Concepts: Foundational Issues

Contents

Preface

This doctoral dissertation has three parts. **Part one**, comprising chapters 1 and 2, addresses some basic commitments which must be presupposed in theorizing about concepts. Most basically, they are constituents of thoughts, and, like thoughts, they are mental representations. They are, in other words, subpropositional mental representations. Chapter 1 tries to clarify the notion of representing, after attempting to impress upon the reader that more rigour is required of initial explications of the representation relation than we are generally accustomed to. Chapter 2 dissects the arguments which I have been able to make out in the work of the modern arch-abstractivist Frege against the mental nature of thoughts and, by the same token, against the mental nature of concepts. His arguments are shown to be confused or, in the case of the argument addressed in the last section, at least to miss their target, leaving the notion of concepts as mental representations untarnished.

Part two, comprising chapters 3 to 5, pursues the aim of closing in on the concept posit in light of the widely shared understanding that concepts form a more specific class than just any kind of subpropositional mental representations. I would regard this part as having succeeded if the reader not merely saw merit in the criticisms advanced and in the positive line taken, but if she also emerged from this part with a heightened sense of clarity concerning the interrelations among a certain cluster of ideas which are customarily connected to the property of being a concept.

Chapter 3 advocates storage in memory as the basic criterion of delimiting concepts from nonconceptual mental representations. This criterion is clear-cut and well-motivated, having an explanatorily principled basis if any criterion has: Mental representations stored in memory, other than the mental representations of sensory-perceptual events, can appropriately occur in the absence of causal contact with their representata. In leading up to this criterion, the chapter subjects to methodological critique the typical manner in which philosophers justify their assumption of a conceptual/nonconceptual-distinction.

Chapter 4 shifts our gaze to the familiar dictum that concepts are, in some sense, mental representations under which we categorize things in the

world. Two versions of this idea of categorizing mental representations are distinguished. One, metaphysical in kind, is the notion of multiple applicability. The other, psychological in kind, is the notion of mental representations resulting from feature abstraction from perceptual representations, being correspondingly capable of classifying across different perceptual representations. For different kinds of reasons, which are relatively straightforward, both of these ideas cannot be regarded as capturing concepthood as such.

Chapter 5 continues on the positive side, reconnecting to an earlier introduced idea by which concepts, in contrast to sensory-perceptual events, are placed in the realm of so-called "high-level cognition", an idea for which storage in memory began to give a solid-looking basis. Concepts are not only stored in memory, they are also capable of entering into cognitive processes of a sort that is not "hardwired". "Flexible" cognitive processes is what I call them. I hope the subdistinctions and explanations which I provide make these complementary notions workably precise.

Part 3 of the dissertation moves on to hands-on contributions to current research. In one way or another, all three of its chapters take as their points of departure the familiar compositionality arguments against prototype theory, which have been incessantly reiterated over a number of years by Fodor, until recently the agenda-setting concept theorist in the philosophy of psychology. Together, the chapters expose and correct a number of deep-reaching confusions: in the basic, generally accepted, presuppositions of these arguments (chapter 6), in the topography of concept-theoretical positions assumed by Fodor and adopted by subsequent commentators (chapter 7), and in his followup comments made in alleged support and supplementation of these arguments (chapter 8). Most of these confusions have failed to be identified in the surrounding discussion so far. Unfortunately, there are many of them. If this sounds like a bold claim, I ask the reader to only consider the plethora of errors, incoherences, red herrings, and gratuitous misattributions by Fodor which are observed alone in chapter 8, or alone in chapters 6 and 7.

Overviews of the points through which we progress within chapters 7 and 8 have been laid out in a way that I am reasonably content with in the opening passages of these chapters, and there would be little point in repeating them here. But the core of chapter 6 can be profitably summarized in one argument, and in doing so below, chapter 8 is inevitably placed in context. I close with an attempt to briefly describe what awaits us in chapter 7.

The compositionality arguments—recapitulated within the text but assumed to be familiar for now—are aimed at "prototype theory" in the most general possible sense of this term, subsuming any position which identifies concepts with probabilistic representational structures. Probabilistic representational structures are ones involving individually defeasible feature constituents, thereby usually thought to effect a gradability of degree of concept satisfaction (which gives prototypes their name). Nowadays almost all psychological theories of concepts posit such kinds of structures. The compositionality arguments thus have broad potential import. Chapter 6 argues that it is trivial that compositional phrasal compounds themselves do not have prototype-theoretical natures. Its main line of argument can be recapitulated as follows.

The complex concepts which the compositionality arguments test for their conformity with prototype theory are, by assumption, phrasal concepts: concepts whose structures mirror phrasal expressions of actual language, and whose contents are expressed by these. Also by assumption, it is phrasal concepts posited in analogy with compositionally understood phrasal expressions which are thus considered. But the idea that prototype theory might possibly stretch to these compositional phrasal compounds in the first place, and that you might have to test for prototype assignability in order to find out that they do not do so, is incoherent. For the compositional constituents of phrasal structures are obviously not probabilistic feature constituents. This, however, is precisely what they would have to be if prototype theory were to stretch to compositional phrasal compounds, given the elementary fact that prototype theory is explicitly understood by everyone, including Fodor, to postulate concepts which immediately supervene on probabilistic structures ("bundles of feature representations", if you will).

What this implies is not that prototype-theoretical structures are incapable of combining compositionally. Rather, it simply shows that, in combining into compositional phrasal compounds, probabilistic structures combine into structures which are not themselves probabilistic, but rather "definitional" or "classical". There is nothing surprising or dubious about the idea of prototype theory's taking such a less-than-universal scope, entering into structures other than its own kind. The idea of lexically confining prototype theory has, naturally enough, been pointed to by a number of other philosophers already, albeit in a less principled manner than that resulting from the argument just sketched. This is where chapter 8, the longest and final one of this work, picks up the thread. It systematically addresses, for the first time, all the rejoinders which Fodor has put forth in scattered loci against this

confinement of scope, or "lexical confinement", as I loosely call it for ease of reference.

Meanwhile, chapter 7 argues for the ill-conceivedness of Fodor's notion of a recognitional concept, which occupies centre stage next to prototype theory as a target of his compositionality arguments, and which other authors have adopted in their discussions without protest. It turns out to conflate the idea of perceptual concepts, which is claimed to be targeted—concepts which are in some sense based on perceptual representations—with the idea of co-variationally constituted concepts. Moreover, the only even prima facie defensible basis for Fodor's application of the compositionality arguments to "recognitional concepts", in analogy with their application to prototype theory, rests on a connection of them to favourable circumstances of recognition. But, self-defeatingly for Fodor, the covariational atomist, this connection is one which he forges in light of what is really only the covariational aspect of his notion of a recognitional concept. Notice that these errors—and others, which I have not explicitly mentioned here—come in addition to the fact that the compositionality arguments, which these "recognitional concepts" are supposed to fall prey to, rest on confused presuppositions in the first place.

Quite an amount of written work previously earmarked for this dissertation was kept out of it for the benefit of delivering a reasonably stringent unit. This applies, for example, to a critique of conceptual atomism which was presented at the last congress of the Gesellschaft für Analytische Philosophie in Bielefeld, September 2003. In one way or another, the desire to provide such a critique formed the driving force behind most of what I felt I was doing in the earlier stages of my doctoral research. To mention another example (more interestingly, as I nowadays find), a chapter rejecting the received view on the status of mental compositionality was excised from this dissertation too. A distant precursor of that work was presented at the Joint Session of the Aristotelian Society and the Mind Association in Manchester, July 2005, and I am hoping to present a final version of it at the upcoming meeting of the GAP in Berlin, September 2006. I would like to emphasize, though, that the dissertation remains officially neutral on the view of compositionality maintained there.

Abbreviations

Apart from the two exceptions indicated below, the only abbreviations used in the dissertation are introduced in chapter 8.

C/NC: conceptual/nonconceptual (ch. 3)
PPC: postperceptual categorization (ch. 4)
LC: lexical confinement
CC: content constitution
CP: concept possession
DD: decompositionality denial
IPC: intraphrasal confinement
UA: Uniformity Assumption
DCPC: definitionality of compositional phrasal compounds

Notational conventions

Small capitals are used to form names of concepts (e.g., 'TABLE' refers to the concept of a table). Single quotation marks function as ordinary mention quotes. Double quotation marks are used for citing other authors in the body of the text, for scare-quoting expressions and, more generally, for the frequent hybrid cases where an expression occurs grammatically indistinguishably from object-level usage, yet with the simultaneous intent of conveying metalinguistic information. Italics fulfill the same function, where highlighting appears appropriate.

Passages cited from other authors have been assimilated to the above notation where necessary. Intrasentential omissions in quoting from others are marked '...'; omissions of suprasentential portions of text are marked '[...]'.

§1
A few basics

_____

1. Introduction
2. Pseudo-elucidations of the representational
3. The representation relation
4. Concepts, words, language
5. Concepts as constituents of thoughts

_____

1. Introduction

What are concepts? It would be nice if we could begin with an answer to this question which is fairly informative and simple, and which every theorist of concepts could agree upon. Unfortunately, no such answer is available that would not require a number of non-obvious clarifications and distinctions. This chapter begins to clarify some of the basic theoretical categories that put the notion of a concept in place.

Theorists of concepts usually waste little effort on attempts to characterize concepts in anything like a fairly rigorous manner. Ordinary linguistic usage is not of any apparent help here; Christopher Peacocke is right to note that the term "does not have in English a unique sense that is theoretically important".[1] What is more remarkable is how impressionistically even the theoretical notion of a concept is introduced in much of the most advanced philosophical and psychological work on the subject. Moreover, in those cases where clarifications of the notion of a concept are actually offered, the characterizations often suffer from even more serious shortcomings. One typical sort of exposition mentions features which do not plausibly reflect what the author himself takes to be distinctive of concepts. On another way of laying things out, putatively generic criteria are offered which are in fact so tendentious as to be unacceptable for many colleagues.

_____

[1] Peacocke (1992: 1).

There is a minimal characterization of concepts, however, which is universally endorsed among philosophers. It says that concepts are *constituents of thoughts*. In using the term 'thought' here, one does not intend to refer to a complete psychological state like a belief or a desire. What is understood by the term 'thought' in the minimal dictum that concepts are constituents of thoughts is a truth-valuable representation which determines what a psychological state represents and which is independent of the mode component of that representational state. I return to this with a few words of explanation later on in the chapter. In saying that concepts are constituents of truth-valuable representations, one also intends to imply that a concept, taken by itself, is not the sort of representation that is evaluable as true or false. It is a *subpropositional representation,* as one might also say.

For the time being, I speak of "concepts" and "thoughts" in a way that remains neutral with regard to two perspectives: On the one hand, there is the view of concepts as abstract representations. On the other hand, there is the view of concepts as types of concrete internal representations, or "mental representations", as one also says[2]. Internal representations are events or states. They are involved in the psychological events or states whose representational content they determine by being concrete parts of those concrete entities. They are also the elementary bearers of representational content in the world, not being related to their representata in virtue of any other, more basic kinds of representational states. In other words, they exhibit underived, intrinsic intentionality, as one also says.

I will shift to speaking of concepts as mental representations later on, with chapter 2 justifying the shift. In the meantime, I will try to remain neutral in my formulations with respect to the two ontological alternatives just mentioned. Given how categorially different from each other these two notions of a concept are, an attempt to say something substantial about concepts as such, while at the same time remaining metaphysically neutral, would inevitably result in convoluted and artificially disjunctive formulations. But I shall not say anything in the introductory sections about the nature of concepts qua concepts anyway. The fundamental notion we will need to start

---

[2]   I actually prefer the expression 'internal representation', which correctly suggests the spatial, intracranial location of a concrete representation inside an organism, to the expression 'mental representation'. The 'mental' in 'mental representation' is not part of the analysans offered by theorists positing these entities. The representational states in question are understood to be "mental" to begin with (whatever that is). Nothing is therefore added by the qualification of the representations in question as "mental". If I nevertheless employ the term 'mental representation' in what follows, this is due to the existence of a widely entrenched habit of understanding this term as referring to concrete internal representations.

out from is that human beings and other animals *represent* to themselves objects, or properties, or compounds of objects and properties in the world. A first main objective, correspondingly, will be to get a bit clearer about what it is for a state to count as a *representational state* at all.

An acute reader might warily ask here whether starting with representational states actually still leaves any room at all for the abstractivist view. A state, one would think, is eo ipso something temporally extended and spatially located, hence a concrete entity. In light of this fact, it might be complained that, in accessing the topic of concepts via the general notion of a representational *state*, there is already an inbuilt decision against the view of concepts as abstracta. After all, any part of a state would itself seem to be categorially incapable of being an abstract entity. So the concepts of the abstractivist could not be constituents of representational states. But is that not what tying concepts to representational states would lead up to? No; what I claimed to be the minimal common denominator of the understandings of 'concept' of interest to us is the idea that concepts are constituents of thoughts in the sense of 'thought' briefly sketched above. This sketch refrained from equating thoughts themselves with representational states. It was compatible with viewing thoughts as abstract entities. So, for all that has been said so far, constituents of thoughts might be abstract entities too.

Nevertheless, the ontologico-methodological complaint here might be this: An exclusion of the abstractivist view, though it is not immediately entailed by any of my explanations so far, is something that one arrives at inevitably if one take as one's basic starting-point representational states. "If ascribing the possession of concepts to somebody ipso facto is to ascribe representational states to her", someone might ask, "then how should any sense be made of this connection other than by assuming that concepts stand in a part-whole relation to representational states?" This would be a good question. I can only emphasize here that it is for the abstractivist to answer. If the rhetorical question points to an implausibility in the view of concepts as abstract objects, then it is an implausibility that the abstractivist is not confronted with as a consequence of any unfavourableness in the way I am laying things out. On the contrary, it goes without saying for the abstractivist too that she acknowledges the existence of representational states. Similarly fundamentally, she too cannot avoid acknowledging that the ascription of *concept possession* to someone constitutes an ascription of a representational *state*. The crucial difference, seen from this perspective, is just this: The abstractivist must take such a representational state to count as a concept-involving state in virtue of its standing in some sort of relation to ab-

stractly understood concepts other than the impossible relation of having them as spatial parts.

Besides these preliminary remarks on the metaphysics of concepts, another general remark is in place concerning the relative status of the concept posit in psychological explanations: One is tempted to introduce a work on concepts with the dictum that concepts are *the basic entities* within the domain of psychological explanations. Indeed, this is a standard piece of introductory rhetoric in writings on concepts.[3] But even the assignment of this status to concepts would be highly misleading, if intended as an uncontentious description. This is because it is at least arguable that organisms also entertain more basic forms of representational states than conceptual ones. If this is so—if organisms can be in states that non-conceptually, *pre*conceptually, represent features of the world—then the cognitive sciences surely deal with countless states of this sort. Later on, I will try to circle in on the more specific requirements to be imposed on internal representations in order for them to qualify as concepts.

2. Pseudo-elucidations of the representational

As I said, in order to get a grip on concepts, the first thing we need to understand is what it is to ascribe representational ("intentional", "content-bearing", "semantic")[4] states to a creature in the first place. Unfortunately, the uninformativeness often encountered in characterizations of concepts is surpassed by yet more thoroughly uninformative characterizations of the more general phenomenon of intentionality. It would be no exaggeration to say that few other central technical notions in philosophy get customarily explained as inadequately as the notion of intentionality or representing does.

There are different ways to diagnose the deficiencies of typical elucidations of the intentional. Whichever way one turns it, however, what is wrong with most alleged elucidations one is likely to come across is that they fall far

---

[3] A few samples of this, picked almost at random: "Concepts are the most fundamental constructs in theories of the mind." (Laurence & Margolis 1999: 3). "Concepts seem to be the very stuff of which cognitions are made." (Rey 1983: 237). "Concepts are the basic timber of our mental lives." (Prinz 2002: 1).
[4] Of all the terms familiar from ordinary use, I believe the term 'representational' is the one which is intuitively most naturally regarded as subsuming both psychological and nonpsychological cases of the phenomenon in question. I also should remark that, in calling a state "representational", I am taking myself to engage in talk that is no less innocuous than talk of "intentional" states or the like. I am assuming that talk of states representing the world is not yet committed to the bona fide representationalist assumption of distinct internal representations corresponding to the constituents of thought (Crane 1995: 134-35, emphasizes this distinction).

short of telling us, in any noncircular, literal terms, what would count as an instance of the phenomenon of intentionality and what would not. The habit of providing an obscurum per obscurius prevails. Often, the explanations are deeply metaphorical. In the case of metaphorical characterizations, the whole burden of defining the theoretically relevant notion of representing is just relocated to the task of naming the shared property which *grounds* the metaphor's comparison. And if one tries to thus spell out in literal terms what an author may have meant by his metaphorical characterization of intentional phenomena, one can expect to find oneself immediately stuck with a circularity. At least this is the case with attempted elucidations which maintain any recognizable connection to what most authors actually say in characterizing the intentional.

The following introductory passage from a philosopher of psychology provides a representative example of half-hearted attempts to delimit the theoretical notion of intentionality.[5]

> Human persons act as they do, and thus often behave as they do, because some aspect of the world is presented to them in some manner. The term 'content', as I shall use it, refers in the first instance, to the way in which some aspect of the world is presented to a subject; the way in which an object or property or state of affairs is given in, or presented to, experience or thought.

The example, which is by no means untypical, is up to its neck in metaphoricity and circularity. Talk of being "presented with" an aspect of the world, or of being "given" an aspect of the world, is simply a poetic rephrasal of representation talk. The author's reference to "experience or thought", in addition to this, is similarly inappropriate. What counts as a psychological state or process of thought crucially depends on what counts as representational, and the same is true for the notion of experience which the author is relying on in the quote.

The quote also illustrates that, to the extent that metaphors employed in explications of representational content do allow plausible restatement in literal terms, they are misleading. The suggestion with the metaphorical idea of something "being presented to a subject" is that representational states are consciously experienced states of a system. But the overwhelming majority of theorists (including the author of the passage above) assume in pos-

---

5   Cussins (1990: 381).

iting representational states, and plausibly so, that their existence does not depend on their being consciously experienced.

Absence of clarity also pervades the traditional characterization which has it that representational states are states which are "about" something, states exhibiting "aboutness" or "directedness". It is, again, surprising that such talk of a state being "about" something else seems to be seriously regarded by many as being of help in understanding the notion of representing. If you are a philosopher specializing on questions of representation and you have to explain to someone outside your discipline what you are working on, then she is likely to be more bemused than she was before if you tell her, by way of explanation, that you are interested in states that are "about" things. Especially if you take your talk of aboutness to be as basic a description of your subject matter as you can uncontentiously give.

In conjunction with the deficiencies mentioned, many characterizations of the intentional do not even pretend to be of the generality which one should require in order to get a better understanding of the notion: The descriptions rely essentially on examples, rather than employing examples merely for purposes of illustrating something independently delimitable. To get an idea of this, read the following passage from an important philosopher of language and mind. Again, I take this passage to be highly representative of a practice worth criticizing.[6] Again, it has been picked almost randomly and countless other such characterizations could be given:

> Some things—for example, names and utterances, memories, mental images and feelings of anger, pictures, charts and graphs—are said to represent, or be about, or be directed at other things. Intentional relations are relations that hold between such things and what they represent, or are directed at.

It lies in the nature of examples that they do not by themselves determine the classes they are members of. When it comes to the task of delimiting what a term applies to, examples can only serve to illustrate independently understood criteria, they cannot play any essential role. They fulfill highly important heuristic functions, but descriptions of the nature of theoretical posits must not rely on them essentially. And where could an explicit and general characterization of a theoretical posit possibly be more important than when it comes to identifying the fundamental subject of philosophy of language and mind?

---

[6]   Stalnaker (1994: 561-62).

In sum, taking a pseudo-definition and a couple of examples together does not make for a general characterization of a theoretical posit that is any more respectable than the sum of the characterization's inadequate parts. Philosophers introducing the intentional tend not to honour this methodological truth. The lack of an informative general characterization of the intentional is especially objectionable when the same authors who fail to come close to it take themselves to be in disagreement with each other about all sorts of fundamental questions connecting to intentionality. The disagreeing parties had better be able to specify beforehand, in a clear and non-metaphorical way, some distinctive features of their subject matter, especially when that subject matter is a theoretical posit with a specific explanatory function and specific explanatory properties suited to fulfill that function.

However, it is not as if philosophers in the philosophy of mind and language could not tacitly be seen as sharing a few crucial convictions as to what is involved in bona fide cases of representing. It is just that, too rarely, all of the crucial features are simultaneously made explicit and put together to mark off the subject of intentionality.

3. The representation relation

I take it that the conditions displayed below this paragraph are necessary and sufficient for any state *S* to count as a *representational state* in the sense that interests us here. States of creatures are the basic bearers of representational content. Any representational *object*—a street sign, a flag, an inscription of a word, etc.—is ontologically dependent upon the existence of representational states. Any state falling inside the causal-explanatory domain of psychology (any psychological state, in other words) is a representational state. Any representational state is either identical to a state fulfilling the three conditions below, or it is genetically or ontologically *dependent upon* the existence of such states, whatever form this relation of dependence may exactly take.

> **1.** *Causal-explanatory status:* Ascriptions of state *S* are capable of causally explaining behaviour.
>
> **2.** *Condition of distal relata:* Any explanatorily relevant specification of state *S* makes essential reference to some distant thing, or to some property instantiated at a distance.

3. *Conditions of correctness/error:* State *S* is, or involves as a component, a state that is capable of occurring correctly or incorrectly.

Condition 1 tells us what the fundamental task of intentional ascriptions is.[7] If something is not the sort of thing capable of exhibiting any behaviour, then it is not the sort of thing we ever have any reason to ascribe intentional states to. Two comments on this: Firstly, this does not already entail a metaphysical linkage between behaviour and intentionality in the sense that the specific content of a representational state is determined by the state's relations to subsequent behaviour. Secondly, one might wonder what exactly is meant here by 'behaviour'. The notion of behaviour is usually treated as not requiring explication. And indeed, the notion of behaviour involved here is not supposed to be an especially technical one nor is it particularly mysterious. What is important here is that a notion of behaviour is involved that does not itself depend on the idea of something occupying intentional states. For example, the notion of behaviour must not be spelled out in terms of whatever observable movements result from a creature's intentions to act upon certain of its beliefs and desires. If behaviour were understood thus, then condition 1 would be relying on the very notion to whose clarification it is supposed to contribute.

The notion of behaviour on the one hand, must be understood here in a way that is independent of the notion of intentionality. On the other hand it needs to be understood in a way that is not so innocuous as to be completely trivializable. Trivialization threatens insofar as there is a perfectly general sense of 'behaviour' according to which any situation in which a given object undergoes a state change can count as behaviour: How does a stone behave when you hold it in the air and let go? It falls to the ground. What behaviour does water exhibit when it gets heated? It evaporates into steam. How does a wall behave when you empty a bucket of red paint on it? It turns red. And so on.

The following restriction on the notion of behaviour seems to help here: Let something count as behaviour if and only if it is an internally caused move-

---

[7] I have noticed with curiosity that some authors place another function of ascriptions of intentional states on the same level as the function of explaining behaviour: the function of acquiring knowledge about the world by attributing beliefs to others. (See Field 1978: §5; Perry 1986: 103; Devitt 1996: 58-59; Field 2001: §6. These works do not cross-reference each other on this point, but there is a chance that Field's influential 1978 article is where the other philosophers mentioned here got the idea from.) I do not, however, think that it is really contentious that the task which I have mentioned is the one basic task that intentional ascriptions are there to fulfill.

ment or surface change of a system. This specification makes no use of intentional idiom, yet it avoids the trivialization of the term 'behaviour' in the condition that expresses the purpose of intentional ascriptions. For a person to suddenly change her colour because someone dropped a can of paint on her would, if this restriction is adopted, fail to count as any behaviour exhibited by the person. Contrast this with the case of a person whose face changes its colour because she blushes. This would count as behaviour because the person changes her colour in virtue of some internal, physiological process.

Condition 2, essential reference to distant features in state specifications, is where the most obvious part comes in of what it is for a state to actually *represent* something. If you want to specify, e.g., a biochemical event at a neural synapse, then you do not need to mention any condition which is far away in order to do that. Most importantly, you do not have to mention a condition that obtains somewhere outside the boundaries of the organism within which the described event takes place. If, by contrast, you want to specify an intentional state, then you do—for the basic cases of intentional states, that is. The condition of distal relata is what the idea of representational content is rooted in.

It is the condition of distal relata where the disjunctive proviso most clearly applies which was prefixed to the triad above: that of representational states either being ones which fulfill this triad *or* being ones dependent upon a state which does so. It is not seriously doubtale that representational states entered nature in order to enable creatures to successfully maouevre themselves through their environment. Representing successively loftier entities , representing one's own representational states, representing inexistents: these are representational states which must be assumed to have entered nature's scene later than representational states about properties, relations and objects in causal reach within a creature's environment.

Condition 3 is the only place where intentional idiom is implicated in the analysans. To give an analysis of the representation relation which managed to wholly do without intentional idiom cannot be the objective above; for to do so would be to provide a reduction of the intentional to the nonintentional already—a "naturalization of mental content", as it is also sometimes said. Such a reductive analysis would occupy a different metatheoretical status than that which is supposed to be occupied by the analysis above. We could not offer it without a substantial amount of vindicative work, and even then it would inevitably constitute a highly contentious thesis. Whereas, by contrast, the interest of the analysis offered above is supposed to lie in its

making explicit assumptions about explanatory role which are highly tacit, yet nonetheless very widely relied upon. The employment of intentional idiom does not mean that the analysis is circular to the point of vacuity. It does seem clear enough that the notion of having conditions of correctness—and thereby, complementarily, conditions of incorrectness too—is one which goes a long way towards regimenting and narrowing down what might be meant by unmodified talk of representing, or of being about something.

4. Concepts, words, language

Besides the characterization of concepts as the constituents of thoughts, there is another characterization which one is apt to encounter similarly often in expository remarks on concepts. It connects concepts to words. Concepts, according to this characterization, are those entities that are expressed by words. This formulation is unspecific enough as to be employed by proponents of both of the two metaphysical positions with regard to concepts mentioned earlier. The following opening passage of a paper by a cognitive psychologist and a philosopher of psychology provides a good example of the word-based characterization of concepts:[8]

> When one says 'dog', or 'yawl', or 'junta', there is the strong impression that discrete ideas correspond to each of those words. Cognitive science, following common sense, calls such ideas 'concepts'. Typically at least, we reflective common folk think of a word as the expression of a concept, of concepts as the constituents of larger mental units, such as thoughts, and thus of concepts as central to our mental life.

It is unlikely that the authors themselves would, strictly speaking, admit talk of "discrete ideas corresponding to words one utters" as a description of what concepts minimally are. It is not introspective appearances correlated with our own speech acts that motivate postulating concepts. Neither is essential reference to words tolerable in a characterization of our subject matter. The specification of any theoretically interesting notion of a concept will have to avoid reference to linguistic expressions altogether. Thought is ontologically prior to language; concepts are ontologically prior to linguistic expressions.[9] More precisely stated, this *priority assumption* amounts to the

---

[8]   Keil and Wilson (2000: 308).

[9]   Some philosophers care greatly to speak of "content" with respect to thoughts and concepts, while speaking of "meaning" only in the context of language, words and communication. If my usage ever oscillates here, let me remark at this point that it is adequate to speak of "meaning" in both cases and simply to specify which sort of meaning we have in mind whenever the need arises to do so. Nothing is lost by this uniform usage.

following: A creature can possess concepts without having any verbal abilities, any means of using or interpreting language in communication. Consequently, a creature can possess a concept without knowing any word that expresses the concept.

The intuitive plausibility of the priority thesis is very strong. This comes out in considering what someone who denied it would be committing herself to. Consider the case of an adult human being who has grown up on a lone island, in isolation from any conspecifics, and who consequently never had the opportunity to communicate with others and to acquire a public language. Can the view really be taken seriously that such person could not entertain a thought? Such a person will not be capable of contemplating something like the upsides and downsides of the European Economic and Monetary Union, but surely she will have concepts. If, for example, it is a tropical island she lives on and bananas form a central part of her daily diet, we may reasonably expect her to have the concept of a banana. Another case is that of a deaf and dumb person who has not learnt any language (be it conventional or sign language). Can the view really be taken seriously that such a person cannot entertain any thoughts? Furthermore, even though many non-human animals lack linguistic abilities, there nevertheless seems to be no explanatory alternative to ascribing representational states to them in order to explain their behaviour. But quite apart from these considerations about the need for ascribing representational states to infraverbal creatures, there is another straightforward argument for the priority assumption.

Pace Wittgensteinian doctrine to the contrary, public language serves the function of communicating the beliefs or desires of a speaker. Language is in this sense subservient to pre-existing representational states; it is not what brings basic representational states into being. The first linguistic acts of communication in hominid history *must* have been done to satisfy certain intentions on behalf of the communicator—presumably, her intention to express to attendant conspecifics a desire for something, or her intention to share with her conspecifics a piece of knowledge which was relevant to a situation they were in. There does not seem to be a remotely plausible alternative to this view of the genesis of language.

What would the alternative be? Presumably, it would have to be the picture that those of our ancestors on the verge of acquiring language began, at some point, to emit and exchange sounds in some regular fashion, in the absence of any intentions to communicate, and even in the absence of any beliefs to be communicated in the first place. How should that have happened? The picture would be one of a community whose members, at some point,

underwent speech-act-like events, much like the way in which prehistoric people sneezed, grew hair or emitted burps. Then, uninfluenced by any intentions of those primordial speakers (since they didn't yet have any), somehow, certain uniform uses would have had to have evolved for the different sounds which were emitted. There would not have been any representational states up to that point. Representational states would simply have come into existence by virtue of non-intentional creatures participating in certain habits of emitting noises.

This does not look like a picture worthy of being taken seriously. Importantly, however, it is not clear what other picture an opponent of the priority assumption could sketch if asked to address the question of the purpose and genesis of language.

Endorsing the priority assumption does not prevent one from acknowledging that our actual abilities to form concepts—at least of those concepts that are available to introspectively accessible reasoning—are immensely *enhanced* by linguistic abilities. Language does enable humans to form far more sophisticated and abstract concepts and thoughts at the introspectively accessible level than we would have without language. Presumably this is because, in subvocalized speech and surely in other, less obvious ways, linguistic forms are reimposed on the medium of reasoning that they originated from, guaranteeing an economy which enables a creature with highly restricted capacities of working memory to represent successively more complex properties and states of affairs. Higher degrees of cognitive sophistication ensue. The priority assumption allows this. Similarly, it allows that language significantly enlarges our inventory of concepts. But it assumes that, firstly, such enhancement can only take place on the basis of prelinguistically available concepts. And secondly, it assumes that, if linguistic abilities are needed to enhance our concept-forming abilities, then this is not a metaphysical necessity, but only due to the actual laws of cognitive neuropsychology in beings like us. What an opponent of the priority assumption, by contrast, would insist on is that concept possession is simply impossible without social, linguistic abilities.

I have just emphasized the priority assumption as a primarily negative claim, as one that denies any metaphysical principle according to which concept possession requires competence with a corresponding expression of public language. Conversely, however, the entailment running from concept possession to a corresponding piece of linguistic competence is, of course, a very different matter. Certainly any plausible account of linguistic understanding must assume competence with a word to require possession of an

underlying concept. Understanding a given word requires the possession of a certain concept, plus the linkage of that concept to a representation of the word's form within long-term memory. By definition, that concept will be the concept "expressed" by the word in question.

5. Concepts as
constituents of thoughts

I return to the minimal dictum that concepts are constituents of thoughts, and to the explanation that 'thought' is meant to refer here to a truth-valuable representation which determines what a psychological state represents and which is independent of the mode component of that representational state.

Let me briefly explain what the mode component of a representational state is. The idea behind asserting mode-independence is that any representational state involving a thought can be factorized into exactly two components, one component of which is the representation, the other one of which is the psychological mode. A given thought can be recombined with any psychological mode you wish to form a new representational state.

Psychological modes are, by common consent, individuated by their functional roles. More specifically, they are individuated by functional roles they occupy in relation to perceptual uptake, in relation to other representational states, and in relation to the initiation of action. One can think of the functional roles of the basic kinds of psychological modes as providing the slots into which, in our ascriptions of representational states, we insert postulated thoughts in order to achieve the explanatory and predictive successes of psychology. Let me say a few words about this contrast.

Identifying the psychological mode of an ascribed representational state is a trivial matter. At least it is as long we do not strive to surpass the degree of theoretical depth exhibited by the respective state ascriptions under consideration. For then, the psychological mode can be read off the verb used to ascribe the representational state. If Ballack believes that the lawn is green, then the psychological mode of the representational state he is thereby in is the mode of believing. If Miss Cutting wants the sun to shine, the psychological mode she is in is the mode of wanting. The list could be continued indefinitely: The verb used to ascribe the representational state is equally indicative of its psychological mode in ascriptions of intending something, imagining something, remembering something, hoping something, assuming something, fearing something, and so on.

We help ourselves to a relatively fixed and limited inventory of expressions in order to refer to the different modes of representational states. And we get by with this quite well. Indeed, even the comparatively small inventory of words which we employ in everyday psychological descriptions arguably provides us with far more terms than would be required to capture the inventory of psychological modes whose distinct postulation is explanatorily required. We acknowledge, corresponding to the verbs used in ascribing representational states, the existence of such psychological states types as belief/judgment, desire/volition, remembering, imagining, or intending. These state types by themselves already cover a very large part of common-sense psychology. Many states picked out by other verbs of representing—for example, expectations or hopes—can be analyzed in terms of such states. And many terms seeming to pick out other psychological modes present little more than stylistic variants of terms referring to modes as central as the ones just mentioned.[10] We have good reason to believe a very limited number of psychological modes to require distinguishing on explanatory grounds, at least with regard to explanations with the limited theoretical depth of commonsense psychology.

The understanding of the count noun 'thought' which underlies its employment in the minimal dictum needs to be kept apart from neighbouring but sharply distinguishable uses of this term in the same theoretical vicinity.[11] On one of these uses, 'thought' is simply understood as an umbrella term tantamount to 'propositional attitude'. In effect, the propositional attitudes are all those representational states that lend themselves to a factorization into a psychological mode and a truth-valuable representation. Thus, they are the psychological states that thoughts, in the presently relevant sense, are components of; they are not to be equated with them. On the other use to be distinguished from the presently relevant one, the count noun 'thought' designates the specific type of propositional attitude which involves the judgmental, belieflike psychological mode. It is the sense according to which, for example, the sentence 'Lahm thinks that he must tackle Camoranesi' is correctly translatable into the sentence 'Lahm believes that he must tackle Camoranesi'. I only mention these uses so as to put them aside. When I speak of "thoughts" henceforth, it will only be with the sense in mind that is relevant to the minimal dictum that concepts are constituents of thoughts.

---

[10] For example, it would be very pointless as regards psychological theory construction to insist on any difference in kind between wishes, desires, wants, and longings.

[11] As a non-count noun, to be sure, 'thought' has another theoretically important use, referring to *processes* of thinking (reasoning, cognition).

It is easily witnessed from examples of the sort just given that propositional attitudes are ascribed by means of a transitive verb that takes as its direct object a complement clause (typically a 'that'-clause, in English). This complement clause is the result of nominalizing a declarative sentence, and it behaves grammatically just like a singular term. In particular, it is a singular term which appears to refer to something truth-valuable. Philosophers, when they focus on so-called de dicto readings of belief ascriptions, often write as if thoughts were denoted by the complement clauses in such ascriptions and as if concepts, correspondingly, are explicitly referred to within these complement clauses. This is implausible for several reasons, but nothing depends on criticizing it for our purposes. Whether one posits concepts or not, and which kinds of concepts one posits, does not hinge upon what one's semantics of belief ascriptions looks like. Much ink has been spilt on the semantics of belief ascription in a powerful tradition within modern philosophy of language. My own view, for what it is worth, is that the tradition which engages in this practice tends to underestimate the variance among the ways in which complement clauses in belief ascriptions are routinely but efficiently enough employed by ordinary speakers.

§2

Refuting Fregean anti-mentalism about thoughts and concepts

When philosophers consider the possibility that concepts might be abstract representations rather than mental representations, their customary reference point is the work of Gottlob Frege. In what follows, I carve out what I take to be Frege's central moves in arguing against the mentality of thoughts and hence, by implication of their constituency, against the mentality of concepts.[1] It will be seen in section 2, the longest section of this chapter, how seriously flawed Frege's major argument to this effect is, involving several basic confusions. This argument, which pivots on a non-sharability claim about mental states, is spread across a passage of a few pages in the middle of Frege's famed paper "Der Gedanke" (pp. 40-43), published in 1918. Section 3 considers a number of side arguments in Frege's work to the same effect. It is explained that they are at least equally flawed, with some of them

---

[1] The objections to mentalism treated are the ones that I am aware of on the basis of my own studies of a range of Frege's writings. I hope that the list addressed is reasonably close to complete. If Frege scholars are aware of any further, interestingly different arguments, I would be glad to hear them.

making points which do not have the slightest thing to do with the negative thesis he is maintaining. These objections cannot even initially be taken seriously as attempts to call the psychological nature of concepts into question. Section 4, finally, turns to a pair of arguments involving some more respectable substance. They are, however, only directed against a specific representationalist idea, one that identifies concepts with mental images.[2] Moreover, they are only directed against a naïve and implausibly simple version of this idea.

Two preliminary clarifications are in order concerning the terms under which Frege discusses the questions of interest to us here: Firstly, what Frege explicitly discusses is the ontological status of what we have called "thoughts". It is not the concepts which form their constituent representations that he usually formulates his claims over. But his remarks apply in parallel manner to concepts. Secondly, Frege was concerned with candidate entities for what we would regard as concepts only qua entities the relations to which constitute understandings of linguistic expressions. Since, however, we are assuming that word understanding is necessarily based on concept possession, we can be comfortable with this way of approaching concepts.

In advance of considering Frege's actual arguments, we also must be clear about a tacit assumption of his regarding his very conception of the subject matter of the mental. This goes hand in hand with the first critique to level against Frege.

1. The phenomenally conscious:
Frege's unwitting distortion of the subject

It does not ever occur to Frege that psychological states might be anything other than *conscious* states. That Frege must be making this assumption is clear from practically each passage in which he poses the question of the possibly psychological nature of concepts. In "Der Gedanke", Frege enumerates some kinds of states of our "internal world" and subsumes them under the heading 'Vorstellungen'.[3] What we would nowadays formulate as the question of whether thoughts are psychological states or events is framed as the question of whether thoughts belong to the *"internal world"*,

---

2   See also the argument addressed in subsection 3.1.
3   I am assuming that we can ignore Frege's statement in the following quote to the effect that "decisions" form a separate kind of entity in the "internal world" which would resist subsumption under the heading of 'Vorstellung'. It is not clear what separate category he might mean, and in any case it does not play any role in what follows.

or tantamout to this, whether they are *"Vorstellungen"*. Witness the passage in which Frege lays down what he means by 'Vorstellungen':

> Auch der unphilosophische Mensch sieht sich bald genötigt, eine von der Außenwelt verschiedene Innenwelt anzuerkennen, eine Welt der Sinneseindrücke, der Schöpfungen seiner Einbildungskraft, der Empfindungen, der Gefühle und Stimmungen, eine Welt der Neigungen, Wünsche und Entschlüsse. Um einen kurzen Ausdruck zu haben, will ich dies mit Ausnahme der Entschlüsse unter dem Worte 'Vorstellung' zusammenfassen. (Frege 1918: 40.)4

Frege's unquestioned identification of the mental with the conscious may be typical of his days, but it lacks the perspective of scientific psychology. What Frege appears to be doing in this passage, as elsewhere, is to give us an inventory of phenomenally conscious states. But it is a basic assumption of the science of cognitive psychology that psychological states are posited and individuated largely independently of the phenomenal qualities of experience.

Even if Frege's metaphorically so-labelled "inner world" were understood as being only contingently associated with phenomenal qualities, and as being necessarily associated only with the functional property of being accessible to introspection—even in this case the domain of scientific psychology stretches far beyond, and the criticism goes through practically to the same extent.[5]

One of the points of scientific psychologist theorizing is that it posits states and processes which are inaccessible even in principle to the creatures that occupy them. This forms one of the features that put scientific psychology on a different level from that of commonsense psychology. The psychological states that do the causal explaining in scientific theory construction are ones that get hypothesized from an outside perspective, with the usual uncertainty that accompanies the postulation of any theoretical entities. They are not captured by any method of closing one's eyes and directing one's attention to the forum of one's inner world. What introspection tells us is only

---

[4]   Cf., however, the characterization of Vorstellungen in Frege (1897: 42) which, in effect, equates them with copies of visual perceptual representations retrieved from memory.

[5]   It is not, however, plausible that Frege is aiming at introspectively accessible states in a functional sense that could include both phenomenal and non-phenomenal states. One reason for this is that the distinction is not immediately obvious, and seems to have only been brought out in full clarity within the philosophy of mind of the 1990s (see, for example, Block 1995, or the lucid presentation of the distinction between two concepts of consciousness in Chalmers 1996).

of heuristic use for psychology. Contemporary philosophy of mind acknowledges this, too. It treats as a default position the assumption that psychological states are only contingently conscious, if they are conscious at all. The conscious is but a tip of the psychological iceberg.

This should not be surprising, given the role of psychological states that I tried to make more precise in chapter 1, stating the general nature of representational states. The conditions we specified for representational statehood did not involve any reference to consciousness. As a matter of fact, the very substantialness of the task of delimiting the intrinsically representational states of psychology rested on a tacit understanding of the fact that we were not permitted to help ourselves to the feature of consciousness in accomplishing it.

