinal images, (...).

The problem is that, as it stands, 1 does not solve the chain problem. As we said, not only monarchs covary with C, but also butterflies, certain activations in the retina, neuronal activity in the optic array, and so on.

Prinz has outlined an original solution to this problem (which, interestingly enough, go beyond INCIPIENT), but it is unsufficient. Prinz (2002, p. 242-3) claims that whether a concept refers to a natural kind, an individual or an appearance is determined by a further condition, which he calls a 'semantic marker'. If, had the appearance X changed, X would still cause tokenings of concept C in the most proximate worlds, then X refers to a kind. If, instead a change in the appearance had stopped X to cause C in the most proximate worlds, then C is a concept of X-looking things. Of course, there are two serious problems with this view: First of all, *monarchs, butterflies* and *insects* are all natural kinds. So semantic markers are not fine-grained enough for the task at hand. Secondly, Prinz is inverting the order of explanation; it seems that the conditionals stated are true precisely *because* what concept C means rather than establishing the conditions for a concept to means anything. This is a general problem for his view that will be discussed below (3.4).

Indeed, it seems that even if we exclude states in different levels of distality (e.g. neuronal firings) and general properties (e.g. being a butterfly, being an animal) Prinz cannot explain why my concept MONARCH refers to monarchs rather than things that in the actual world resemble monarchs (like many other butterflies) because nothing ensures that the first thing that cause my MONARCH concept was a monarch rather than a similar butterfly. This problem will be extended and several consequences will be considered in 3.3.

[1] Indeed, in some cases the connection is much stronger. If monarchs are butterflies *necessarily,* then in all metaphysically possible worlds where a monarch causes MONARCH, a butterflies does.

So, pace Prinz, it is not easy to see how nomological covariance and the incipient cause can solve any of the problems of indeterminacy that affect other prominent theories of content.

3.2 Method of cases

The second objection is that the notion of nomological covariance appealed to in condition 1 causes INCIPIENT to attribute the wrong content to some mental states. On the one hand, condition 1 can be satisfied by the wrong entity playing the role of X. Suppose that John lost part of his visual capacities due to an extremely unlucky traffic accident when he was a child. Due to this impairment, he fails to distinguish oranges from tangerines. He applies the same concept to all of them. I think we would intuitively claim that his concept means something like *orange or tangerine*. Nonetheless, 1 and 2 might still hold in respect to oranges; it might happen that his first tokening of the concept was (by chance) caused by an orange. Furthermore, in all proximal worlds he has not had a traffic accident (remember that in the actual world he was extremely unlucky), so in these worlds he can perfectly distinguish oranges from tangerines and token this mental state only when confronted with oranges. So, it follows from INCIPIENT that *in the actual world,* his mental state means orange. But that cannot be right of John's actual concept.

Secondly, there seems to be cases where a subject has a concept even if condition 1 is not satisfied. Suppose John won the lottery. For this reason, he cancels a trip to Morocco and travels to China, where he bumps into an exotic fruit. He wonders how people call this fruit, how they would cook it,.. so John develops a well-formed concept of this fruit. However, in all proximal possible worlds, John does not win the lottery, so he travels to Morocco where he finds a different exotic fruit and wonders how do people call it, how they cook it,... So, again, INCIPIENT has as a consequence that in the actual world John lacks the concept that refers to the fruit in China because condition 1 is not satisfied.

Now, I think there is a plausible reply available to Prinz in support of the necessity and sufficiency of INCIPIENT.[2] Prinz could respond that the concept in the actual world and the concept in the counterfactual condition are different; since, according to COVARIANCE, in order to assess whether there is nomological covariance between the concept and its referent we must consider the most proximal worlds where a subject has *the same concept,* these counterexamples can be dismissed (this answer seems to be suggested in Prinz, 2002, p. 253) So, on the first example I gave, the concept applied to oranges and tangerines in

[2]One could claim that the 'ceteris paribus' clause in INCIPIENT is supposed to deal with this sort of cases, but it not easy to see how this clause should be interpreted (indeed, in Prinz (2002, p.13) there is no mentioning of 'ceterius paribus'). If 'ceteris paribus' is supposed to mean something like 'in normal conditions', it is hard to assess whether in the scenarios I present normal or abnormal conditions hold (without begging the question, of course).