The main reason why the conditions for the scientific notion of psychological statehood do not involve reference to phenomenal consciousness is that psychological states as such play causal-explanatory roles (namely, in explaining behaviour), whereas phenomenal properties do not. Acknowledging phenomenal, felt properties is not a matter of explanatory hypotheses at all. Phenomenal qualities are immediately known by their experiencers. They are experienced, but they are not hypothesized. One of the things that are so special about experiential qualities is that their appearance *is* their reality.

Someone unfamiliar with the literature on this topic might protest that phenomenally conscious states surely count as "psychological" states if anything does. I share the spirit behind this protest; it is only terminology that is at issue here. The term 'psychological' is often by definition intended to refer to an exclusively causal-explanatory type of property. I have been adhering to this use of 'psychological' in the delimitation of psychological states and elsewhere so far. We might also have settled with using the term 'psychological' in a wider sense. However, like other authors, I prefer to reserve the more generic term 'mental' for this potentially wider sense.

The so-called "mental" is a truly twofold category. Any demarcation of the mental which aims to be at once theoretically uncontentious and sufficiently non-metaphorical is necessarily two-pronged along the lines just indicated. An adequate characterization of these two prongs would have to mention, besides the phenomenal/theoretical-distinction, a related difference between perspectives of epistemic access—sometimes also referred to as the difference between a "first-person" and a "third-person" perspective: On the one hand, there is the phenomenal conception of mental states which rests on the unmediated knowledge one has of one's own experience. On the other hand,

there is the psychological, causal-explanatory conception of mental states which primarily rests on the indispensability of ascribing representational states to others for purposes of causal explanation. Both aspects are central to how we think of ourselves as mental creatures. No fusion of these two aspects can be performed without a non-obvious and theoretically highly contentious argument to back it up.

The relevant point of this is that, in the general case, once they are understood mentalistically, *thoughts* and their constituent *concepts* are paradigms of mental states which are individuated over psychological, causal-explanatory properties. In the general case, any accompanying phenomenal qualities they might have are inessential to their individuation. The occurrence of thoughts, mentally understood, may be phenomenally laden in virtue of their occurring in such states as hopes, desires, or fears, but such phenomenality-ladenness is not what is distinctive of their semantic identity. Mutatis mutandis for concepts. To be the thoughts they are, thoughts must have the constituent concepts that they do; and for concepts to be what they are, they must represent what they do. And for concepts to represent what they do, their occurrences need not involve any concomitant phenomenal qualities.

It is surprising that, when the abstractivist option is considered by contemporary representationalist theorists about concepts, these theorists hardly even mention either that Frege's addressal of the phenomenal in place of the psychological largely misses the topic, or that concepts are theoretical posits of a merely contingently phenomenal nature, if at all.

Nonetheless, we had better take a close look at Frege's arguments. We must take seriously the possibility that Frege's arguments for the abstractness and against the mental nature of concepts might crucially contain elements that are independent of his unreflectively phenomenological perspective. Indeed, since Frege's abstractivist, anti-mentalist position does continue to be taken seriously by present-day authors, one should be very much inclined to think that there are such prima facie debatable elements in his arguments which survive a scientific understanding of the psychological. It will become clear in the course of the following that there are not. Not only that; Frege's considerations would have to be viewed as baseless and confused even if the realm of psychological states were delimited in accord with the de facto phenomenal.

2. Frege's (1918) multiply confused argument from bearer-dependence and non-sharability of mental states

After having made clear in the earlier-quoted passage of "Der Gedanke" that he means, in effect, phenomenally conscious states by 'Vorstellungen', Frege immediately goes on to enumerate, within two pages, four characteristics which he takes to distinguish mental states from "objects of the external world". These four characteristics lead up to his major argument against the mentality of thoughts.

2.1. Frege's contrastive claims about "Vorstellungen". Before I say anything further, let me first draw together Frege's contrastive claims about "Vorstellungen" in contrast to physical objects. (Occasionally, I will allow myself to use this term cross-linguistically, without the scare quotes.)

Claim 1: First, Frege briefly declares that "Vorstellungen" cannot be seen, felt, smelt, tasted, or heard, in contrast to physical objects. ("Zuerst: Vorstellungen können nicht gesehen oder getastet, weder gerochen, noch geschmeckt, noch gehört werden", p. 40.) **Claim 2:** Frege's second contrastive claim is that Vorstellungen are "had", and that they are had in one's consciousness. ("Zweitens: Vorstellungen werden gehabt. Man hat Empfindungen, Gefühle, Stimmungen, Neigungen, Wünsche. Eine Vorstellung, die jemand hat, gehört zu dem Inhalte seines Bewußtseins", p. 41.) **Claim 3:** Frege's third claim is that Vorstellungen "require a bearer", as he says. Physical objects, by contrast, are assumed to be bearer-independent, in some sense. ("Drittens: Vorstellungen bedürfen eines Trägers. Die Dinge der Außenwelt sind im Vergleiche damit selbständig", p. 41.) **Claim 4:** Frege's fourth and final contrastive claim about Vorstellungen is that they are not sharable by different people. Each Vorstellung, as he says, has only one bearer; no two people can have the same Vorstellung. ("Viertens: Jede Vorstellung hat nur einen Träger; nicht zwei Menschen [sic] haben dieselbe Vorstellung", p. 42.)

This list is best seen as falling into two parts. The first part of the list is formed by the first three characteristics given by Frege. They lead up to the feature of *bearer-dependence*. We do well to treat them in advance of presenting Frege's arguments. The second part of Frege's list is constituted by the fourth characteristic, that of *non-sharability*. This is the feature that receives by far the fullest justification by Frege within the brief space on which he expounds the four claims. Frege appears to think that non-sharability is entailed by the earlier attributed feature of bearer-dependence. We will see further down that his argument for this conditional is not par-

ticularly transparent, and that the conditional itself is clearly false. This will be important because, in effect, the whole substance of his argument from non-sharability against the mentality of thoughts/concepts will have to be located in whatever justification he gives for the non-sharability claim. Before we get to this, however, I turn to clarifying Frege's first three claims.

2.2. Bearer-dependence and statehood. Throughout the following explanations of Frege's considerations, one thing needs to be observed: Even though it is conscious states which are termed 'Vorstellungen' by Frege, we cannot treat him as being fully aware of the fact that he could replace 'Vorstellung' by 'conscious state' throughout. My emphasis here in on 'state'. It will become plausible in what follows that, while the notion of a conscious *state* plausibly *governs* Frege's use of the term 'Vorstellung', he *nonetheless lacks an explicit ontological awareness of the involvement of the property of statehood*.

Claim 1 by Frege amounts to the claim that conscious states, contrary to physical objects, are not perceivable by the external senses. The idea that claim 1 is naturally taken to express, when taken by itself, is simply that the conscious states of *others* are not observable. This claim is clearly true. When we are perceiving a conscious creature, our perceptual experience of that creature obviously does not tell us which conscious experiences it is undergoing. This is a familiar wisdom about access to conscious states.

The remark that Frege makes in support of claim 1—in fact, the only remark he makes in this regard at all, before asserting claim 2—expresses something quite different. Frege gives the following example by way of explanation: When you see a green meadow, the experience of greenness which you are thereby entertaining is not the object of your visual experience. It is rather the case that you are *having* an experience of greenness in seeing a green meadow. ("Ich sehe eine grüne Wiese; ich habe dabei den Gesichtseindruck des Grünen. Ich habe ihn, aber ich sehe ihn nicht.") This is clearly true too. What Frege expresses here is another platitude. An outer-directed perceptual experience is not its own perceptual object.

Claims 2 and 3, it can be seen, form spurious items on Frege's list. Claim 2 reduces to an attempt to restate what we must assume to form our subject matter from the outset, namely conscious statehood. Claim 3 reduces to a restatement of claim 2, more specifically, of that part of it which effectively asserts the property of statehood. As indicated already, the general metaphysical category of a state is crucial in what follows. Frege never mentions it. My critique will be headed in the following direction: Frege's unaware-

ness, at a fully explicit level, of the involvement of statehood in a mentalist notion of concepts, makes him run into some serious confusions in his discussion of mentalism.

Claim 2 falls into two subclaims. Frege's focus lies on claiming that Vorstellungen are "had"; the additional point is that they are had as parts of the "contents of someone's consciousness". It is clear that the latter, added subclaim just amounts to an ornate way of saying that Vorstellungen *are conscious*. This removes any remaining uncertainty someone may have had concerning Frege's restriction of the psychological to the conscious. The interesting part is in Frege's special emphasis on the feature of *being had*. If you look at claim 3, which asserts that Vorstellungen "require a bearer", you will notice that it merely constitutes a repetition—albeit in clearer terms—of the preceding claim's assertion that Vorstellungen are "had". From now on, we can refer to the idea which Frege expresses by *both* of the descriptions encountered in claims 2 and 3 as that of the *bearer-dependence* of conscious states.

Frege's only remark in support of claim 2 is constituted by the same cursory remark with which he supports claim 1 too, and bridges his statements of claims 1 and 2. In this briding remark, he says of an experience of greenness that "I have it, but I do not see it". Frege's intended support of claim 3, in turn, is provided by the entire passage that spans between claim 2 and claim 3 in his text. This vivid passage is worth quoting, for it gives us the best impression, apart from the actual claims which Frege highlights in his text, of what he is heading at in emphasizing bearer-dependence:

> Die Wiese und die Frösche auf ihr, die Sonne, die sie bescheint, sind da, einerlei ob ich sie anschaue oder nicht; aber der Sinneseindruck des Grünen, den ich habe, besteht nur durch mich; ich bin sein Träger. Es scheint uns ungereimt, daß ein Schmerz, eine Stimmung, ein Wunsch sich ohne einen Träger selbständig in der Welt umhertreibe. Eine Empfindung ist nicht ohne einen Empfindenden möglich. Die Innenwelt hat zur Voraussetzung einen, dessen Innenwelt sie ist. (1918: 41.)

This is true enough as far it goes, but the status that Frege assigns to this truth in his present claims reveals serious confusions.

The first thing that one is apt to be puzzled by in Frege's discussion is this. By 'Vorstellungen', as noted, we must assume that Frege means conscious *states*. But states are entities that are bearer-dependent as a perfectly trivial and general matter of fact. The notion of a bearer-independent state is obvi-

ously incoherent because the notion of a state just *is* the notion of a state *of* an entity—in other words, of a bearer—having a certain property during a given period of time. Of course there are no free-floating experiences. But this is because experiences are states, and there are no states without something they are states of.

It might be illuminating here to consider other examples of creature states for a moment. Compare pregnancies: Is there any such thing as a state of being pregnant without a creature that is *in* that state of pregnancy? Of course there is not. There are no free-floating pregnancies. Similarly, there are no free-floating malaria infections. And so on, for any state of an organism. The same is true for any physical state. It is incoherent to think there could be a particular state of having a mass of three kilograms without any object having that mass. And so on.

Frege, curiously, does not appear to regard the bearer-dependence of Vorstellungen as a trivial matter at all. Throughout claims 1, 2, and 3, what he is after is a rejection of a potential view on which a creature's "Vorstellungen" might occur independently of a creature's being related to them. How could Frege be concerned about this—how could he take such a demonstrandum to be worth arguing for? It is hard to believe that Frege might be overlooking the triviality that bearer-dependence is immediately entailed by statehood. And indeed, given such additional facts as that Frege does not ever mention the notion of a state in "Der Gedanke", it seems, rather disturbingly, that, from the outset, Frege has not clearly noticed that he is addressing states with his talk of "Vorstellungen". In particular, he has not clearly recognized that he is trying to contrast kinds of states with physical objects.

The complaint that Frege has not recognized the subject of his own discussion is confirmed by the presence of a second, more specific confusion in passages like the quoted one. This confusion, in fact, is the more immediately relevant one for us, given that we want to learn whether Frege has any interesting argument on offer against the mental (for him: conscious) nature of thoughts/concepts. In the quoted passage, Frege thinks he has noticed something interesting about conscious states qua *conscious* states. Frege clearly suggests that the bearer-dependence of the "interior world" constituted by Vorstellungen is specifically due to their being conscious. He *mistakenly puts the feature of bearer-dependence down to the specific property of being conscious,* when in fact, as noted, it is not the feature of being conscious, but the general feature of being a state of a creature, which is responsible for the bearer-dependence that Frege is so keen to emphasize here.

One might, I suppose, wonder for a moment whether Frege might not be construed as setting out from a minimal understanding of 'Vorstellungen' that corresponds to an ontologically less specific category than that of conscious *states*. On such an understanding, claims 2 and 3 might be taken to reflect an attempt by Frege to *circle in on* the specific property of conscious statehood, rather than as reiterating a triviality about it. However, we certainly did not go too far in assuming that Frege's 'Vorstellungen' minimally had to be interpreted as designating conscious states. That Vorstellungen are conscious states must indeed be construed as Frege's basis of departure, rather than as a resultant insight of his discussion about Vorstellungen. Let me briefly explain.

First of all, on the negative side, the metaphorical notion of being an inhabitant of an "inner world" is a completely useless one; it is none that we can seriously work with, even provisionally. Secondly, our interpretation receives full justification by Frege's introduction of the term 'Vorstellung', where he enumerated a number of conscious states and nothing besides. Similarly, within his very statement of claim 2, Frege enumerates "sensations, feelings, moods, inclinations, wishes", assuming these to be representative of "Vorstellungen". It is clear that he is aiming at types of states here. Thirdly, the idea that Frege might be aiming at a more general ontological category than that of states with 'Vorstellungen' is a non-starter. It does not get any more general than that if you are looking for subjects of instantiation for the property of being conscious.[6] Fourthly, however, and most importantly, the fact that Frege means states by 'Vorstellungen' is simply the basic presupposition of what he is taking himself to be doing here, namely opposing the mentalist conception of thoughts/concepts. To oppose the mentalist conception of concepts is to oppose the conception of concepts as mental states. Our very assumption that Frege's discussion bears upon the question of whether concepts are abstract or mental would appear to depend on assuming 'Vorstellungen' to designate mental states.

In contrast with this, we have at the same time had to assume that Frege lacks explicit awareness of the fact that he is addressing states with 'Vorstellungen', since he would otherwise not have cared to emphasize the trivial feature of bearer-dependence.

A relatively clear explanation seems to be available as to how the categorial nature of the entities which Frege is addressing with the term 'Vorstellun-

---

[6] As far as categories at the same level of generality are concerned, of course one could run the present discussion over events instead of states, if one preferred to. But everything would remain the same.

gen' could have come to be concealed from himself. His employment of the term itself seems to be a crucial factor in the etiology of his confusion. 'Vorstellung' could be viewed to be subject to a mental act/object-ambiguity, and Frege's adherence to talk involving this term intermittently traps him in a misleadingly hypostatizing perspective. Frege appropriately relies on an act notion of a Vorstellung when he groups emotions, moods, desires, and other such states under 'Vorstellungen'. But on the other hand, an object notion of a Vorstellung, which is at least not obviously tied to the statelike, exerts a strong pull on Frege's treatment too. It must be the hypostatizing pull exerted by this object understanding of 'Vorstellung' which makes Frege fail to clearly recognize the statehood of the entities he is considering concepts/thoughts to be candidates for, and which makes him regard the property of bearer-dependence as being so much worth emphasizing.

2.3. The non-sharability claim and the dark path leading up to it. Let us turn to claim 4 and the buildup to it. This, recall, is the claim that Vorstellungen are *not sharable* between different individuals. It is Frege's final and most important contrastive claim about Vorstellungen. Upon first and second readings, one still feels quite in the dark about what Frege's reasoning for this claim is supposed to be. This is no coincidence, as we will see. The substance of Frege's considerations here is thoroughly confused, and it is not an easy task to find a basis for charitable reconstruction. In the following dissection of Frege's relevant considerations, we will see that they do not include even an initially plausible-sounding reason for Frege's non-sharability claim about mental states. This simultaneously forms the major portion of work in refuting Frege's argument from non-sharability against the mental nature of concepts/thoughts.

Frege's non sequitur: From bearer-dependence to non-sharability. Frege takes the immediate warrant for his claim of the non-sharability of conscious states to be provided by an inference from bearer-dependence to non-sharability. The centrality of this conditional to Frege's argumentation is clear enough. Here is one of the most explicit statements:

> Wenn jeder Gedanke eines Trägers bed[ürfte], zu dessen Bewußtseinsinhalte er gehört[e], so [wäre] er Gedanke nur dieses Trägers ... . (1918: 43.)

In addition, here are remarks from the last passage preceding Frege's highlighted claim 4 in the text. Anticipating the assertion of the non-sharability of Vorstellungen, he expresses a view on which the alleged non-sharability

of Vorstellungen is intimately related to their dependence on individual
bearers:

> Kein anderer hat meine Vorstellung; aber viele können dasselbe
> Ding sehen. Kein anderer hat meinen Schmerz. Jemand kann
> Mitleid mit mir haben; aber dabei gehört doch immer mein
> Schmerz mir und sein Mitleid ihm an. Er hat nicht meinen
> Schmerz, und ich habe nicht sein Mitleid. (1918: 42)

Frege also often expresses the conditional from bearer-dependence to non-
sharability in contrapositive form. The first of the following quotes (in
which I have contracted Frege's claim 4, as displayed by him, with his im-
mediately ensuing remark) considers, per impossibile, what would be the
case if Vorstellungen *were* sharable, claiming that in this case Vorstellungen
would have to be bearer-*in*dependent. The second passage, which occurs a
paragraph further down in his text, explicitly states the same general impli-
cation from sharability to bearer-independence:

> Jede Vorstellung hat nur einen Träger; [keine] zwei Menschen
> haben dieselbe Vorstellung. Sonst hätte sie unabhängig von die-
> sem und unabhängig von jenem Bestand. (1918: 42.)

> Wenn der Gedanke, den ich im pythagoreischen Lehrsatz aus-
> spreche, ebenso von andern wie von mir als wahr anerkannt
> werden kann, dann gehört er nicht zum Inhalte meines Bewußt-
> seins, dann bin ich nicht sein Träger ... . (Loc. cit.)7

Now, the question of course is why the bearer-dependence of conscious
states is supposed to entail their non-sharability. We do not find any straight
answer to this in Frege's surrounding text; but it seems clear enough that
Frege takes himself to be providing the required justification within the pas-
sage on pp. 41-42 which leads up to the non-sharability claim. This dark
passage I now turn to.

**The non sequitur underneath: From epistemic asymmetry to cross-per-
sonal inapplicability.** Frege asks you to imagine the situation of yourself—
a person with ordinary colour vision, we may assume—and a colour-blind

---

7   This latter sentence ends on the added clause "und kann ihn trotzdem als wahr anerken-
    nen", which is clearly misplaced and which I therefore omit. With it, Frege not only re-
    dundantly conjoins to the consequent of his asserted conditional its own antecedent, he
    even employs the inappropriate concessive adverb 'trotzdem' in doing so. If it is not too
    presumptuous, allow me to register the impression that Frege's attention sometimes
    shifts within one and the same sentence as he composes his texts, making him link his
    clauses with function terms which express rational relations other than the ones he is at
    a given point focusing on.

companion looking at the same meadow. Differently from yourself, your companion is unable to distinguish red from green. The colour experience that he has of a strawberry is the same kind of colour experience that he has of the leaves surrounding the strawberry: "Der Farbeneindruck, den er von der Erdbeere erhält, unterscheidet sich nicht merklich von dem, den er von dem Blatt erhält." Having described this scenario, here is the chain of thoughts which Frege connects to it:

> Sieht nun mein Begleiter das grüne Blatt rot, oder sieht er die rote Beere grün? [O]der sieht er beide in einer Farbe, die ich gar nicht kenne? Das sind unbeantwortbare, ja eigentlich unsinnige Fragen. Denn das Wort 'rot', wenn es nicht eine Eigenschaft von Dingen angeben, sondern meinem Bewußtsein angehörende Sinneseindrücke kennzeichnen soll, ist anwendbar nur im Gebiete meines Bewußtseins; denn es ist unmöglich, meinen Sinneseindruck mit dem eines anderen zu vergleichen. Dazu wäre erforderlich, einen Sinneseindruck, der einem Bewußtsein angehört, und einen Sinneseindruck, der einem andern Bewußtsein angehört, in einem Bewußtsein zu vereinigen. Wenn es nun auch möglich wäre, eine Vorstellung aus einem Bewußtsein verschwinden und zugleich eine Vorstellung in einem andern Bewußtsein auftauchen zu lassen, so bliebe doch immer die Frage unbeantwortet, ob das dieselbe Vorstellung wäre. Inhalt meines Bewußtseins zu sein, gehört so zum Wesen jeder meiner Vorstellungen, daß jede Vorstellung eines andern eben als solche von meiner verschieden ist. (1918: 41.)

In other words, Frege is asking of your colour-blind companion which unitary colour experience he is having of things that you yourself perceive distinctly as red and as green. Frege calls this sort of question "unanswerable". More astonishingly, Frege even calls this question "nonsensical".

The connection with the non-sharability claim is as follows. If we assume that conscious states are interpersonally sharable, then, by semantic ascent, it follows that terms designating conscious states are applicable across different people. The question that Frege imagines being asked with respect to the envisaged situation is one that takes this cross-personal applicability of phenomenological vocabulary for granted. By rejecting the question as nonsensical, Frege rejects this very semantic assumption. He is, in other words, claiming the non-sharability of conscious states as a matter of semantic necessity: as a matter of a *non-applicability* of phenomenal terms across persons.

Another important observation to make concerns the way in which Frege argues, in the quoted passage, for his semantic claim of the interpersonal

non-applicability of phenomenal terms. It is clear from the above passage that Frege rests his justification for this claim *solely on the asymmetry between first-person and third-person access to conscious experiences*.

How do these moves by Frege fit in with the already established fact that Frege relies on an inference from bearer-dependence to non-sharability? The overall picture to be extracted is that Frege attributes the following features to conscious states, where each of these features is taken by him to entail the ensuing feature:

> (i) Bearer-dependence of conscious states
> → (ii) Asymmetry of access to conscious states
> → (iii) Cross-personal inapplicability of terms
> designating conscious states
> → (iv) Non-sharability of conscious states

Practically all that can be said in Frege's favour here is that platitudes (i) and (ii) are true enough of conscious states, and that the target claim of non-sharability, (iv), indeed would follow from the inapplicability claim of (iii). The rest is confused in one way or another.

For one thing, as we have already seen, Frege fails to recognize the bearer-dependence expressed by (i) as the special instance of a metaphysical platitude that it is. Another error, which is of relatively minor importance in the present context, is this: The epistemological wisdom expressed by (ii) concerning privileged private access to conscious experience is of course not entailed by the bearer-dependence of statehood. Presumably, this epistemological wisdom would not even be entailed by the specific property of being a *conscious* state. It simply forms an independent fact about conscious states, albeit a similarly platitudinous one.

More interest lies in the following points:

- The inference from (ii) to (iii) is a non sequitur. Frege is confusing the epistemological wisdom about the private nature of phenomenological access for the erroneous semantic claim that phenomenal state terms are inapplicable to others.

- Equally importantly, the cross-personal inapplicability thesis (iii) is absurd itself. So is the non-sharability thesis expressed by (iv).

- The epistemological wisdom of privileged first-person access only applies to phenomenal consciousness, but not to the psychological posits of concepts.

The last point takes off where my earlier section about the psychological and the phenomenally conscious ended. We could, as a matter of fact, simply leave the rest aside, jump to this point straight away, and assert that Frege has indeed missed the topic all the way through. However, not least because Frege's considerations supporting abstractness in "Der Gedanke" have been such an extraordinarily influential reference point, I think it is important to dissect the rest nevertheless. Also, perhaps we should want to tie *some* concepts to phenomenal experiences (though I do not want to go into this here), and in this case, it would be good to have refuted Frege's claim that phenomenal states are never shared.

I will begin by establishing the fact that Frege's inference from (ii) to (iii)— from asymmetry of access to phenomenal states to a cross-personal inapplicability of phenomenal-state terms—is a non sequitur, and I will then try to explain what confusion made Frege get there. Does Frege predicate his inference from epistemic asymmetry to cross-personal inapplicability on any more basic principle? No. As far as Frege's text goes, we arrive at rock bottom here. No justification is given for this inference, though Frege strangely appears to regard it as compelling. I want to distinguish two construals of how Frege might be intending the inference to be understood. The inference is fallacious on both construals. Then I will make a suggestion as to which more basic fallacy is responsible for Frege's fallacious inference from (ii) to (iii).

Prima facie, it would seem that the general principle behind the inference from (ii) to (iii) is this: One can sensibly apply a predicate to a person only when one is absolutely *certain* that the person satisfies that predicate. This principle would be patently false. It is a platitude that things can fail to satisfy the terms that we ascribe to them. There does not seem to be anything that would not fall under this platitude, given that we are less than omniscient beings. Also, this is something which we do typically acknowledge in our ascription of predicates. We constantly take ourselves to be engaging in mere hypothesis rather than quasi-papal declarations of infallible insights.

A more charitable interpretation of the inference license which Frege is relying upon—the most charitable one I can come up with—would be this: One can sensibly apply a predicate only when there have existed at least *some situations* in which one has been absolutely certain that the predicate ap-

plies. The applicability of mental predicates to other people would fail, then, given that there are no possible situations in which we can be absolutely certain that someone else satisfies a mental predicate.

It is unclear to me why somebody should want to hold such a position, but it is probable that this is what Frege has in mind. It fits together with his brief, thought-experiment-like consideration in the last-quoted passage, his consideration of a scenario of experience transfer between people. Frege regards the scenario of experience transfer as one that might, prima facie, provide support for the third-person applicability of phenomenal terms, via the certainty which it might provide concerning the conscious experiences of others. So he does seem to regard it as a sufficient condition for third-person applicability if *some situation can be conceived*, for a given set of terms, in which we can be absolutely certain that the terms apply. (Frege dismisses the scenario for the irrelevant reason that it too involves hypothetical assumptions, namely, in virtue of the hypotheses which, as it were, go into the construction of an experience-transferring device.[8])

Nevertheless, for the same sort of reason as the one mentioned with regard to the earlier construal, the latter construal too fails to provide Frege with any plausible backing principle for his inference from epistemic asymmetry to cross-personal inapplicability. Certainty as to when a phenomenal term applies to somebody else need not even obtain in *some* conceivable cases of applying that term.

Explaining Frege's underlying non sequitur. It seems to me that a plausible etiological explanation can be given for what makes Frege slip into the erroneous inference from epistemic asymmetry to third-person inapplicability. On this etiological diagnosis, the fallacious inference is due to a more basic fallacious inference of Frege's which remains fully implicit. To see this, consider the following. It is plausible that, in learning a phenomenal term, each of us, guided by a publicly observable causal role, associates with the term a type of own, private experience, the presence of which is something we can indeed feel certain about when we have the experience. Frege, one might suspect, infers from the acquisitional fact that we *learn* phenomenal terms on the basis of associating them with our own, infallibly ascertainable conscious experiences, the assumption that phenomenal terms

---

[8] At first, it might seem that what Frege is saying in the last-quoted passage is that the question whether the experiences of the "donor" and of the "recipient" are identical is one that we, the *imaginers* of the scenario, can never be sure of. However, in imagining the scenario of experience transfer, one is *by hypothesis* imagining the same experience to be received that was transferred. Questioning the identity of donated and received experience would simply contradict the stipulated description of the scenario.

are *only applicable under circumstances of absolute certainty*. Given this and the fact that one's own phenomenal states are the only ones whose phenomenal qualities we can know with certainty, Frege is led to the idea that phenomenal terms are *only applicable to one's own experiences tout court*.

It seems plausible that a conflation of the cases grounding the acquisition of a term and the cases sensibly subsumable under that term has prevented Frege from recognizing what the perfectly natural and correct thing is to believe about the meaning of phenomenal terms: True, the only conscious experiences we observe in learning phenomenal terms are our own. But the concept formed in learning a phenomenal term is one that represents *this type of experience as such, irrespectively of which individual happens to be its experiencer*.

This individual-transcendence of phenomenal terms is, in fact, uncontentious in the philosophy of mind, as far as I know. And, to transpose matters into the material mode in which the non-sharability claim itself is couched, the *cross-individual type-individuation of phenomenal states* is not remotely contentious either. One wonders how these facts could really be regarded as contentious by anybody at all. Types of mental states are individuated by mental properties. These properties are not co-determined by the respective individual essences of people that happen to instantiate them. But if this is so, then how else could we think of the way we type conscious experiences but as being such that the experience types are at least *possibly* instantiated in other people too?

Frege is committed to the idea that phenomenal states, and thus phenomenal properties, are individuated individual-specifically. Nobody would be seriously prepared to believe that phenomenal properties are individuated individual-specifically, and rightly so. To mention just one consequence of such a view, it would be committed to any phenomenal term being infinitely ambiguous. For each different individual, it would designate a potentially different, unknown phenomenal property. This flies in the face of the plain idea that, for example, with a phenomenal term such as 'red', we understand a specific phenomenal colour property other than, for example, phenomenal blueness, or phenomenal greenness.

2.4. Dismissing the argument from non-sharability. We have reviewed and assessed Frege's claims 1 to 4 as to how Vorstellungen contrast with physical objects. With claim 4, Frege's considerations seamlessly blend over into his statement of the argument from non-sharability against the mental nature of concepts/thoughts. More precisely, they only blend over

into the completion of this argument. For his remarks in the context of his contrastive claims about "Vorstellungen" already have provided most of the substance of his major argument:

> **Argument from non-sharability:**
> (I) Mental states are bearer-dependent.
> (II) Bearer-dependence entails non-sharability.
> (III) Mental states are non-sharable. (I, II)
> (IV) Thoughts must be sharable.
> ∴ Thoughts cannot be mental states. (III, IV)

We have already seen Frege make the claims which constitute premises (I), (II), and (III). The completion of this argument lies in claim (IV), that thoughts must indeed be sharable between thinkers.

The assumption that concepts/thoughts are sharable is supported by a consideration that is still often encountered in present-day philosophy of mind. We find the consideration repeatedly in Frege's work too, but it is generally treated, understandably, as a point that is so basic as to not require crediting Frege with its discovery. The consideration is simple: The sharability of thoughts and their constituent concepts is required in order for the idea of *agreement or dissent* between people to make any sense at all. Agreement or disagreement presuppose a relation to one and the same thought whose potential truth the agreement or disagreement is about. Clearly, this is right. Since we can agree or disagree with respect to the truth of any thought, any thought must be sharable. The following quote from "Der Gedanke" is representative. It is embedded in a passage forming the closest that Frege gets to stating the whole argument.

> Wenn jeder Gedanke eines Trägers bed[ürfte], zu dessen Bewußtseinsinhalte er gehört[e], so [wäre] er Gedanke nur dieses Trägers, und es [gäbe] keine Wissenschaft, welche vielen gemeinsam wäre, an welcher viele arbeiten könnten [...]. Jeder von uns beschäftigt[e] sich mit Inhalten seines Bewußtseins. Ein Widerspruch zwischen beiden Wissenschaften [wäre] dann nicht möglich; und es [wäre] eigentlich müßig, sich um die Wahrheit zu streiten ... . (1918: 43.)

Much as this is right, we have seen that Frege's argument from non-sharability is radically defective. Mental states obviously *are* sharable, and they would be sharable even if the mental really were exhausted by phenomenally conscious states.

3. Strange and stranger anti-mentalist objections in Frege (1897)

Though the argument from non-sharability which has just been extracted from "Der Gedanke" forms the major argument against the mental nature of thoughts by Frege, there are a number other passages in his work which reject the mental nature of thoughts on different, only intermittently related-sounding grounds. Frege's posthumously published writing "Logik", dating from 1897, is a rich source of such loci, and I will draw on it in the following.[9] At least within this work, Frege cannot seem to decide on which grounds he rejects the mentality of thoughts. Let me distinguish the more saliently offered reasons by Frege and let me make clear, in not too many words, how they are likewise off the mark.

3.1. An objection from mental images and derived truth-valuability. In his first take on the matter within (1897), Frege begins by conceding that one does also speak of true "Vorstellungen". Frege introduces Vorstellungen as mental images here ("ein Phantasiebild, das … aus den wiedererweckten Spuren vergangener Empfindungen oder Tätigkeiten besteht", p. 42) and denies that mental images might be the sorts of things that can be intrinsically truth-valuable.

> Eine Vorstellung ist, wie jedes andere Bild, an sich nicht wahr, sondern nur in Hinsicht auf etwas, dem es entsprechen soll. Wenn gesagt wird, ein Bild solle den Kölner Dom darstellen, nun gut, dann kann man fragen, ob diese Absicht erreicht ist; ohne Hinblick auf eine Absicht, etwas abzubilden, kann von keiner Wahrheit des Bildes die Rede sein. Daraus ist zu entnehmen, daß eigentlich nicht der Vorstellung selbst das Prädikat wahr zuerkannt wird, sondern dem Gedanken, daß sie einen gewissen Gegenstand abbilde. Und dieser Gedanke ist keine Vorstellung und nicht aus Vorstellungen irgendwie zusammengesetzt. (1897: 42-43.)

Now, a great many concept theorists simply refuse to regard mental images associated with a word as concept-constitutive to begin with. Moreover, those concept theorist who do incorporate mental images into their views of conceptual essences do not do so across the board.[10]

---

[9]   Henceafter quoted from the selection of posthumous writings *Schriften zur Logik und Sprachphilosophie: Aus dem Nachlaß,* edited by Gottfried Gabriel ([4]2001).

[10] One of the ground rules of proper concept-theorizing should be never to start out from generic claims such as 'Concepts are Xs', not even *roughly,* not even to a *first approximation*. This is simply not the way to approach a field of empirical inquiry. However platitudinous the maxim may seem, it is hard to exaggerate how gravely it has been neglected in philosophical discourse surrounding concept-and content-theorizing and how

Still, let us focus on what Frege has to say about mental images qua mental images. Frege mixes up some entirely different issues in this passage and, as in other passages we will turn to, makes it impossible for us to discern anything that even initially resembles a sensible argument for non-mentality. Let me restrict myself to the following comments.

The observation which Frege takes as the basis of his objection is not expressed accurately by himself, but it is clear enough what he is aiming at. He is expressing a familiar criticism of the idea that the resemblance relation—typically considered as a relation of pictorial resemblance—might provide a sufficient condition for the instantiation of the representation relation. Pictures, so the objection goes, count as representing something only derivatively upon someone's intention to depict something with them.[11] Thus, they do not have intrinsic intentionality; they do not represent something in- and of themselves. A similar situation is thought to obtain with mental images. This is what Frege is aiming at, even though his actual statement involving explicit reference to mental images is beside the point. (He is not really aiming to emphasize the non-intrinsic, correspondence-like nature of truth, as that remark suggests; he is rather aiming at the non-intrinsic nature of a picture's counting as truth-conditionally evaluable, hence as intentional.)

Now, to begin with, for a given mental representation to have merely derived intentionality would not at all show that the mental representation cannot be bona fide truth-valuable. Sentences of natural languages have intentionality which is derivative upon our communicative intentions, but this obviously does not make their truth-valuability spurious. This is the first respect in which Frege's objection is confused.

More importantly still, consider the following. Let us grant, for the sake of argument, the familiar objection to mental imagism which attacks it via its traditional ally of a resemblance-based view of content-constitutivity. Let us thus grant for the sake of argument that mental images as such can only be derivatively intentional. The fact of the matter, then, is that Frege has not of-

---

incessantly philosophers start out from perspectives ignoring it. I will not embark on a methodological campaign of exposing and criticizing this approach within this work. But it does not seem too bold a claim to say that the climate has begun to change. Edouard Machery (2005) or Daniel Weiskopf (unpublished) are cases in point. Medin, Lynch and Solomon (2000) already address questions of concept-theoretical heterogeneity and parameters of variation from the perspective of cognitive psychologists.

[11] For a classical, take-no-prisoners assault on the idea of reducing representation to resemblance, see Goodman (1976: 3-5).

fered the slightest reason as to how taking a step down to the more basic intentional state which lends the mental image its content is supposed to constitute any progress towards an *anti-mentalist* conclusion, let alone how this conclusion is thereby supposed to be established. And in fact, there is no such connection, as I want to point out now.

What we can see at this point is that Frege's exclusive focus on mental images in his consideration of mentalistically understood concepts has more profoundly detrimental effects than just the immediate consequence of capturing only one segment of the relevant reality at a given level. Frege altogether fails to see that the more basic, content-conferring intentional state which the allegedly derivedly-intentional mental image would have to depend on is a state involving an *amodal mental representation,* i.e., one belonging to a mental representational medium which is not dependent on the formats and resources of a specific sensory-perceptual modality.

From this angle, it can be seen that, with his anti-imagist objection, Frege has only kicked the burden of argument for his non-mentality claim one level down from the level of conscious mental images, namely, down to the level of an amodal mental representational state which contains as a component the "Gedanke, daß [die Vorstellung] einen gewissen Gegenstand abbilde". *This* thought is the candidate whose non-mentality Frege would now have to argue for. Instead, he offers nothing to this effect. Without having made any rational progress at all, he acts as if the non-mental nature of the more basic thought underlying the mental image must be clear: "Und dieser Gedanke ist keine Vorstellung und nicht aus Vorstellungen irgendwie zusammengesetzt."

3.2. An objection from objectivity of truth. On several occasions, Frege reasons against the mentality of thoughts on the basis of considerations having to do with wholly orthogonal issues of objectivity and metaphysical realism. These sorts of considerations are at least equally bewildering as the one just encountered.

In (1897: 45), Frege pursues the platitude that thoughts are *true independently of whether we acknowledge them or not,* combatting a mad relativism which would assert "[daß] etwas nur für den wahr wäre, der es für wahr hielte". Although he is not explicitly criticizing a mental conception of thoughts in the passage at hand, the context is rather clear about the fact that Frege intends the platitude to count this way indeed. If so, however, I have no idea why Frege mixes up the assumption that there are truth-valuable mental representational states with the patently silly relativistic notion that

whether something is true depends, in the general case, on whether it is believed or not. I hope it is needless to say that one has nothing to do with the other. Frege goes on to protest against the irrelevant relativistic notion on grounds which we are familiar with from his justified insistence on sharedness in "Der Gedanke", namely, on grounds of ensuring the possibility of disagreement.

3.3. An objection from mind-independence (propositionalism). On the following page, Frege connects to the critique just considered, this time explicitly turning against the mentalistic conception of thoughts. Perhaps encouraged by the objectivity (the mind-independence) of truth and falsity he has just insisted upon, Frege moves on to decree here that thoughts must *themselves be mind-independent:*

> Gedanken—z.B. Naturgesetze—bedürfen nicht nur unserer Anerkennung nicht, um wahr zu sein, sie brauchen dazu nicht einmal von uns gedacht zu werden. Ein Naturgesetz wird nicht von uns ersonnen, sondern entdeckt. Und wie eine wüste Insel im Eismeer längst da war, ehe sie von Menschen gesehen wurde, so gelten auch die Gesetze der Natur und ebenso die mathematischen von jeher und nicht erst seit ihrer Entdeckung. Wir entnehmen hieraus, daß Gedanken nicht nur, falls sie wahr sind, unabhängig von unserer Anerkennung wahr sind, sondern daß sie überhaupt unabhängig von unserem Denken sind. (1897: 46)

But remarks like these miss the target of mental thoughts. Frege is simply aiming here at posits playing the role of *facts* (for example, laws of nature). Nothing could be further from the truth than the idea that mental representations are rivals to the metaphysical role played by facts. Facts are what true mental representations are supposed to represent. How on earth could mental representations be thought to *replace* them?

To be sure, I am taking certain liberties in describing Frege as aiming at facts here, or more generally, at states of affairs—constellations of objects and properties or relations—with facts forming the subclass of actual existents among states of affairs. In faithfulness to his metaphysical intents, and employing standard philosophical terminology, we would have to say that Frege is aiming at entities which philosophers have been calling *propositions* since sometime after the early works of George Edward Moore (1899) and Bertrand Russell (1903).[12] Propositions, to cut a long story short, are

---

[12] The earlier explained notion of a propositional attitude is, notabene, not committed to propositions. Talk of "propositions" is metaphysically more loaded than talk of "propositional attitudes". The expression 'propositional attitude' has long ago acquired independence, becoming a metaphysically innocuous term of art in philosophy of mind.

posits which can be seen to play precisely the same metaphysical role that states of affairs do, except for the delicate twist that they are by definition truth-valuable entities, hence on the other side of the representational divide than states of affairs are. Now, for what it is worth, I have preciously little sympathy for the idea that one should include propositions in one's metaphysical inventory. They are plausibly not needed and the postulation of propositions over and above states of affairs rather seems to be based on a misguidedly literalist understanding of the complement clauses in propositional attitude ascriptions. But I do not want to argue for this here. Taking a position on propositions is not relevant to our concerns. If we paraphrased Frege's aforequoted insistence on mind-independence with reference to propositions, this would invite just the same objection as the one made.[13]

The point here is that, if you wish to count propositions among your inventory of entities in addition to states of affairs, and if it is propositions which you paraphrase Frege's statement over, then the objection just made reapplies. Positing propositions in no way *impugns* the existence of truth-valuable mental representations. All it does is to introduce a further metaphysical player besides *mental representations* and *representata,* namely, *abstract representations*. Whether one acknowledges propositions or not has nothing to do with whether one can acknowledge the existence of mental representations or not. It just entails a metaphysics on which one assumes that truth-valuable mental representations express propositions.[14]

As it happens, in immediate intra-paragraph continuation of the passage just quoted, Frege moves on to anticipate the main argument from (1918).[15]

---

[13] An earlier draft of this dissertation included two chapters arguing that propositions are entities whose alleged specialness simply derives from grafting truth-valuability onto entities otherwise identical to states of affairs, and providing an argument against these propositional entities. Such considerations are irrelevant to concept-theorizing, however.

[14] Notice, for example, in accord with this, that Jerry Fodor, the most influential champion of mental representations, soon shifted from early descriptions of the nature of propositional attitudes in dyadic terms (Fodor 1978: 514), which related an organism to a mental representation, to triadic descriptions (Fodor 1987: 17; Fodor 1994: 47), which related an organism to a mental representation expressing a proposition. The shift only consists in making explicit an additional metaphysical commitment to propositions. What lies on the other side of the representation relation will in any case have to be the represented states of affairs, be it actual or possible (in other words: the truthmakers of the truth bearers).

[15] "Der Gedanke ist den Denkenden nicht so besonders zu eigen, wie die Vorstellung den Vorstellenden, sondern steht allen, die ihn auffassen, in derselben Weise und als derselbe gegenüber. Nie verbänden [sonst] zwei Menschen denselben Gedanken mit demselben Satze, sondern jeder hätte seinen eigenen: und wenn nun z.B. jener 2 · 2 = 4 als wahr hinstellte, während dieser es verneinte, so wäre das kein Widerspruch, weil das,

Mentality is claimed to entail non-sharability, sharability is insisted upon with reference to the possibility of disagreement, and mind-independence is therefore thought to be called for in order to satisfy sharability. I should like to point out in passing that this way of continuing on the already asserted claim of mind-independence makes no sense. If Frege *already* thinks that he can decree thoughts to be as mind-independent as an undiscovered island in the Arctic Sea (similarly, if he already thinks that laws of nature are good examples of thoughts), then thoughts would ipso facto not be mental representations. Further complaints about an unpalatable entailment of non-sharability by mind-dependence would be beside the point.[16]

3.4. An objection from infinite regress. Immediately following the long paragraph which includes the considerations just commented upon, a short paragraph shows up in which Frege take a completely new shot at refuting the notion of truth-valuable mental representations. Similarly as the objection we have considered in subsection 3.1, this objection purports to uncover a fault in the mentalistic conception of thoughts in a direct manner. It has to be said that the objection is again appallingly off topic. Frege commences with the following unobjectionable remark:

> Wenn der Gedanke etwas Inneres, Seelisches wäre, wie die Vorstellung, so könnte seine Wahrheit doch nur in einer Beziehung bestehen zu etwas, was kein Inneres, Seelisches wäre.

This statement, for once, has Frege saying something uncontentiously true about mental representations. In particular, there is no trace here of the attribution of an absurd relativism; quite the contrary. But just look at what Frege seamlessly connects to this:

> Wenn man also wissen wollte, ob ein Gedanke wahr wäre, so müßte man fragen, ob diese Beziehung stattfände, mithin, ob der Gedanke wahr wäre, daß diese Beziehung bestände. Und so wären wir in der Lage eines Menschen in einer Tretmühle. Er macht einen Schritt vorwärts und aufwärts; aber die Stufe, auf die er tritt, gibt immer nach, und er sinkt auf den vorigen Stand zurück. (1897: 47.)

---

was jener behauptete, verschieden wäre von dem, was dieser verwürfe." (1897: 46.) See also Frege's own recapitulation in his extended table of contents on p. 36.

[16] Quite apart from the fact that the inference from mind-dependence to non-sharability is entirely unwarranted anyway, as we established in §2. Notice, by the way, that in "Der Gedanke", Frege at least offered a reason to support the inference, however fallacious it was. Here, we are offered no reason at all, not even a fallacious one.

Frege is claiming here that assuming the mentality of thoughts leads to the reductio of an infinite regress. Why? To ascertain the truth of a given mental thought, so Frege claims, one would have to ascertain whether it corresponds to the facts. But ascertaining whether the thought corresponds to the facts would, he reasons, require ascertaining whether a meta-thought about that target thought is true—a meta-thought to the effect that the target thought corresponds to the facts. And so on.

It is beyond me how Frege could come to think that, even if he had a point here, this would have anything to do with his demonstrandum that thoughts cannot be mental. In any case, it does not. If Frege is worried about an infinite regress, he might as well turn his worries against propositions. In fact, if anything, it is only them he should be worried about. For it is obvious that entertaining a true mental thought does *not* require entertaining an infinite hierarchy of distinct meta-thoughts each of which ascribes truth to a respective lower-level thought. But this is what Frege must be assuming proponents of mental thoughts to be committed to if he criticizes mental thoughts for posing an infinite regress to anyone wishing to ascertain their truth.

Moreover, the regress worry is baseless in the first place, quite independently of mental reality. If one wants to ascertain whether a truth bearer is true, one need only ascertain whether the state of affairs obtains which would make it true. There is nothing over and above this for one to accomplish—just as nothing is required of a truth bearer itself in order for it to *be* true, over and above the obtaining of the state of affairs which it represents. We should not have to be hammering in such points here, though. This is not the sort of territory we should have to be treading in the context of our investigation.

4. Frege's arguments from mental specificity and mental variation

Let me finally turn to what is still the most respectable-looking anti-mentalist passage I can detect in Frege's work. It occurs later on within the same posthumously published manuscript just drawn upon, and it argues that understanding a word cannot consist in connecting to it a specific mental image. Since understanding a word consists in connecting a concept to it, this amounts to the claim that concepts cannot be mental images. Since Frege is again taking for granted here that any mental representations would be conscious mental images, his intent is that of arguing that concepts cannot be mental representations tout court.

Frege takes the example of the word 'horse', and of which mental images we connect to it. There are two kinds of arguments here to the effect that understanding 'horse' cannot consist in connecting a mental image to the word. The first is an argument from mental specificity, as we might call it. The second is an argument from mental variation.

The passage I have in mind begins with the following remark, which is exhaustive of what Frege says by way of arguing from *mental specificity:*

> Wer das Wort 'Pferd' mit Verständnis hört, dem wird wohl alsbald das Bild eines Pferdes vor die Seele treten. Dieses Bild ist aber nicht mit dem Sinne des Wortes 'Pferd' zu verwechseln; denn über die Farbe des Pferdes, über seine Haltung in Ruhe oder Bewegung, über die Seite, von der es gesehen wird, und dergl. ist kein Wink im Worte 'Pferd' gegeben. (1897: 55.)

Frege's implicit claim here is that the mental images which we connect to the word 'horse' in linguistic comprehension contain information which is obviously inessential to an understanding of the word 'horse'.

The argument from *mental variation* immediately ensues within the same paragraph. There are two subclaims to Frege's claim of mental variation. One is that of the *interpersonal* variation of mental images associated with a given word. The other goes beyond it, claiming the *intrapersonal* variation of mental images which are associated with a given word, dependent upon varying employments of the word in descriptions of different situations.

> Wenn verschiedene Menschen ihre beim Worte 'Pferd' auftauchenden Vorstellungen etwa auf eine Leinwand sofort projizieren könnten, so würden recht verschiedene Bilder zum Vorschein kommen. Aber auch bei demselben Menschen wird das Wort 'Pferd' nicht immer dieselbe Vorstellung hevorzaubern. Viel wird dabei auf den Zusammenhang ankommen. Man vergleiche z.B. die Sätze '[W]ie reitet er so freudig sein mutiges Pferd' und '[E]ben sah ich ein Pferd auf dem nassen Asphalte stürzen'.
>
>   Also davon kann [sic] keine Rede sein, daß mit dem Worte 'Pferd' immer dieselbe Vorstellung verknüpft sei. (Loc. cit.)

What we have here could appropriately be called an argument from the sharability of concepts too, but it is quite different and more respectable than Frege's major argument in "Der Gedanke", to which we have already attached this label. For there, Frege held the purported modal fact of a *necessary non-sharability* of mental states against a requirement of concepts' sharability. Here, by contrast, he is effectively only complaining about an

actual non-sharedness of mental images in cases where we do want to acknowledge the presence of a shared concept.

Frege's observation about variation between mental images connected to a given word is correct enough, and it is an important fact to keep in sight in other contexts of investigation. But, pace Frege, it does not have the anti-mentalistic consequence of concepts not being mental representations.

To begin with, as noted earlier on already, mental imagism is only a specific position encountered among concept theorists. Mental images are simply representations derived from the sense modality of vision. In how far concepts can be identified with perceptually derived representations, visual or otherwise, is a deep, moot question among representationalist theorists of concepts. A majority of concept theorists do not discernibly rely on mental images at all in their theories of concept essences, and those who do practically invariably refuse to do so without the aid of non-imagistic mental representations. The former kind of theorists would not need to answer in the first place to the objection of mental images being unduly specific or unduly heterogeneous, when compared with the unique concept whose possession is required to understand a given word, on a given meaning.

In order to get a feeling for the non-dependence on mental images, consider, by comparison, the following examples which I can produce on an introspective basis. Quite reliably, when I think of the 1970s, an illuminated band logo of the pop group ABBA or images of long-haired youths with flared trousers pop into my head. When I hear the word 'freedom', I often entertain a faint mental image of broken chains and of an Afro-American slave on the run. Nobody would claim that these mental images themselves form my concepts THE 1970S or FREEDOM. That is, the non-concept-constitutivity of associated mental images is uncontentious enough in these kinds of cases. It is so already for the reason that the images in question represent entities of categorially altogether different kinds (people, objects, inscriptions) than a time span such as the 1970s, or the state of being free. But the non-concept-constitutivity of associated mental images might be likewise maintained for cases where there is categorial agreement, such as in the case of horse images and the concept HORSE. And this is indeed maintained by countless theorists who either believe that concepts are constituted by content-conferring nomic relations to their representata, or who believe that concepts are constituted by content-conferring dispositions to perform certain inferential transitions involving these concepts, regardless of whether the mental representations involved in the transitions are mental images or not. These latter strategic options point to the respects in which concept

theorists can deviate from taking concepts to be constituted by mental images or, more generally, by perceptually derived representations.[17]

However, even for theorists advocating the identification of some concepts with mental representations qua perceptually derived ones, Frege's observations of specificity and of heterogeneity are not insuperable. In my remaining comments, I will continue to focus on the case of mental images. The first comment to make will be critical. It counts against the argument from specificity, but it also counts against the argument from variation, to the extent that the latter depends on the former. My second comment concedes a point of substantial theoretical import to Frege's argument from variation, while simultaneously maintaing that this point does not count against mental representations as such.

There is a flaw in Frege's argument from specificity. Identifying a concept with a mental image does not require the identification to be performed in virtue of *every* aspect contained in the mental image. Frege is overseeing the option of identifying concepts with mental images such that mental images are type-individuated in a way that abstracts from a certain amount of visual representational detail. Relatedly, he oversees the option of letting relevantly similar mental images count as tokens of one and the *same type* of mental image. Instead, Frege is taking for granted a scheme of type-individuating mental images which is, for all that we can say, maximally fine-grained: an individuation scheme according to which two mental images are of the same type just in case they are exactly similar in all introspectively ascertainable respects. There is no reason why one should be forced to adopt such an individuation scheme.

All the same, this rejoinder will not suffice to fully answer the present kind of misgivings about identifying concepts with mental images. For indeed, if somebody, say, associates with one utterance of 'horse' the image of a brown horse, observed from the side as it vibrantly gallops through the forest with a rider on its back, while associating with a different utterance of 'horse' the frontal image of a grey horse as it stumbles to a wet tarmac ground while pulling a cart, the intrinsic similarities in the visual representations will be slight. And other examples could be imagined in which the intrinsic similarities are slighter still, in fact, arbitrarily small. This does refute a naïve version of the mental imagist idea. For the situation is this: There is a strong intuition that the word 'horse' is understood in the same way on the occasion of hearing both of the embedding utterances mentioned by Frege,

---

[17]   Later chapters will be returning to this idea in slightly more detail.

and we can grant that it is desirable to respect that intuition. This entails commitment to the assumption that the word 'horse' gets connected to the same concept in both events of understanding. If the two events of understanding involve mental horse images which are sufficiently dissimilar to go beyond what is plausibly bridgeable by a more abstract image subsuming the two, then the unique HORSE concept which is granted to be involved in the understanding of the two 'horse' utterances cannot be identified with a mental image.

Exactly which consequences examples of the sort discussed require someone to draw who wishes to incorporate imagistic elements into the essences of some concepts seems to be quite an underdetermined theoretical question which we cannot begin to settle here. Suffice it to say that the naïve imagist idea of identifying a concept simply with an occurrent mental representation with a static pictorial content will certainly not do. Clearly, an embedment of the occurrent visual representation in a broader concept-constitutive data structure must be assumed. Dynamic perceptual representations must be taken into account. Recombinable perceptual representations of various degree of abstractness must be taken into account within the data structure.[18] Traditional mental imagism begins to look simplistic to the point of caricature once one begins to take such elements into consideration. What exactly, if anything, "the" concept of a horse is to be identified with on such a picture, forms an important foundational issue which still awaits a more detailed investigation at the time this work is written. Exciting philosophico-psychological research can be hoped to lie in this direction. But I will not explore these matters further within this work, even less so within the present chapter, for the aim of this chapter appears to have been accomplished.

We have found nothing wrong about the idea of identifying concepts with mental representations in Frege's work. Moreover, contemporary theorists who are reluctant to identify concepts with mental representations add little to the arguments stated quite clearly by Frege in the quote within this subsection (see Peacocke 1992: 3, or Rey 1998: 507). This goes a long way towards vindicating the mentalist notion of concepts. It is with regard to concepts understood as mental representations that I conduct my investigation from here on.

---

[18] See, for example, Jesse Prinz (2002: ch. 6), for an exposition of a number of resources which concept-theoretical neo-empiricists can avail themselves to.

§3
Concepthood vs. nonconceptual representing: Storage in memory

_____

_____

A majority of theorists, both in philosophy and in psychology, do not understand concepts to form just any kind of subpropositional mental representation, but rather quite a special such kind of representation. It is widely assumed that there is a lot of *nonconceptual* representing going on in our minds. Any elucidation of what we posit when we ascribe a concept to a creature would be seriously incomplete if it did not make clear how concepts contrast with nonconceptual representations.

The rationale behind this and the following two chapters is that there is something substantially more to be said than what we have said so far about the theoretical role which concepts must be assumed to occupy by anyone understanding them to be mental representations. The present chapter leads up to a characteristic necessary feature of concepts that goes beyond the minimal dictum about concepts being constituents of thoughts, a feature that is elementary enough to count as neutral vis-à-vis any of the various positions that go through as theories of concepts. It is the feature of being stored in memory.

Somewhat surprisingly, philosophical work which is dedicated to the nature of concepts without a special focus on the contrast with the nonconceptual tends to give little or no indication of the centrality of storage in memory to the postulation of concepts. And even where the question of demarcating the realms of nonconceptual and conceptual representing is discussed, the fea-

ture of storage in memory is too rarely made explicit. What does tend to happen, rather, is that the distinction between a concept and something counting as a nonconceptual representational state is drawn on the basis of unargued, explanatorily wholly unprincipled intuitions. Relatedly, the conceptual/nonconceptual-distinction (C/NC-distinction, henceforth) is often drawn in an arbitrary fashion.

The central objective of chapter 3 is to carve out a non-arbitrary foundation for the notion of a concept and to show how this leads us straight to the characteristic feature of memorization. Later on, chapter 5 will continue the undertaking begun here, adding characteristic features subsumable under what I will refer to as the "flexibility" of conceptual states. In the meantime, chapter 4 will consider what are, at the most general level, two other popular candidate notions for demarcating the realm of the specifically conceptual, notions connected to the label of "classifying", "categorizing" representations. It will be argued that those two different notions of categorizing representations cannot serve as basic ways of demarcating concepts from their nonconceptual representational counterparts, though for very different reasons.

In embarking on a discussion of how the distinction between the conceptual and the nonconceptual ought to be drawn, I think that we have to begin by driving two stakes into the ground. The *first* claim represents a basic starting point, however imprecise. Whatever more precise content this first claim is given, it is plausibly regarded as an overarching, general idea behind any understanding of concepts that is not all-inclusive with respect to the realm of mental representing. The *second* claim is closely connected and looks at the reverse side of the conceptual realm. It is a somewhat stronger claim and provides the foundation for the rest of the investigations in the chapter.

1. A hazy first location:
Concepts, high-level cognition, intelligence

The most obvious starting point for circling in on concepts is this: Concepts are mental representations that are involved in *high-level cognition*—we also tend to call "thinking", "reasoning", or, on a more restricted sense of the term, "cognition" (though the latter term is one whose conventions of use vacillate far more in degree of inclusiveness than others). Propositional attitudes and their truth-valuable component states which we have been referring to as "thoughts" are also understood to be restricted to this domain of high-level cognition where concepts are located. The present problem is that

it is not an easy task to say in principled terms what exactly marks off high-level cognition, or "thinking", from other cognitive processes.

Identifying the realm of the conceptual with high-level cognition might seem to commit one to a location of conceptual versus nonconceptual processes at different levels of nature within the realm of representational states. If so, however, the notion of a level of nature cannot simply be spelled out here in terms of levels of aggregation. That is, more specifically, it could not simply be spelled out over concepts, situated at the higher level, being constructions out of nonconceptual representations, situated at the lower level. This is far too tendentious as a basic criterion of concepthood as such, for reasons which will be made clear later on, towards the end of chapter 4.

Possessing concepts is also a central prerequisite for a creature to possess *intelligence*. This idea is vaguely related to the idea that concepts are situated in high-level cognition, yet quite distinct. If possession of concepts is necessary for an organism to merit the dignifying description as "intelligent", then concepts are not just situated at higher levels of nature than nonconceptual representations. In some sense, concepts also form a mark of *higher cognitive sophistication* than would be exhibited by organisms which represent without concepts.[1] In what respect concept possession can be said to entail a higher degree of cognitive sophistication is still an open question, the answer to which would point to what is distinctive of conceptual representation itself.

A similar thing is true for the relation between concept possession and high-level cognition. We are locating the notion of a concept relative to other notions here, but those other notions are not more basic ones, and we cannot rely on them essentially in defining concepts, on pain of circularity. The present subsection is rather more of heuristic use in getting at concepthood.

---

[1] It might be regarded as a dubiously pleonastic addition when I say here that concept possession is tied to a higher degree of *cognitive* sophistication, since, on a widespread use of the term, 'cognitive' is interdefined with the realm of concept-involving processes. If this were the use intended here, then indeed the reference to the cognitive in talk of "higher cognitive sophistication" would be circular, failing to restrict the respect in which the conceptual is more sophisticated. However, 'cognitive' might be understood to mean any psychological processes involving states that aim at truth (at correctly representing the world). In this sense, the characterization given is appropriate, though exactly wherein its appropriateness lies remains to be seen.

2. A resolute second step:
Nonconceptual as sensory-perceptual representing

We have firmer ground underneath our feet once we perform the second step which I advise us to take: The basis of any principled opposition between concepts and nonconceptual representations must lie in the opposition between concepts and sensory representations. If there is any nonconceptual representing at all, then its *basic instances* are *sensory-perceptual representations*, i.e., the representational states that are directly involved in sensing or perceiving something. In conformity with psychological terminology, we can call the mental representations immediately involved in sense perception *percepts*. Sensory-perceptual experiences are ones that are generated by input to our *senses*. The senses are an organism's specialized systems of stimulus uptake.

This connects to the preceding subsection. The higher degree of cognitive sophistication that is supposed to be tied to concept possession comes into play once a creature is capable of transcending the possibilities of sensory representing. Before I move on now, some notes of clarification are in order on talk of "sensory" states, and on the contrast between 'sensory' and 'perceptual'.

Whether an input system is outer-directed or inner-directed makes no difference to whether it qualifies as a sense or not. To be sure, the habitual talk of "the five senses" of sight, hearing, smell, taste, and touch—all of which are outer-directed—dominates commonsense understanding of the senses from our childhood education onwards. The same five senses are distinguished already in the work of Aristotle, and it is not seldomly the case that philosophers understand by 'sense' a capacity that is ipso facto outer-directed. This restriction, and the exhaustiveness suggested by talk of "the" five senses of vision, audition, olfaction, gustation, and tactition, are mistaken. Neuroanatomy teaches that there are several further senses in the human species alone. That is, there are several further systems of stimulus uptake with their distinct kinds of receptor cells sensitive to their respective kinds of magnitudes of physical energy. What to say about exactly how many distinct sense faculties exist is a partly interest-relative matter, depending on the fineness of one's individuation scheme of choice. But there is no question that most of the senses which remain ignored by the fivefold pre-school taxonomy are ones that supply the brain with information about states of the body. Among these are, for example, the faculty of pain perception (nociception), the faculty of perceiving position of parts of one's body (proprioception), or the faculty of perceiving temperature (thermoception).

Concerning the terms 'sensory' and 'perceptual', I do not wish to deny that these terms tend to be used with different things in mind. The basic difference here looks similar to the one mentioned in my first, preliminary remark on the conceptual and the nonconceptual: Within the class of representational states resulting from receptor stimulation, the *sensory* is identified with more primitive states, while the *perceptual* is broadly identified with higher-level states. This could do with a bit of explanation. What tend to be regarded as sensations are brute, primitive experiences: the feeling you have when cool water runs down your throat; the feeling you have when a needle stings your fingertip; the visual experience of a patch of red right in front of your eyes. Complex experiences constructed out of such more elementary experiences are much more likely to pass through as "perceptual" experiences. For example, hearing a movement from Bruckner's Seventh Symphony is surely more appropriately called a "perceptual experience" than a "sensation" (or a "sensory experience", for that matter). However, it is not *complexity* of representational content as such that plausibly distinguishes perception from sensation; this would make it wholly arbitrary where to draw a line between sensation and perception, if any. The point is rather that, in some sense or other, perception is always meant to involve an *integration of sensory information* too.

In contrast with this, for what it is worth, a suggestion which I formerly felt inclined to make is that one can draw a principled difference between sensation and perception on the basis of the assumption that sensation and perception are capable of representing different ontological types. While sensations are only capable of representing *properties,* perceptions represent sensed properties in a way that they are represented as belonging to *objects*, or constellations of objects, or events involving these. In fact, this is practically equivalent to the way in which psychologists appear to draw the distinction. For example, Colman informs us in his *Oxford Dictionary of Psychology* (2001) (interestingly enough, referring us back to a philosopher): "In psychology a distinction is conventionally drawn between sensation, the subjective experience or feeling that results from excitation of sensory receptors, and perception, sensory experience that has been interpreted with reference to its presumed external stimulus object or event, this distinction having first been made in 1785 by Thomas Reid ... who pointed out that the agreeable fragrance of a rose is merely a sensation inasmuch as it can be experienced without thinking of a rose or of any other object, whereas the perception of a rose or of anything else always refers to the external object that is its cause."

The objection that can be justifiably raised against this take on the sensory/perceptual distinction is that, if this is the way that perception is understood, then it is hard to see how perception would not ipso facto involve the application of concepts to what one is sensorily representing. But assuming that the involvement of concepts delimits perception from sensation would be inconsistent, in light of the assumption that perception forms a case of *non*conceptual representing. One way to respond to this would be that perceiving a rose involves its perception as an object, hence would be a concept-involving state, only insofar as the highly generic concept of an external object is involved here, not the concept of a rose. But this would not be a plausible option. Even if one did make do with the much weaker requirement of the concept of an external object being involved, a radical over-intellectualization would seem to be involved in this assumption about what is required of an activity in order for it to count as perceiving, rather than as merely sensing.

This brings me back to the claim introduced two paragraphs back as to what really distinguishes perception from sensation, namely, processes of *integration*. By talk of "integration" I mean that perceptual representations represent scenes as having overarching properties and relations that reach beyond the sum of spatially arranged or temporally ordered individual sensory representations of a given scene.[2] Phenomenologically, in fact, we cannot even begin to imagine how our experience would be like without the integrational activity of perception. What I just want to point out here is that, though untenable, the attempt to describe perception as necessarily involving object representations is at least close to the truth insofar as perception, in the course of integrating sensory information, successively chisels out higher-order *invariances* within the scenes that our sensory receptors take up. Where such invariances form an appropriate configuration, the immediate basis within the perceived scene is provided for a creature to bring to bear an object concept.

Sensation, then, is raw representing, while perception is representing that results from the operation of complex integrational processes. However, these differences do not matter particularly much for present purposes. What matters is that they are both, as I said, *sense-dependent* representational states in the obvious sense that they immediately depend on the activation of our sensory receptors. If it were not for readability, I would more often speak of "sensory or perceptual states" in what follows. Apart from readability, the reason why I often simply speak of perceptual states (like many others in-

---

[2] The idea is best appreciated when seen in action in perceptual psychology. I trust that it is clear enough to the reader without us going into examples.

terested in the C/NC-distinction) is that percepts, given their higher-level nature, come closer to concepts to begin with, so that anything said in delimiting them from concepts applies all the more to the rawer representations of sensory events.

We can now turn to the C/NC-distinction. I begin with what I take to be the usual way in which it is introduced and, initially at least, defended. I then offer a different way of looking at things which is very straightforward. What emerges is a straightforward way of grounding the C/NC-distinction which immediately leads to the feature of storage in memory.

3. The usual approach:
Perception without thinking

Take the perceptual state you are in when you look at a meadow on a spring day and your visual system is functioning normally. An immense amount of sensory and perceptual information is contained in your visual experience. Many theorists treat as basic and intuitively compelling the assumption that your sensory-perceptual experience of a meadow represents a vast array of properties, spatial relations, and configurations, without representing these things in virtue of involving any *concepts* (let alone in virtue of involving such a correspondingly vast construction of concepts). Theorists of this stripe often set out to justify the C/NC-distinction by taking such a perceptual situation and claiming, without much of an argument, that perceiving all that detail does not count as entertaining beliefs or thoughts about all that detail. Let us call this the *perception-without-thinking claim*.

The following passage by the philosopher Fred Dretske, one of the preeminent and persistent champions of the C/NC-distinction since the late 1960s, contains a typical statement of the perception-without-thinking claim. Though I hasten to add that Dretske himself has far more to say on the non-conceptual nature of sensory-perceptual experience,[3] the quoted passage,

---

[3]   I should take this opportunity to mention, paying no little deference to his contributions, that we will come across Fred Dretske's work in each of the three chapters of this middle part which revolves around the demarcation of concepthood. Even though none of chapters 3, 4, or 5 developed out of the intention of providing running comments on a specific philosopher's work, I am impressed by how often Dretske's work contains the most explicit formulations of an idea in this field by a philosopher that I am aware of.

  Dretske is, of course, perfectly aware of the possibility that one might take perception to be conceptual. For example, he correctly acknowledges in a footnote to a different work: "Some (e.g., Armstrong 196[8]; Dennett 1969, 1991; Pitcher 1971) try to analyze experience in cognitive terms, as 'suppressed', 'potential', 'micro', or otherwise unconscious judgments that things are (or seem) a certain way". (Dretske 1995: p. 170, n. 8.)

when taken by itself, is nicely representative of the sort of justification of the perception-without-thinking claim (and thus of the C/NC-distinction) which many of its expounders, I maintain, prematurely rest content with. (I hasten to add, though, that it also involves a converse claim, which I think is more valuable. We turn to that in the next section.)

> Clyde's performance [of playing the piano] can [not only cause one to see or hear Clyde play the piano, but] also cause a belief—the belief that he is playing the piano. A perceptual belief that he is playing the piano must be distinguished from a perceptual experience of this same event. A person (or an animal, for that matter) can hear or see a piano being played without knowing, believing, or judging that a piano is being played. Conversely, a person (I do not know about animals) can come to believe that Clyde is playing the piano without seeing or hearing him do it—without experiencing the performance for herself.
>
> The distinction between a perceptual experience of *x* and a perceptual belief about *x* is, I hope, obvious enough. (Dretske 1993: 113).

This passage is very typical in how it considers an example of perception and intuitively denies to it the status of a high-level factive psychological state (in particular, of belief). On the basis of an uncashed linguistic intuition as to what deserves to be called a "belief", it is then claimed that perceptually representing a scene does not entail conceptually representing it.

What must be emphasized here is the following: It is a perfectly real option, for all that has been said so far, that sensory-perceptual representing and conceptual representing simply lie along a *continuum* of situatedness at different levels of cognition (where, as noted, the notion of levels intended would still have to be spelled out precisely). It is uncontentious that sensory and perceptual representations are located at *lower* levels of processing stimuli and that conceptual representations are at the *highest* level, i.e., what I referred to as "high-level cognition" earlier. The challenge, for anyone wishing to defend the claim that perception as such is nonconceptual, is to come up with a clear and untendentious criterion that allows us to draw a *nonarbitrary dividing line* between the conceptual realm and the putatively nonconceptual. In the absence of any nonarbitrary principle to draw the line, concepts and perceptual representations would have to be acknowledged to grade off into each other. But if this were so, then perceptual processing could just as well count as an *early stage of concept application.*

Here is a different way of seeing the present challenge: If we are to claim that perception as such is nonconceptual, we must be able give a principled reason why sensation and perception are *not* properly described as involving myriads of (low-level, introspectively inaccessible) applications of *micro-concepts*. Of course—to adhere to the example of perceiving Clyde as he plays the piano—the low-level concepts involved would not be concepts such as that of an individual person, or the concept of a piano, or the concept of playing an instrument. The low-level concepts would simply consist in the countless component representational states that go into the perceptual representation entertained in looking at someone playing a piano. For all we have seen so far, this is a very real possibility. Concepts can represent *any* sort of entity; at least this must be assumed in the absence of a very unexpected argument to the contrary. So why should not all the properties and relations constituting a scene depicted in a perceptual representation count as being represented by concepts too? Why are the features represented in one's rich perceptual experience thereby not to be counted as conceptually represented?

That there is a challenge to be met here before one can help oneself to the perception-without-thinking claim bears emphasis because it is too rarely acknowledged by philosophers in making the claim. We are usually only presented with solid reliance on unexplained linguistic intuitions about the applicability of such terms as 'thought' or 'belief' or simply 'concepts'. This is bad methodology, and a few words about it are still in order.

We are not engaging in the practice of simply asking how *comfortable* you feel about, e.g., calling sensorily generated representations 'concepts'. We must attempt to trace back disagreements about the applicability of a word to an underlying difference in the kinds of nonlinguistic domains under debate. If we want to maintain that there is a difference in kind between percepts and concepts, then we had better be able to say, in fairly precise terms, what explanatorily relevant features the posit of a concept has which percepts lack. If theorists are sure that perceptual representations do not qualify as special types of concepts, then we should reasonably expect them to be able to tell us which features it is in virtue of which this is so, and we can reasonably expect them to be similarly sure about what those features are. They should be able to name the distinctive features of specifically conceptual representations with a unanimity that matches their degree of confidence in the nonconceptual nature of percepts.

Rejecting blind reliance on unexplained intuitions is not to deny that strong intuitions about the nonapplicability of the theoretical term 'concept' and its

cognates may play a role in this investigation. Someone who has been exposed to the theoretical discourse involving uses of the term 'concept' even should go after such intuitions, if she has them. But rather than unquestioningly adhere to them, she should consider discarding these intuitions when they cannot be rationally grounded. The intuitions are only of heuristic use in pointing to a possible underlying difference in kind which itself has yet to be made precise.

One possible criterion is provided by the most prominent feature in customary characterizations of concepts: that of being a category, of being involved in acts of classification. I will argue later that this criterion cannot be given any clear content that would not be either trivially satisfied by sensory mental representations as well, or that would be too strong to provide a minimal common ground among concept theorists. This will be argued in the following chapter, which also scrutinizes a number of other notions that are widely encountered in distinguishing concepts from the nonconceptual. What I want to do now is to pursue the positive line of thought that leads to the criterion of memorization announced earlier.

4. The proper starting point:
Thinking without perception

One reason I have quoted the passage above by Dretske is that it does not merely state the perception-without-thinking claim, but that it also states, equally explicitly, the *converse* claim. That is, Dretske also makes a *thinking-without-perception claim* here.

What I would like to draw your attention to now is that there is an *asymmetry of clear cases* concerning these directions of dissociation between perception and thinking. Contrary to the hazily intuitive idea that perception does not count as thinking, the idea that you can think about something without perceptually experiencing it is *highly obvious indeed.* While we cannot safely claim, at a pretheoretical level, the perception-without-thinking claim, we can indeed safely claim, at a pretheoretical level, the thinking-without-perception claim. I am right now entertaining the occurrent belief that Neil Armstrong landed on the moon in 1969, but I am obviously not thereby perceiving Armstrong's lunar landing.

Strangely, the platitude that one is able to think about something without perceiving it appears to get far less attention than the perception-without-thinking claim. This is not as it should be, I believe, since the fact which lies at the heart of the thinking-without-perceiving platitude is more important

than tends to be noticed. It is this potential direction of dissociation, rather than the reverse one, that plausibly provides us with the entrance wedge into a C/NC-distinction.

I maintain that, if we want compelling cases to give initial justification to the C/NC-distinction, then we must start from the obvious truth that thinking of something is possible without perceiving it, becoming clear from the perspective of this nonentailment about the fundamental difference in types of causal role between perceptual states and thoughts. Once *this* difference has been made apparent, we have a basis for C/NC-distinction and can go back to the claim that perceiving is possible without thinking. Proceeding in this order, the perception-without-thinking claim can be seen to be true indeed: It follows from a principled distinction that we reach in light of the thinking-without-perceiving platitude. This is the path I will be pursuing on the following pages. Trying to introduce the C/NC-distinction via the perception-without-thinking claim, by contrast, would be to put the cart before the horse.

5. Stimulus-autonomy

Facts like the one that I can think about the first lunar landing without perceiving any lunar landing illustrate the difference between the general roles played by states of perceiving and states of thinking.

For a perceptual state, there is necessarily only one way in which it can correctly occur, namely that of its *being caused by its intentional object* (its representatum, its representational content), and more specifically, for the perceptual state to be caused by its intentional object *by way of sense receptor stimulation*. For me to perceive the moon is for me to be in a psychological state representing the moon, where this state has the causal etiology of having been caused by the moon via the stimulation of my sense receptors and subsequently, via the pathways of my sensory organs (specifically, those of vision, in this case). We can say, then, that perceptual states are characterized by the feature of *stimulus-dependence*. We can think of the two aspects just distinguished as constituting, on the one hand, the *distal* stimulus-dependence, or representatum-dependence, of perceptual states (moon perceivings causally depending on the moon) and, on the other hand, their *proximal* stimulus-dependence (moon perceivings causally depending on appropriate sensory stimulation).

Concepts, by contrast, are independent in both of these respects: They can appropriately occur without being caused by their intentional objects. And

they can appropriately occur without any concomitant activity of the sense receptors. In other words, conceptual states are in principle *stimulus-autonomous* representational states. The preceding contrast squares nicely with the traditional way of regarding perceiving as being passive, "receptive" in nature and thought as "spontaneous" cognitive activity.

Being capable of stimulus-autonomous representing—*stimulus-autonomy*, for short—is quite a spectacular property for an organism to have. The step from stimulus-dependent to stimulus-autonomous representing brings a whole new level of complexity into nature through the organisms that have it. Think of a hierarchical progression along successively higher stages of complexity in natural kinds (a "chain of being", to put it in the grander terms of the history of ideas). Such a progression would start with lifeless matter, proceeding over self-replicating macromolecules, on to bona fide organisms. At some point, closely tied to the development of locomotive abilities, organisms enter nature's stage that are endowed with a capacity to represent features of their environment. It is a similarly significant step when, at some later point, some of these organisms evolve into creatures that have the capacity to appropriately represent their environment independently of any particular stimuli to trigger the occurrences of these representations. Put differently, locating the basis of the C/NC-distinction in the difference between stimulus-autonomous and stimulus-dependent representing goes quite a long way towards making more precise sense of the hazy notion that concepts are representations which are only present in higher-developed cognitive organisms—that they somehow form a mark of *intelligence*.

The basic respect in which intelligence requires stimulus-autonomous representing lies in the representational powers to which stimulus-autonomous representing opens doors. An organism which is only capable of representing the contents that it presently perceives is a stupid creature. Notice that most things we ever think about are things we do not perceive in the moment that we think about them. Looking at this more closely, different respects can be distinguished in which the representational powers of thoughts transcend our present sensory-causal reach. This should make clear how entirely new prospects arise for organisms which can represent things without requiring sensory-causal contact to do so.

Firstly, thoughts can represent concrete events or objects that lie beyond the thinker's sensory-causal reach as a thoroughly contingent matter of fact, insofar as it is only the thinker's *spatiotemporal location* that prevents her from perceiving what she is representing in thought. A thought can repre-

sent past or possible future events which the thinker would be capable of perceiving if appropriately located. Representing possible future events is crucial to even the simplest acts of *planning* or *deliberating* in creatures. By the same token, a thought can represent a simultaneously occurring event that is outside the spatial range of the thinker's appropriately oriented sensory organs. Spatial remoteness is often pointed out as a common feature of intentional objects of thinking, but notice that it is not spatial remoteness as such that counts here, only spatial remoteness that transcends causal-sensory reach.

Furthermore, thoughts can represent concrete events or objects that lie beyond the thinker's sensory-causal reach as a nomologically necessary matter of fact, insofar as the thinker is not *equipped* to sensorily detect them, no matter how appropriately situated she is. We can think about ultraviolet radiation, even though it is outside the part of the electromagnetic spectrum that is visible to us. We can entertain thoughts about protons, even though they are far too small to detect by means of our sense organs. This is the familiar distinction between theoretical and observable entities.

Finally, there are the sorts of intentional objects that are, as a matter of logical necessity, not the sorts of things that could be sensorily perceived. These include nonexistent or fictional objects (the Easter Bunny, or the goddess Aphrodite), logically impossible objects (round circles), mathematical objects, logically complex facts, nonactual states of affairs.

Wherever representational states have entities (or non-entities) of the sorts mentioned in the preceding paragraphs as their intentional objects, those representational states must be stimulus-autonomously tokenable.

6. Memorization

Given the decision to take sensory-perceptual representing as the basic case of nonconceptual representing, and given the subsequent identification of nonconceptual representing with stimulus-dependent representing, a nonarbitrary root of the C/NC-distinction is not at all difficult to locate. It lies very close to the surface. All that we have to do is to specify, in the most general terms available from psychology, the condition which a representational state must fulfill in order to be stimulus-autonomous. The condition is that of being stored in memory. The memory criterion forms a basic part of the minimal common denominator of what philosophers as well as psychologists are prepared to count as specifically conceptual mental representations. I will positively expand on this common denominator in chapter 5.

The inference from stimulus-autonomy to memorization is plausibly not just a case of an inference to the best explanation out of several conceivable ones. It rather seems that the sort of stimulus-autonomy we are aiming at is nothing less than constituted by the feature of memorization. What we are aiming at is the reliable availability of internal representations to an organism independently of sensory stimulation, something that is deserving of ascribing the *possession* of those representations to the organism. What is intended, therefore, is those representations' *storage* in some structure where they are usually not activated, and where that structure is *functionally independent of the sensory pathways,* enabling on an independent basis appropriate retrievals/activations of those representations. Whatever more specific notion of memory a cognitive scientist might wish to delimit, it seems that the general psychological notion of memory simply is that of some structure that fulfills this sort of role. If this generic notion of memory is what we are aiming at with stimulus-autonomy, and if concepts are necessarily stimulus-autonomously tokenable, then the inference from the stimulus-autonomy to the memorization of concepts is a deductively valid one.

It should be added that there is one important way of entertaining a representation stimulus-autonomously that we have not explicitly considered so far: One can token concepts that have been constructed on the spot, by combining other concepts. This possibility of entertaining a representation stimulus-independently is derivative upon memory-based retrieval. The building blocks of representations which get assembled spontaneously themselves *are* stored in memory. So the additional aspect that spontaneously assembled representations bring to stimulus-autonomous representational activity does not violate the effective equivalence of stimulus-autonomy and memorization of representations. It only points to a difference between representations that are *immediately* memory-based and such that are *mediately* memory-based.

The path we have taken to circle in on concepts involves the reduction of the C/NC-distinction to successively more specific contrast pairs. Let me recapitulate these contrast pairs in their order of reduction, with the conceptual being captured by the left side of the respective contrast pairs:

> **The path of reduction taken:**
> Conceptual *vs.* nonconceptual repr. states
> → Conceptual *vs.* sensory-perceptual repr. states
> → Stimulus-autonomous *vs.* stimulus-dependent repr. states
> → Memory-dependent *vs.* stimulus-dependent repr. states.

Aligning the C/NC-distinction with the question of memorization is not an altogether new idea, but the connection is a remarkably often overlooked one in philosophy, as I have pointed out in this chapter's introduction, even among those philosophers whose work is expressly dedicated to the theory of concepts. Conceptual representations seem to form a posit in philosophy that is uniformly connected to a criterion of existence which is, curiously enough, as clear and informative as it usually remains tacit.

§4
On the idea of concepts as categorizing representations

_____

_____

The classical characterization of concepts that would appear to give a more specific criterion for concepthood than that of being a subpropositional mental representation is that concepts are the representations which we employ in *categorization*, or equivalently, *classification*. The characterization of concepts as classificatory or categorizing representations is the single dominant one in psychology.[1] It is similarly widely encountered there as the constituent-of-thought dictum is encountered in philosophy. This is not to say that it takes the place of the latter. The constituent-of-thought notion which we have explained earlier must be understood to be presupposed when theorists claim concepts to be categorizing representations.

I do not think that the idea of concepts as categories is of much help in getting at a minimal criterion to distinguish concepts from nonconceptual representations. The following seems to be underappreciated. The notion of "categorizing" or "classifying" is not a univocal one. Once we attempt to spell the notion out, we either obtain a feature that is at least equally descriptive of nonconceptual representations, or else we obtain a feature that is too strong to reflect a minimal consensus about concepthood. Arguing to this effect forms the bracket which holds this chapter together.