A more general worry is that 'ceteris paribus' clauses are usually not accepted in theories of content determination without explicit analysis for a very good reason: these clauses seem to be introducing what has to be shown, namely what are the *normal* conditions for content determination (Fodor, 1990; Neander, 2006; Millikan, 2004)

the actual world is different from the concept applied to oranges in the counterfactual condition. Secondly, the concept I apply to an exotic fruit in China and the concept I applied to an exotic fruit in Morocco in the counterfactual condition are different concepts. So it seems Prinz has a satisfactory reply to all the cases I just presented.

However, I think this reply is utterly flawed. First, we may reasonably ask what grounds the claim that they are different concepts. In order for the reply not to be ad hoc, Prinz is required to provide some justification this assertion. The only way I see he could justify the claim that they are different concepts is either by appealing to the fact that they have different prototypes, proxytypes or functional roles or to the fact that they have different contents.[3] For instance, taking the similarity of content as a criterion, he could argue that in the first example the concept in the actual world (let us call it 'A-concept') means *orange or tangerine* and the concept in the counterfactual situation (C-concept) means *orange*. Since the only counterfactual condition that matters for content determination according to INCIPIENT is the one where the same concept is involved (the reply runs), and in the counterexamples there are always different concepts involved because they have different content, this is not a valid counterexamples to INCIPIENT. Unfortunately, this reply will not do for obvious reasons: Prinz cannot merely assume that the content of the two concepts differs, since what we are trying to settle is what determines the content of A-concepts. So he cannot individuate concepts across possible worlds by appealing to their content (at least, not when assessing whether a given concept satisfies 1 of INCIPIENT).

On the other hand, appealing to functional roles is also unsatisfactory, since in all the counterexamples we can stipulate that A-concepts and C-concepts share functional role in the mental economy of the subject: he is supposed to make the same inferences, perform the same actions,....Indeed, that gives us a good reason for thinking that the A-concept and the C-concept are indeed the same concept.

Prinz could adopt a different strategy. He could reply that the functional roles he appeals to in order to individuate concepts include wide dispositions (Harman, 1990); so, while in the actual world John is disposed to apply A-concepts to orange and tangerines, in the counterfactual world, he is disposed to apply it only to oranges. Since there is a difference in wide dispositions, there is also a difference in the functional role of A-concepts and C-concepts, and hence it seems Prinz could appeal to these dispositions in order to justify the claim that A-concepts and C-concepts are different. The problem, however, is that dispositions do not distinguish between right applications and mistakes, since we are also disposed to make errors. So, we can merely stipulate that in the counterfactual world, while he prominently applies 'orange' to oranges and orange has been the incipient cause, once in a while he makes mistakes and applies a C-concept to tangerines. If we add this condition, then the wide

[3] Prinz would probably opt for identifying concepts across possible worlds by appealing to something like proxytpe, prototypes or functional role (Prinz, 2002, p.7, p. 270)

functional roles of A-concepts and C-concepts are identical, and there is not reason to believe they are different. So the objection still holds.

3.3 Vagueness

The third problem is that, while it is usually thought that concepts can progressively change their meaning, Prinz cannot accommodate this fact without abandoning the key insight of his theory.

First of all, notice that many contentful concepts fail to satisfy 2. As Papineau (2006) points out, requiring that the content of the mental state has to be the first cause of the mental state seems too strong. If, for instance, one of our concepts is systematically tokened by a certain item, it is plausible to think that at some point it will come to represent this item, no matter whether it was the incipient cause or not. For instance, if the first time I saw a caiman I tokened the same concept that for the rest of my life I have used when I wanted to think about crocodiles, it seems very plausible to claim that I have been using the concept CROCODILE. But INCIPIENT entails that if when I created the concept it was caused by a caiman, then it represents caimans, and so I have been using the concept wrongly all my life. To say the least, that looks very implausible. Again, in this case Prinz suggests that the concept originally used for caiman and the concept I use most of the time are different concepts (Prinz, 2000, p. 253). Since they are different concepts, he seems to be able to accommodate the intuition that the concept I have used all my live in order to refer to CROCODILES in fact refer to crocodiles. Furthermore, in this case the he is not appealing to counterfactual worlds, so the problems raised earlier in identifying concepts across possible worlds do not apply.