1. Categorization and concepthood, I:
Multiple applicability

In a minimal but robust sense that can be given to talk of "classification" or "categorization", a representation need only satisfy one requirement in order to count as a classifying or categorizing representation: It must have multiple applicability; it must be a general representation in the sense that it must be capable of subsuming different particulars under it. My talk of generality here is adopted from the terminology of the traditional distinction between "general terms" and "singular terms". Concepts, in line with the notion addressed in this section, are understood to be mental analogues of general terms in natural language.

A few words of clarification first. The traditional distinction between general terms and singular terms is somewhat less clear than one might think, once one tries to define it in terms that avoid tacit reliance on more specific linguistic-syntactic categories such as those of an adjective or a noun. What

---

[1] Psychologists also often say, to the same effect, that concepts represent *categories*. They likewise like to say that concepts *are* categories. Often the same authors do both within the same work. In fairness, however, we should note that the term 'category' indeed lends itself to both sides of the representational divide and that it is convenient to employ it both ways. Context usually leaves little doubt as to which sense is intended.

it seems we can say for certain, however, is that this distinction is supposed to rest on a distinction between two different metaphysical types which the referents of the respective kinds of terms are supposed to belong to. Singular terms refer to *particulars*. Particulars are supposed to be singular entities. They exist exactly once. General terms, by contrast, refer to *universals*. Universals are properties or relations. They are supposed to be entities that can be instantiated by different particulars, particulars which can be arbitrarily greatly separated in space and time. The distinction is usually trusted to be clear enough on the basis of the examples used to illustrate it.[2] For example, it would be said that, in the sentence 'Odonkor is fast', the term 'Odonkor' is intended to refer to exactly one thing, the unique person Odonkor, while the term 'fast' refers to the property of being fast, something that can be present in many particulars. To avoid tedious phrasings, I will sometimes reduce my focus on general representations to the case in which a representation refers to a property *F*. Nothing will be said that could

---

[2] The intuitive support provided by examples here actually must be more heavily relied upon than is desirable. I have strong reservations about the characterization of the distinction between universals and particulars along the traditional lines just given. For consider what is done in the usual way of defining the distinction: In being told that particulars are unique entities while universals can be multiply instantiated, the emphasis is supposed to lie on the contrast between uniqueness and multiplicity. But in fact there is no contrast between uniqueness and multiplicity *as such*. After all, each universal is a singular entity just as much as each particular is a singular entity. The property of being a tennis ball, the relation of spatial contiguity, the property of being humourous: these entities, if they do exist, are all as uncontentiously singular as particulars are—for example, Paul Breitner, the planet Jupiter, Lahm's opening goal of the 2006 World Cup, or the number 3. It cannot therefore be uniqueness or multiplicity simpliciter that distinguishes particulars from universals.

What becomes apparent here is that the present definition of the distinction crucially depends on talk of universals being *multiply instantiable*. Conversely, the feature that would really have to be said to set off particulars from universals is that the former are *not* multiply instantiable. But if this is so, then we really would not seem to have the beginning of a criterion here. For a definition of universals that makes essential use of the term "instantiation" is circular. "Instantiation" is a metaphysical term of art which is *tailored to* universals. Only universals are the sorts of things that can be said to be instantiated. Moreover, it is patently not the case that we have any understanding of what the instantiation relation is prior to an understanding of what properties and relations are. Particulars do not violate multiple instantiability. Rather, since particulars are not universals, they are not the sorts of things that could count as being instantiated at all.

If these considerations are on target, the traditional definition of two preceding paragraphs show the traditional definition of the universal/particular-distinction over multiple instantiability as opposed to singular existence is a bogus definition. There may well be other, more adequate definitions available that I am not aware of. But we need not explore the different paths which one might take from here. The present objective in the text was to clarify the idea of generality that the notion of classificatory representations commits one to; but solving metaphysicians' deepest questions cannot be our intention here. However bad we are at defining the distinction between properties and individual objects, we seem to have a firm and reliable idea of which is which. This is something we must fall back on in this context.

not be straightforwardly reformulated over representations that refer to a relation.

The event of classification or categorization itself is commonly described in a number of equivalent ways, some of which have already been exemplified in my own usage. One says of representations of *F*-ness that they get applied to entities which are *F;* that entities that are *F* are classified or categorized as having *F*, or that they are subsumed under the representation of *F*-ness. Different notions of classification could be made precise in accord with different specifications of this relation between the categorized particular and the categorizing predicative representation. Where the general representation is a mental representation, there are at root two options as to what classification can be. One is that the event of classification consists in a system's representing the combination of *a* and *F,* or, put differently, in representing *a* as instantiating *F*. The other variant of classification is more primitive. It would not require specific particulars instantiating *F* to be themselves represented, but only that an *F*-representation is somehow selectively sensitive to instantiations of *F*-ness in the world.

Now, the idea is simply false that being classificational in the sense of being general (applicable across particulars) might be sufficient for a mental representation to be a concept. Generality does not even provide an initial step into the direction of a characteristic, sufficient feature for concepthood. Any sensory representational state can qualify as a being classificatory in kind too. The most straightforward way to see that the criterion of classificational nature does not exclude sensory, nonconceptual states, is via the platitude that any sensory-perceptual state is repeatable. Suffice it to consider here the raw sensory state that consists in representing a certain hue of red somewhere in my visual field. A sensation of redness can reoccur innumerably often; and it does. Once it is taken to be a bona fide representational state with conditions of correctness and the possibility of error, its different occurrences *ipso facto classify* various points in different surroundings that I happen to be in as being red. In other words, if sensory states are granted to be representational states, then they clearly are general representations, and hence classificational states, in one bona fide sense of the term. Each elementary sensory state of a given type classifies some point or expanse in the environment as instantiating a specific property rather than any other.

The following is a similarly straightforward rebuttal of the idea that being a classificational qua general representation might be a feature which is suited to characterize concepts to the exclusion of nonconceptual, sensory states. Sensory events represent instantiations of certain properties, as indicated al-

ready in comparing the sensory and the perceptual (notwithstanding the previously ascertained fact that perceptual states need not differ from them in this respect). After all, it is certain properties that the sense receptors are, as a matter of nomological fact, causally sensitive to. So sensory nonconceptual states are anything but less characteristically classificational representations than concepts, if being a classificational representation is understood to amount to being a general representation.

In fact, the suggestion of concepts as distinctively classificational rather seems to get matters the wrong way round, if we understand classification to consist in a property representation being tokened in application to a particular. For this is precisely the sort of event that sensory states are restricted to of necessity. Sensory states are factive states whose appropriate occurrences are restricted to registering the presence of a given property in the environment or the body. By contrast, conceptual states are precisely not restricted to this kind of context of appropriate occurrence.[3] Given their essential capacity to occur stimulus-autonomously, they can occur other than in a classificational event of being applied to a particular. The simple notion of classification, then, is actually more closely tied to that of sensory representations than to that of concepts. Sensory representations exhibit a necessary tie with it, while concepts do not.

Though I do not wish to make too big a meal about the notion rejected in the present section, I do think I should point out that this assumption of a criterial tie between concepts and generality (or representation of universals) is not a strawman I have brought into play. Witness, for example, Bertrand Russell as he firmly states, on the first page of a paper entitled "On the relations of universals and particulars", that "[t]he first distinction that concerns us is the distinction between percepts and concepts, i.e., between between objects of acts of perception and objects of acts of conception. If there is a distinction between particulars and universals, percepts will be among particulars, while concepts will be among universals." (1911: 105) This is straightforwardly tranposable into talk of mental representations, yielding the untenable idea just critiqued.

A word is still in order concerning the traditional conception on which concepts are understood in analogy with general terms. Being classificational in the sense of having general applicability is not merely, as argued in the paragraphs back, non-sufficient for a mental representation to be a concept.

---

[3] This is why conceptual states span the other families of psychological state types too besides that of factive states, namely, those of volitional states, and states of pure thinking-about.

It is also non-necessary. For, no less than there must be conceptual mental representations corresponding to general terms, there must be conceptual mental representations corresponding to singular terms—be it concepts expressed by proper names (e.g., the concept HORST KÖHLER) or concepts expressed by definite descriptions (e.g., THE NUMBER OF PLANETS).

What this points us to is the whole extent to which the notion of generality (viz., applicability to different particulars) is concept-theoretically irrelevant. In particular, you cannot spell out an interesting notion of a classificational representation over the notion of generality, or applicability to different particulars. Psychologically speaking, singular concepts can equally well be said to be employed in "classification" as general concepts can. The application of these types of concepts would seem to involve preciously little psychological difference. Take the case of applying concepts to things perceived. I have a stored representation of Horst Köhler in my memory and, very often, when I come across a press photo of Horst Köhler, or his voice on the radio, or merely his printed name, my mental representation HORST KÖHLER is retrieved from memory and is applied to what I am perceiving. Similar processes occur with my concept DOUBLEDECKER BUS when I perceptually encounter doubledecker buses. If applications of the latter count as classification, so must applications of the former. A uniqueness of the representatum has nothing to do with this.

More generally, there is surely no difference in the basic kinds of psychological processes underlying the application of a singular mental representation as such, in comparison to the application of a general mental representation to things I perceptually encounter. At any rate, there is no such difference which would begin to motivate a dividing line between explanatorily different kinds of mental representations—let alone a dividing line as basic as the rift between concepts and nonconceptual representations. In order for my HORST KÖHLER mental representation to be a singular one, it will have to be accompanied by the assumption that it picks out a unique individual. But the attachment of this collateral knowledge does nothing to diminish the appropriateness with which the mental representation HORST KÖHLER itself is to be regarded as conceptual in kind.

In fact, the matter could be carried further, if we wished to. If we remind ourselves of how concepthood is customarily connected to a higher level of cognitive sophistication, then the general/singular distinction among representations not only fails to agree with an explanatorily relevant C/NC-distinction, it rather seems to point into opposite directions. Plausibly, having singular mental representations stored in memory requires precisely a

minimum degree of cognitive sophistication in a creature which general mental representations as such do not require. Singular mental representations arguably must be derived from general mental representations, rather than the other way round. But I will not go into this here. Rather, let us now shift to a different notion of classificational representationhood which innumerable theorists connect to concepthood—and without doubt, more relevantly so. This one is based on post-perceptual processing.

2. Categorization and concepthood, II: Classification across different percepts

Especially when psychologists define concepts over categorization, what they mean is not the bare sort of categorization that we have just been subjecting to criticism, but really the more specific notion of *imposing categories upon the world in the sense of partitioning one's perceptual experience of it.* Derivatively upon this, it is the notion of recognizing a given entity, or type of entity, across different perceptual representations of it. I will try to explain these and related notions in the course of the present section. Since the categorization performed by concepts here is assumed to operate on perceptual representations, we can also call it *postperceptual categorization.* I shall sometimes speak of the thus grounded notion of a concept as the "PPC notion" of concepts.

To get an impression of how psychologists characterize concepts over the PPC notion, let me offer a sample—hopefully representative, as I believe—of pertinent passages from the literature. Like many other psychologists, the authors quoted show an acute awareness of the need to introduce concepts by means of addressing the purpose of having categorizing abilities.[4] Also notice that, from the outset, what is under consideration here are creatures with an ongoing perceptual experience. Consider first the expositions of the subject matter by two of the leading psychological researchers on concepts:

> We are forever trying to carve nature at its joints, dividing it into categories so that we can make sense of the world. If we see a particular child pet a particular dog at a particular time and a particular place, we code it as just another instance of children liking dogs. In so doing, we reduce a wealth of particulars to a simple relation between the categories of children and dogs. Not only do we free our mental capacities for other tasks, but also we can see the connection between our current situation and our past experience. [...] The major functions of cate-

---

4 It must be lamented that one all too often misses such awareness in philosophical perspectives on concepts.

gorization are coding of experience and licensing of inference. (Smith 1995: 3 and 5.)

When we walk into a room, try a new restaurant, go to the supermarket to buy groceries, meet a doctor, or read a story, we must rely on our concepts of the world to help us understand what is happening. We seldom eat the same tomato twice, and we often encounter novel objects, people, and situations. Fortunately, even novel things are usually similar to things we already know, often exemplifying a category that we are familiar with. Although I've never seen this particular tomato before, it is probably like other tomatoes I have eaten and so is edible. If we have formed a concept (a mental representation) corresponding to that category (the class of objects in the world), then this concept will help us understand and respond appropriately to a new entity in that category. Concepts are a kind of mental glue, then, in that they tie our past experiences to our present interactions with the world, and because the concepts themselves are connected to our larger knowledge structures.

[...] (If I were more melodramatic, I could spin a fairy tale in which a person who has no concepts starves while surrounded by tomatoes, because he or she had never seen *those particular* tomatoes before and so doesn't know what to do with them.) (Murphy 2002: 1 and 3.)

In a similar vein, here is the explanation of concepts offered in the relevant section of a general psychology textbook.

Stellen Sie sich eine Welt vor, in der Ihnen jeder Gegenstand und jedes Ereignis neu vorkommen, ohne Bezug zu irgendetwas, das zuvor geschehen ist. Sich gestern verbrannt zu haben, hätte Sie nicht gelehrt, heute auf einen heißen Gegenstand anders zu reagieren. In einer solchen Welt der immer neuen Wahrnehmungen, ohne eine Möglichkeit der Klassifikation der Informationen, könnten Sie nicht auf vergangene Erfahrungen zugunsten effektiveren künftigen Verhaltens bauen. Glücklicherweise müssen wir auf Reize nicht so reagieren, als handle es sich um singuläre, bezugslose sensorische Ereignisse; wir ordnen sie vielmehr als Fälle in *Kategorien* ein, die wir aufgrund unserer Erfahrungen gebildet haben.

Diese Fähigkeit, individuelle Erfahrungen zu kategorisieren—sie gleich zu behandeln oder ihnen ein gemeinsames Etikett zu verleihen—wird als eine der grundlegenden Fähigkeiten der meisten lebenden Organismen betrachtet. Die Kategorien, die wir bilden, werden *Begriffe* oder *Konzepte* genannt. (Zimbardo 1995: 363.)

A wealth of different descriptions are offered by the quoted authors to describe the purpose and nature of employing concepts. On the following

pages, I want to sort out which distinct ideas are in play in psychological researchers' statements of the above kind, to state these ideas precisely, and to identify the relations obtaining between them.

2.1. Perceptual feature abstraction. The idea which captures the essence of concepts as postperceptual categorizations is that of a representation which *results from abstracting from perceptual detail*. It is often said that experience is "divided" or "organized" by imposing categories on it. The basic operation in play here is really that representational states with significantly different sensory-perceptual contents are grouped together, on the basis of relevant similarities, as representing the *same* entity. Different but relevantly similar percepts effect the tokening of one and the same concept. Representational differences among the percepts cease to matter for subsequent stages of processing. This core fact about abstraction, the *unavailability* of more detailed information beyond that which is captured by a given partitioning of perceptual input, is also often described as a case of making a transformation from an "analog" to a "digital" format of representing.

Perhaps the preeminent philosopher to have described the partitioning of information delivered by perceptual experience, and the ignoring of differences within different perceptual inputs of a given category representation, is Fred Dretske (1981: ch.6; 1995: ch.1, §3). Here is an illuminating passage from Dretske's book *Knowledge and the Flow of Information:*

> To describe a process in which a piece of information is converted from analog to digital form is to describe a process that necessarily involves the *loss* of information. [...] Until information has been lost, or discarded, an information-processing system has failed to treat *different* things as essentially the *same*. It has failed to classify or categorize, failed to generalize, failed to "recognize" the input as being an instance (token) of a more general type. (Dretske 1981: 141.)

In immediate continuation of this passage, Dretske applies the idea of abstraction—or equivalently, to mention a few terms, of "feature extraction", "analog/digital-conversion", "categorization", "classification", "stimulus generalization"—to the distinction between sensory, preconceptual processes and concept-involving processes:

> The contrast between an analog and a digital encoding of information ... is useful for distinguishing between sensory and cognitive processes. Perception is a process by means of which information is delivered within a richer matrix of information (hence in *analog* form) to the cognitive centres for their selec-

tive use. Seeing, hearing, and smelling are different ways we
have of getting information about [a source] *s* to a digital-
conversion unit whose function it is to extract pertinent infor-
mation from the sensory representation for purposes of modify-
ing output. It is the successful conversion of information into
(appropriate) digital form that constitutes the essence of cogni-
tive activity. [...] The traditional idea that knowledge, belief,
and thought involve *concepts* while sensation (or sensory ex-
perience) does not is reflected in this coding difference. Cogni-
tive activity is the *conceptual* mobilization of incoming infor-
mation, and this conceptual treatment is fundamentally a matter
of ignoring differences (as irrelevant to an underlying same-
ness), of going from the concrete to the abstract, of passing
from the particular to the general. It is, in short, a matter of
making the analog-digital transformation. (Ibid.: pp. 141-142.)

What should be borne in mind when we talk about the analog/digital-
distinction in the context of the theory of concepts is that it only draws a
relative distinction concerning degrees of representational detail. By defini-
tion, a "digital" representation incorporates *less* information about a given
representatum than the "analog" representations that *it* has been extracted
from. But there is, in principle, no upper bound to how much detail can be
encoded by a digitally generated representation. Similarly, there is, in prin-
ciple, no lower bound to how little detail can be encoded by an analog rep-
resentation. (Just consider, for example, the perceptual representation of the
most minute patch of red that your eyes can discern.)

There is a cluster of related notions in this vicinity. The analog/digital-
distinction is customarily thought of as being tantamount to the distinction
between "continuous" and "discrete" formats of representing. It is also often
said that nonconceptual representations are "homogenous", that they are
"grainless", or that they are "seamless". For the reason given, I do not see
anything substantially new in these descriptions that is not captured by the
property of resulting from extraction of features represented at a prior stage.
Also, certain perils attach to the utilization of these notions. Firstly, the
properties are typically ascribed to representations in a non-relativized man-
ner, strongly suggesting that the authors are taking them to be absolute, in-
trinsic properties of a given perceptual representation, in contrast to con-
cepts. Secondly, the talk of "homogeneity" or "grainlessness" occupies a
central position in the debate about phenomenal properties and might
merely reflect a phenomenal property of perceptual representations, bring-
ing into play an irrelevant aspect in the comparison between perceptual and
conceptual representations.

Turning back to cognitive psychology proper, here is a good example of how postperceptual categorization, understood as the abstraction from perceptual detail, provides the basis for the psychologist's notion of concepts. In his well-known cognitive psychology textbook, John R. Anderson locates concepts as follows in the scheme of things.

> In diesem Kapitel befassen wir uns mit Wissensrepräsentationen, die das Bedeutsame eines Ereignisses herausfiltern und die meisten der unwichtigen Details weglassen. Wissensrepräsentationen, die in der genannten Art vom Unwesentlichen abstrahieren, nennt man *bedeutungsbezogene Repräsentationen,* im Gegensatz zu den wahrnehmungsbasierten Repräsentationen in Kapitel 4. (Anderson 1996: 133.)

> Die grundlegende Charakteristik [bedeutungsbezogener Wissensrepräsentationen] besteht in einer bedeutsamen Abstraktion, die von den (sensorischen) Erfahrungen, die ursprünglich zum Aufbau des Wissens geführt haben, wegführt. [...] Eine Art des Abstrahierens besteht darin, von den spezifischen Erfahrungen abzusehen und stattdessen die Merkmale und Kennzeichen der jeweiligen Erfahrungsklasse allgemein zu kategorisieren. Das Resultat derartiger Abstraktionen nennt man konzeptuelles Wissen. (Op. cit.: p. 147.)

In this case, concepts are presented as types of "semantic" representations, as opposed to "perceptual" representations. One might wonder what this should mean, given that representations, after all, are ipso facto semantically evaluable entities. The label is not very adeptly selected and simply aims at the already mentioned fact that concepts are formed in line with what is relevant/significant for a creature. Understood thus, the label basically just repeats our description of feature extraction/abstraction as occurring with respect to relevant similarities in percepts. Cognitive psychologists in fact often speak of "semantic memory" as the functional unit within which concepts are stored.[5]

---

[5]   Here as elsewhere in psychological texts, there is some taxonomizing terminology that cannot be given a firm independent sense and that is best ignored in tracking the perspective on concepts I want to illustrate here. For example, the talk of "knowledge representations" as opposed to simply "representations" is highly dubious, for more reason than one. Also, Anderson is introducing conceptual representations only as one type of "semantic representation", the other type, which he introduces earlier in the chapter, consisting in "propositional representations". The contrast is misleading, as is the term 'semantic representation' itself. But I am bracketing such complaints here. A great number of psychological texts exhibit severe unclarities in their ways of drawing foundational distinctions. We must allow ourselves to talk about basic constants in the psychologists' scheme of things without being overly distracted by such unclarities.

2.2. Perceptual recognition. Notice how practically everything that the quoted authors say assumes the involvement of recognitional processes, in other words, how it is straightforwardly paraphrasable into talk of recognizing entities. Thus, Edward Smith speaks of "seeing the connection between our current situation and our past experience", and of "the coding of experience", thereby meaning someone's coding of a perceptual scene "as just another instance of" a certain type of scene. Gregory Murphy speaks of "novel things often exemplifying a category that we are familiar with". Philip Zimbardo employs several descriptions which he appears to regard as equivalent for his purposes: avoiding to have all new experiences "seem new" to one; being able to notice the "relatedness" of events perceived now and in the past; being able to "categorize individual experiences".

It is a bit surprising that the authors do not actually use the term 'recognition' here. The term has quite the same meaning in scientific psychology as it does in everyday usage. In what follows, when I speak of "recognition", I ipso facto mean perceptual recognition, i.e., recognition as a process operating on perceptual input.

Initially, one would be inclined to describe recognition something along the lines that recognition is the process of identifying a presently perceived entity as a specific (kind of) entity which one is already familiar with. But a more substantial decomposition into component processes can be given on a relatively a priori basis, in advance of considering any specific psychological theories. It seems that, on any representationalist theorist's account, at least the following facts must obtain in order for someone to count as recognizing an entity *x:* One has been storing in memory a mental representation of *x, C.* Some perceptual encounter with *x* has caused one to token a percept of *x, P.*[6] Consequently upon this, some sort of correspondence (some "match", whatever exactly that amounts to) is obtained between *P* and *C.* And as a result of obtaining this match, one's representation *C* is retrieved from memory and applied to *x* in thought.[7]

---

[6] Actually, the occurrence of the percept need not necessarily be a consequence of having perceptually encountered the real-world entity which is eventually recognized. It might also, for example, be a consequence of perceptually encountering a pictorial representation of that entity (you can recognize a giraffe in the course of looking at a drawing of a giraffe). These niceties can be prescinded from here, though.

[7] From the folk perspective, I suppose one might also want to include in the analysis of recognition an occurrent state of meta-knowledge to the effect that one is familiar with *x* (that one has already perceptually encountered other instances of *x* before). This cognitive event's phenomenal correlate would be that feeling of familiarity in moments of recognizing which we are all acquainted with, commonly captured by metaphors of vision and illuminating flashes. But to include this cognitive element in the basic psycho-

Why should the notion of recognition hang together with the notion of a representation that results from categorizing (abstracting from) perceptual input? Well, nothing really merits being called 'recognition' that is not recognition of *x* across *differing* perceptual representations of *x*. Recognition of *x* is usually taken to be based on the recognition of certain features in the percept whose presence is independent of most of the details that are usually represented in percepts of *x*. That is, the matching between the percept and the concept of *x* that is required in order to achieve recognition is one for which great amounts of informational detail in the percept can be ignored, with matching between percepts and one and the same memory representation (i.e. concept) being possible across considerable variations in the percepts. So, for example, the ability to recognize the goalkeeper Lehmann could amount to the ability to correctly match one's occurrent percept of Lehmann with one's memorized representation of Lehmann across such situations as seeing Lehmann from the front and close-up, seeing Lehmann lying on the ground, seeing Lehmann in the dusk, seeing him dive for the ball, and so on.

2.3. The presupposition of memorization. Since storage in memory was offered as a minimal criterion of concepthood in the preceding chapter, it bears emphasis that both of the reviewed notions which are knit into the PPC notion of a concept—that of *perceptual feature abstraction* and that of *perceptual recognition*—crucially presuppose the idea of concepts being mental representations stored in memory. The connection of recognition with memory has already been made explicit in the analysis offered above. Roughly, recognition is simply the process of bringing the right memorized representation to bear on a perceptual object, given a specific percept. The connection of perceptual feature abstraction with memory is equally indisputable. What makes a feature count as having been abstracted/extracted from perceptual representations simply *is* that is has been encoded in memory. For representations of features are counted as having been extracted from perceptual representations only insofar as they are available for perception-independent processes, and this availability amounts to nothing other than storage in memory.

2.4. Contrast with perceptual discrimination. What perceptual recognition contrasts with, in light of questions about memory, is *perceptual discrimination*. The experimentally well-established, simple example of colour discrimination nicely illustrates that our ability to perceptually discriminate

---

logical notion of the process of recognition would constitute an explanatorily unreasonable constriction.

perceivable properties from each other reaches down to a level of detail that far surpasses our ability to memorize distinct perceivable properties—in other words, that it far surpasses our ability to recognize perceivable properties as the ones they are.[8] Take a specific hue of green ('Green$_{66}$', we might call it) and take one of the most similar colour hues to it which you can still reliably make out to be distinct from it (call it 'Green$_{67}$'). By assumption, you will be able to notice a difference between Green$_{66}$ and Green$_{67}$ when you have both of these colour hues side by side in front of you. But when you encounter one of the colour hues by itself, you will in the overwhelmingly likely case not be able to tell with certainty whether it is Green$_{66}$ or Green$_{67}$ that you have in front of you, or perhaps some other, similar hue of green.

Put in terms of the distinction I have emphasized already, we can represent the property of being Green$_{66}$ and the property of being Green$_{67}$ in a stimulus-dependent manner, but we cannot represent them stimulus-autonomously. The first ability is what enables the undoubtedly behaviourally highly relevant event of perceptual discrimination to occur—of noticing a difference between perceived properties along given dimensions of variation. The second ability is what constitutes a recognitional ability with respect to a perceivable property and thus the possession of a concept of this property.

2.5. The functions of postperceptual categorization. Enabling object recognition across perceptual variations is the *basic* function of postperceptual categorization.

The existence of PPC concepts is also often justified with reference to an *economical* purpose it serves. The economical effects are clear enough, given that the process of abstracting from perceptual detail reduces the workload on memory and on memory-dependent processes. And these economical effects are indispensable, given the severe memory limitations of actual creatures. Notice, however, that economy itself does not form the basic raison d'être of postperceptual categorization. For the ability to assign identical postperceptual categories to one's perceptual input is a requisite no matter how unlimited your storage and retrieval capacities are. You must be able to recognize identicals as identical either way, when perception provides you with the requisite information.

---

[8] See, for example, some of the references to vintage psychological literature assembled in Diane Raffman (1995: 294-296). As she puts it, "[i]t is a truism of perceptual psychology and psychophysics that, with rare exceptions, discrimination along perceptual dimensions surpasses identification."

Moving on, we can note that practically nothing in the general statements given above by psychologists is about the process of recognition simpliciter. The very same phrases by which the authors bring into play obviously recognitional notions are phrases which are highly laden with an idiom of *epistemically valuable* states. What is considered is understanding, knowledge structures, licensing of inferences. Nevertheless, the psychological authors are invariably interested in object recognition as an ability relative to these epistemic notions.

Thus, Smith speaks of "making sense of the world" and of "seeing the connection between our current situation and our past experience". Murphy describes concepts as "helping us understand what is happening", as "helping us understand and respond appropriately to a new entity in that category", as "being connected to our larger knowledge structures" and, relatedly, of concepts as fulfilling the function of "licensing inference" and as "tying our past experiences to our present interactions with the world". Furthermore, in striking similarity to the other authors, Zimbardo asks us to imagine someone who lacks the ability of postperceptual categorization, aiming to indicate in an ex negativo fashion the benefit of being capable of postperceptual categorization. The benefit is described as being that of enabling one to "build on past experience in favour of more effective future behaviour". For example, one is in a position to let harmful experiences of the past "teach" one to react differently in the future to situations of the sort that caused these harmful experiences.

Notice that these epistemic and behavioural benefits are distinct from what the event of recognition guarantees in and of itself. More specifically, no states of understanding are entailed by the bare process of recognition which was analyzed above in general outlines. Consider yourself looking at a previously unencountered instance of an object *x:* A correspondence has been achieved between your percept of *x* and a stored concept of *x* and, as a consequence, you have tokened the concept in application to the perceived *x*. But this event by itself does not give you an understanding of what you are looking at. What is required for you to count as enjoying some understanding of what entity you are perceiving is for you to have some knowledge about *x* stored in connection with the concept of *x*.

By way of illustration, take the situation of perceptually encountering new instances of tomatoes. A piece of general knowledge which tomato-perceiving people who retrieve tomato representations will likely share is the knowledge that these objects are edible. Many of the further characteristic pieces of knowledge which would be activated in conjunction with catego-

rizing a perceived object as a tomato are based on perceptual experiences of tomatoes which one has made in different sense modalities and from various perspectives in the past. For example, when I myself think of a tomato, I typically token a mental image of a tomato. Sometimes it depicts the tomato from the outside. When I think of the tomato from the outside, I recall the tomato's characteristic shape, its skin's bright redness, its skin's smooth, shiny texture, and the circular little portion at its centre top where the tomato fruit was connected to the tomato plant's stem. I think of the consistencies of a ripe tomato, or of a less ripe tomato, both of which I have tactually experienced in the past by gently pressing tomatoes with my thumb. Sometimes I think of a tomato from the inside, recalling the gelatinous consistency of the pulp of a tomato, the pulp's fairly transparent colour, the characteristic pattern formed by the seeds in the tomato's pulp, and the pale red flesh of the tomato. I recall that the flesh has a coarse texture, that it encloses the pulp and also extends into various portions of the pulp. I know, based on memories of cutting and peeling tomatoes in the past, that the outer peel of the tomato immediately covers the flesh and is much thinner than the fleshy layer beneath. Similarly, thinking of a tomato typically makes me recall the gustatory qualities of tomatoes. I recall the taste of eating fried tomato slices on toast, but I also recall the decidedly less agreeable taste of raw tomatoes.

Some of the aforementioned sort of knowledge about tomatoes must, in the case of perceiving a tomato, be assumed to be retrieved together with the representation of a tomato from memory if cognitive psychologists' talk of being in a position to *understand what one is perceiving* is to make any sense. Put differently, one does not understand that one is looking at a tomato if one just applies to the perceived object an isolated "label" from memory, and remains otherwise incapable of judging such elementary things as whether this thing is edible or made of stone on the inside, or whether it has any of the perceptual features which one has observed upon earlier encounters with tomatoes. Even if we assume the recognitional process to include an acknowledgment of the fact that one has already seen this sort of object before, we still would not have a realization of how one's present experience is related to past experience in the sense that the psychologists quoted above have in mind.

The capacity of connecting to something, upon perceptual recognition of it, data which one has collected about it is ultimately subservient to the capacity of initiating appropriate behaviour. At least this is clearly so in the phylogenetically more basic cases of cognitive abilities. Two of the quoted passages strongly emphasize the enabling of appropriate behavioural reactions

as a function of postperceptual categorization. In the general case, the behavioural responses will be mediated by the performance of inferences over stored knowledge. This is evidenced, for example, by the fact that, when you are perceiving a tomato and recognize it as such, you could act otherwise than to eat it, even when you are hungry. The quoted passage by Murphy is explicit about resting the behavioural capacity enabled by postperceptual categorization on the prior epistemic capacity which he connects to it. Hence his comment that the application of a concept in categorizing perceptual objects is supposed to "help us understand and respond appropriately" to newly encountered tokens. Hence his description of concepts being a "mental glue" insofar as "they tie our past experiences to our present interactions with the world". Hence his consideration of someone who starves in the presence of tomatoes because, failing to categorize them appropriately, "she doesn't know what to do with them".

The quote given by Zimbardo is heavily focused on behaviour and might, when taken by itself, also allow the construal that the categorizing event of perceptual recognition immediately serves behavioural output, without the mediation of inter-representational transitions. Zimbardo describes categorization as enabling one to learn from past experiences of certain situations how to react effectively to future situations of the same type. This, taken by itself, is consistent with the possibility that the classificational event of recognition initiates these reactions without exploiting knowledge structures embedding the category representation. This is not what Zimbardo envisages, however. And indeed, more generally speaking, once one shuns a behaviourist picture by making the representationalist commitment of positing postperceptual categorizing representations, then it is hard to see why one should want to assume, for the general case, that these representations directly cause motor responses to their instances in perceptual encounters.

2.6. Not a criterion of concepthood as such. I hope the preceding distinctions and explanations have shed light on the PPC notion of a concept and on the cluster of notions surrounding it. But we are still in the business of considering whether there is some notion of a concept that provides us with a further untendentious, characteristic feature of concepthood than the memorization criterion by which concepts have been set off against other mental representations in chapter 3. We have seen the inadequacy of the idea that a mental representation's multiple applicability might be criterial of concepthood. Would the idea of a mental representation that results from postperceptual feature abstraction be any more viable? The idea, appropriately enough, presupposes the feature of memorization, as we have seen.

But the way in which the idea goes beyond that make it dubious as a general criterion of concepthood.

There are two respects in which the PPC notion of a concept—the notion of concepts as mental representations which classify across different percepts—carries a too heavy theoretical burden, when considered as an untendentious criterion of concepthood. Both of these respects are straightforward.

Firstly, the PPC notion is restricted to a certain subclass of concepts, pending a highly non-obvious extension of the idea. We are certainly still putting matters mildly if we say that there are many kinds of concepts which do not obviously result from a feature extraction from perceptual representations. There are countless concepts of unobservable concrete objects, concepts of abstract entities, or concepts corresponding to evaluative and normative terms, whose reduction to postperceptually categorizing concepts would constitute *a highly substantial theoretical feat*. One might, in light of this, say that the PPC notion of a concept only captures one type of concept, namely, unsurprisingly, that of concepts of perceivable objects or properties. But this would be to change the subject, for we were after *criteria for concepthood as such*. On the other hand, one might say that the PPC notion of a concept is indeed the basic notion of a concept and that any concept must ultimately allow an analysis in its terms. The commitment would be one to a basis of perception-based concepts to which the PPC notion straightforwardly applies, with all other concepts being ontologically dependent on these basic concepts, being, in some sense (see chapter 6, and partly, chapter 7) constructed from them. What this would be, in other words, is the extremely substantial commitment to an *across-the-board concept empiricism*. Now, as far as I can tell, I happen to have stronger sympathies with this idea than most philosophers do. But to offer this idea as an untendentious *starting point* to mark off the realm of the conceptual from the nonconceptual is not an option.

Secondly, the PPC notion fails to constitute a viable option for delimiting concepthood insofar as *concept atomism* is accepted as a contender in the arena of concept-theorizing. Concept atomism does not simply deny to concepts constitutive relations to perceptual representations; it claims in all generality that very many concepts (typically, concepts expressed by words, see chapter 6)[9] do not entertain any constitutive relations to other mental

---

[9] Concept atomism is only claimed for "lexical concepts", i.e., concepts expressed by syntactically simple expressions. This characterization is deplorably language-relative and loose, not least because you can coin a word for any concept you like. But Fodor

representations at all; in other words, that there are very many concepts which are representationally primitive tout court. This is a minority view, not without reason. It has hardly even been in the purview of discussions about concepts before Jerry Fodor openly endorsed it.[10] But it is generally accepted to be an initially coherent option in concept-theorizing—there is not even a generally accepted story among its opponents as to wherein exactly its incoherences lie. I will not begin to go into a critique of concept atomism within this work, but I take it to be a clear rule of the game that we cannot exclude atomistic views of concepts by fiat of our characterization of concepts as such.

---

has assured me in the past that the language-relative characterization is only a heuristic device for him. What the thesis is at root about is how large the basis of representationally primitive concepts is.

[10] The idea dates back at least to Fodor et al. (1980) and Fodor (1981), but it was not until the very explicit endorsements of concept atomism in the book-length treatments of Fodor & Lepore (1992) and Fodor (1998-a) that it began to have somewhat wider reverberations in the philosophical community.

§5
The flexibility of conceptual states:
How concepts transcend hardwiredness

_____

_____

This chapter looks further into the issue of marking off concepts from non-conceptual mental representations. There is more to be said about this on the positive side than we have said so far. The essential embedment of concepts in higher, more sophisticated cognitive capacities has a broader basis than that of storage in memory which was carved out in chapter 3. I want to make clear that concept-involving capacities are also more sophisticated insofar as they transcend the rigid, hardwired nature of sensory and other non-conceptual states, and I will try to make precise in how far they do.

1. Flexibility versus hardwiredness

Sensory processes are obviously rigid, hardwired processes; their occurrence in the presence of a specific stimulus is determined by a fixed biological pathway (a neural pathway leading from the stimulus receptors to the brain). Conceptual processes, by contrast, are not.

In what follows, I want to distinguish two aspect of hardwired cognitive processes: firstly, the *mechanicity* of hardwired processes and, secondly, the *cognitive impenetrability* of the cognitive states embedded therein. This prepares my ensuing focus on the respectively opposite features of concept-involving processes: their *nonmechanicity* and their *immediate cognitive modifiability*. I hope that, in addition to the more basic prerequisite of stimu-

lus-autonomy, these two aspects adequately capture the basis of conceptual flexibility.

1.1. The hardwired nature of sensory processes. Looking at sensory processes, to start out with, it helps to notice a fundamental difference between the manner in which false beliefs usually occur and that in which erroneous sense-dependent states do. False beliefs are far more ordinary occurrences than erroneous sensory states are. Our concepts constantly get deployed in false beliefs, tentatively or less tentatively held. This is not the case with sensory states. Someone searching for bona fide examples of erroneous sensory representing must resort to truly extraordinary cases. For example, if you are looking at something and all of a sudden you see a big patch of black spread across your visual field, even though there is nothing black to perceive in your visual field, then something must be seriously out of order with your visual system.

It is in some sense intuitively clear that the altogether different degree of abnormality that attaches to a sensory error (a sensory misrepresentation) in contrast to a doxastic mistake (a false belief) is that *some mechanical malfunction* occurs in the first case but not in the second. That is, in the case of a visual-perceptual error such as the one mentioned, some part of a fixed mechanism of intercellular activation has locally broken down, and now the visual system is not delivering all the right representations of the visible properties in the appropriate part of your visual field that you should be getting. I will try to get at the nonmechanicity of concept-involving states more precisely in the following subsection. Right now, let me continue by conveying the opposing intuition concerning the corresponding concept-involving case of a false belief: In the case of misbelief,[1] no mechanical malfunction is usually involved and, in very many cases, not even in any wider sense can malfunctioning be properly said to be involved. Many false beliefs are blamelessly, sometimes even justifiedly held. And where false beliefs are unjustifiedly held, it is some less-than-optimal (though not even necessarily less-than-average) performance at an intuitively nonmechanical, psychological level of description that is responsible: weaknesses in someone's reasoning abilities, deficits in memory, wishful thinking, and similar kinds of factors.

---

[1] An important difference which I wholly leave aside here is the fact that beliefs admit of degrees of strength with which they are held. A sense perception, by contrast, does not admit of any corresponding variation along such a dimension of "assertive force", so to speak.

Lest we forget, there also exist erroneous sense-dependent representational states which do not result from breakdowns in the neural circuitry. I am alluding to the case of perceptual illusions.[2] But perceptual illusions only complete the picture begun with the observation just made. Whereas pathological occurrences of sensory errors evidence, ex negativo, the hardwiredness of veridical processes of sense perception which are encountered across the members of a species, perceptual illusions evidence the hardwiredness of certain *non*veridical processes of sense perception uniformly encountered in a species. The situation with perceptual illusions is familiar. We might put it this way: One cannot but represent a stimulus array *a* to be *F,* even when one has conclusively falsified the impression that *a* is *F*. For example, one cannot but represent the parallel lines of the Müller-Lyer illusion to be of unequal length, even after one has measured them for oneself, coming to know that they are in fact of equal length. We are incapable of *correcting* this perceptual information we seem to be getting in light of what we know.

The property which is nicely illustrated in this way by perceptual illusions has been described as that of the *cognitive impenetrability* of sense processes (Pylyshyn 1980; Pylyshyn 1984: 130-145) or, equivalently, as that of their "informational encapsulation" (Fodor 1983: 64-86). Cognitive impenetrability forms another aspect of the rigid, hardwired nature of sensory-perceptual processes. It contrasts with the *immediate cognitive modifiability* of concept-involving states, in particular, of beliefs. We constantly correct previously held beliefs in light of new beliefs acquired.

The hardwired nature of sensory processes is also illustrated by the fact that their operation, once initiated, is *mandatory:*[3] Once our sense organs are directed upon the sorts of stimuli whose representation they are keyed to, we cannot exercise any volitional control over whether we go into the appropriate sensory-perceptual representational states or not: It happens to all functioning members of our species, whether we like it or not. Notice, for example, that this is the reason why we are susceptible to physical torture. When our pain receptors are stimulated, we undergo pain experiences independently of whether we like it or not. I suspect that we can regard this as something of a volitional equivalent to cognitive impenetrability. While the operation of sensory processes is not volitionally penetrable, the operation of

---

[2]   Perceptual illusions typically involve the erroneous attribution, at a relatively late stage of stimulus-processing, of some relational property or other to a perceived array of stimuli.

[3]   I am recruiting here another term from Fodor's already quoted essay *The Modularity of Mind* (1983: 52-55). The basic observation, however, ranks as a commonsense platitude.

conceptual processes is, at least to an important extent. We need not go into this aspect further in the context at hand, though.

Let us now take a closer look at the two central aspects of the flexible nature of concept-based cognitive states, as already distinguished: their non-mechanical occurrences and their immediate cognitive modifiability. I address these aspects in the order of mention.

1.2. Nonmechanicity of concept-involving processes. To be sure, for many concepts of perceivable properties and objects, there may well be a bona fide hardwired causal pathway from percepts of those entities to the activations of those concepts: It seems quite plausible that our concept of a dog is invariably tokened when we entertain a perceptual representation of a creature looking the way that dogs typically do. But this is only a small part of the truth about concepts, even about concepts relatively close to perception. Even for concepts to whose appropriate activation there does exist some hardwired path or other, such a path always only remains one of many possible paths. A concept is capable of being appropriately tokened as a consequence of a potentially infinite variety of different inferential processes, and it can itself enter into causal interactions with a potentially infinite variety of different representational events. *Its occurrence is not rigidly embedded in a specific causal pathway.*

If you have a concept of a dog, you can invariably think of dogs as a consequence of perceiving dogs, but there are also countless other chains of thought and contexts within which dog-thoughts can be made to pop up. For example, you might think of dogs as a consequence of being asked to think of things that are referred to by expressions beginning with 'd'. Or you might be stuck in the middle of a snow-covered landscape in Alaska with a sled and have the thought that it would be useful to have some dogs to pull your sled. And when you think of dogs, the dog-associated knowledge which you exploit in the respective cases can differ from case to case, just as the ensuing paths of thinking which are partially caused by the tokening of your dog-concept also can differ from situation to situation.

A common way of describing this nonmechanical property of thought processes is to say that we are dealing here with processes that are *sensitive to the contents* of the states involved, or, as one might say, that the processes are "semantically governed". When philosophers talk about types of causal processes and describe something as "mechanical", then this is usually the complementary class that they intend to set off mechanical processes against.

It is high time to make the notion of mechanicity more precise now, and I will try to do so before elaborating on the notion of a semantically governed process. We have been heavily relying on the notion of mechanicity so far. One might aptly wonder, demanding an explanation, what it would be for a causal process not to count as "mechanical". After all, it is not a particularly contentious thing to say that any processes in the causal order of the world are ultimately decomposable into events that fall under the laws of physics. If this is so, then they are ipso facto mechanical, on one natural understanding of 'mechanical'. Fair enough; but to see the philosophical point of ascribing mechanicity to processes, what one must take into account is that such ascriptions are always made *relative to a given level of description.* Ascriptions of mechanicity are typically made with a higher level of description in mind than that of physics. Now, what lies behind the ascription of mechanicity to a process is the assumption that, *at the level of description in question,* the process can be seen to consist in *a sequence of distinct steps* each of which effects the ensuing event in the causal chain *as a matter of lawlike necessity.* As soon as we gain this perspective on a process, we are prepared to call the process 'mechanical' (by the way, this fairly well mirrors people's readiness to call a process 'physical'). To put it concisely, processes which we term 'mechanical' are ones that are *recognizably deterministic.* To the extent that such deterministicness is not apparent at a given level of description, one refrains from attributing mechanicity to the process. The hardwired process of stimulus transmissions through fixed intercellular constructions is a fair enough example of a mechanical process.

Given this clarification of the notion of mechanicity, it is easier to explain what it is to say that high-level cognition (reasoning, thinking) is a semantically governed process. The point is that, *at the semantic level* of describing cognitive processes, there are no laws of nature that would allow one to *causally explain* why one thought leads to the next, let alone to *predict* which thought leads to which. Perhaps this is not surprising. Perhaps it is to be expected from the fact that causal powers reside in the local, intrinsic properties of causal relata, while intentional properties, by contrast, involve relations to distant objects and properties.

Another noteworthy contrast pair corresponding to that between mechanical and nonmechanical, semantically governed processes is this. When we are talking about semantically governed processes between representational states, we regard the transitions between these states as *inferences* rather than as *merely causal* transitions. Furthermore, the contrast between mechanical processes and semantic processes corresponds to the contrast be-

tween *mechanicity* and *rationality*. Rational processes are ones that lead an organism from true beliefs to other true beliefs, or from true beliefs and appropriate volitional states to other true judgments, and ultimately to actions that have a utility for the organism. Our ascriptions of semantically governed processes are largely guided by considerations of rationality. There is substantial disagreement among philosophers as to how much rationality is required for intentional ascriptions, and as to what modal status the tie between intentional explanations and assumptions of rationality occupies.[4]

What at least seems undeniable is that the point of an organism's being capable of entertaining semantically governed processes lies in these processes being rational ones. And it is hard to see how we could possibly dispense with very wide-ranging presumptions of rationality in our psychological explanations other people's behaviour. Suppose that Hans needs to travel abroad in order to attend a conference. He packs his suitcase, calls a cab to his place, goes downstairs, waits for the cab to arrive, mounts the car, and asks the cab driver to drive him to the airport. You do not explain this by attributing to Hans the desire to smoke a cigar and the belief that seven is a prime number. Rather, in the absence of specific pieces of evidence to the contrary, you explain this by assuming that Hans had the desire to attend the conference, by assuming that he believed his actions to be required in order for him to attend the conference, by assuming that these beliefs by Hans made him form intentions to realize those actions, and by assuming that he successively performed these actions as a consequence of having these intentions. Other familiar contrast pairs come up in the context of the opposition between purely semantic explanations and mechanistic explanations, most importantly perhaps the distinction between *reasons* for forming a belief or performing an action, and (again) mere *causes*. We need not delve into this family of notions here further.

But let me still mention one question which I feel is bound to come up here. Given what I have just said about the contrast between semantically governed processes and machinelike, mechanical processes, and given a location of concept-involving representational states on the former side of this divide, one might wonder what place is left in this picture for nonconceptual representational states and processes in the first place. One would seem to be in trouble if one wanted to hold onto a belief in nonconceptual representations in spite of taking nonconceptual representations to be located at a mechanical level of nature—in fact, even *necessarily* located at such a level, if what we are aiming at in the present context is correct. For a state to be

---

4 See, for example, chapters 1-3 of Stich (1990).

located at a mechanical level would seem to entail that explanations involving such a state precisely *need not* take recourse to the semantic level of description. Whereas one would think that *representational* states are ipso facto states of organisms that must be semantically described in order to capture relevant generalizations.

I think such a worry would have to be defused as follows. Most fundamentally, even where representational processes are directly mirrored by processes that are biologically determined, one cannot describe these processes in a way that is explanatorily relevant to the organism's behaviour (or potential behaviour) in its environment if one refuses to describe them representationally. One would miss the explanatorily interesting point of having the states in question if one refused to describe them with reference to the properties and relations whose presence they indicate. Relatedly, one would not even be capable of capturing the relevant generalizations, since for any description of a fixed representational pathway in terms of the specific neurobiological components at work therein, there would be numerous alternative descriptions in terms of other implementations. The fact that a given, hardwired sequence of biological states mirrors a nonconceptual representational process within one species does not mean that it must do so in other species too. This is what you would expect if the representational description captures explanations which intrinsic, mechanical descriptions do not. To conclude the present point: The embedment of a state in purely mechanical push and shove need not count against that state being a bona fide representational states.

1.3. Immediate cognitive modifiability of beliefs. Cognitive states involving concepts are cognitively modifiable. They are permanently capable of being revised and reshaped in light of other beliefs we have or acquire. The ubiquity of belief revision by itself already contrasts with the fact that it is in general impossible to control what sensory-perceptual representations one entertains when one's sense organs are active. But the contrast with the hardwiredness of sensory states goes deeper. What we are talking of here, more specifically, is the fact that beliefs are *appropriately revisable as a consequence of being in other cognitive states*. When I have perceptual representations of redness in seeing something red, I cannot make myself token representations of blueness instead. By contrast, when it comes to the belief that grass is red, this is something which I can abandon without much ado (taking into account obvious evidence to the contrary) in favour of the belief that grass is green.

Relatedly with their modifiability by cognitive states, beliefs are modifiable *immediately*. If you sincerely claim in Laurie's presence that she is 20 years old and if she corrects you, pointing out that she is 21 years old, then you can modify your belief about Laurie's age within a split second of having understood the utterance by which she corrected you.

Some form of modifiability (or "plasticity", as it is also often called) is not too difficult to find in the neural wiring underlying nonconceptual representing. Many sensory neural pathways, and thus sensory representations, are capable of reforming after crucial connections have been previously severed. But they do not get reshaped as a consequence of one's being in certain other cognitive states. And a gradual neural rewiring of cognitive states over time, whatever its causes, forms a completely different kind of process from that of instantaneously correcting assumptions in one's web of belief.

1.4. Holistic sensitivity of modifications of beliefs. There is still a crucial feature of concept-involving cognition to be mentioned, one which we can call its *holisticity*. It is for good reason that we turn to this last. For the feature of holisticity in question *presupposes* cognitive modifiability. It is supposed to be a property of the way in which cognitive modifications of concept-involving states proceed—paradigmatically: *belief revisions*. The holisticity of cognitive modifications of concept-involving states also provides particularly clear cases of nonmechanical processes. The holisticity of concept-involving cognition surely forms a great part of what theorists are impressed with when they emphasize the nonmechanical, semantically governed nature of reasoning (thinking, belief revision).

What does the holisticity of reasoning consist in? The question is perhaps best answered by drawing together some illuminating observations from part IV of Fodor's *Modularity of Mind,* which forms a locus classicus in the foundations of cognitive science when it comes to higher-order cognition and holistic processes. Taking the main job of higher-level cognitive systems—or "central systems"—to consist in the nondemonstrative formation of beliefs,[5] Fodor likens this activity of central cognitive systems to the nondemonstrative fixation of beliefs in science. He distinguishes two interesting properties in the confirmation procedures of science. The first of these he calls the property of *being isotropic*. The second of these he calls the property of *being Quinean* (in reference to Quine 1951). The two properties are intimately related. Both constitute paradigmatically global, holistic properties of confirmation procedures.

---

[5]   "Nondemonstrative" inferences can simply be understood to be inferences that could possibly fail to lead from true premises (beliefs) to a true conclusion (belief).

The term 'isotropic', taken literally, designates an object's property of exhibiting identical physical magnitudes in all directions of measurement. Fodor appropriates the term to designate an hypothesis' property of being confirmable and disconfirmable by beliefs which can be drawn from anywhere in the embedding belief system. As he puts it, confirmational isotropy obtains just in case "the facts relevant to the confirmation of a scientific hypothesis may be drawn from anywhere in the field of previously established empirical (or, of course, demonstrative) truths. Crudely: everything that the scientist knows is, in principle, relevant to determining what else he ought to believe." (1983: 105.)

By the property of being "Quinean", Fodor means "that the degree of confirmation assigned to any given hypothesis is sensitive to properties of the entire belief system" (107). An additional specification is required to express the essential idea here. What Fodor has in mind by those properties are the metascientific properties with which in view we select between conflicting, equally data-compatible hypotheses: properties such as simplicity, conservatism, and perhaps a few others. The point here is that these properties are ascribed to whole webs of assumptions, not simply to isolated hypotheses. If, for example, we had a generally accepted method of measuring simplicity then our assignment of a certain degree of simplicity to a system of belief could only be reached at by means of a global consideration of the system, rather than by assigning degrees of simplicity to each of the system's beliefs in isolation and adding those measurements up.

Fodor wants to keep the properties of being isotropic and of being Quinean separate. The rationale behind this separation is best summarized as follows: The claim that confirmation is isotropic says something about the set of assumptions that form the *potential provenance* of a target belief's confirming or disconfirming assumptions. The claim that confirmation is Quinean says something about the set of assumptions with respect to which a target belief's *degree of confirmation* (or disconfirmation) is evaluated. In both cases, nothing less than one's *whole* system of assumptions is claimed to form the relevant set. This is what the holistic character of hypothesis formation and revision in science consists in. Assuming that confirmation procedures for scientific beliefs are representative of confirmation procedures for beliefs in general, we have in our hands a description of the holistic character of belief formation and revision in the general case.

The holisticity of intelligent processes is nicely illustrated by the observations underlying the most notorious problem in the philosophy of artificial

intelligence, the so-called *frame problem* of AI.[6] Let me try to give a rough, generally applicable description of what the frame problem is. It is the problem one confronts in trying to bring about that artificial agents, or the programs underlying them, speedily take into account numerous entirely obvious consequences of the actions whose performance they are considering, and equally obvious collateral conditions of new situations which they are placed in. Accomplishing this sort of task is much more difficult than is prima facie suggested by the cognitive effortlessness with which we actually *do* take such consequences into account. I will briefly try to explain a bit further.

Let me borrow the example from Dennett (1984: 181-82) of a robot whose only objective it is to survive and which learns that its survival-relevant spare battery is locked inside a room with a time bomb attached to it. The robot knows where the room is, and it knows where the key to the room is. It also knows that the battery is placed on a wagon inside the room. It hypothesizes that the action of pulling out the wagon from the room will remove the battery from the room. The robot successfully performs this action. But the time bomb is still attached to the battery, so the robot pulls out the battery complete with the time bomb, and the battery explodes. The important point of this rudimentary example is that the robot actually *knew* that the time bomb was attached to the battery on top of the wagon. The robot just failed to infer all of the relevant consequences of the action it planned. What the robot deduced (correctly enough, to be sure) was merely that the battery would be pulled out of the room which contained the bomb. But the robot failed to deduce that the timebomb would be pulled out of the room together with the battery.

Now, what the designers of such a robot have to do in order to make it improve upon such stupid behaviour is to make the robot deduce not only the immediately intended consequence of its planned action, but also the *collateral consequences*. But then the problem arises how to make the robot dedi-

---

[6]   The connection is particularly stressed by Fodor (1983: 112-117). The term "frame problem" was coined in a paper by the two influential AI researchers John McCarthy and Patrick Hayes (McCarthy & Hayes 1969). Dennett's 1984 paper "Cognitive wheels: The frame problem of AI" provides the most prominent attempt by a philosopher to get clear about what the problem is. For further reading, see the collections edited by Pylyshyn (1987) and by Ford & Pylyshyn (1996), both specially dedicated to this problem.

        The frame problem got its name from the "frame axioms" that were programmed into so-called planning systems (the programs supposed to simulate artificial agents). These frame axioms constituted background assumptions (in addition to axioms/assumptions about the situation at hand) which, as Dennett puts it, "describe general conditions and the general effects of every action type defined for the system". (1984: 195.)

cate its processing capacities to the deduction of the *relevant* collateral consequences of its planned action. The most central aspect of this problem is that there are innumerable irrelevant consequences that must be *ignored,* such as the fact that pulling the wagon out of the room would not change the colours of the room's walls, or the fact that the wagon's wheels perform less than a thousand revolutions when the wagon is pulled out. Once this challenge is faced, the problem poses itself how to design a thinking agent that *swiftly and efficiently identifies the relevant implications* of its considered actions in contrast to the irrelevant ones. One cannot make a robot go through the potentially infinite number of consequences of its planned action in the way it might go through a list, having it decide for each of the consequences separately whether they are relevant or irrelevant to achieving its aim. This, then, is the frame problem. One might also call it the problem of *updating one's belief system* in light of new descriptions of the world. Not too surprisingly, given what I've just said, it is sometimes also called the "problem of relevance".

Given the holistic nature of belief fixation, the cluster of problems known as the frame problem is quite unsurprising. The frame problem arises insofar as *the potential range of the revisions within one's belief system* which a new piece of knowledge may require is in principle *unlimited,* and this is precisely what Fodor means by the "isotropic" nature of belief revision. The related property that belief revision adheres to holistic confirmation principles (its property of being "Quinean") similarly serves to explain how the frame problem could arise.

The involvement of mental representations in cognitive states that are holistically sensitive to modification by other cognitive states probably constitutes the paradigm of nonmechanicity in cognition. The alternative description of nonmechanicity as a cognitive process's feature of being semantically governed fits together with the feature of holistic sensitivity. After all, no principle other than that of following relevant semantic relations is recognizable behind the holistic sensitivity in question.

We can now return to what I said in chapter 3 about concepts being generally acknowledged to be mental representations of a more sophisticated sort than nonconceptual representations, the possession of concepts being intimately connected to *intelligence*. It is noteworthy that, in some foundational remarks made with regard to cognitively holistic processes, Fodor emphasizes precisely the sort of connection between holisticity and "paradigmatically intelligent" processes which we have been leading up to, and that he

does so in the same stroke as he denies a mechanical ("reflexive", "routine") nature to the processes in question:

> ...the closer we get to what we are pretheoretically inclined to think of as the "higher", "more intelligent", less reflexive, less routine exercises of cognitive capacities, the more such global properties as isotropy tend to show up. I doubt that this is an accident. I suspect that it is precisely its possession of such global properties that we have in mind when we think of a cognitive process as paradigmatically intelligent. (Fodor 1983: 107.)

1.5. Conclusion. It seems that we have arrived at a fairly informative and discriminating understanding of what was meant earlier by the intuitive location of conceptual representations on the high grounds of intelligence. As usual, the term 'intelligence' had the import of dignifying a cognitive process in ways that are difficult to make precise. Looking back now, much of this precisification seems to have been achieved. Back in chapter 3, the universally acknowledged assumption was made explicit that concepts are not stimulus-dependent, but appropriately tokenable in the absence of their representata, and that they must therefore be mental representations stored in a memorylike unit. The criterial considerations offered in the present chapter do not call this foundation into question. They build upon it. It is psychological states involving stimulus-autonomously tokenable representations that are assumed to be cognitively modifiable by each other, and it is their occurrences which exhibit a highly nonmechanical nature, culminating in holistically sensitive nonmechanical cognitive modifications.

Acquiring stimulus-autonomous representational states can be analytically distinguished as the most basic step in an organism's development away from a purely receptive, mechanical embedment in its surroundings. But if concepts are mental representations that occupy a more exclusive role than the generic role of mental representations, this is also in virtue of the ways in which they transcend the limitations of hardwired, rigid processes, in addition to the simple fact that they transcend the limitations of stimulus-dependent representing. While stimulus-autonomy enables a system to entertain at least minimal forms of representational autonomy from its surroundings, the aspects of flexibility layed out in the present chapter provide the basis of a system's becoming an altogether autonomous cognizer and agent, a truly *thinking* organism with concepts at its disposal.

2. Learnedness versus innateness
is not the issue

Relatedly with hardwiredness, one could think that processes involving nonconceptual states are innately determined. Opting for innateness as the basic criterion for distinguishing the nonconceptual from the conceptual is a path taken in at least the more recent work by Fred Dretske (1995: 15 and 19). As Dretske says, he takes the contents of sensory/nonconceptual states to be "phylogenetically determined", or "systemic", while the contents of conceptual states are "ontogenetically determined", or "acquired". This is, in other words, the contrast between innateness and learnedness.[7]

The tie between concepthood and flexibility (non-hardwiredness) differs in modal status from the tie between concepthood and learnedness (non-innateness). While being hardwired is essential to the privileged class of states that representationalists aim at when they posit concepts, being learned is not. On the following pages, I will make some largely critical points about the relations between the sensory/conceptual-distinction, the innate/learned-distinction, and the hardwired/flexible-distinction.

To begin with, how does *innateness* relate to *sensory statehood?* It seems clear enough that our potential repertoire of sensory-perceptual states is innately determined. Sensory representational states are innate because they depend on the sense organs, and the sense organs are obviously innate. They are, after all—like any other organs, presumably—fixed biological structures that we are born with. This is not to say that their development does not unfold or develop after birth (which it obviously does). The point is that this process of unfolding itself is genetically predetermined, much as, e.g., the maturational changes of puberty which humans undergo more than a decade after their birth are predetermined. Eyes, for example, complete with the retina, its two types of photoreceptor cells, the layers of neurons below and the neural pathways leading from there to the visual cortex, need not be acquired by individuals of our species in the course of growing up, on the basis of any individual history.

However, contrary to the criterion of division provided by Dretske, the link between sensory representing and innateness is a contingent one. The property of innateness does not capture even part of the essence of sensory representing. It is coherent to imagine a creature that acquires sense organs

---

[7] The term 'learned' is used here in a theoretical, wider way, according to which it subsumes ontogenetic acquiredness by noncognitive means no less than ontogenetic acquiredness by cognitive means ("learnedness" in the folk sense).

complete with sensory transducers by genetically non-predetermined means. Simply imagine an organism which, within its lifetime, recapitulated the evolutionary process that natural selection required hundreds and thousands of successive generations to perform. Such an organism would have to be endowed with an amazingly malleable physiology and cellular structure. Having an amazingly long lifetime would surely help too. But, though it is highly unlikely that organisms like this should occur in nature, there is nothing incoherent about them.

That sensory states are only contingently innate should not seem surprising given the basic point of the contrast described in chapter 3 between the sensory and the conceptual. Sensory states are states that are involved in sensory processes, and the elementary characteristic of sensory processes is that they fulfill the function of stimulus pickup. To require anything more by way of a sufficient condition for sensory statehood would be implausible. But this criterion of sensory statehood is logically compatible with ontogenetic acquiredness. Since sensory representational states are nonconceptual if any representational states are, nonconceptual representational statehood is compatible with being ontogenetically acquired, or "learnt".

Conspicuously enough, even if we ignore thought experiments showing that innateness is not necessary for sensory statehood and, hence, that learnedness is *not sufficient* for concepthood, learnedness when considered all by itself can hardly seem to be more plausibly viewed as a *necessary* condition for concepthood. To be sure, it is a central fact about concepts that very many of them have to be learnt. This contrasts with how we came to have the sensory repertoire that we do.[8] But it is generally taken for granted by philosophers of psychology that innate concepts are at least possible, and many believe that quite a number of innate concepts do in fact exist (hot candidates being concepts the possession of which had a selectional advantage for a species). Debates about concept innateness turn around such questions as the question of how many of our concepts are plausibly innate. The coherence of the idea of innate concepts as such is presupposed in such debates.

Of course the mere fact that compatibilism about concepthood and innateness is generally taken for granted does not already constitute an argument

---

[8] Learnedness as an at least frequently encountered feature of concepts occupies the status of a general constraint upon theories of concepts even in the work of Jerry Fodor (1998-a: 27), whose extreme endorsement of concept innateness is notorious.

against the proposal that learnedness might be necessary for concepthood.[9] But a great burden of argument is on the side of somone who denies the compatibility of concepthood with innateness, and, as far as I can see, there simply is no reason to regard learnedness as necessary for concepthood. In particular, a potential impression to the contrary would appear to be based a fully explicable error: Such an impression, I submit, would be traceable to the confusion of being non-innate (learned) with the closely related feature of being non-hardwired (flexible), which, as argued, truly *is* essential to the representationalist's privileged notion of concepthood.

I think it could be argued, as regards Dretske, that an intention to give flexibility its due provides the tacit driving force behind his criterial connection of concepthood to learnedness. To establish this appropriately would require going through some of the more difficult passages of his work (in particular, Dretske 1988: ch. 4), however, and I will not do that here.

I do, however, still want to offer the following consideration to the effect that learnedness is plausibly not even a necessary condition for concepthood. The earlier layed out aspects of flexibility are of *intrinsic causal-explanatory value*. They had better be so, for this metatheoretical characteristic is certainly a prerequisite for something to be a candidate criterion of concepthood. But having been learnt, having been ontogenetically acquired, does not constitute such a property, in and of itself. It merely tells us something about the historical origin of a representational state. Learnedness can only be *of derivative interest* for someone who looking to find a criterion whose power of distinction captures a causal-explanatory difference in the here and now. It is hard to see how this derivative interest in the property of being ontogenetically acquired could rationally owe itself to anything but an intrinsic interest in the property of being *flexible* in something along the lines I have spelled it out in this chapter. In particular, it would seem to be derivative upon an interest in the property of being immediately cognitively modifiable.

It is understandable how one might be led to assign to the innate/learned-distinction the concept-demarcating role which would really have to be occupied by the hardwired/flexible-distinction. The grain of truth involved is that for a cognitive process to be innately determined indeed entails that it is

---

9   The denial of innateness compatibilism need not even be as radical in its implications as it might at first seem. Perhaps a discovery that the notion of innate concepts is incoherent would merely call for the need to rephrase questions couched in terms of the innateness of concept-involving structures in terms of the innateness of nonconceptual representational structures, with just about everything else remaining equal.

hardwired. But the connection does not go beyond this, and I still want to remark on this in closing.

Firstly, and conversely, hardwiredness can be seen not to entail innateness, even when viewed more generally than from the perspective of the hardwiredness of sensory states. In fact, in contrast to thought experiments about ontogenetic acquisitions of entire sense organs, the mere idea of some cognitive process becoming hardwired into an individual organism over time is anything but outlandish; it rather would seem to be actually realized more or less often. It will not be necessary to go into this further here.

Secondly, though I have just granted that the innateness of *cognitive processes* entails the hardwiredness of these processes, notice that I have refrained from saying that the innateness of a given state of *concept possession* entails the hardwiredness of the processes involving that concept. I have refrained from saying this because indeed it does not follow. But this is precisely what would be relevant in attempting to characterize concepthood. Let me explain. We need to keep a clear head about the types of entities that innateness and hardwiredness are ascribed to. Hardwiredness is a property that can, strictu dictu, only be ascribed to *processes*. Saying that a state is hardwired can sensibly only be shorthand for saying that the state is part of a hardwired causal chain of states. Innateness, on the other hand, is a property which is sensibly ascribable *both* to states (e.g., of concept possession) *and* to processes (e.g., of concept-involving reasoning). One can have the innate endowment to token a concept in certain situations, just as one can have the innate endowment to entertain a specific sequence of cognitive states. Now, if a cognitive *process* is innately determined, then it is ipso facto biologically hardwired. However, if a cognitive *state* (a concept) is in one's innate repertoire, what follows from this concerning the state vis-à-vis hardwiredness? Not, of course, that the cognitive state itself is hardwired, since only processes could count as hardwired. What I want to point out now, however, is that the innateness of a cognitive state does not even entail that it is limited to occurrences within hardwired cognitive processes. For example, maybe the concept CAUSE is innate. But it would be absurd to claim that all (or even most) of our reasoning about causes consists in innately determined, hardwired routines.

There is, in conclusion, not even a single direction of entailment in which an innate/learned-distinction potentially relevant to the C/NC-distinction could permissibly be thought to coincide with the hardwired/flexible-distinction which lies at the heart of the C/NC-distinction. I have claimed that innateness owes any initial appeal it may have as a criterion of the nonconceptual

to the logical bond which innateness exhibits with hardwiredness in virtue of innate processes being ipso facto hardwired processes. But, quite apart from the fact that the converse of this implication does not obtain, the implication in question misses the issue which would be aimed at in an innateness-based attempt to capture the nature of the C/NC-distinction. The relevant issue would be that of what is entailed by the innate possession of a representational state. And the innate possession of a representational state does not entail its confinement to hardwired processes.

Bear in mind that these critiques come in addition to the general points already made before: The innate/learned-distinction does not have the intrinsic causal-explanatory relevance which is desirable for the purpose of grounding the C/NC-distinction. The hardwired/flexible-distinction does, and it has ample independent motivation. I conclude that we can rest content, for purposes of precisifying the C/NC-distinction, with the results obtained by the end of the previous subsection.

§6
Concept structure, prototype theory, and compositional phrasal compounds

---

1. Concept structure
    1.1. Hypothesized concept structure
    1.2. Concept definitionism vs. concept probabilism
    1.3. Supervenience of prototype representations on
        supra-constituentially gradable structures
2. The compositionality arguments against prototype theory
3. The Scope Assumption about prototype theory
4. On admissions of noncompositionality
5. The confusedness of the debate about
  prototype theory and compositionality
6. The obvious definitionality of compositional conjunctive compounds
7. On the very constituent structure at issue
    7.1. On incoherently shifting the constituent level at issue
    7.2. A tacit notion of non-preservation of compositional determiners?
8. Concept definitionism, properly generalized, and a generalized
  refutation of Compatibilism
9. The error of having prototype tails wag the dog of probabilistic structures

---

The nature of concepts as subpropositional mental representations of the sort clarified up to here is the subject of a lively subdiscpline on the borders of philosophy and cognitive science. The vast majority of theorists of concepts in cognitive science heavily rely upon some variety or other of the idea that concepts have probabilistic structure. At the same time, there is a widespread perception—strongly promoted by Jerry Fodor and expressed in two well-known arguments within his work—that concept probabilism, or "prototype theory", faces serious problems in doing justice to the requirement of compositional concept combination.

This chapter aims to correct several elements in what I argue to be a confused picture that is shared by all sides in the debate about compositionality and prototype theories of concepts. It aims to establish some interlinked ne-

cessities immediately following from notions which must be operative by assumption in the debate about prototype theory and compositionality, but whose open recognition has been egregiously lacking in the field so far. The dialectical strategies performed within this debate in the foundations of cognitive science rest on confused presuppositions.

Foremost among these confused presuppositions, I argue, is the idea that prototype theory might coherently be taken to stretch to actual compositional phrasal expressions and their corresponding phrasal concepts in the first place. Actual compositional compounds preclude the very idea of a supra-constituential gradability of concept satisfaction which is constitutive of prototype theory, this preclusion being most glaringly obvious in the prominent case of conjunctive compounds. Complementarily, I will bring out the strangely overseen fact that actual compositional phrasal expressions are necessarily defined by their constituent structure.

The unifying theme behind this chapter's central points could be expressed thus: Once a compositional compound with a specific constituent structure has been postulated, its concept-theoretical nature at the level of the definitionism-probabilism opposition is thereby fixed, and any statement as to which of these opposing theory schemes the compound falls under can only be trivially true or trivially false. Following this insight through, the chapter's last section envisages the possibility of probabilistic phrasal expressions and rebukes the idea that such compounds could have to face any separate test of intuitive prototype assignability for their prototype-theoretical nature over and above their already ascertained type of structuredness.

It will take a little while before I get to the main contentions and develop the arguments for them. We will need to start out with some clarifications and distinctions which are vital to the discussion to follow. In the course of this, connecting to the exposition of the compositionality arguments against prototype theory, I will already be making a few substantial comments on the types of reactions encountered with respect to the compositionality arguments against prototype theory.

1. Concept structure

The point of departure for any discussion of compositionality with respect to concepts is that, mirroring the distinction between syntactically simple and syntactically complex linguistic expressions, there exist corresponding types of concepts thereby expressed.

Following widespread practice, let us call those concepts *lexical concepts* which are conventionally expressed by syntactically simple expressions within our language. Lexical concepts are the default topic in theorizing about concepts. They are what talk of "concepts" standardly aims at, in the absence of any modification. Let us define as a *phrasal concept* a syntactically complex concept that is isomorphically structured as a syntactically complex expression in our language, where the complex concept, each of its constituents, and its construction mode bear the same contents as their linguistic counterparts. A stipulatively phrasal concept is conventionally named by means of the isomorphic phrasal expression which expresses its content, suitably transposed into a concept-naming notation (the concept expressed by 'red car' is RED CAR).

1.1. Hypothesized concept structure. Against the representationalist background, the first branchings of a decision tree concerning the nature of concepts are easily identified. At the highest level of generality, theorists need to take a position on whether lexical concepts are representationally structured or not; in other words, whether their contents (by which they are individuated) are determined by a representational structure or not. *Concept atomism,* already brought into play contrastively at the close of chapter 4, denies that lexical concepts have representational structure. Mainstream theorizing about concepts, by contrast, assumes that concepts do have representational structure. To such an extent does mainstream theorizing about concepts take this generic view for granted that it lacks a label for it. Let me call the positive view that atomism is set against *concept structurism*.

When I say that a concept is "representationally structured", this is supposed to mean that it is metaphysically impossible for the concept to represent what it does without standing in certain relations to other representations which one must possess if one is to possess the concept. There are two ways of understanding these concept/content-constitutive relations and the talk of "representational structure" here. The literal understanding of 'concept structure' is the mereological understanding. On this understanding, a structured concept contains its content-constitutive representations as proper parts. The mereological understanding of concept-constitutive interrepresentational relations is that of a necessary co-tokening relation.[1] On a non-mereological understanding of concept-constitutivity, a structured concept

---

[1] See Fodor (1996: 114-115). There, as in many other parts of Fodor's work, talk of "constituents" is eo ipso taken to be mereological. But cf. Fodor & McLaughlin (1990: 93), who reserve talk of "classical" constituents for this specific understanding of concept-constitutivity.

can be tokened without its constitutive representations being tokened. The relation then will still have to be a causal-dispositional one.

For lack of any other simple term that could be used to refer to content-constitutive representations regardless of whether they are mereologically understood or not, I will, in what follows, often loosely use the term 'constituent' to speak of content-constitutive representations generically.

1.2. Concept definitionism vs. concept probabilism. There are different ways to cut the pie of concept structurism, and I believe it is by no means clear that the main positions distinguished in standard pictures of concept theorizing do in fact capture what are, at the highest level of generality, the potentially most interesting dimensions of variation. But in the decision tree suggested by much of actual theorizing about concepts, the most elementary branching within concept structurism is the question of whether to posit *definitional structures* or *nondefinitional, probabilistic structures* for concepts. This dichotomy will lie at the focus of the following investigations in this chapter. Notwithstanding a widespread habit of giving unprincipled enumerations of concept-theoretical positions, the distinction between definitionism and probabilism is an exhaustive one, relative to the postulation of representational structures for lexical concepts.

Let definitionism about a given concept *C* be the position that *C* is definitionally structured, while probabilism about a given concept *C* is the position that *C* is probabilistically structured. Definitional structures and probabilistic representational structures crucially differ in what modal status they assume the respective properties to have that are individually represented by the concept's hypothesized constituents (also called 'feature representations').

Definitionism assumes that each of a concept's constituents represents a *necessary* property of things that fall under the concept. This entails that nothing can fall under a concept *C* (nothing can satisfy *C*) that lacks any of the properties individually represented by *C*'s constituents. In other words, for an object *x* to be an instance of a definitionally structured concept, it is necessary that *x* satisfy *each* of *C*'s constituents. Moreover, for *x* to satisfy a definitionally structured concept, it is also *sufficient* that *x* satisfy each of *C*'s constituents.

The probabilistic idea of concept structure shares this semantic sufficiency claim. Concept probabilism, however, loosens the condition upon what is necessary for an *x* to satisfy a concept. Probabilism assumes that a concept's

constituents individually represent properties which things satisfying the concept are *statistically likely* to exhibit. Accordingly, *x* can satisfy a probabilistically structured concept *C* while falling short of satisfying each of *C*'s constituents. For ease of exposition, let us help ourselves to the simplifying assumption that the constituent representations of a probabilistic concept all have an equal degree of diagnosticity. We can then say that falling under a probabilistically structured concept only requires satisfying *sufficiently many* of *C*'s constituents.

Thus, each of the opposing theory schemes is characterizable by a *generic content-determining rule* for any concept that is claimed to fall inside their scope. These content-determining rules, as given below, state in the most general possible terms how the theory schemes take the content of a given concept to be determined by those of its constituents.

> **Definitionist content-determining rule:**
> An individual *x* falls under a definitionally structured concept *C* just in case *x* has *each* of the properties individually represented by *C*'s constituents.

> **Probabilistic content-determining rule:**
> An individual *x* falls under a probabilistically structured concept *C* just in case *x* has *sufficiently many* of the properties individually represented by *C*'s constituents.
> Depending upon how many of these properties *x* has, *x* is a more or less typical instance of *C*, satisfying *C* to a greater or lesser degree.

In the debate that we will be concerned with, the term 'prototype theory' is, in effect, used interchangeably with 'concept probabilism'; it is thereby used in the widest possible sense. I will follow this practice in this paper. A few words are in order to relate the notion of a prototype to the probabilistic idea just defined.

1.3. Supervenience of prototype representations on supra-constituentially gradable structures. What is meant by 'prototype' here is a "best", or "most typical" instance of a concept. The notion of a best instance is supposed to be a psychologically real one. It is supported by intersubjectively reliable judgments of typicality and several other kinds of psychological effects in categorization tasks. What counts as a concept's prototype is determined by that concept's probabilistic representational structure. A prototype is that instance of a probabilistic concept which satisfies each of the feature

representations in the concept's constituent set; it is, as we might also say, the "maximal satisfier" of a probabilistic concept structure.

In other words, in the most general but quite precise sense of 'prototype representation' which we are aiming at here, the availability of a prototype representation *comes for free* once you have a probabilistic structure. Conversely, probabilistic structures are necessarily presupposed in psychological discourse about prototype representations. To stress this equivalence is not to deny that the different terms 'prototype theory of concepts' (in even the most generic sense of the term) and 'probabilistic theory of concepts' naturally lend themselves to the expression of different perspectives on one and the same theory. Whereas talk of "probabilistic" theories is more likely to occur in contexts that emphasize the involvement of representational constituents and the contingent modal status of their represented features relative to the target representatum, talk of "prototype" theories is more likely to occur in the context of typicality judgments in categorization.

Categorization is where the term 'prototype theory' is usually introduced. What is emphasized from this perspective is that an object is subsumed under a probabilistic concept just in case it is sufficiently similar to the concept's prototype. When probabilistic concepts are employed in categorization, the concept's prototype and the potential instance of the concept are compared by means of a similarity metric applied to the features that constitute the prototype. All other things being equal, the degree of similarity is measured in terms of how many constituents of C are satisfied by x. This echoes what is already expressed above by the content-determining rule for probabilistic concepts. The prototype theorist assumes the property of being sufficiently similar to the prototype of C only qua property that supervenes upon the possession of sufficiently many of the features represented by the constituent representations of C.

The use of 'prototype theory' as a catch-all term for concept-probabilistic theories, to be sure, abstracts away from numerous differences which make for big differences from the point of view of actual theorizing in cognitive psychology.[2] Most importantly of these, cognitive psychologists would usually take 'prototype theory' to designate a position which is distinct from what is known as "exemplar theory". But, from the vantage point of the ar-

---

[2]   Edward Smith and Douglas Medin, in their classical book-length survey of the field, comment with regard to the prototype-based wave of theorizing in the 1970s that "no two researchers seemed to mean the same thing by [a] prototype" (Smith & Medin 1981: vii). In the years since the publication of their book, more clarity seems to have been achieved over the main theoretical options and parameters of variation within probabilistic theories of concepts, but the diversity of approaches still appears to obtain.

gument against prototype theory that we will be considering, what are strategic-level oppositions to the cognitive psychologist[3] present themselves as details of little relevance. So, for example, it does not matter for these purposes whether the best instances are represented by means of "exemplar" representations, i.e., representations of actually encountered instances, or by means of more abstract, so-called "summary" representations. From the general psychosemantic perspective that we are interested in, this difference doesn't make any difference.

It bears emphasis that this is precisely the view of the geography which the central proponent of the argument we will be considering assumes too: It is Fodor's own understanding that probabilistic theories are simply theories that soften up the modal ties between a concept's content and the contents of its constituents (see Fodor 1998-a: ch. 5, the fullest exposition to be found in his work of this theory family). We have seen that this equation of the basic idea of prototype theory with that of concept probabilism is well-founded, but perhaps it is reassuring to see that it is not idiosyncratic. Laurence and Margolis (1999: 27), while in the business of giving a classification of theories of concepts, perform the same equation with a remark that perfectly sums up the position we are assuming here:

> What we are calling the Prototype Theory is an idealized version of a broad class of theories, which abstracts from many differences of detail. But ... the core idea can be stated plainly. According to the Prototype Theory, most concepts ... are complex representations whose structure encodes a statistical analysis of the properties their members tend to have.

In considering prototype theory, we can, for example, abstract away from such matters as the "weighting" of different features in prototype theory, or from subdivisions made within representational constituent sets, corresponding to various attribute dimensions and to various values which each of these attributes can take. Such theoretical elements presuppose prototype theory, rather than criterially setting it against definitionism.

Here is an example of a no-frills probabilistic bundle of feature representations. I like to think of my concept of an analytic philosopher as being a prototype-theoretical concept if any concept is. For all I know, the concept which I connect to the adjective 'analytic' in 'analytic philosopher' exhibits

---

[3] Witness Gregory Murphy's joking warning to prospective researchers in the theory of concepts that "the risk of becoming [mentally] unbalanced is clearly greatest in the prototype-exemplar debate, which shows little signs of abating after many years of controversy" (Murphy 2002: 8).

graded satisfiability, owing itself to the concept's decomposability into the following individually defeasible feature constituents.

ANALYTIC of a philosopher
- ARGUES RATIONALLY FOR AND AGAINST CLAIMS
- AIMS AT CLARITY OF EXPRESSION
- AIMS AT PRECISION
- MAKES THEORETICALLY FRUITFUL DISTINCTIONS
- AVOIDS UNEXPLICATED METAPHORS
- VIEWS PHILOSOPHY AS ATTEMPTING TO SOLVE OR DISSOLVE PROBLEMS AND PUZZLES
- VIEWS PHILOSOPHY AS A SYSTEMATIC, SCIENCELIKE RESEARCH ENTERPRISE
- ENDORSES AN AHISTORICAL METHODOLOGY
- PUBLISHES PAPERS IN PEER-REVIEW JOURNALS
- DESCRIBES HIMSELF AS DOING "ANALYTIC PHILOSOPHY"
- FORMULATES HYPOTHETICAL, DISCONFIRMABLE CLAIMS

Notice, most importantly, that the idea of graded satisfiability *as such* does not mark a difference between definitionism and probabilism, and is appropriately cancelled out in comparisons of the theory schemes. What matters is the more specific idea expounded earlier. It is (and this a point to which significant relevance attaches later on) the idea of *graded satisfiability in virtue of different ways of satisfying the concept at the supra-featural level*. In other words, it is the idea of a *supra-featural gradability of concept satisfaction*. Nothing that is essential to definitionism according to its canonical understanding prohibits it from allowing *graded satisfaction that owes itself to the individual featural level*. This point has already been urged by Eric Margolis (1994: 84) against the still widespread idea that prototype theory deserves to be favoured over definitionism because of the vagueness (or "fuzziness") of concepts. To illustrate this by means of an example, SKY-SCRAPER is a pretty good candidate for a definitionally constituted concept. It might well be defined by VERY TALL BUILDING. Clearly, the predicate 'very tall' is probably as vague (and context-relative) as a predicate gets. But this changes nothing about the definitionist requirement that anything satisfying SKYSCRAPER *must* satisfy VERY TALL to a sufficient degree, whatever that sufficient degree is taken to be in the context at hand.