However, when we consider the details of such an account, some tensions appear. Consider again the example in which the concept I have always been applying to crocodiles was incipiently caused by a caiman. Suppose at t_1 my concept C is caused by a caiman and at t_2 it is caused by a crocodile. Does the concept at t_2 mean caiman (and hence, it is wrongly applied to a crocodile) or is it the first tokening of a new concept (and hence it is rightly applied to a *crocodile*?) How can we know whether a concept is wrongly applied to an entity or whether it actually means something different? There are only two replies available to Prinz and none of them seems to be satisfactory. Prinz faces a dilemma.

On the one hand, Prinz can argue at t_2 John is correctly applying a new concept. The problem, of course, is that this account fails to account for cases of misrepresentation: if any case where the concept applies to a different item, this item counts as its incipient cause and it is considered a new concept, there will be no case where a concept is wrongly applied to a certain entity.

On the other, he can argue that that at t_2 John is misapplying C to a crocodile. Similarly, we can imagine that at t_3 John is confronted with a crocodile as well, and at t_4, and so on. As we saw, Prinz's answer is that after many tokenings of the concept being caused by crocodiles, at some determinate time t_n a different concept arises. Hence, (assuming INCIPIENT), there must be

a time $t_n$such that at $t_{n-1}$the concept was wrongly applied to a crocodile, and at t_n it suddenly becomes a new concept, whose incipient cause is a crocodile. I think this claim is very implausible.

The standard reply to this sort of cases is that at t_2 John is wrongly applying C to crocodiles, and there is no determinate point at which a new concept is created. Instead, there is a gradual change of meaning and, after a large number of times John has used C to refer to crocodiles, C gradually comes to mean *crocodile*. Unfortunately, this reply is not available to Prinz, since it contradicts the main insight of INCIPIENT, namely the appeal to an incipient cause. In a nutshell, the objection I am trying to raise is that INCIPIENT cannot account for progressive change of meaning. So, if condition 2 was unmotivated, now we see that we also have some reasons for rejecting it.

A related problem is that, according to INCIPIENT, non-deferential concepts can never have ambiguous contents (for an account of deferential concepts- see Prinz (2000)). Following INCIPIENT, if my concept JADE had not been deferential, it would either mean jadeite or nephrite, depending on the entity that first caused it. That is an implausible result since, as a matter of fact, some of our concepts are ambiguous (Millikan, 2000). So neither vagueness nor ambiguity can be accommodated within the theory.

3.4 Circularity

Finally, I would like to raise a general worry concerning this sort approach. A striking problem with INCIPIENT is that (as Fodor's Asymmetric dependence theory) we lack a (non-intentional) justification of why 1 should hold. Of course, it is true of many of our concepts that in the most proximal worlds the referent still causes them, but this is usually explained by appealing to the fact that concepts mean why they mean. In other words, Why do monarchs in most proximal worlds cause my concept MONARCH? precisely because MONARCH means monarch. The intuition that 1 is on the right track, comes from the fact if MONARCH means monarch, it seems the former will usually covary with the latter.

The root of the problem is that the truth of counterfactual statements is usually thought to be grounded in relations that hold in the actual world. For instance, consider the following counterfactual: *If Obama had not won the elections in 2008, McCain would have been the U.S. president.* We think this counterfactual is true because of certain causal relations holding in our world. The general problem with counterfactual accounts of content is that there is always the worry that the truth of the counterfactuals might be grounded on the intentional relations they are trying to explain. So, in order to provide a full characterization of a concept and its content, one should specify in virtue of what non-intentional property this nomological relation holds. The fact that no such characterization is provided, I think lends support to the suspicion that these accounts are merely assuming what they are supposed to show.

References

[1] Dretske, F. (1981) *Knowledge and the Flow of Information.* MIT Press.

[2] Dretske, F. (1986). Misrepresentation. In R. Bogdan (ed.), Belief: Form, Content, and Function. Oxford University Press

[3] Fodor, J (1991) *A Theory of Content and Other Essays,* MIT Press.

[4] Millikan (2000) *On Clear and Confused Ideas,* MIT Press.

[5] Papineau, D. (2006). Phenomenal and Perceptual Concepts. In Torin Alter & Sven Walter (eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism.* Oxford University Press.

[6] Prinz, J. (2006) Beyond Appearances: The Content of Sensation and Perception. *In Tamar Gendler & John Hawthorne (eds.), Perceptual Experience. Oxford University Press*

[7] Prinz, J. (2002) *Furnishing the Mind: Concepts and Their Perceptual Basis.* MIT Press

[8] Rupert (2008) Causal Theories of Mental Content. *Philosophy Compass*, 3: 353–380