2. The compositionality arguments against prototype theory

The biggest threat for prototype theory is commonly taken to be posed by the requirement that concepts have to be compositional. The principle of

compositionality is a claim concerning the derivability of a compound representation's meaning: The meaning of a compound representation is derivable from the meanings of its constituents together with the meaning of its syntactic mode of construction. Hardly anybody working in the theory of concepts wishes to deny the capacity of concepts being compositional, and, to be sure, neither do I.

It would be difficult to think of a negative argument from any philosophy of a special science which has of recent been reiterated similarly persistently and vigorously as the argument from compositionality against probabilistic theories of concepts has been in the hands of Jerry Fodor. Fodor has been pushing two sorts of violations of compositionality by concept probabilism, yielding, correspondingly, two kinds of compositionality arguments.[4] The first is the argument from missing prototypes. The second is the argument from noncompositional prototypes.

The *argument from missing prototypes*[5] rests on the observation that a vast, potentially infinite number of phrasal concepts lack anything that one might intuitively regard as a particularly good, typical instance of that concept. A central family of examples for concepts lacking prototypes is that of concepts formed by Boolean connectives, for example, negational concepts (like NON-CAT) or disjunctive concepts (like STRAWBERRY OR PROTON). Another family of examples is that of concepts which involve a long chain of modifying concepts, e.g., FRENCH PHILOSOPHER OF SCIENCE WHO LIVED IN BERLIN BEFORE LEAVING TO COACH THE SQUADRA AZURRA. The longer the chain of modifications on the head concept is, the smaller the chances are that people would share the association of a specific type of good instance with that phrasal concept; indeed, the smaller the chances are that people would intuitively associate any specific type of instance with it at all.

The *argument from noncompositional prototypes* has been more extensively discussed among cognitive scientists. It maintains that, in those cases where phrasal concepts *can* be intuitively assigned a prototype which has some claim to an interpersonally reliable association with the concepts, this prototype is usually not derivable from the respective prototypes intuitively assignable to the phrasal compound's constituents (together with the meaning of the phrasal compound's syntactic construction mode); in particular, such a failure of prototype derivability occurs even where the content of the phrasal compound clearly appears to *have* been compositionally generated.

---

[4]   The distinction between two kinds of arguments is explicitly made in Fodor (1998-a: 100-107) and in Fodor & Lepore (1996). See also already Fodor (1994: 108).

[5]   Fodor (1981: 296-97) is the original locus of this.

Take the literature's favourite example, the phrasal concept PET FISH. The expression 'pet fish' has its reference compositionally determined from the references of 'pet' and of 'fish', in accordance with the reference rule for predicate conjunction (or similarly, set intersection, if you prefer). That is, 'pet fish' refers to something just in case it is both a pet and a fish. This, accordingly, is how the content of PET FISH is determined by its corresponding constituents. But, the argument from noncompositional prototypes complains, if prototype theory were true of PET, of FISH, and of PET FISH, then compositionality would be violated. Let, for example, the prototypical kind of pet be a dog. Let the prototypical kind of fish be a trout. Now, the prototypical case of a pet fish—of something that is in the intersective class of pets and fish—appears to be a guppy, or something much like it. Guppylike fish are small, they are brightly coloured, and they usually live in bowls or aquariums located indoors. But these features of the PET FISH prototype are not predictable on the basis of the features of the dog-centered prototype structure of PET, plus the features of the trout-centered prototype structure of FISH, plus the ability to conjoin concepts. Put differently, knowing what counts as a prototypical pet and knowing what counts as a prototypical fish, together with the ability to conjoin concepts, does not put you in a position to know what counts as a prototypical member of the intersection of these two classes of entities. In light of this result, the conclusion is drawn that prototype theory cannot apply to PET FISH, nor to its constituents PET and FISH.

Like the argument from missing prototypes, the argument from noncompositional prototypes takes the form of a reductio argument. Rigorously stated, we might render it as follows:

Argument from noncompositional prototypes:
1. Combining the concepts PET and FISH by means of predicate conjunction produces the compound concept PET FISH. (*Compositionality of* PET FISH.)
2. The concepts PET and FISH consist in prototype-centered representational structures. (*Lexical-level prototype theory.*)
3. The conjunctive compound PET FISH consists in a prototype-centered representational structure. (*Phrasal-level prototype theory.*)
4. Combining the prototype-centered structures of PET and FISH by means of predicate conjunction does not produce the prototype-centered structure of PET FISH. (*Intuitive typicality judgments.*)

5. Combining the concepts PET and FISH by means of predicate
conjunction does not produce the compound concept PET
FISH. (*4,2,3.*)

6. ⊥ *(1,5.)*

3. The Scope Assumption
about prototype theory

Both of the compositionality arguments obviously rest on a presupposition
without which they would instantly collapse. They crucially presuppose that
if prototype theory is true, it is true of compositional phrasal concepts no
less than of lexical concepts which, by default, form the home domain of
theories of concepts. I want to call this the "Scope Assumption" concerning
prototype theory.

> **Scope Assumption:** Phrasal concepts that are compositionally
> constructed from lexical concepts satisfying prototype theory
> satisfy prototype theory themselves.

Without the Scope Assumption, prototype theory is immune to the alleged
threat posed by the compositionality arguments. Unsurprisingly, one of the
major reactions to the compositionality arguments in defence of prototype
theory is that of abandoning the Scope Assumption.

I will claim that the Scope Assumption is wrong. Nevertheless, one could
not be farther from the truth if one thought of the present chapter as chiming
in with the position just anticipated, which dispenses with the Scope As-
sumption only in an ad hoc reaction to the compositionality arguments.

Concerning the argument from noncompositional prototypes, denying the
Scope Assumption has been the dominant reaction among philosophers who
regard it as possibly true. The argument from missing prototypes is usually
reacted to in one and the same move, and the reaction here is, more pre-
dictably still, of exactly the same kind.[6] To the extent that the verdict of
missing prototypes is conceded for compositional phrasal compounds, the

---

[6]   See, e.g., the list of philosophers mentioned in Fodor 1998-II: n. 5. Here, as elsewhere
in print, philosophers react to the compositionality argument with regard to Fodor's
analogous application of it to the alleged notion of a "recognitional concept". This is a
thoroughly ill-conceived notion, however, and it will be subjected to critique in the next
chapter. The best we can say of it is that it adds nothing to the parameter of variation
addressed in this chapter concerning probabilistic versus definitional structuredness. It
does, however, introduce numerous confusions besides. Nothing more needs to be said
about it here.

falsity of the Scope Assumption about prototype theory must ipso facto be conceded, since compositional phrasal compounds without prototypes are nothing but direct counterexamples to the Scope Assumption.

The theorists who deny the Scope Assumption in defence of prototype theory against the compositionality arguments do so in a *purely ad hoc* fashion. They take the compositionality arguments to reveal a contingent state of affairs which would have remained undiscovered if it were not for the compositionality arguments. Typically admitting that a certain loss in theoretical elegance is suffered, they take the concessive stance that a restriction of prototype theory's scope to lexical concepts (and in fact, also noncompositional phrasal concepts) must be imposed in order to do justice to the observations of the compositionality arguments. They understand their reaction as a renunciation of a desirable element of theoretical simplicity in this domain, whose preservation could have been rationally hoped for, had it not been for the observations of missing or noncompositional prototypes corresponding to compositional phrasal compounds.

In central respects, these commentators occupy a viewpoint which is far more akin to that of someone upholding the Scope Assumption than to the picture which I will argue for. The rejection of the Scope Assumption that I will advocate is principle-based and independent of any alleged insights made in the compositionality arguments. It assigns a different modal status to the denial of the Scope Assumption. Relatedly, it assigns a different epistemic status to this denial. These are some basic respects in which my picture contrasts with the ad hoc rejection of the Scope Assumption just sketched.

Though Fodor's original statements of the compositionality arguments take the Scope Assumption for granted, a number of his more recent comments on the compositionality arguments, scattered across his work, explicitly address its denial and, correlatively, the notion of what I will loosely refer to as "lexically confined" prototype theory later on. It will be shown in chapter 8 that the arguments with which Fodor purports to exclude a lexically confined prototype theory are without exception fallacious. Equally deplorably, Fodor's arguments usually do not even succeed in addressing this simple idea, even where he correctly quotes it. Establishing these matters all the way through is a considerable critical undertaking of its own (tellingly enough, making for what is by far the longest chapter of this work). For the purposes of the contentions and arguments put forth in the present chapter, however, the defence of the philosophical integrity of lexically confined prototype theory is not relevant. It will altogether be reserved for chapter 8.

One thing that still deserves mention is that, differently from denials of the Scope Assumption, many cognitive scientists have tried to hold onto the Scope Assumption in light of the argument from allegedly noncompositional prototypes. They have tried to let the operation of concept combination make the phrasal concept's prototype come out as a result of some systematic, predictable process after all. What is really involved here, however, is an abandonment of the assumption of compositionality—e.g., in the case of a compound like PET FISH, of the assumption that the content of the target conjunctive compound is derivable on the basis of the constituents' prototype structures and a conjunctive operation alone. The critique is familiar from Fodor and Lepore (1996), and we needn't go into it further here. If the belief that the Scope Assumption could *possibly* be true is very seriously confused already, as I argue, then we needn't begin to comment separately upon the a fortiori confused belief that the Scope Assumption is *actually* true.

4. On admissions of noncompositionality

Apart from ad hoc rejections of the Scope Assumptions, another popular sort of reply (though perhaps not one that is often encountered in print) is that of replying to the argument from noncompositional prototypes that the target compound, e.g., PET FISH, is indeed not compositional. I find this sort of objection surprising, since it plainly misses what is at issue.

Syntactically well-formed expressions always have *one* reading which is compositional. This is the basic reading which noncompositional readings build upon. The existence of a noncompositional reading is only additional over that of the compositional reading. The noncompositional meaning is often one that predominates in actual usage, but it is never one that takes the place of the compositional reading. But of course, *only the compositional reading of a phrasal expression is the relevant one* if it is compositional concepts that prototype theory is tested against.

In particular, imagine someone who claims that 'pet fish' (and, analogously, PET FISH) does not have its content exhaustively determined by the joint contents of PET, FISH, and conjunction. What would that be supposed to mean? Could the claim be supposed to mean that there is only one permissible understanding of 'pet fish' in English, and that this unique permissible understanding of 'pet fish' is one on which 'pet fish' fails to refer to all and only

those things that are both a pet and a fish? Surely not.[7] Once it is acknowledged that the syntactic construction involved in 'pet fish' is one meaning property conjunction, then of course there is at least *one* understanding of 'pet fish' on which the expression simply refers to what the conjunctive reference rule, applied to its syntactic constituents, says that it refers to, whether the compositional reading is the one that is predominantly intended in our uses of the expression 'pet fish' or not. Similarly for any other complex expression and the syntactic construction type exhibited by it.

The situation is similar to that with the use of an outright idiomatic expression. More frequently than not, an idiomatic expression is used/understood in the non-literal, compositionally not predictable sense. But its compositional, literal reading remains available nonetheless.[8]

Correspondingly, there are two sorts of concepts we might conventionally ascribe by means of the concept-naming expression 'PET FISH'. One is the concept that is syntactico-semantically isomorphic to the compositionally understood expression 'pet fish'; the other is a quasi-idiomatic concept that is not. The quasi-idiomatic concept would in fact more properly be called 'PETFISH'. But the relevant point here is that, *even if* it is granted that there is a widely shared quasi-idiomatic concept PETFISH, there *nevertheless also exists* a widely shared compositional concept PET FISH, constructed from the conceptual constituents PET and FISH, and only these, in a construction mode with a fixed functional meaning (in this case, that of conjunction).

The point I am leading up to with respect to the prototype-compositionality debate is this: *No matter how many cases of noncompositional phrasal concepts are claimed to exist, this does not make any of their compositional counterparts go away*—and *if* the compositionality arguments are to be seen as posing any challenge, then it is the challenge posed by the question of what the prototype theorist says in reply to the criticism that *the compositional compounds* cannot be accounted for within prototype theory. It is *those* phrasal concepts that prototype theory is intended to be measured against in the context of the argument from noncompositional prototypes.

---

[7] It is partly explicable how someone could get caught in this idea, since philosophers of language do usually describe putative cases of noncompositional expressions in a way that suggests the compositionality-violating understanding to be the uniquely permissible one in a given case.

[8] The same is true for any other claim stating that an expression is noncompositional. This point about the only plausible true understanding of claims of noncompositionality constitutes, I believe, an importantly underappreciated platitude. Its implications for discussions of compositionality perhaps remain to be treated elsewhere.

By the same token, it is them that prototype theory is measured against in the argument from absent prototypes.

This is also the right place to mention a seemingly novel type of reply to the argument from noncompositional prototypes which has been put forward by Jesse Prinz (2002: 291-295). Prinz concedes that prototype theory yields a noncompositional phrasal concept corresponding to 'pet fish'. His take on the compositionality principle, however, allows him to occupy a particularly relaxed and unapologetic position with regard to the acceptance of noncompositional concepts. He does accept the imposition of the compositionality requirement, but he claims the standard insistence upon it by Fodor and others to be confused about the modality involved therein. What the principle amounts to, Prinz points out, is not the requirement that concepts must actually *be* compositional, but merely the requirement that concepts must be *capable of* combining compositionally. The compositionality requirement is standardly motivated by its explanations of the phenomena of systematicity and productivity, but these are only representational capacities of the conceptual system, and their explanation requires no more than the postulation of a compositional capacity in turn. This capacity he does postulate; it just occupies the position of "a fallback system, not a mandatory mode of operation" for combining concepts (293).

Prinz's observation about the modal status of the compositionality requirement is straightforward enough to make one wonder why no one in cognitive science or philosophy appears to have thought about it earlier. Nevertheless, when we look at what I have just said in reply to the denial of compositionality for "the" concept expressed by 'pet fish', it becomes clear that his observation changes nothing about the situation described there. Let me repeat this point in the context of a summary description of the dialectic.

Prinz does not deny our possession, or implicit possession, of any the compositional concepts that a theorist like Fodor wants to posit. For any noncompositional combination of concepts, Prinz would accept the parallel formability of a compositional combination. He explicitly agrees that we are in possession of general, compositional concept-forming mechanisms. Most importantly for present purposes, he always takes for granted against the background of this presupposition that wherever prototype representations *are* combined compositionally, the phrasal concepts they combine into are themselves of a prototype-theoretical nature. That is, he too shares the Scope Assumption.

What the disagreement between opposing positions like Prinz's and Fodor's comes down to is simply a question of the following sort: Just how large a proportion of actual cognition is occupied by compositional processes of concept formation? The positions here agree upon the existence of both compositional and noncompositional phrasal concepts, and they agree about the Scope Assumption. Their point of dissent, however, hinges upon a question which I do not want to begin to adjudicate here. What I do care about in the present paper is that both sides are equally mistaken about how the varieties of phrasal concepthood relate to the idea of a prototype-theoretical, probabilistic conceptual structure.

5. The confusedness of the debate about
prototype theory and compositionality

All of the reactions to the compositionality arguments encountered in the literature share with the proponents of the compositionality arguments the presupposition that the Scope Assumption concerning prototype theory and actual compositional phrasal compounds is *coherent* to begin with. I have emphasized the fact that this coherence is taken for granted already in the context of ad hoc denials of the Scope Assumption. For the purposes of this work, I will sometimes also refer to this idea as "Compatibilism".

Let me now make the contentions explicit which I have already begun to mention in commenting upon reactions to the Scope Assumption. The general claim to start with in my critique of the prototype-compositionality debate is this:

- The Scope Assumption is incoherent.

Compositional phrasal concepts could not possibly be prototype-theoretical compounds in the first place. Participants on all sides of the debate about concepts and compositionality are oblivious to some elementary relations obtaining between the properties of representational structuredness discussed. In particular, within the next few sections I am going to make the following claims:

1. Compositional phrasal compounds postulated to be conjunctive in kind trivially satisfy the content-determining rule of concept definitionism.

2. Concept definitionism plausibly permits a far more general formulation than that over the content-determining rule for conjunctive phrasal compounds.

3. On the generalized formulation of concept definitionism, all the compositional phrasal compounds ever discussed in the compositionality arguments come out as having a trivially nonprobabilistic, and a trivially definitional, representational structure.

The Scope Assumption thus is not only incoherent, it is trivially so too. Before moving on with my central claims, let me remind us of the following fact. The notions of a lexical and of a phrasal concept—and with them the domain of the Scope Assumption—were defined with reference to the expressive resources of English. This is appropriate, since debates about concepts and compositionality only speak of syntactically complex concepts that mirror actual compositional phrasal expressions of English. But it is easy to imagine syntactically complex expressions that are drawn from conservative extensions of English, even though they are not available in actual English. One of the further things I will point out is that, in spite of what I am saying about *actual* compositional phrasal compounds being trivially nonprobabilistic in kind, it is perfectly *possible* for there to exist compositional phrasal compounds which do have a probabilistic structure.

4. Compositional phrasal compounds with a probabilistic representational structure are easily conceivable.

It is just that no such phrasal compounds are ever considered in the compositionality arguments, given that consideration of concepts and compositionality only stretches to compounds mirroring actual linguistic expressions of English.

Again, however, *if* a phrasal compound is a probabilistic compound at its postulated phrasal level of structure, it can only be trivially so, just as *if* a phrasal compound is a definitional compound at its postulated phrasal level of structure, it can only be trivially so too. Both of these insights flow from one and the same general insight forming the unifying undercurrent of this chapter's main line of argument:

- Once a concept with a specific compositional constituent structure has been postulated, its concept-theoretical nature at the level of the definitionism-probabilism opposition is

> thereby fixed, and a statement as to which of these sides it
> falls on can only be trivially true or trivially false.

Which of the two concept-structurist schemes a compositional compound with a postulated constituent structure falls under is immediately ascertainable by comparing that constituent structure against the general content-determining rules that define the structurist theory schemes.

Against this background, one reaches the following doubly negative verdict concerning the compositionality arguments' idea of testing compositional phrasal compounds for prototype-theoretical structuredness against intuitively assignable prototypes. This amounts to the verdict that the compositionality arguments against prototype theory are red herrings:

- Once a concept has been posited as having a specific compositional phrasal constituent structure, then
  (a) answering the question of whether the compound has a probabilistic nature could not possibly require discovery by the method of checking for intuitive prototype assignability;
  (b) the ascertainment of the compound's probabilistic nature could not possibly be defeated by intuitions of allegedly absent or allegedly noncompositional prototypes.

I will say more about the last idea towards the end of this chapter. This will bring us back to the point emphasized early on in the chapter, that the assignability of a prototype comes for free given a probabilistic content-determining structure.

Back to the more concrete agenda of this chapter: In the following short section, I will restrict myself to claim 1 about conjunctive compounds (together with making a side observation), which is very straightforward. The conjunctive compounds which claim 1 refers to coincide with the domain of the argument from noncompositional prototypes. Later sections will move on to claims 2 and 3.

Along the way, I will engage in another round of ancillary considerations, trying to conjecture whether any kind of view about the compounds at hand could make the Compatibilist about conjunctive compounds rationally think that she does not fall prey to the refutation expressed by claim 1. Such options will be shown to be unavailable. Doing so will not only corroborate the argument about conjunctive compounds, but also the more general point

which is to be made with respect to all compositional phrasal compounds in the generalized claim 3.

6. The obvious definitionality of
compositional conjunctive compounds

Here is a frighteningly simple observation: Conjunctive compounds like the compositional phrasal concept PET FISH ipso facto preclude the individual defeasibility of constituent representations characteristic of probabilism. Besides, even disregarding this observation, binary compounds anyway would not allow the supra-constituential gradation of concept satisfaction which concept probabilism envisages.

The canonical content-determining rule for definitionist concepts, highlighted earlier, is identical to the content-determining rule of compositional conjunctive compounds. The content-determining rule for definitional concepts said that a definitionally structured concept *C* represents whatever satisfies each of its constituents, and nothing else. But of course, for a concept to be such that it is satisfied just in case each of its constituents is satisfied just is for that concept to be a compositional conjunctive compound out of those constituents. Definitionally structured concepts, according to their canonical characterization, simply *are* conjunctively structured concepts. By the same token, since the general content-determining rule for probabilistic concepts allows a concept *C* to be satisfied even if less than all of its constituents are satisfied, concept probabilism necessarily cannot be true of a conjunctive compound. Since conjunctive compounds are obviously definitional, prototype theory necessarily and trivially cannot be true of conjunctive compounds.

I should like add that this is not the only respect in which prototype theory cannot apply to conjunctive compositional compounds. Even if, for a moment, we ignore the fact that conjunctive compounds do not tolerate any failure of satisfaction among their constituents, there is another observation to make about the binary conjunctive compounds which almost invariably provide the examples for the compositionality argument.

Prototype theory assumes a gradation of degrees of concept satisfaction which is determined, all other things being equal, by how large a proportion of that concept's feature constituents are satisfied. But the discussed examples of compositional phrasal concepts with a noncompositionally determined, associated prototype are practically always conjunctive compounds that correspond to a sparsely modified head noun, typically even a merely

binary conjunctive compound, such as PET FISH or BROWN COW. And there is obviously no way that a supra-featural gradation can be put to work if all that you have are a meagre two constituents.

7. On the very constituent structure at issue

Now, one would think that theorists in the field must be tacitly holding a picture of the application of prototype theory to conjunctive compounds on which the simple counterarguments just given—primarily, the one from identical content-determining rules—somehow would have to regarded as missing their target. It is difficult to believe that Compatibilists have all along committed such an elementary blunder as the one which the above observations would force us to attribute to them. Surely, one would think, it *must* be the case that Compatibilists would regard these plain and conclusive arguments against the probabilistic nature of PET FISH as misdescribing their view of how prototype theory would apply to a conjunctive compound, if it did. If this is so, however, standard theorists have entirely failed to explain in what way it is so (a hardly surprising omission given that they do not even mention the possibility of raising these objections in the first place). Let me conjecture.

How might conjunctival Compatibilists think that prototype theory might potentially apply to conjunctive compounds in a way that avoids the obvious objection raised? Preciously few types of options offer themselves for answering this question. I will consider two possible construals of the conjunctival Compatibilist's position below. Both of the potentially underlying conceptions will be shown to be confused for similarly straightforward reasons as the idea of Compatibilism about prototype theory and compositional conjunctive compounds itself.

What we will reassure ourselves of with reasonable certainty is this: The compositionality arguments are indeed committed to the idea that the constituents that form the *stipulated compositional determiners* of the putatively possible probabilistic compound are the very constituents that must be considered as the *probabilistic feature representations* of the target compound in question. This point continues to remain of crucial importance when, further down, we broaden our view beyond conjunctive compounds.

7.1. On incoherently shifting the constituent level at issue. If we refuse to believe that the Compatibilists about conjunctive compounds could be the victims of such an obvious refutation as the one above, the assumption which, by way of conjecture, we can most naturally attribute to Compati-

bilists is what we might call the *level-shifting assumption* concerning the question of what representational constituents are at issue in debating the possibly probabilistic nature of compositional compounds. Nobody ever seems to speak about this explicitly. But let me speculate and imagine a Compatibilist as she explicitly replies along these lines to the argument from obvious definitionality and non-probabilistic constructedness just given:

> **A speculative rationalization of the Compatibilist's view:**
> "In considering whether a compositional phrasal concept has a prototype-theoretical nature, we are not considering its overtly posited constituents as the potential probabilistic feature constituents. We are tacitly assuming that the potential probabilistic feature constituents would be *non-overt, yet to be postulated ones*. More specifically, it is indeed obviously true that both of the overtly posited constituents of the compositional conjunctive concept PET FISH would have to be satisfied in order for the compound to be satisfied. But in considering PET FISH as a potential probabilistic compound, we are not considering a situation in which its immediate constituents PET and FISH are individually defeasible, but rather one in which *those constituents' respective constituents themselves* are defeasible. The set of constituents which would be responsible for the probabilistic structuredness of the conjunctive compound PET FISH, thus, would not be the pair of stipulated conjuncts PET and FISH, but the union set of those conjuncts' *underlying constituents*. Which constituents these are is up for grabs."

This, I suppose, must express the implicit picture behind Compatibilists' consideration of compositional phrasal compounds as potentially prototype-theoretical concepts.

However, if the level-shifting picture imagined in the above statement is what a Compatibilist presupposes when she assumes the structurist nature of the compositional conjunctive concept PET FISH to be open to debate, then her position is just as confused as it is without it. The respective internal structures of PET and of FISH are simply *not located at the structural level that is at issue* in the compositionality arguments.

By assumption, it is the concepts PET and FISH that are the compositional determiners of PET FISH. But of course, crucially, *any probabilistically structured concept's compositional determiners ipso facto form its set of probabilistic feature representations.* This should actually not be in need of

much explanation. After all, the concept-structurist theory schemes are defined over content-determining—i.e., compositional—rules obtaining between representational constituents and their target concept. Representational structures, whether probabilistic or definitional, are only posited for a target concept qua compositionally determining constituent structures in the first place. Or, to put the point this way, *if* a concept that is assumed to be probabilistically/definitionally structured is thought to be compositionally structured, then its compositional determiners must ipso facto be its probabilistic/definitional constituents.

Therefore, we were right: Checking the allegedly prototype-theoretical nature of PET FISH for its compositional determinedness by the conjunction out of {PET, FISH} indeed entails the idea that this binary "feature set" is all that PET FISH is probabilistically structured from. The orthodox, Compatibilist view of PET FISH remains stuck with an obvious incoherence.

7.2. A tacit notion of non-preservation of compositional determiners?
Another idea which we could speculate to lie behind Compatibilism about conjunctive compounds does not even at first glance constitute a potentially viable option. I mention it nevertheless because it may be the case that this is an incoherence that many theorists tacitly slip into. The picture I have in mind is one in which the prototype structures of PET and FISH, when conjoined, *generate* the compound PET FISH without being *preserved* as discrete elements within that compound. Perhaps, that is, the reason for the fact that the compositional concept PET FISH is not treated as the binary conjunctive compound that it is, is that the constituent structure of PET FISH is tacitly conceived as a reorganized, unitary structure out of the elements of PET and FISH.

The present picture would still more obviously deviate from what is at issue than the level-shifting assumption by itself already did. Notice that, in trying to evade the argument against Compatibilism about conjunctive compounds, it shares the previous strategy's fatal incoherence of relying on the distinctness of a probabilistic concept's set of compositional determiners from its set of content-constitutive probabilistic feature constituents. That is, again, {PET, FISH} is supposed to constitute the set of compositional determiners, while the set of probabilistic constituents would have to consist in the union set at the level below.

As if this were not enough, however, the presently imagined picture, in addition, abandons the subject of compositional phrasal compounds altogether, since compositionality is, by common definition, compositional determina-

tion by *parts,* and parts of the alleged compound PET FISH are exactly what PET and FISH are not assumed to be. The nonarbitrariness of the canonically acknowledged presupposition of mereological structuredness by the notion of compositionality is particularly easy to make clear in the case of conjunctive compounds. Possessing a conjunctive compound requires both the ability to deduce the truth/satisfaction of the conjunctive compound from the joint truth/satisfaction of its respective constituents, and it requires the ability to deduce the separate constituents' respective truth/satisfaction from that of the conjunctive compound. The second half of this, which corresponds to the elimination rule for the meaning of 'and', is abandoned with the idea that PET and FISH are not preserved as parts of PET FISH.

The most straightforward way to notice the change of topic in the putative strategy of evasion just imagined, however, is this: The phrasal compound PET FISH under consideration is, by definition, the syntactically complex concept that is isomorphically structured to 'pet fish'. And generally, the compositionality arguments, as clarified earlier, only consider the multitude of familiar, compositional formations known from actual linguistic expressions. How could one consider a compositional phrasal expression, consider the syntactico-semantically isomorphic phrasal concept, and then *consider its openly postulated phrasal constituents as its compositional determiners,* while at the same time turning around and denying that these compositional determiners are constituents of the phrasal concept after all? An inconsistency could not be more obvious than this.

We do well to remind ourselves that the literature is for the most part perfectly clear and explicit about assuming the discrete recombination of the lexical concepts into the target phrasal compound considered (i.e., their preservation within it). In other words, they are perfectly clear about the fact that the compositional determiners form the syntactic constituents of the target compound (in parallel with linguistic phrasal expressions). Take the following quote from a classic 1981 paper by Osherson and Smith, "On the adequacy of prototype theory as a theory of concepts":[9]

---

[9]   I am inserting 'prototype structure', in this quote, for Osherson and Smith's more specific talk of "representations of quadruples," which they identify concepts with in §1.1 (the respective quadruples consist of a "conceptual domain", a "distance metric", a "prototype", and a "characteristic function"). Their definition of prototype theory is an attempt to give a rigorous characterization of the position, but they make clear that it only presents one variant of prototype theory, directly contradicting other bona fide variants of prototype theory. Since I think there are problems with the way they define prototype theory, and since we are only interested in the general idea here, I am leaving the specific formulation over their quadruples away in the quote.

> Suppose a given concept C has concepts C_1 and C_2 as constitu-
> ents. For prototype theory to be compatible with this case of
> conceptual combination, principles must be available to specify
> the [prototype structure] associated with C on the basis of the
> [prototype structures] associated with C_1 and C_2. (Osherson &
> Smith 1981: §2.2)

Or take just about any pertinent passage you wish from the works in which
Fodor, or Fodor and Lepore, push the compositionality arguments. Pars pro
toto, here are the core statements of the alleged problems, as stated in Fo-
dor's *Concepts* monography:

> For indefinitely many Boolean concepts, there isn't any proto-
> type even though their primitive constituent concepts all have
> prototypes and the complex concept itself has definite condi-
> tions of semantic evaluation. (Fodor 1998-a: 101. Original ital-
> ics and displaying omitted.)

> [P]rototype theories can represent conceptual repertoires as be-
> ing compositional only if (barring idioms) a thing's similarity to
> the exemplar of a complex concept is determined by its similar-
> ity to the exemplars of its constituents. However, this condition
> is not satisfied in the general case. So, for example, a goldfish is
> a poorish example of a fish, and a poorish example of a pet, but
> it's a prototypical example of a pet fish (Ibid.: 102.)[10]

Let me conclude: The present section took its cue from my claim of the ob-
viously definitional nature of conjunctive compounds, and from the suspi-
cion that the Compatibilist about conjunctive compounds must have some-
thing else in mind than the assumption that the probabilistic constituent set
of the allegedly candidate prototype-theoretical concept PET FISH consists in
its set of conjunctive constituents {PET, FISH}. What we have seen, however,
is that the Compatibilist about conjunctions is indeed committed to this as-
sumption and thus that he is indeed subject to the immediate reductio ad ab-
surdum stated in the previous section.

Perhaps it is better if we just forgot about the idea of finding rational expla-
nations for how proponents of standardly encountered Compatibilism about

---

[10]   Notice the subtle fact that, here and elsewhere, the means by which Fodor casually
       states that the compound's prototype should be tested against its constituents' proto-
       types crucially *presupposes* that the compound is a conjunctive one. That is, an entity's
       similarity to the compound's prototype is to be determined by the entity's similarity
       both to the prototype of the first constituent *and* to the prototype of the second constitu-
       ent. This is obviously not, e.g., how someone reasoning along the lines of the composi-
       tionality argument would test a disjunctive compound's prototype's compositional de-
       termination.

conjunctive compounds could have come to think that theirs was a remotely viable option. One cannot help beginning to feel that the incoherences whose attribution we are led to by way of attempting to explain the mainstream's neglect of the obvious equivalence between the conjunctival and the traditional definitional content-determining rules are themselves equally puzzling as the neglect of this equivalence is itself.

In the following section, I shall move on to a more general consideration of the nature of phrasal concepts, not just of conjunctive ones. However, our establishment of the incoherence of potential strategies of avoiding the equation of a candidate prototype-theoretical compound's compositional constituents with its probabilistic constituents remains equally pertinent to a generalized claim about compositional phrasal concepts as it is to compositional conjunctive concepts.

8. Concept definitionism, properly generalized, and a generalized refutation of Compatibilism

Let us take a step back now from the canonical portrayal of concept definitionism as a position that is committed to the idea that the content of its concepts is determined by the *conjunction* of their content-constitutive representations.

It is a somewhat unhappily neglected fact in concept-theorizing that the assumption of conjunctive structures in the characterization of definitionist structures already focuses on a quite specific form that a definition might take. The familiar notion of a *definition* is a considerably more general one. Nothing about the notion of a definition as such requires the defining representation to consist in a conjunction. And definitionism as such in the theory of concepts does not exploit any other notion than that.

In the philosophy of language and elsewhere, when one speaks of a definition, one simply means a claim stating the synonymy of a given target expression with a defining expression of the same syntactic type as the target expression.[11] What is usually also intended, the symmetry of the synonymy relation notwithstanding, is that, epistemically, the synonyms stand to each other in an asymmetrical relation. The defining expression is one the meaning of which is taken to be in less need of explanation than the meaning of the expression defined. It is thus capable of elucidating the defined expression, and elucidation is what the demand for a definition aims at. One need

---

[11]   Let us disregard the idea of "implicit" or "contextual" definition here.

not understand the definiendum in order to understand the definiens, but one does need to understand the definiens in order to understand the definiendum. Without this epistemic asymmetry, there would not be any point in talking of "definitions" rather than simply of "synonymies". In analogy to the usual use of 'definition' on which, in order to count as a definition, a statement of synonymy has to exhibit an epistemic asymmetry between synonym 1 (the definiendum) and synonym 2 (the definiens), the psychological theorist is interested in "definitions" of concepts only insofar as they relate unequally complex representations to each other—in other words, only insofar as they provide progress in terms of conceptual decomposition.

However, nothing about the notion of a definition requires the defining expression to consist in a conjunction. To be sure, prominent examples of explicit definitions are often very simple conjunctive combinations: If 'bachelor' is definable through 'unmarried man', then 'bachelor' refers to whatever is referred to both by 'unmarried' and by 'man'. If 'vixen' means the same as 'female fox', then 'vixen' refers to whatever is referred to by both 'female' and 'fox'. But, to take a quite different example, 'farthest', on one of its uses, might be defined by the expression 'at the most remote point in space or time'. This candidate definiens is not a conjunction. The constituents of the candidate defining phrase 'at the most remote point in space or time' do compositionally determine its meaning, but they do not do so by being embedded in a conjunctive construction scheme. Many more examples could be adduced.

What I am pointing to is made most obvious, though, if we consider how, in the potentially infinite number of stipulative definitions which we can perform, we take it for granted that all kinds of syntactic constructions are on equal footing, as far as the capacity to serve as definientia for newly introduced terms goes. When we acknowledge somebody (perhaps ourselves) to have performed a stipulative definition, we do not take it to be of the slightest relevance which syntactic construction mode is exhibited by the phrasal expression that the freshly introduced expression has been stipulatively defined over. Take another example of a nonconjunctive construction than the one already mentioned: Imagine that we introduce a term by stipulatively defining it over a disjunctive phrase ("Let 'plonk' designate anything that is blue or green or red"). Obviously, nothing different in kind—let alone anything that is in any way deficient—has been done from what is done in stipulatively defining a term over a conjunctive phrase ("Let 'plonk' designate anything that is blue and loud and fast").

Shifting back to the psychological perspective now, there is nothing whatsoever to prevent a concept definitionist qua concept definitionist from assuming that non-conjunctive structures, too, can provide the syntactic skeleton of a concept posited to be definitionally decomposable. Analyzing things into conjunctions surely makes things easier to handle. And it is probably also intrinsically desirable, from a point of view of explaining how concepts are assembled, to be able to reduce as much as possible to such an elementary combinatorial mode as conjunction. But it does not clash with anything in the spirit of the concept definitionist's view as such to assume that the concept FARTHEST is constructed from the nonconjunctive phrasal concept AT THE MOST REMOTE POINT IN SPACE OR TIME.

An exclusive focus on conjunctive structures in characterizing the idea of definitionism, then, is too specific. The practice of narrowly delimiting definitionism over the content-determining rules of conjunctive structures, altogether disregarding the possibility of other kinds of syntactic construction schemes, is unjustified and prevents us from noticing the properly general idea underlying concept definitionism. What, then, is the properly general content-determining rule that is characteristic of definitional representational structures?

To answer this, we need only remind ourselves of the earlier, introductory explanation that it is supra-featural gradation of concept satisfaction that really separates concept probabilism from concept definitionism. The generic formulation of concept definitionism, then, is simply this:

> **Concept definitionism:** A concept *C* is definitionally structured just in case it has a compositional structure that disallows a supra-constituential gradation of *C*'s satisfaction.

Now, as I have made clear from the beginning, the concepts which the compositionality arguments say something about are concepts which are, by hypothesis, isomorphically structured with phrasal expressions of actual English. But linguistic syntactic constructions, at least English ones, do *not* allow any supra-constituential gradation of satisfaction. Given the generalized essence of definitional concepthood, therefore, *all* of the concepts ever considered in the compositionality arguments are trivially definitional mereological compounds. By the same token, of course, all of the concepts ever considered in the compositionality arguments trivially lack probabilistic structure.

For the sake of perspicuity, let me also expressly state the idea of prototype theory against whose background the complementary formulation of generalized concept definitionism has just been stated:

> **Concept probabilism:** A concept *C* is probabilistically structured just in case it has a compositional structure that allows a supra-constituential gradation of *C's* satisfaction.

It would be rather inappropriate to start enumerating examples of the fact that actual phrasal expressions of English and their conceptual analogues preclude supra-constituential gradation. If someone requested examples of this kind with the intention of requesting supportive evidence, the appropriate thing would be to ask her in return whether she could adduce a single, actually existent example in which a phrasal expression does allow a supra-constituential gradation of predicate satisfaction.

Well, here is what we would have to look for: In analogy with syntactic construction modes which carry the meanings of conjunction or disjunction, there would have to be what we might appropriately call a *cumulative* mode of combination in English. It would be perfectly coherent to imagine—so as to make this more tangible—a connective with a cumulative meaning being introduced into English. We could call this connective 'prand', with this label serving as an aide-mémoire to the fact that the connective constitutes a probabilistically relaxed cognate of 'and'. 'Prand' would connect sentences or predicative expressions in a fashion that would assign the resulting compound a probabilistic reference rule along the lines stated in the first section of this paper. So, for example, one could say, "If something is a bird, then it is such that it has wings, prand it is feathered, prand it flies, prand it lays eggs, prand it chirps, prand it nests in trees", requiring, say, at least four of the predicates to be satisfied by something in order for it to fall under 'bird'.[12]

There is no connective or construction mode with the meaning of 'prand' in English. It would be an interesting question to ask a linguistic typologist whether any languages exist that do perhaps contain such a connective. For

---

[12] To be sure, this toy example disregards, unrealistically in the general case, different feature weightings etc. Also, the meaning of 'prand' would have to allow a wide variation of probabilistic construction modes, of course. There is only one type of conjunctive construction mode, but there are many ways in which compounds can be probabilistically determined by their constituents. This plurality of conceivable probabilistic structures, however, doesn't make their subsumption under a single function word inappropriate. Think, by comparison, of a word like 'most'. The situation here is much the same. Imagining a connective like 'prand' is not to imagine an outlandish scenario.

our purposes, however, this does not matter in the least, since, to repeat, all the compositional phrasal concepts ever considered in the compositionality arguments are ones that are postulated in analogy with phrasal expressions of English.

9. The error of having prototype tails
wag the dog of probabilistic structures

For all that has been said so far, the following question still has not been addressed: What if there *were* any phrasal expressions that consisted in probabilistic constructions from their constituents? Could the argument from noncompositional prototypes and the argument from missing prototypes be constructed for them in analogous ways as for all the concepts that really were trivially definitional? The answer to this must be negative. A probabilistic representational structure compositionally entails a prototype as a trivial matter of fact. Therefore, *even if we stuck to truly probabilistic compounds, there could not be any problems of the sort suggested by the compositionality arguments to worry about.* Even for coherently envisaged probabilistic compounds, the idea that they might have *some test to stand* as to their conformity with intuitively assignable prototypes is confused.

Remember the fact, made clear within section 1 of this chapter, that which prototype (best instance) is associated with a probabilistic concept is entirely determined by the concept's probabilistic representational structure. This is not an arbitrary tenet. It simply follows from the fact that, for any supra-constituentially gradable concept, there will be exactly one combination of properties instantiating which will make something the probabilistic concept's maximal satisfier. That maximal satisfier, the entity instantiating that combination of properties, thereby is the prototype of the probabilistically structured concept.

What I want to emphasize with this in mind is that the prototype representation which supervenes upon a probabilistic compound's stipulated structure could not possibly be contravened or even overridden by another, different, prototype that someone's intuition might make one connect with the compound. Of course, if the job you are facing is to work your way back from the concept expressed by a lexicalized linguistic expression (typically, a word) to some hypothesized probabilistic representational structure underlying it, then intuitions about prototypes are indeed the place to start. But if, on the other hand, a probabilistic structure has already been *fixed* relative to the purposes of an investigation—as it has in the case of the compositionality arguments—then the potential prototype-determining combination of

properties has thereby been fixed too and there is *no tribunal of intuitive prototype associations for the posited probabilistic structure to face*.

We have thus been lead to the last of the central respects, anticipated in section 5, as to how the compositionality arguments and the surrounding debate are confused: They seem to *presuppose a picture on which prototypes are to be nontrivially assigned to a probabilistically structured compound*. If someone knew nothing about prototype theory and she looked at how Fodor or others present the compositionality argument, then she could almost get the impression that prototype theory is some sort of "dual-factor" theory of content. The mistaken picture conveyed is one on which, in a first step, one has a certain type of compositional probabilistic constituent structure, and in the next step, one had better find some prototype to fit the bill of that constituent structure.

The main critical line of argument of this chapter explains this mistaken impression well. We have seen that, contrary to what the compositionality arguments purport to be addressing, they never really address probabilistic concepts to begin with, but only what are really definitional concepts. This helps to explain why discussants of the compositionality arguments have effectively separated prototypes from probabilistic compounds, instead of relying on the correct prototype-supervenience picture. If one is only considering compounds with definitional structures to begin with, then assigning these compounds prototypes is *bound* to be a non-automatic procedure, requiring an extra step of consulting one's intuitions. One obviously will not, in this case, be able read a prototype off from the constituent set, as I maintained one should be able to.

Take, for example, the compound PET FISH: What theorists always do is to consider the intersection of the class of all pets and the class of all fish, and then, relying on experience-based intuitions to give them an idea of what this intersective class's prototypical members are, they come up with something like guppies, or coloured small fish living in bowls. By contrast, if one really does consider a probabilistically structured compound in the first place, then the assignment of a prototype to the compound is a matter of brainless routine. Take the probabilistic compound constructed from the constituent set, {HILARIOUS, GOOD AT BAKING APPLE CRUMBLE, FOND OF DOING TAX DECLARATIONS, SLOW RUNNER, THIN, CAPABLE OF STAYING UNDERWATER FOR MORE THAN THREE MINUTES, CAT-LOVING, FORMER TEST PILOT, INCAPABLE OF READING MUSIC, TRILINGUAL}. The prototype of this compound can simply be read off from this set of constituents. The compound's best instance is the type of person that is hilarious, good at baking

apple crumble, fond of doing tax declarations, a slow runner, thin, capable of staying underwater for more than three minutes, cat-loving, a former test pilot, incapable of reading music, *and* trilingual.

The notion that a fitting prototype has to be found for a probabilistic compound is no more sensible than the frivolous notion that someone who understands a conjunctive compound must still know, over and above that, "which type of entity" fulfills that conjunction, or that someone who understands a disjunctive compound must know, over and above that, "which type of entity" fulfills that disjunction.

The present point, interestingly, also provides us with a general diagnosis of what has gone wrong in the debate about phrasal compounds and prototype theory. In the context of the compositionality arguments, prototype tails have come to wag the dog of probabilistic structurism. The notion of a best instance seems to exert a powerful grip on the imagination, blocking principled considerations about the types of structure under consideration.

But how could it have come to do so? Well, the more specific error that seems to be committed by theorists taking the compositionality arguments seriously is that the role of a best instance is implicitly assumed to have to be filled out by an actually existent kind of object. Often, it even seems to be the case that a lexicalized expression for this type of object is preferably sought for. But this has nothing to do with what the probabilistic concept-structurist idea which we have discussed requires. The concept-structurist idea underlying prototype theory says something about how the satisfaction condition of a compound concept relates to the potential satisfactions of its constituents. Nothing about this precludes probabilistic constructions that are maximally satisfied by "non-natural" jumbles of properties. Nor, in fact, does anything even preclude probabilistic constructions involving constituents whose joint satisfaction is logically impossible. How could it?

It is quite illuminating to consider how we take for granted that matters lie with the comparison class of defined representations here: Counting as a definitional compound obviously does not require representing a natural and actually instantiated category in addition to fulfilling the requisite content-determining rule—so how could such a requirement pop into existence when all one has done is to shift to a probabilistic content-determining rule allowing degrees of supra-constituential satisfaction? It is in fact supposed to be one of the very *points* of the definitional construction of concepts that we can form concepts of fictional entities or even of logically impossible entities by means of them. For example, the concept representing round

squarehood quite obviously cannot be represented by atomic concepts. Even the generally extreme atomist Jerry Fodor makes *this* concession in his theory of concepts for logically or nomologically impossible entities.[13]

To be sure, it is more than understandable that there should be a focus of attention on "natural" categories in the context of concept probabilism, given the facts that it is, firstly, inquiries into the hidden potential structures of lexical concepts that theoretical interest lies with and, secondly, that words predominantly do encode natural categories. But a fallacy occurs where this focus of attention is treated as capturing the whole range of entities which can count as prototypes represented by probabilistic structures.

A much more elementary error still which may perhaps be at work among some philosophers is that of obliviousness tout court to the probabilistic structurist basis of prototype theory. Where operative, such an oversight would easily explain the imposition of the erroneous requirement of natural categoryhood. However, I suggest this only tentatively. As a generalized verdict with regard to philosophers—to say nothing of psychologists—diagnosing ignorance of the structurist nature of prototype theory is out of the question. The probabilistic-structurist nature of prototype theory forms elementary textbook lore. Some philosophers' expositions of the compositionality arguments are preceded by expositions of prototype theory which are perfectly explicit about the structurist nature of the prototype-theoretical idea. Some commentators even speak explicitly of the compositional determination of probabilistic feature representations in their statements of the arguments.

To wrap up these last points: It is plausible that the crucial role in distracting from the trivially probabilistic or non-probabilistic concepthood of stipulated phrasal constituent structures is played by the misguided notion that, for a compound to fall under concept probabilism, the compound must be assignable some prototype in a *separate, nontrivial, intuitive step*. The crucial factor responsible for the misguided quest for nontrivially assignable prototypes must lie in a tacit, unwarranted inclination to commit the generic notion of a prototype to a notion of *natural categoryhood*.

---

[13] Thus, Fodor (1990: ch. 4, p. 101) sees himself committed to "quite a strong consequence of the asymmetric dependence story [i.e., of his specific atomist theory of mental content, author's note]: *no* primitive symbol can express a property that is necessarily uninstantiated. (There can't, for example, be a primitive symbol that expresses the property of being a round square.)"

We can add to this that the misguidedly exclusive focus on natural categories in thinking about prototypes is in turn significantly facilitated by the actual absence of probabilistic construction modes which we have already observed in the syntax of English and other languages. If there really were expressions such as the 'prand'-connective, actual language would provide us with clear examples of probabilistically structured phrasal expressions. Under such circumstances, one might reasonably expect a swifter recognition of the non-probabilistic nature of actual phrasal compounds which has been argued for, and a swifter recognition of the fact that the alleged problem of having to "find a prototype" for an allegedly prototype-theoretical/probabilistic compositional structure is a spurious one.

§7
The red herring of
recognitional concepts

In the course of reiterating his compositionality critiques in the theory of concepts, Jerry Fodor has created the impression that more than just openly declared prototype-theoretical, probabilistic approaches to concepts violate compositionality. In particular, he has also criticized what he calls "recognitional concepts" for violating compositionality. He has claimed this violation only relative to a claim that belief in so-called recognitional concepts implicitly commits one to the association of a prototype, or of favourable conditions for recognition, with the concept. Philosophical commentators responding to Fodor have tended to comply with his assumption that recognitional concepts form an important notion, unquestioningly adopting both the notion itself and the assumptions which Fodor has connected to it.[1] The

---

[1] For example, Schiffer (unpublished), Peacocke (2000), Recanati (2002). In fact, the latter goes further in even sharing Fodor's talk of concept-constitutive "epistemic properties" (prominently featuring, e.g., in Fodor 2001) which is still unclearer, and which runs even more non-pertinent issues together than the focus on recognitional abilities already does. Philosophers firmly rooted in cognitive science, such as Laurence and Margolis (1999) or Prinz (2002), have tacitly refrained from addressing Fodor's notion of a recognitional concept in the compositionality context. This may be due to a more principled picture of the relevant concept-theoretical options which one might expect from these quarters.

tendency evinces a surprisingly uncritical stance, given what I want to argue.

The present chapter's general aim is towards topographical clarification and correction in a concept-theoretical area where it is wanting. It will be argued that the notion of a recognitional concept, with which Fodor has identified the notion of a perceptual concept without noticeable protest, is a confused one. Most importantly, it runs percept-concept relations together with a direct concept-world relation, when in fact only the former are immediately content-constitutive of perceptual concepts. Fodor's emphatically reiterated claims about so-called recognitional concepts are baseless too—both as interpreted with regard to the strawman of recognitionalism and with regard to the true notion of perceptual concepts.

The first task, approached in section 1, is to get clear about Fodor's severely underdescribed notion of a recognitional concept itself. It is then argued, in section 2, that Fodor's recurrent pronouncements to the effect that believers in so-called recognitional concepts are committed to the notion of a good instance is baseless and fallacious. More importantly still, it is argued in section 3 that Fodor's notion of a recognitional concept is a red herring that results from conflating the proper notion of a perception-based concept with the entirely distinct idea of a direct concept-world relation. Section 4 argues, in light of the preceding results, that Fodor's association of so-called recognitional concepts with favourable circumstances is, in its own way, equally unfounded as the previously rejected assocation of recognitional concepts with good instances. Section 5 independently makes clear that, contrary to Fodor's suggested topography, the real concept-theoretical notion of a perceptual concept bears no intrinsic connections to prototype theory anyway and does not lend itself to an alleged anti-skeptical undertaking which Fodor persistently attributes in the context.

1. Fodor's notion of a recognitional concept

Fodor stipulatively defines his term 'recognitional concepts' to refer to concepts that are constituted by "recognitional abilities". What exactly these recognitional abilities are supposed to be is something he does not explain. All that Fodor repeatedly stresses is that the recognitional ability in question is that of recognizing a concept's instances *as such*. Take, e.g., his first paper employing the notion:

> …a concept is *recognitional* if and only if ... [a]mong its possession conditions is the ability to recognize at least some

> things that fall under the concept *as* things that fall under the
> concept. (Fodor 1998-I: 35, emph. his.)

An explanation that does not itself employ the term 'recognizing' is never
given. Let me briefly say something about the "recognizing" part of this ex-
planation, then something about the "as such" part of his explanation.

Given any remotely available sense of 'recognition', and given Fodor's uses
of the term throughout his relevant works, recognition can only be under-
stood as perceptual recognition here. But Fodor noticeably desists from talk
of perceiving or perceptually recognizing, employing the term 'recognizing'
almost only generically. Initially, it would be natural to think that we can
ignore this fact as having no deeper meaning, and that we can straightaway
substitute the clear notion of perceptual recognition in place of the superfi-
cially more generic-looking talk of recognizing that Fodor relies on.
Strangely enough, however, Fodor indeed explicitly states the view that the
relevant notion of recognizing is more generic than that of perceptual recog-
nizing. That is, when Fodor once does mention perceptual recognizing in the
vicinity of his first introduction of recognitional concepts, it is presented
merely as *a species* of the genus of recognizing: "Lots of philosophers are
sympathetic to the claim that there are recognitional concepts. For one thing,
insofar as *recognitional* capacities are construed as *perceptual* capacities,
the claim that there are recognitional concepts preserves the basic idea of
Empiricism" (1998-I: 35, emph. his). Fodor's supplementing comment
shortly afterwards confirms that the implied distinction in terms of general-
ity is not accidental: "[T]he claim that there are recognitional concepts is a
bastion of last resort for philosophers who think that semantic facts are con-
stituted by epistemological facts, a doctrine that includes, but is not ex-
hausted by, various forms of Empiricism." (Loc. cit.) What an ability to rec-
ognize objects might consist in that is not a perceptual recognitional ability,
is something Fodor says nothing about. Apart from a joking side remark
about "criteria", nothing is provided.[2]

The suggestion of the availability of a relevant non-perceptual notion of
recognizing is best ignored altogether, since it plays no role in Fodor's ac-
tual discussion of his examples—and we clearly must ignore this if we are
to make any progress in understanding what Fodor's target could be here.
Let us do so.

---

[2] "If you have ever, even in the privacy of your own home among consulting adults,
whispered, hopefully, the word 'criterion', then probably even you think there are rec-
ognitional concepts." (Loc. cit.) The extreme looseness that prevails here as elsewhere
in Fodor's later writings could be more fun if it did not go hand in hand with elementary
conflations and confusions.

As for Fodor's occasionally emphatic 'as such' talk, here as elsewhere, its explicatory-looking emphasis is misleading. It does not serve to characterize the application of the type of concept that Fodor is trying to characterize. In particular, it does not, as one might easily think, mean that the recognitional acts described involve metacognitive thoughts to the effect that the instance recognized falls under the concept under which one is subsuming it (which is clearly what the literal reading of "recogniz[ing] things … *as* things that fall under the concept" would suggest). Fodor's 'as' talk only serves to indicate which concept he is ascribing, respectively. It is used to indicate that a de dicto ascription is intended: that the expression chosen to designate the object of perceptual recognition is the very expression that expresses the concept under which the perceptual object is subsumed by the perceiver. For example, in saying '*S* recognizes a house as such', all that the 'as such' is supposed to mean is that *S* not only perceives a house but also applies the concept HOUSE to it on this basis—in contrast, for example, to perceiving the house and subsuming it under the concept LARGE OBJECT WITH WIN-DOWS, or under the concept MY AUNT'S HOME, or under no concept at all.

What we should do, especially as Fodor has not done it, is to pay explicit attention to the general structure of perceptual recognition and to the potential concept-theoretical import of the notions involved. In chapter 4, it was already indicated that perceptual recognition, in the most general outlines, would at least seem to involve the following:

> (1) One perceptually encounters an object *x*.
> (2) One veridically entertains a perceptual representation *P* of *x*.
> (3) One obtains some sort of correspondence between *P* and a
>   concept *C* stored in memory.
> (4) One correctly applies *C* to *x* in thought.

With this, we at least have something in our hands that is significantly more informative than anything Fodor explicitly says about "recognitional concepts". Fodor must be assuming that recognitional concepts are concepts which are constituted (/whose contents are constituted) by their possessor's disposition to utilize them with respect to their instances as specified by the above quadruple of steps. He is, furthermore, assuming that thus defined recognitional concepts are what just about anybody aims at who wishes to acknowledge the existence of perception-based concepts.

For the duration of the following section, I will be suspending judgment as to whether recognitional concepts are really what theorists with concept-

empiricist leanings rely on or not. Accordingly, I will not speak of recognitional concepts as *perception-based* or *perceptual concepts,* neither will I refer to their imagined proponents as *concept empiricists.* I will simply use the term 'recognitional concept' to refer to the putative kind[3] whose nature we have just made precise and I will call somebody who believes that there are recognitional concepts a *recognitionalist.* The following section occupies a perspective that remains internal to Fodor's treatment of the notion of a recognitional concept. It shows Fodor's claims to be fallacious that recognitionalism is committed to the notion of a good instance.

2. The fallacious limitation of
recognitional abilities to good instances

In spite of leaving the notion of a recognitional ability unanalyzed, Fodor feels confident enough to assure us that the ability to recognize instances of a concept as such can only be assumed to be an ability to recognize *good instances* of the concept, or to recognize instances of the concept as such under *favourable ambient conditions.* It is *only* on the basis of their limitation to good instances/conditions that Fodor asserts recognitional concepts to violate compositionality. (This also happens to be reflected in the fact that, in more cases than not, the examples of recognitional concepts which he gives in reapplying the compositionality arguments are the same ones as those he gives in the context of prototype theory.) Here is a representative quote. It is as explicit as matters can get in terms of stating the connection—though not in terms of justifying the assertion of this connection:

> [F]or all intents and purposes, the notion of a recognitional concept is hostage to the notion that concepts are constituted by their stereotypes.[4] Here's why. Nobody could (and nobody does) hold that the possession of a recognitional concept requires being able to identify *each* of its instances as such; if that *were* the requirement, then only God would have any recognitional concepts. So, the doctrine must be (and, as a matter of fact, always is) that possession of a recognitional concept requires the ability to identify good instances as such in favourable conditions. […]. But now, unsurprisingly, the ability to recognize *good instances* of *F*s doesn't compose, and this is for

---

3   Or "putatively putative", to be more precise, but I do not want to verbally complicate matters.

4   Sic. The use/mention oscillation exhibited in this passage with respect to the term 'prototype' (or 'stereotype', in this case) has to be endured both in Fodor's work and in much of the psychological literature. However, this does not constitute one of the more harmful loosenesses in the present context, and we do well to ignore it. On the use of 'prototype' that I try to adhere to, a prototype-theoretical concept is not constituted by its prototype, but by a structure that is centered on a representation of that prototype.

exactly the same reason that *knowing the stereotype of F doesn't compose* […]. (Fodor 1998-I: 38-39, emph. his.) [5]

Notice that Fodor is not, neither here nor elsewhere, talking about limitations on recognitional abilities that are due to a *malfunctioning* of the mechanisms implementing perceptual recognition, be it the hard-wired process that leads to the tokening of a percept, or be it the process of retrieving the concept whose satisfaction the veridical occurrence of such a percept would entail. All that Fodor is doing is to assert, as an obvious fact, that recognitional abilities cannot stretch to all instances of a concept, nor to all kinds of perceptual conditions. He does not give us any explanation as to the kinds of cases he might have in mind here.

It seems intuitively clear that there are limitations on recognitional abilities which are due to *bad ambient conditions* for recognizing a perceptually encountered object. An example which we can assume to be representative of Fodor's rationale here is that you cannot be expected to recognize cats crouched together at a distance in a dark night.[6] The idea would have to be that a percept generated by looking at such a temporal cat slice is thoroughly dissimilar from the percepts whose generation allows one to recognize that cat. However, the two very general options of organismic malfunctioning and unfavourable ambient conditions already would just about seem to exhaust the respects in which concept-constitutive recognitional abilities might be limited.

2.1. Unjustifiedness of the limitation. Just what makes Fodor think his single most salient claim to be justified that you cannot be able to apply a perceptual concept to *each of its instances?* We are left completely in the dark about this. Notice that the truth of the claim that concept-constitutive recognitional abilities are limited to good instances of those concepts is *not even initially suggested* by a limitation of recognitional abilities to favourable ambient conditions (nor even with proper organismic functioning,

---

[5] Here are some more of Fodor's bridging statements to this effect (all emphases his): "The notion of *a recognitional capacity* … is connected to the notion of a *good* (in the sense of a typical, or an epistemically reliable) instance; the best that a recognitional capacity can promise is to identify good instances in favourable conditions." (Fodor 1998-I: 42.) "That *goodinstancehood* doesn't compose is, I think, the ineliminable fly in the empiricist's ointment." (Fodor 1998-II: 51.) "Epistemic capacities are internally connected to the notion of a *good instance:* Except maybe [for] God, having a concept clearly does not require being able to recognize its instances *come what may*." (Fodor 2001: 8-9.) Fodor (1998-a: 105) includes almost literally the same wordings as the one displayed in the text.

[6] For example, Fodor (2004: 38) uses the example that favourable circumstances for recognizing a blue thing being "in broad daylight".

which Fodor does not mention anyway). The question of whether one is dealing with an untypical, fairly typical, or prototypical instance of a cat has nothing to do with, e.g., whether a given cat crosses your visual field in broad daylight or in the darkness. Fodor's treatment is seriously defective in suggesting otherwise.

It is not just that we have not been given a justification in the present, characteristically casual assertion of the connection by Fodor. It is rather, as I want to show now, that Fodor's limitation of concept-constitutive recognitional abilities to good instances of the concept is positively fallacious. Let me first address a consideration hinted at elsewhere by Fodor in support of the present limitation. The consideration can hardly even be taken seriously, as we shall see. I will then make the more principled point that the limitation of concept-constitutive recognitional abilities to good instances of a concept is, to all intents and purposes, incoherent.

A textually supported interpretation as to what cases Fodor might have in mind with his limitation of recognitional abilities to good instances is implied by the parenthetical remark within the following explanation: "I guess I know what 'triangle' means. But it's certainly not the case that I can tell, for an *arbitrary* object in an *arbitrary* situation, whether 'triangle' applies to it. (Consider triangles outside my light cone.)" (1998-II: 51.) From the thus stated fact, he immediately infers that one can at best be required to be able to recognize "a good instance of triangles [sic] in circumstances that are favourable for triangle recognition", alleging that nobody disputes this and that nobody should.

The connection which Fodor forges here is absurd. For the parenthetic explanation which he gives rests on the platitude that there are instances of a recognitional concept the perception of which is impossible due to their *spatiotemporal remoteness*. But examples of triangles that lie beyond the reach of how far light can travel from my present position within the universe's lifetime are beside the point. Contrary to what Fodor insinuates, it is obvious that no intrinsic difference, and, a fortiori, no difference as to their intrinsic *typicality,* is entailed by the difference between triangles which are too far away for me to perceive them consistently with the laws of physics being as they are, triangles which are just around the corner but which I simply happen not to be looking at, and triangles which I do happen to be perceiving.

In fact, it has to be said that cases like these are clearly not indicative of any limitation of recognitional dispositions for triangles at all. The platitude that

most of the countless instances of a given perceptual concept in the universe are never perceptually encountered does nothing to diminish the counterfactual reliability of the perceptual-recognitional *disposition* in question. Fodor's present explanation of the limitation of recognitional abilities is fallacious tout court, because it depends on a wrong understanding of the antecedent of the counterfactual that is intended in ascribing a perceptual-recognitional disposition. In ascribing to someone the ability to recognize instances of her perceptual concept, what is meant is her ability to recognize *perceptually encountered instances* of that perceptual concept. What is certainly *not* meant is this ability *plus* an *ability to bring about a perceptual encounter* with whatever instances of the perceptual concept happen to exist in the universe. It is a patently silly idea to run together an organism's locomotive powers for eventually encountering a triangle in another galaxis with the reliability of the organism's cognitive-perceptual disposition to recognize triangles.

Put in terms of the earlier quadruple, an ascription of a recognitional ability with respect to *x* ascribes a disposition to the effect that, given the perceptual encounter stated in 1, one can get to the concept application stated in 4, via a reliable causal mechanism that leads there over steps 2 and 3. It is the reliability of the transitions starting from the perceptual encounter that the reliability of a recognitional ability is to be evaluated against. The fact that one is perceptually encountering the perceptual concept's instance is *presupposed* in the ascription of the disposition to perform these transitions; in particular it is presupposed already in ascriptions of sensory-cognitive malfunctioning or in ascriptions of perceptually unfavourable ambient conditions.

2.2. Incoherence of the limitation. The deeper embarrassment for Fodor's much-reiterated claim that recognitional abilities are limited to good instances is that the limitation is incoherent. Let us take for granted, as Fodor does, that we can move freely between talk of possession conditions for concepts, of concept-constitutive conditions, and content-constitutive conditions for concepts—the latter two's equipollence being warranted by the fact that concepts are simply typed by their contents, in contexts of investigation such as the present one.

Now, *if* someone assume a concept's involvement in a recognitional disposition to determine the content that it has, then she thereby assumes this disposition to determine whatever counts as an instance of the concept—*whether good or bad*. After all, bad instances are instances too. But then it is incoherent to think that, even relative to ideal ambient conditions and flaw-

less perceptual-cognitive functioning, you might perceive a bad instance which you are simply not able to subsume under your recognitional concept. Because it is, by assumption, this very ability that makes the bad instances count as instances of the recognitional concept too. You can limit a concept-constitutive recognitional ability to recognition of good instances in unfavourable conditions. You can limit it to recognition of both good and bad instances in favourable conditions. But you cannot coherently limit the concept-constitutive recognitional ability only to good instances perceived in favourable conditions, because in that case you would not be able to recognize bad instances at all—which, by assumption, you have to if the concept subsuming them is supposed to do so in virtue of a recognitional ability.

I suppose one might protest here that recognitional abilities need not be exhaustively constitutive of a recognitional concept's content, on Fodor's definition of the notion, and that one might try to find a loophole in the exact wording of the first quote I gave from Fodor. He said there that a recognitional ability with respect to some of a concept's instances is only *among* a concept's possession conditions, not that they form the sole possession condition for a recognitional concept. Might Fodor, then, mean to say that a recognitional ability is non-exhaustive of a concept's content-constitutive conditions insofar as it determines what count as good instances but leaves it to a different factor what count as non-typical instances? Surely not, one should think—what kind of idea would that be anyway? Surely, the non-typical instances are not supposed to fall under the recognitional concept in virtue of content-constitutive conditions of an entirely different sort than the one that is responsible for the typical instances falling under the concept. Surely, the non-typical instances fall under the recognitional concept because they are, in optimal conditions of broad daylight, *perceivably similar to the typical instances*.

It is relatively clear, then, to repeat, that recognitional concepts are in no way hostage to the notion of a prototype. The upshot is that Fodor has at best pointed to the fact that recognitional abilities are limited to favourable perceptual conditions. His central, massively advertised claim about the logical topography—in effect, the claim that recognitionalism is some sort of crypto-prototype-theoretical position—is simply confused.

3. Perceptual concepts are not recognitional concepts, and recognitionalism is an inherently unstable strawman

I proceed to argue that the notion of recognitional concepts is a strawman which Fodor has installed in place of the real thing. The way that Fodor

characterizes the notion of a recognitional concept, and the explanations which he connects to it, only add confusion to a notion which is fundamentally defective to begin with. Fodor claims to capture perception-based concepts with his addressal of recognitional concepts. But recognitional concepts bear little resemblance to anything one might rationally wish to aim at in endorsing perception-based concepts. Recognitionalism is a strawman in that it conflates the idea that is truly essential to perceptually based concepts with a commitment to a content-constitutive causal relation between a concept and its instances. Let me explain.

Take the relatively a priori statable steps into which perceptual recognition was broken down earlier. Let us speak of a sensibly maintainable notion of a perception-based concept as that of a *perceptual concept,* as opposed to the notion of a recognitional concept which Fodor is assuming to be germane. Something is a perceptual concept just in case it is constituted by relations to sensory-perceptual representations. More specifically, following Jesse Prinz (2002: 108-09), we can assume that perceptual concepts are stored copies of perceptual representations, or stored instructions for producing copies of perceptual representations—whatever this exactly amounts to. The point is that, whatever exactly this relation is, it is *this* relation that is constitutive of possessing a perceptual concept. Similarly, it is *this* relation that endows a perceptual concept with its content, derived from the most basic intentional content of sensory-perceptual representations.

The present characterization of perceptual concepts contrasts heavily with the suggestion that perceptual concepts have their contents constituted by their role in the comparatively sprawling causal-dispositional chain of perceptual recognition which is characterized by (i) through (iv). Perceptual concepts simply have their contents constituted by whatever causal relation underlies the percept-concept correspondence referred to within (iii).

In the first instance, assuming the existence of perceptual concepts is tantamount to assuming the existence of appearance-based concepts—concepts of objects typed according to their superficial appearances, or concepts of perceivable properties themselves. Examples of these are legion. By way of concepts of perceivable properties, take the concept RED, the concept BUMPY, the concept LOUD, or the concept RIGHT-ANGLED. As for appearance-based concepts of objects, consider HOUSE, or TREE, or CAT, or FISH, where these are not understood to be concepts of categories the membership in which is determined by hidden properties, but where these are simply concepts of objects which have certain outwardly observable properties. Superficialist classification matters enormously to our everyday classification

of things in the world. Let us assume, to flesh out an example, that someone's FISH concept is constructed from the following perceptual feature representations. The respective feature descriptions in small capitals are merely convenient linguistic ways of approximating what are not phrasal concepts, but (in this case) visual representations which have been copied into memory—be it visual representations of a characteristic motion pattern, of salient body parts, or of the overall shape. I add a subscript reminding us of the fact that we are dealing only with the perceptual, appearance-based FISH concept here.

FISH$_p$
- HAS FLATTISH BODY WITH VERTICALLY ORIENTED EDGES
- GLIDES THROUGH WATER
- HAS FINS
- PROPELS ITSELF AHEAD BY FLAPPING TAIL FIN
- HAS LIDLESS EYES
- HAS PROTRUDING ROUND LIPS
- HAS SLICK SKIN

If a concept is assumed to be decomposable into stored copies of modality-specific mental representations with the respective contents listed here, then it is these constituents that form the source of FISH$_p$'s content, full stop.[7] The appearance-based concept has its content from the representations listed, which in turn have been extracted from occurrently entertained percepts. The percepts themselves have to get their contents from some sort of nomic relation to the features in the world that are represented. But when it gets to how the appearance-based concept gets its content, *there is no additional, direct concept-world relation to attend to over and above the content-transferring chain from represented properties to percepts, percepts to stored copies, and from stored copies to the concept embracing them*. Perceptual, appearance-based concepts are representations with derived intentionality; they derive their contents from the intrinsic intentionality of sensory-perceptual representations.

The notion of a recognitional concept which Fodor attributes in place of perceptual concepts *runs together two distinct ideas of content constitution under the umbrella of the unanalyzed notion of perceptual recognition*. Per-

---

[7]   Near enough, at least. These feature representations are what gives the concept FISH$_p$ its specific content. Since FISH$_p$ is, superficiality notwithstanding, supposed to be a concept of a mind-independently individuated type, the concept would have to be framed by implicit ontological assumptions to the effect that fish$_p$ are persisting objects in the world which exist independently of our perceiving them, and that fish$_p$ can appear quite differently to us depending on different angles or lightings.

ceptual recognition necessarily involves percept-concept causation as embedded into instance-concept causation. But the idea of perceptual concepts does not require this embedment. Fodor grafts the idea of a directly content-constitutive concept-world relation onto the self-sufficient concept-empiricist idea of content-constitutive percept-concept relations. One of his first remarks on the notion of a recognitional concept documents this misidentification very clearly. Please give another moment of consideration to a remark I had already quoted earlier. This time, I expand the portion quoted by a remark from Fodor about content-constitutive concept-percept relations which I had held back before (emphases added this time; Fodor's own omitted):

> [I]nsofar as *recognitional capacities* are construed as perceptual capacities, the claim that there are recognitional concepts preserves the basic idea of Empiricism: that the content of at least some concepts is constituted, at least in part, by their *connections to percepts*. (1998-I: 35.)

Look at it this way. Either you think that the fact that FISH$_p$ represents the superficially individuated kind that it does is due to a (lawful) causal instance-concept relation obtaining. Or you think that the fact that FISH$_p$ represents the superficially individuated kind that it does is determined by a construction out of the above-listed visual representations, which result from storing in long-term memory certain saliencies and invariances from a certain range of percepts. But not both. This, however, is precisely what Fodor's ascription of concepts with contents constituted by recognitional abilities entails.

We have all but mentioned by name the family of mental content theories to which the causal instance-concept relation which we have just been talking about leads. Recognitionalism's component of a content-constitutive causal relation between instances of a concept and concept occurrences is tantamount to the base component of any *causal-covariational theory of mental content*. Needless to say, Fodor's past attempts to naturalize content belong to this familiy of theories (Fodor 1987: ch. 4; Fodor 1990: chs. 3 and 4). Now, as no one has emphasized more than Fodor himself, once one assumes the causal-covariational relation itself to be content-constitutive, it does not matter how this relation gets mediated—Fodor and Fodorians like to say that the putative lawful relation between a concept and the property it expresses allows for a multiplicity of different *sustaining mechanisms* (e.g., Fodor 1998-a: 75-80, Margolis 1998). In particular, it does not matter whether the sustaining mechanism (the "mechanism of semantic access", as Fodor also calls it) is mediated by perceptual representations or by theoreti-

cal inference, or by something else. All that matters is the obtaining of the lawful causal instance-concept relation itself, as these authors emphasize. But if this is so, then why should the alleged believers in recognitional concepts assume perceptual feature representations to be content-constitutive of a concept at all, in addition to the disposition for instance-concept causation?

The situation seems to be this: Either the alleged recognitionalist insists that the concept-world relation is only content-constitutive as mediated by certain percepts (percepts representing features within a certain range of similarity). In this case, though, it would have to be maintained that the relation of instance-concept causation which brackets the crucial percept-concept relation is not the true source of content at all. In this case, recognitionalism collapses into *concept empiricism,* i.e., the assumption of percept-based concepts previously distinguished from recognitional concepts. Or the alleged recognitionalist maintains that the relation of instance-concept causation is indeed the only properly general relation that can be appropriately regarded as content-constitutive, in which case, however, it is the percept-concept relations that turn out not to be content-constitutive at all. In this case, recognitionalism collapses into *causal-covariational semantics.* These are the two distinct ideas of content constitution which Fodor runs together under the umbrella of the unanalyzed notion of recognitional concepts.

To summarize: First of all, recognitional concepts are not perceptual concepts. Secondly, the notion of a recognitional concept is not even a clear idea; it collapses either into the notion of a perceptual concept, or into the notion of a covariationally constituted concept.

4. Recognitional concepts, covariationism, and favourable circumstances

Now, it is, embarrassingly enough, precisely the *covariational aspect of recognitionalism* in light of which Fodor forges the connection between recognitional concepts and *favourable conditions* of (perceptual) recognition—the connection which makes his notion of recognitional concepts prima facie amenable to the compositionality arguments.[8] Recall how we saw that Fodor's reasoning for the commitment of recognitionalism to good instances is fallacious. The claim that recognitionalism is committed to favourable ambient conditions for perceiving, by contrast, did not appear to be similarly confused. But the ironical fact of the matter is that, if recognitional concepts

---

[8]   If not secunda facie, given considerations of the kind put forth in chapter 6.

are committed to favourable circumstances for recognition, they are no more or less so than causal-covariational theories of content themselves are.

In fact, this will remind the informed reader of something. Whether and how to deal with references to favourable situations in the statement of content-constitutive causal concept-instance laws is the recurrent leitmotiv of the literature on causal-covariational theories of content (see, for example, Cummins 1989). In fact, it is so right down to the favourableness of conditions for *perceptual* recognition, since it is in practice only applications of concepts to perceivable objects that discussions of covariational theories are run over.[9] In the context of the covariationism literature, favourable circumstances are brought into play as a potential means of dealing with what Fodor (1987: 101-02) calls the "disjunction problem", or the "problem of misrepresentation": the challenge to the causal covariationist which consists in the fact that *not only* its instances cause occurrences of a concept, but that other perceptually encountered distal objects which it does not represent do so too. Favourable conditions, here, are postulated conditions in which *only* instances of a concept cause it to get tokened. In the context of the notion of recognitional concepts, by contrast, Fodor brings favourable conditions into play as a means of dealing with the fact that *not all* perceptually encountered instances cause occurrences of a concept. Favourable conditions, here, are postulated as ones in which *all* instances of a concept cause it to get tokened. But this just fits the other direction in which a covariationalist might want to secure her content-constitutive biconditional, which says that a concept represents all and only those things that it is caused by.

Let me draw the first conclusion from the last couple of paragraphs: It is only in virtue of its causal-covariationist side—which strangely goes unnoticed as such by Fodor—that Fodor ascribes to recognitionalism a commitment to favourable conditions for perceptual recognition. Given the previous exclusion of a connection to goodinstancehood, commitment to favourable perceptual conditions forms the only remaining basis for Fodor's application of the compositionality critique to the alleged notion of recognitional concepts. But given the way that the ascription of this commitment is motivated, what Fodor consistently would have to do is simply to openly criticize causal covariationists who are committed to favourable conditions of perceptual recognition. Another perspective you could take on this rests on the fact that Fodor happens to have maintained a covariational theory of content in the past which he claims does not rely on an idealization to fa-

---

[9] The reason being, I have to add, that the idea is so outlandish when applied to anything other than perceivables that one is happy to deal with the problems that concepts of perceivables bring up alone. But nothing more of that here.

vourable conditions.[10] If there really is such an option, then the imagined recognitionalist, whose position partly consists in a fully-fledged crypto-covariationism, should be offered the option too.

Quite apart from this, though, there is another, even more fundamental point to make here: Since Fodor's critique really rests on the causal-covariationist aspect of the alleged notion of recognitionalism, and since covariationist content constitution is inherently limited to conceptual primitives—it is in its nature that it does not involve content-constitutive relations to other representations—, there is really nothing here for any compositionality argument to apply to anyway. Covariationist content-constitutive relations are ipso facto representationally atomistic, as one might put it. They thus cannot be understood to apply to compositional compounds in the first place, since compositional compounds, by definition, get their contents from constituent representations. Covariationally constituted representations do not have content-constitutive representational parts. Thus, Fodor's idea of directing arguments from the accomodation of compositional compounds against recognitionalism seems to rests on incoherent presuppositions.[11] Its connection of recognitionalism with a commitment to favourable conditions is made in light of the covariationist aspect of recognitionalism, covariationist theories of content are atomistic, and atomistic theories are ipso facto not meant to capture compositional compounds.

If the arguments of the foregoing sections are correct, Fodor's notion of recognitional concepts and his treatment of them are a complete shambles. The aim of this chapter has been almost accomplished in the course of showing this up to here. But there is still a thing or two to set straight about perceptual concepts.

5. Perceptual concepts, reprise

Shifting back to the notion of a perceptual concept which recognitional concepts were really supposed to capture, in some sense or other, two further clarifications are in order. The first point is a reminder of the fact that properly conceived perceptual concepthood bears no intrinsic connection to prototype-theoretical concepthood anyway, and a reminder of the fact that this

---

[10]   Though see Boghossian (1991), who argues that this theory of Fodor's, the "asymmetric dependence theory" of content, does.

[11]   A surprising actual case where the presuppositionally ill-founded move of applying the compositionality argument against Fodor's atomistic/covariationist theory is performed is the tu quoque in Laurence & Margolis (1999: 67). As argued in chapter 6, however, the arguments from compositional phrasal structures hardly make any more sense in their application to probabilistic, prototype-theoretical constituent structures.

is uncontentious. The second point is that, contrary to another one of Fodor's bizarre and persistent conflations to the contrary, the positing of perceptual concepts is entirely independent of pursuing an epistemological, anti-skeptical agenda.

5.1. Independence of perceptuality and probabilistic constituency. Perceptual concepts do not bear any intrinsic connection to prototype theoretical concepthood anyway. It is one of the many unclarities of Fodor's treatment that he always speaks of concept-constitutive connections to prototypes without ever making explicit whether he takes this to imply bona fide prototype-theoretical concepthood or not. Whether this implication is intended is unclear particularly in light of the fact that prototype-theoretical concepthood is prototype-theoretical *structuredness* (as everybody knows, and as Fodor is one of the quickest to emphasize), but that Fodor nonetheless never even mentions concept structure in connection with recognitional concepts. It is not really clear how he could regard recognitional concepts as something other than prototype-theoretical concepts. For one thing, he seamlessly transfers examples from the discussion of prototype theory to the discussion of recognitional concepts. But perhaps this does not mean much. For another thing, it is not clear how a prototype association is supposed to be content-constitutive if not in the way of a structured representation whose satisfaction is gradable across individually defeasible constituents, centered on a prototype. Surely the prototype has to be a psychologically represented one, and surely it cannot be unstructured?

In any case, the point to make about perceptual concepts is that they bear no intrinsic connection to the notion of a prototype because *the question of whether a concept is structured from perceptually derived representations or not is independent of the modal status of the relations that the concept-constituent relations encode.* That is, the composedness of a concept like FISH$_p$ from perceptually derived representations is, of course, equally compatible with the definitionist idea that these constituents encode individually necessary conditions for the satisfaction of FISH$_p$ as it is compatible with the prototype-theoretical, probabilistic idea that these constituents encode merely statistically probable conditions for the satisfaction of FISH$_p$. In fact, it is standard textbook wisdom, when it comes to the history of the predominantly philosophical "classical theory", or concept definitionism, that it was held by empiricists no less than by other theorists.[12] This universally accepted assumption of the mutual independence of conceptual constituents'

---

[12] The logical empiricism of the Vienna Circle is a prime example. John Locke is implicitly attributed a definitionist empiricism by Laurence and Margolis (1999: 14); Prinz (2002: 34-35) prefers mentioning David Hume to this effect.

properties of perceptuality and individual defeasibility should in itself suffice to make a reader of Fodor's works extremely suspicious about the fact that he criticizes an empiricist idea by means of its allegedly essential association with a prototype.

5.2. Fodor's misattribution of an anti-skeptical agenda. As a final point in the progressive disclosure of Fodor's tangle of confusions, consider the metatheoretical perspective from which Fodor introduces recognitional concepts, and from which he persistently describes the import of his discussion. From this perspective (e.g., 1998-I: 35-36), the postulation of recognitional/perceptual concepts is described as being epistemologically motivated: as being motivated by an *anti-skeptical argument* for the existence of these concepts' instances. This is a breathtakingly sweeping conflation of entirely distinct projects in epistemology and in the philosophy of psychology. Concept theorists deserving of this description are in the business of telling us what kinds of concepts there are and what their nature is. They are in the business of trying to say something about the subpropositional mental representations which we have stored in long-term memory, utilizing them in non-hardwired cognitive processes. What they are not doing, however, and rightly so, is to pursue epistemological projects. In particular, they do not, qua theorists of concepts, care about whether the appearance-based concepts they posit are of any help to the epistemologist in launching arguments against skepticism about external objects, or whatever exactly Fodor has in mind.

And how could a concept theorist think so anyway? The postulation of concepts of external objects, whether appearance-based or theoretical concepts, merely *recaptures,* among other things, general background assumptions to the effect that there are mind-independent, stable objects with these appearances, existing independently of whether we perceive them or not. It is not supposed to *justify* this claim. Moreover, to reapply a pattern familiar by now: Why does Fodor not likewise pursue the idea that causal-covariationist theories of content provide an argument against skepticism? He could with equal justification. Which is only to say, with none.

§8

The innocence
of lexically confined
prototype theory

---

1. Lexical confinement
and the path ahead

Contrasting with the Scope Assumption rebutted in chapter 6, anyone sympathizing with prototype theory should properly maintain what we might, loosely labelled, call the position of *lexical confinement* (LC) for prototype theory. Let me define this position in a generic way:

> **(LC)** Concept theory X captures concepts expressed by lexical expressions without capturing the phrasal concepts expressed by compositionally interpreted phrasal constructions out of these expressions.[1]

Please note that my label 'lexical confinement' sacrifices accuracy for the sake of brevity. Prototype theory still can provide the concepts expressed by *non*compositional interpretations of phrasal expressions.[2] One other terminological note: Whenever I speak only of "lexical-level" or "lexical" prototype theory in this work, and, analogously, whenever I speak of "phrasal-level" or "phrasal" prototype theory, I mean to refer to prototype theory as applied to the lexical level or to the phrasal level, independently of whether it is applied to the respectively other level too.

---

[1] At least in this canonical statement of LC, I deliberately avoid speaking of compositional *concepts* themselves, because this contrasts with noncompositional concepts, and, contrary to the situation with linguistic expressions, I have to confess at this point that I do not know what a noncompositional *concept* should be. This a story of its own, which I develop in a paper still to be finalized, but I should briefly say something about this here. It is not clear how a concept that is expressed by a noncompositional phrasal expression should, or even could, *itself* be noncompositional. For to be noncompositional, by definition, that concept would have to have a content that is not determined by its component states. But clearly one should assume a structured, intrinsically contentful conceptual state to have as component states simply whatever states are needed for it to have the intrinsic content it does; not only *need* it not mirror the constituent structure of a noncompositional phrasal expression expressing it, it plausibly *cannot*.

Though I take this matter to be important and to call for rather deep-reaching revisions in the received perspective on compositionality in the philosophy of mind, I have tried to steer clear of it altogether within this dissertation and will continue to do so. The only reason why I mention it here is to preempt protest at the convoluted-sounding way of talking of "concepts expressed by phrasal expressions on their compositional/noncompositional understandings," and to avert the misunderstanding that this description can accurately be replaced by talk of "compositional/noncompositional concepts". If I ever happen to employ the latter talk, this will be rather out of fear of coming under too much pressure of having to explain myself, for which there is no space here apart from this footnote.

[2] Plus, as noted earlier, it is in principle possible that a kind of prototype theory provides the concepts expressed by compositionally understood phrasal expressions in some conceivable language. But I do not wish do complicate matters too much here.

LC is where the agenda of this chapter sets in. It was pointed out earlier that philosophers, in reply to the compositionality arguments, have typically, but erroneously, denied the Scope Assumption in an ad hoc fashion. This is tantamount to an ad hoc limitation of prototype theory's scope in accord with LC. Advocacies of LC have accordingly lacked an appropriately principled spirit. Equally importantly, however, LC has never begun to be systematically and comprehensively defended against a host of elliptical counterarguments which Fodor has emphatically presented as refuting it, in various loci scattered across his work. The present chapter's aim is to fill this considerable gap. Reconstructing and dissecting Fodor's various arguments both against what is in fact LC, and against the lexical-level concept theory it presupposes, will bring to the surface an astonishing accumulation of red herrings, confusions, and incoherences.

As we shall see in sections 3, 4, and 5, the central constant throughout most of Fodor's otherwise heterogenous lot of arguments is that he does not address or even recognize LC as the thesis which it is, irrelevantly identifying it instead with the denial of a property which I will call 'decompositionality'. Matters are complicated, among other things, by the fact that Fodor's employments of the notion of decompositionality involve substantial tacit shifts and inconsistencies. Once the gratuitous decompositionality confusions have been cleared up and removed from the arena, LC emerges unscathed. So does the lexical-level concept theory which it presupposes, for anything that is argued in this context.

The fact that no remotely cogent or even pertinent consideration against LC can be produced is just what is predicted by principled considerations against a uniformity assumption (UA) about content-constitutive properties, which would entail the falsity of LC. In the later parts of this chapter, sections 6 and 7, I turn predominantly to UA itself. First though, subsection 6.1 rebuts a UA-independent argument which Fodor offers against lexical prototype theory in the context of one of his rare addressals of UA as such. The longer subsection 6.2 rebuts one of the most staggeringly confused arguments which we will come across in this chapter, an argument which constitutes the only one offered by Fodor in support of UA. Section 7 gets down to a principled level in rejecting UA: It gives a reductio argument against UA, then moving on to some more general considerations to the effect that UA is entirely baseless.

The groundwork for the dismissal of UA in the final section 7 has been laid already with our principled and modally strict repudiation of the Scope As-

sumption. But the insights involved in this repudiation will be of occasionally recurrent relevance already beforehand in this chapter. They are relied upon for the first time in section 2, which criticizes a one-off argument of Fodor's. In this argument, by way of exception, he really does address the idea of LC for what it is.

For all that is argued in this work, prototype theory, or probabilistic concept structurism, can be criticized on other grounds, though I most urgently advise anybody to do so on a case-by-case basis. The paramount philosophical objection to prototype theory, however, and the support which Fodor has subsequently provided for this objection, will have been shown to be confused to the bone by the time we finish.

Before we move on, one important preliminary still. I have already pointed out that Fodor has directed the compositionality arguments not only against openly declared probabilistic theories of concepts, but also against what he calls "recognitional concepts". One reason why it was important to expose this notion as a red herring in the preceding chapter is that the majority of Fodor's rejoinders to LC, in his followups to the compositionality arguments, are indeed stated with regard to this radically ill-conceived notion. Given the defectiveness and irrelevance of Fodor's notion of a recognitional concept, we should leave it to the side wherever we can. It only adds obfuscation to a discussion by Fodor which will be seen to be up to its neck in confusion already without it. Accordingly, I will often speak of prototype-theoretical representational structures or concepts in the following, even where the quoted passages discussing LC are nominally about "recognitional" abilities or concepts. That said, in certain cases where Fodor's considerations really do depend on his notion of a recognitional concept itself, I will indeed address them and refute them from within that internal perspective. In particular, I do so in subsection 5.3.

2. Fodor's (1998-I) side argument
from evidential parity against LC

"Simply not credible", Fodor begins his earliest rejoinder to the lexical confinement of prototype theory (1998-I: 41), continuing to reason thus:

> After all, people who have the concept PET FISH do generally have a corresponding recognitional capacity; for example, they are generally good at recognizing goldfish as pet fish. And, surely, being able to recognize (as it might be) a trout as a fish stands in *precisely* the same relation to having the concept FISH as being able to recognize a goldfish as a pet fish does to having

the concept PET FISH. So, how could that relation be constitutive
of concept possession in the one case but not in the other? (Fo-
dor 1998-I: 41)

The claim of evidential parity that Fodor makes here could not be farther
from the truth. It is false that having the compositional phrasal concept PET
FISH stands in exactly the same relation to the ability to recognize good in-
stances of pet fish as having the concept FISH stands to the ability to recog-
nize good instances of fish. Interestingly enough, though he has not taken
back the claim, Fodor has never made the argument again in his work.

Fodor is accusing LC proponents of an untenable double standard concern-
ing their treatments of FISH as compared to PET FISH. He is, in effect, claim-
ing that, prior to the argument from noncompositional prototypes, there is a
parity of evidence for assuming a "recognitional" essence with respect to the
constituents and the compound. But there is obviously not, if you think
about it for a moment. The observations made in chapter 6 put us in a posi-
tion to set the record straight. The compositional phrasal compound PET FISH
has its content determined by PET, FISH, and a conjunctive operation. PET
FISH is, as one can also say, definitionally structured from its openly posited
constituents. It cannot be coherently thought to be even a candidate for an
alternative type of content constitution, such as having its content "recogni-
tionally" constituted (even if any independent sense could be given to the
notion of recognitional content constitution). Or rather, to focus on the truly
relevant notion of prototype-theoretical structural essences, it cannot be co-
herently thought to be a candidate for falling under prototype theory. On the
other hand, if someone wished to consider a concept corresponding to a
noncompositional understanding of 'pet fish', *that* concept might have a
prototype-theoretical meaning indeed—but considering that sort of concept
would be to abandon the topic of the compositionality arguments.

Quite apart from this, it is, notice, already highly implausible on the face of
it to suggest that the compositionally understood phrasal expression 'pet
fish' is similarly reliably and intimately connected to a goldfishlike proto-
type as 'fish' might be connected to a prototype of its own. Only people
who have, e.g., encountered actual pet fish in our culture or who have quite
an amount of background knowledge about pet fish, would know that gold-
fish would qualify as prototypical pet fish. Even apart from the principled
considerations just advanced, not everybody who understands 'pet fish' on
its compositional meaning knows that goldfishlike creatures qualify as the
prototypical instances of this expression. Surely countless people understand
the expression as the binary conjunctive compound it is without being able

to assign such a prototype. When it comes to 'fish', by contrast, the suggestion that different people might have to associate with it a specific prototype-centered mental representation in order to understand the word in the same way has an entirely different standing. Since it is not a phrasal expression, it does not have a trivial compositional understanding on which a content-constitutive prototype association would be ipso facto precluded.

3. The decompositionality muddles, part 1

It is practically impossible to introduce the central portions of this paper—those addressing arguments involving "decompositionality", as I will call it—in a fairly precise way without anticipating much of the substantial reconstructive work that is to follow. Let me try to give a general picture. What we now turn to is a loose cluster of largely elliptical, generally perplexing arguments. The arguments are sometimes internally incoherent, often incompatible with each other, and invariably embedded in even more perplexing attributions to Fodor's opponent. Generally speaking, these arguments are supposed to *supplement the compositionality arguments' critique concerning phrasal concepts with a critique that is aimed at the level of lexical concepts.* The arguments exhibit various kinds of overlaps. Their characteristic common element is that, in one way or the other, they revolve around an allegedly non-trivial property which complements compositionality—*decompositionality,* as I will call it.[3] This will be defined presently, in accordance with Fodor.

It is for good reason that, even though the prospect of a critique of LC is what motivates us to look into these matters, I have not described the decompositionality-related arguments as arguments *against LC*. There are two fundamental reasons why this description would be inappropriate, and they will become clear as we move along. For one thing, as we will see, Fodor never succeeds in recognizing LC as the thesis it is. This failure is, in fact, of a piece with his repeated launches of decompositionality-related arguments. For it is a misidentification of LC with a denial of decompositionality that lies at the heart of these moves—though the misidentification is anything but exhaustive of the errors he makes in the course of his arguments.

---

3   Fodor himself does not use a unitary term. In fact he usually does not use a term for this property at all, except in Fodor (2000: 371) and Fodor and Lepore (2001: 59), where he speaks of "reverse compositionality". This term has caught on. But it is a bit too cumbersome, given the various composite nouns we will still have to form which involve reference to this property. More importantly, it is also a misnomer, since it is not supposed to be a form of compositionality at all. It is simply intended as a complementary property that hangs together with compositionality.

For another thing, even as considered under the veil of Fodor's misconstrual, LC is not the only idea that he repudiates. Fodor's decompositionality-related considerations often blend criticism of the LC, which is, misconstrued or not, a *negative claim* stating the confinement of the concept theory in question, with criticism of the prior, *positive claim* of the concept theory for lexical concepts which is presupposed by LC.

The relevant loci which call for examination are Fodor (1998-I: 44-46), Fodor (1998-II: 51-53), Fodor (2000: 371-372), Fodor (2001: 8-9), and, in yet again a very different employment of decompositionality than in the aforementioned places, Fodor and Lepore (2001: 59-60).

Fodor (1998-II) expounds a decompositionality dialectic which is, in some respects, something of a standard for the others to be compared with. It is still the most stringently presented one; its moves are relatively clearly flagged; and they recur elsewhere too. By contrast, practically each of the relevant passages from the other papers involves at least one crucial one-off notion of its own. It is, consequently, Fodor (1998-II) that we are best advised to start out with and which we will take the closest look at.

3.1. Reconstructing Fodor's (1998-II) decompositionality dialectic. Fodor formulates two versions of the inverse principle of decompositionality which he maintains for compositional compounds, versions to which he often assigns importantly different dialectical roles. This difference is encountered most clearly in the dialectic developed in his (1998-II). One of these formulations makes reference to the *content-constitutive properties of concepts,* or—warranted by the uncontentious essentialness of contents to the typing of concepts—to *concept-constitutive properties,* as Fodor also sometimes says. A second formulation of decompositionality makes reference to *possession conditions for concepts*. In faithfulness to Fodor, the principles would have to be defined as follows:

> **Content-constitution (CC) decompositionality:**
> The content-/concept-constitutive properties of a compositional compound's parts are fully "contributed to"/"transmitted to"/"inherited by" the compositional compound.
> **Concept-possession (CP) decompositionality:**
> Possessing a compositional compound entails possessing its parts.

The scare quotes I use in the statement of CC-decompositionality indicate that I am only recording Fodor's own quasi-canonical terminology here. He

does not treat the talk of "transmitting" and "contributing" as being in further need of explanation: He does not utilize any more basic terminology to express the relation between the content-constitutive properties of a compound concept on the one hand and those of its constituent concepts on the other. It is very tempting, in light of this, to give what seems to be the only permissible clarification of Fodor's phrasings to begin with. We cannot yield to the temptation, however, because doing so would immediately transform it into a thesis which Fodor could not consistently treat as the distinct assumption from CP-decompositionality *as which he in fact treats it*. I turn to the matter in a later subsection.

What I will call the *decompositionality dialectic* of Fodor (1998-II) will be reconstructed below in a way that remains faithful to his perspective. Even where I take basic errors to be involved in the story recounted, the objections will be articulated only at a subsequent stage. I mention this to avert misunderstanding. I caution the reader that, as in Fodor's other papers which are to be examined later, we will confront bizarre ways of reasoning. But it is indispensable for us to secure a clear rendition of Fodor's own picture of the situation first. Let me also point out that the fairly continuous provision of quotes in this context is deliberately intended. It is the only way of pre-empting the suspicion that the confusions and misattributions to be made out in Fodor's line of argument cannot have really been made by him, and that they must rather be due to some grossly unfair reading. The opposite is the case. The dissection of Fodor's considerations on these pages follows the hope of finding at least *some* kind of prima facie sensible, pertinent argument in Fodor's rejoinders against LC and lexical-level prototype theory.

The starting point of Fodor's argumentation consists in claiming, early on, the joint principle of CC-decompositionality and its compositionality counterpart (1998-II: 50). Shortly after that, Fodor begins to focus on CC-decompositionality, claiming that, in addition to the violation of compositionality, prototypes are guilty of violating CC-decompositionality. His formulations are clear about the fact that he takes this argument to count specifically against prototype theory for *lexical concepts*. Here is a passage that has Fodor restating his conclusion with special emphasis.

> *The constitutive properties of an expression include only the ones it contributes to its hosts. But, in the general case, expressions don't contribute their good instances to their hosts (being a good instance of a fish isn't necessary for being a good instance of a pet fish).* […] I think that's a pretty damned good argument that the epistemic properties of lexical items aren't

constitutive of their identity qua linguistic. (Fodor 1998-II: 52;
italics his, here and in the ensuing quotes.)

Do not be distracted by the fact that, in this as in later quotes, Fodor speaks
of "constitutive properties of a linguistic expression qua linguistic". This
talk is quite inappropriate, since what we are interested in are the content-
constitutive properties of concepts, rather than what gives linguistic expres-
sions their meanings; much less even are we interested in what makes some-
thing a linguistic expression. The inappropriateness, however, is best ig-
nored right now. We do best to pretend for now that Fodor is speaking of
constitutive properties of concepts rather than of "constitutive properties of
linguistic expressions qua linguistic". (I will return to an expression-concept
difference, though, in considering the status of decompositionality in section
4.) It is concepts that prototype theory is a theory of. And indeed, in other
works, as we will see, Fodor discusses decompositionality directly with re-
spect to concepts. Fodor quite generally takes great liberties in moving
freely forth and back between talk of words and talk of concepts.

Fodor's verdict above relies on an *argument from non-decompositional pro-
totypes,* or *the decompositionality argument,* as I will also call it, which he
never really spells out and takes for granted. (NB: As I am using these
terms, the decompositionality "argument" presents the initial move of the
decompositionality "dialectic".) The intended argument from non-decom-
positional prototypes is analogously structured as the argument from non-
compositional prototypes. I give an explicit reconstruction below, with the
prototype-theoretical idea inserted and stated in appropriate explicitness. I
am using the term 'decompositionality' without modification in these lines.
We could also speak more specifically of CC-decompositionality. I refrain
from doing so because the need for this specification arises only against Fo-
dor's later introduction of the other, CP-formulation of decompositionality,
together with Fodor's assignment of a different role to it.

> **The decompositionality argument:**
> **1.** The compositional compound PET FISH inherits its content-
>   constitutive (essential) properties from its constituents PET
>   and FISH. (*CC-decompositionality fulfilled.*)
> **2.** The concepts PET and FISH have their contents (essences)
>   constituted by prototype-centered representational structures.
>   (*Lexical-level prototype theory.*)
> **3.** The compositional compound PET FISH has its content (es-
>   sence) constituted by a prototype-centered representational
>   structure. (*Phrasal-level prototype theory.*)

4. The prototype-centered representational structure of PET FISH does not include the prototype-centered structures of PET nor of FISH. (*Intuitive judgment.*)

5. The compositional compound PET FISH does not inherit its content-constitutive (essential) properties from PET nor of FISH. (*2,3,4.*)

6. ⊥ *(1,5.)*

As noted already, this reductio argument from non-decompositional prototypes is not presented by Fodor as refuting prototype theory indistinctly. Curiously, Fodor takes entirely for granted that the decompositionality argument targets specifically *lexical-level* prototype theory. This is crucial to the decompositionality dialectic.

Though Fodor takes the decompositionality argument to be fairly decisive against lexical-level prototype theory, he now goes on to turn to a position which he claims to be widely occupied in response to the argument just given. Relative to the picture painted by Fodor, this position would amount to biting the bullet: It consists in a flat-out *denial of decompositionality,* more specifically, in the denial of CC-decompositionality:

> However, I have shown this argument to several philosophical friends who disagree. They think, rather, that it's a pretty damned good argument that [CC-decompositionality] can't be true. In particular, so the reply goes, epistemic properties […] are essential to morphosyntactically *primitive* linguistic expressions (like 'red' and 'fish') but *not* to their hosts (like 'red triangle' and 'pet fish') *even in cases where the hosts are semantically compositional.* If this reply is right, then a fortiori, the constitutive properties of a linguistic expression can't be among the ones that its hosts inherit. (Ibid.)

Fodor counters this alleged rejoinder of denying CC-decompositionality with a renewed attempt at a reductio ad absurdum. The reductio, naturally, is one which he takes to be more conclusive still than the one already attempted. Denying CC-decompositionality would commit one to denying CP-decompositionality, he reasons. But, so Fodor seems to think, though someone might take a denial of CC-decompositionality to be a viable option, a denial of CP-decompositionality must be considered untenable by anybody's lights:

> I was … surprised to hear this suggestion so widely endorsed. In particular, I argued like this: Consider *any* property *P* that is constitutive of [constituent] *E* but not inherited from *E* by its

> hosts; I'll call such a property an "extra". If *E* has such an extra property, then, presumably, it would be possible for a speaker to satisfy the possession conditions for a complex expression containing *E without satisfying the possession conditions for E itself:* Somebody who has learned the linguistically essential properties of 'pet fish', for example, need not have learned the linguistically essential properties of 'pet' or 'fish'. […]
>
> But, I supposed, it is something like true by definition that mastering a complex expression requires mastering its constituents, since, after all, constituents are by definition *parts* of their hosts. (Ibid.) [4]

At this point in the putative dialectic, Fodor *stills* sees coherent room for a further round in the dialectic and equips the prototype theorist's denial of CC-decompositionality with a new rejoinder. The rejoinder he attributes makes CC-decompositionality and CP-decompositionality come undone. It consists in *continuing to deny CC-decompositionality* while at the same time *conceding CP-decompositionality,* doing justice to it by simply stipulating prototype theory for the lexical constituents:

> Now, there is a reply to this reply. […] One could just *stipulate* that recognitional capacities (or, indeed, any other sort of extra that you're fond of) are to count as constitutive of the primitive expressions that they attach to *even though* they are not inherited by the hosts of which such primitives are constituents. (Ibid.)

Fodor's response follows on the heels of this, within the same paragraph:

> To which reply there is a reply once again; namely, that explanation is better than stipulation. Whereas principle P [i.e., the joint principle of CC-decompositionality and compositionality] *explains* why meeting the possession conditions for a complex expression almost always requires meeting the possession conditions for its constituents, the proposed stipulation just *stipulates* that it does; as does *any* account of constituency that allows primitive expressions to have extras. (Ibid.: 52-53.)

---

4   The passage bracketed here in fact even acts as if the denial of CP-decompositionality were the opponent's idea from the outset, which would contravene Fodor's preceding exposition: "For, by assumption [of the suggestion "so widely endorsed"], the mastery of 'pet' and 'fish' requires an ability to recognize good instances of each in favorable circumstances; whereas, again by assumption, the mastery of 'pet fish' requires neither of these abilities (not even if it does require an ability to recognize good instances of pet fish in favorable circumstances)." If we are attempting to at least make initial sense of Fodor's course of argument, we must try to ignore these internally inconsistent oscillations.

In other words, Fodor rejects the idea of an ad hoc stipulation which he has just attributed to the prototype theorist allegedly denying CC-decompositionality, and argues that there is no other way to explain the truth of CP-decompositionality than by assuming CC-decompositionality to be true.

This closes off the main part of Fodor's imagined exchange. To exactly which extent Fodor regards it as reflecting *actual* exchanges, and to which extent it only expresses Fodor's preemption of replies he regards as likely, is left somewhat indeterminate. Here is a summary of the dialectic developed as a sequel to the familiar compositionality arguments by Fodor:

> **Fodor's (1998-II) decompositionality dialectic reconstructed:**
> **I.** *Fodor:* Decompositionality argument.
>   → Lexical prototype theory violates CC-decompositionality
>   → Reductio of lexical prototype theory (provisional).
> **II.** *Reply attributed:* Lexical prototype theory retained,
>   CC-decompositionality denied.
> **III.** *Fodor:* Denying CC-decompositionality requires denying
>   CP-decompositionality. → Reductio (still only provisional).
> **IV.** *Reply attributed:* CP-decompositionality admitted by
>   stipulation; CC-decompositionality still denied.
> **V.** *Fodor:* Denying CC-decompositionality renders CP-decom-
>   positionality inexplicable.

The decompositionality dialectic is confused in every step. The confusions underlying step I and, most of all, step II, are particularly fundamental. Once we have gotten that far, the baselessness of the remaining steps III through V is easily recognizable and can be dealt with swiftly. I address these matters sequentially in the following three subsections.

3.2. The myth of a decompositionality argument against lexical proto-type theory (Step I). Let us make clear to ourselves one thing first. As an attempt to refute LC, the decompositionality argument which Fodor elliptically relies on in his paper would be question-begging. The argument from non-decompositional prototypes involves precisely the same pair of premises assuming both lexical *and* phrasal prototype theory as the corresponding argument from noncompositional prototypes does. But, as indicated, Fodor speaks of the decompositionality consideration as if it specifically counted against *lexical items* being of a prototype-theoretical nature. Thus, it is meant to put the presuppositional basis of LC on the defence, i.e., the idea that the concept-theoretical idea under debate might apply to lexical

concepts in the first place. It is meant to count against the specifically lexical-level component of lexical-cum-phrasal prototype theory.

We must understand Fodor as intending the compositionality and the decompositionality arguments to perform a half-and-half, exhaustive division of labour in criticizing prototype theory at different levels of concept structure. The intended division of labour is that the compositionality argument takes care of prototype theory for the *phrasal* level, while the *de*compositionality argument takes care of prototype theory for the *lexical* level, and that, taken together, the two arguments exclude prototype theory tout court.[5] The following passage from another paper of Fodor's gets very close to stating the assumption of this dialectical constellation explicitly. Following up on a renewed exposition of the compositionality arguments, Fodor (2001: 8) states:

> I conclude that the possession conditions for *at least some* concepts (in effect, the complex ones; the ones that have compositional structure) are *not* constituted by recognitional capacities. […] What I haven't shown, however, is what I promised you that I'd argue for; namely, that *no* concepts are constituted by epistemic possession conditions. […] [A]n epistemic semantics might work for the [lexical concepts] even if compositionality rules it out for the [phrasal concepts].

In his (2001) paper, it is in direct continuation of this passage that Fodor, purporting to show what he has claimed to require showing, turns to *de*compositionality, claiming—in analogy to the move observed in the present paper—that *this* property rules out a prototype-theoretical essence (a constitutive "recognitional capacity") for the *lexical* concepts.

If the decompositionality argument is meant to count specifically against lexical prototype theory, but, inevitably, can only be meant to count against it as embedded within a lexical-cum-phrasal prototype theory, then the question arises: What reason Fodor does have for thinking that the decompositionality argument counts specifically against the lexical-level component at all? The point to be made in the following paragraphs is that he has no rea-

---

[5]   In light of earlier presentations of the compositionality argument by Fodor, the present observations are rather non-trivial already. Whereas his remarks in the context of the decompositionality argument are typically lexically focused to begin with, the compositionality arguments were standardly described by Fodor as falsifying prototype theory tout court, with no level-specific restriction being hinted at. Still, I do not present these observations on a par with the ones to follow, because they are reconstructive, rather than critical.

son for this assumption about direction of disconfirmation and that his assumed direction in fact gets things the wrong way round.

Observation 1. *No justification has been offered for taking the decompositionality violations to disconfirm prototype theory at the lexical level.* There is something dubious about presenting two reductio arguments, both of which rely equally crucially on a premise A&B, and nevertheless taking one of these reductio arguments (in this case, the compositionality argument) to refute assumption A only, while taking the other (in this case, the decompositionality argument) to refute assumption B only. Given that both of the arguments crucially rest on assuming the conjunction of lexical-cum-phrasal prototype theory, what justification does Fodor offer for thinking that the first argument refutes the phrasal-level component, while the second argument refutes the lexical-level component? None. Fodor does not even begin to point to an explanation. For all that Fodor argues, it remains open which level to blame for the alleged semantic constraint violations he claims to observe.

But significantly more is to be said. The *true assignment of blame* for Fodor's alleged decompositionality violations is precisely opposite to the lexically focussed assignment of blame which Fodor, puzzlingly, takes for granted. It is demonstrably the *phrasal component,* rather than the lexical component of a lexical-cum-phrasal prototype theory, that is responsible for its decompositionality violations. As I point out now (observation 2), the assumption of phrasal-level prototype theory suffices all by itself to bring about Fodor's decompositionality violations. For another thing (observation 3), the assumption of lexical-level prototype theory in fact plays no role at all in bringing about those decompositionality violations.

Observation 2. *The phrasal component of lexical-phrasal prototype theory suffices to bring about the decompositionality violations.* Recall the insight pushed earlier concerning the relation between phrasal compositional structures and prototype-centered structures. Phrasal compositional constituent structures are trivially not cross-constituentially gradable structures. They are not, to put it differently, bundles of statistical feature constituents. But, identifying concepts with these is the essence of prototype theory. The assumption of phrasal-level prototype theory is thus incoherent.

Now consider the inconsistent double standard that is at work in Fodor's treatment of his example of choice. On the one hand, *faithful to* the prototype-theoretical idea, Fodor consults his intuitive judgments as to what constituents might or might not be involved in the prototype-centered structure

PETFISH (i.e., the potential prototype-theoretical concept expressed by 'pet fish'),[6] *implicitly relying on the correct picture* of the prototype-theoretical concept as a probabilistic bundle of constituents. His judgment is that a prototype-theoretical understanding of 'pet fish' neither requires PET nor FISH. On the other hand, though, *disregarding* this prototype-theoretical idea, Fodor acts as if a prototype theorist hypothesizing a concept structure faces the empirical challenge of having to capture the constituents of the binary compound that contains PET and FISH. Fodor assumes the prototype-theoretical concept not to require the latter elements as constituents. We can grant this assumption. Fodor's error here is that he understands himself as presenting the discovery of a constraint violation *within* prototype theory, when he is *really* relying on a trivial mismatch between a probabilistic constituent bundle and a *non*-probabilistic, binary constituent structure. Although Fodor must directly rely on a recognition of the respective types of constituent structure in obtaining the mereological non-entailment characteristic of the decompositionality violation, the trivial distinctness of these very constituent structures simultaneously appears to escape his attention, on a more explicit level.

Observation 3. *The lexical component of lexical-phrasal prototype theory is idle in the derivation of the decompositionality violations.* Together with the preceding one, the present observation against Fodor's picture of his own decompositionality argument provides the ultimate refutation of this picture. Lexical-level prototype theory is in fact entirely inessential to the decompositionality argument. For when Fodor hypothesizes which the potential probabilistic constituents of the prototype-theoretical concept PETFISH are, claiming that it does not require PET nor FISH, the respective *internal* natures of PETFISH's probabilistic constituents are entirely irrelevant. It may well be that the prototype-centered structure of PETFISH does not contain the respective prototype-centered structures of PET or of FISH. But if it does not, then this has nothing specifically to do with the latter concepts being constituted by prototype-centered structures. Fodor's non-decompositionality assumption does *not* specifically depend on the prototype-theoretical PETFISH bundle's failing to require knowledge of the *prototypes* of PET and of FISH. It depends on an intuition to the effect that the PETFISH bundle does not contain the fully-fledged concepts PET and FISH *at all,* full stop.

The decompositionality argument thus forms a confused manoeuvre. Just like the compositionality arguments, it is predicated on the incoherence of

---

[6] The fused notation serves the purpose of a convenient reminder that we are not dealing with a binary conjunctive structure isomorphic to the expression 'pet fish' here, but with a probabilistic structure corresponding to a fused reading of 'pet fish'.

the Scope Assumption, the extension of prototype theory to actual compositional phrasal compounds. The decompositionality violations which the decompositionality argument intuits really only go to re-illustrate this incoherence. Fodor nonetheless emerges from the argument blaming the violations on lexical prototype theory, overlooking the independently ascertainable fact that the latter assumption is entirely inessential to the decompositionality violations. Fodor's assignment of blame is not only unjustified; it demonstrably turns the only correct assignment of blame on its head. There really is no decompositionality argument against lexical prototype theory.

Our view shifts now to the rest of the decompositionality dialectic which Fodor describes. In the ensuing exchange, Fodor presents himself as putting defenders of prototype theory successively on the defence: First, they are portrayed as denying decompositionality; more specifically, as denying CC-decompositionality. Then, Fodor introduces the allegedly distinct and allegedly even more undeniable notion of CP-decompositionality. Defenders of prototype theory are portrayed as being forced to the ad hoc concession of CP-decompositionality in spite of continuing to deny CC-decompositionality. We will now see that none of this has anything to do with LC and that the dialectic does not even make sense in its own right.

3.3. Fodor's central strawman: Misidentifying LC with a denial of decompositionality (Step II). Recall that, according to Fodor's opening depiction of the criticized theorist's response to his compositionality/decompositionality argument, "the reply goes, epistemic properties … are essential to morphosyntactically *primitive* linguistic expressions … but *not* to their hosts". This corresponds to LC concerning prototype theory, which is indeed as widespread as Fodor portrays the reply to be:[7] Prototype theory applies to lexical concepts, but not to compositional phrasal compounds containing them. However, while we have reached the point now where Fodor does address LC *de re,* as it were—in fact, even reporting the idea correctly—he never begins to recognize LC as the claim that it is. Instead he understands it as the claim that he attributes in step II of the decompositionality dialectic: the alleged move of *holding onto lexical-level prototype theory while denying CC-decompositionality.*

The open incoherence of a decompositionality denial. LC claims that the constituents of a compositional phrasal compound form a compound *whose own essence is of a different type from theirs*. This is directly transformed by Fodor into the claim that lexical concepts have prototype-theoretical es-

---

[7]   For simplicity's sake, let us, like Fodor, assume for now that prototype theory is applied universally within the domain of lexical concepts.

sences, but that compositional phrasal concepts containing those lexical concepts *do not inherit those very essences,* or conversely, that the parts *do not contribute their essences to the whole*. What can Fodor's attribution of this denial of CC-decompositionality possibly mean? Unsettlingly, only one thing:

> **The alleged denial of CC-decompositionality, clearly stated:**
> Compositional phrasal concepts which contain lexical concepts with prototype-theoretical essences *do not have parts which have such essences*.

This is all that the denial of *CC*-decompositionality amounts to, once we replace the metaphorical talk of "contribution", "inheritance", or "transmission". The attribution of a decompositionality denial constitutes a perversion of LC. It is an out-and-out strawman. In contrast to LC, the notion simply amounts to a mereological contradiction, or to a modal contradiction, depending on which way you look at it. We will adhere to the mereological perspective on this:

> → Compositional phrasal concepts which contain lexical concepts with prototype-theoretical essences *do not contain them.* ⊥
> → Compositional phrasal concepts which contain lexical concepts with prototype-theoretical essences *contain them without their essences.* ⊥

Notice that, with regard to its formulation over *CP*-decompositionality, the contradictory nature of the decompositionality denial (DD) had to be obvious already from its *initial* formulation. Notice now also that it does not make any difference whether we speak of the former or the latter variant of DD. A distinction with respect to their dialectical status cannot be even prima facie upheld.

This brings us to the remaining steps of the decompositionality dialectic which Fodor imagines. They are rebutted on straightforward grounds in the following subsection. I then return, in a subsequent section of its own, to Fodor's LC/DD-misidentification itself, commenting on its inexplicability and on the real status of the property of decompositionality.

3.4. The remaining decompositionality dialectic makes no sense and is fictitious (Steps III–V). The remainder of the decompositionality dialectic which we have reconstructed earlier can now be seen to simply add a con-

fused appendage to the already exposed confusions of both the decompositionality argument and the misattribution of DD, which everything depends upon. It does not constitute any progression in critique. The appendage rests on a strange failure to treat CC-DD already as the obvious incoherence it is. It is very puzzling where Fodor gets the idea from that there is or might be such a remaining dialectic.[8]

Fodor writes as if it were the case that, while CC-DD is widely held, the subsequently described CP-DD is not. A significant difference in degree of rational deniability is suggested with respect to these notions, and this alleged difference is described as being widely acknowledged. But there is no such difference, as I have already pointed out, let alone one that is widely believed in. CC-DD involves just the same mereological contradiction as CP-DD.

Fodor, in his remaining decompositionality dialectic, discernibly only presents us with the situation of someone who addresses an incoherence which nobody maintains, who does not explicitly recognize the obvious incoherence himself, and who nevertheless relies on a fundamental recognition of that incoherence in ultimately rejecting it by implication.

It is likewise puzzling how Fodor could think that his fictitious opponent might still deny the triviality of CC-decompositionality *by stipulation,* in spite of apparently having recognized the triviality of CP-decompositionality already, and how this alleged stipulational move could be conceded to form an at least provisional rescue for the prototype theorist who is alleged to maintain CC-DD in the alleged dialectic.[9] After all, contradictions cannot be suspended by stipulation. Apparently, Fodor thinks that the connection which he exploits in the dialectic between CC-decompositionality and CP-decompositionality obtains with less than the force of logical necessity—otherwise the tie could not be severed by stipulation—and that CC-decompositionality is coherently deniable, while CP-decompositionality is not. It is hard to comprehend how Fodor could point out a seeming commitment of CC-DD to CP-DD, plus CP-DD's incoherence, without appreciating that CC-DD itself expressed that incoherence already.

---

[8]   Fodor, unsurprisingly enough, does not reference any authors in the course of the general attribution of these moves.

[9]   More specifically, a provisional rescue of the alleged CC-DD, which is first countered by a truistic complaint about ad hoc moves by Fodor and, in the course of the main part of Fodor (1998-II), by an argument from the learnability of the lexicon for CC-decompositionality.

Further down, peculiarly enough, we will come across a treatment of CP-DD by Fodor which is incompatible with the dialectical role he has assigned to the principle here. That is, in at least one of his later works, Fodor unwittingly makes explicit that he is addressing a contingent, *spurious* notion of CP-DD—the very principle which, in the present decompositionality dialectic, he is in still assuming to form the ultimate reductio. On the other hand, for precisely the CC-DD which, in the present decompositionality dialectic, Fodor still treats as coherently deniable, we will not even encounter a correspondingly spurious version which could explain such a treatment.

4. On misidentifying LC with DD, and on
the status of conceptual decompositionality

With his astonishing misidentification of LC with DD, Fodor's central and most pervasive error enters the scene. It is a constant throughout most of Fodor's reactions to what is in fact LC.

4.1. Fodor's misidentification as unforced and inexplicable. LC, the statement that lexical concepts have prototype-theoretical natures, but that compositional phrasal concepts constructed from those concepts do not, is in itself perfectly innocuous. Just as it is a plain truth that a property of a whole need not be instantiated by its respective parts, it is a plain truth that a property instantiated by each of a whole's respective parts need not be instantiated by the whole. A round pie which has been sliced into wedgelike segments is made up of wedge-shaped parts, but it is not itself wedge-shaped. A water molecule has a certain internal structure, but an aggregation of water molecules does not itself have this structure. Nothing could be less mysterious than this. We are not even making an ambiguous statement in saying this, nor, similarly, are we doing so in stating LC. We could also state LC with added emphases as follows (though it would not add anything new to the way that LC has already been stated): Even though many lexical concepts have *their* contents prototype-theoretically determined by their constituents, compositional phrasal constructions out of these lexical concepts do not have *their own* contents determined *by them* in a prototype-theoretical manner.

It is likely that Fodor's talk in terms of "non-inheritance" of concept-constitutive properties has had the occasional effect of concealing to readers the actual incoherence of CC-DD. This sort of talk may have occasionally been understood as a strained way of expressing merely the idea that the compound does not have the same *type* of essence as its parts. That would be nothing else than what LC claims. But it is not what Fodor means. Fodor

consistently alleges that the *individual constituent essences* are claimed not to be contributed to the wholes they enter into. That is, it is *the individual essences* which Fodor is claiming not to be inherited by the compound concept on the view ascribed, rather than the *essence type* of the constituents. Besides, if, contrary to fact, he really were only attributing the view that the parts and the whole have different essence types, he could never have begun to launch the critique which he does, deriving an incoherence in the next step.

Elementary though the mistake of Fodor's misattribution of DD in place of LC is, it is one he performs with astonishing perseverance throughout his works. Further textual confirmation of this will be encountered as we progress. The mistake is astonishing in three respects. Taken together, these facts make the mistake practically inexplicable, as far as I can see: **1.** DD is consistently ascribed by Fodor in spite of theorists he is reacting to explicitly stating an obviously different claim with LC. **2.** Most of the time, Fodor does not even seem to notice the perfectly natural and simple option of LC as a move to make against the compositionality arguments. **3.** DD is conjured up by Fodor and taken seriously in spite of its obvious incoherence.

It is a similarly surprisingly fact that, its elementariness notwithstanding, the misattribution has hardly been recognized as such in the literature. The only written piece I know of which points out essentially the same mistake is a note by Francois Recanati (2002).[10] Even this work has been left widely uncited, however—a reflection upon the lack of adequate responses to Fodor's confused followups to LC.

Recanati's note, I should like to add, is more concessive to Fodor's layout of the matters than I believe we can justifiably be. Recanati pictures Fodor as systematically trading upon an ambiguity in his talk of a host compound "inheriting" concept-constitutive properties from its constituents (288). But to suggest, as Recanati in effect does, that Fodor has fallen prey to a pre-existing ambiguity here, seems charitable beyond the stretch of plausibility. For first of all, it is solely Fodor who introduces the potentially obfuscating talk of "inheritance". Actual statements of LC, including those by authors that Fodor explicitly reacts to, do not employ ambiguous talk in stating their position at all. They only require that the reader be able to make the distinc-

---

[10] One misleading aspect of Recanati's note which still merits protest is this. In the ambiguity he identifies, Recanati coins the term 'compositional inheritance' for concept composition with lexically-phrasally uniform essence types (contrasting with with 'simple inheritance'). This runs contrary to the whole motivation for entering into the present discussion. It is one of the points behind this paper's defence of LC that nothing about compositionality counts against it.

tion between attributions of properties to parts and to wholes! Secondly, the ambiguity that Recanati is aiming at can be seen to correspond to the previously made distinction between denying that constituents contribute their individual constituent essences to their hosts (the incoherent claim) and denying that, metaphorically speaking, constituents contribute their types of essence to their hosts (the coherent claim of LC). But, as I have pointed out already, Fodor's talk of inheritance in fact *unequivocally* aims at individual essences of constituents.

4.2. The irrelevance of decompositionality violations in language. If denying decompositionality is an obvious incoherence, then decompositionality itself must be a triviality. It is indeed. If we free ourselves for a moment from the constraint of having to employ categories which allow us to reconstruct Fodor's own lines of reasoning, we can say that the property of conceptual decompositionality simply is that of *conceptual constituency*. Stating decompositionality of a posited compound concept as an allegedly substantial principle is confused because it amounts to asserting nothing but the fact that certain representations are parts of a given compound which already by definition has those representations as parts.

Failing to recognize the triviality of the relevant notion of decompositionality—and conversely, the incoherence of DD—appears to be a comparably widely encountered shortcoming in the literature as the aforementioned failure to recognize Fodor's LC/DD-misidentification for what it is. On one or two occasions where Fodor adverts to what I have been calling decompositionality, he speaks of "reverse compositionality".[11] (This is not the most happy terminology, since talk of compositional determination seem appropriate only for direction of determination that runs from properties of parts to their composed whole. But the term has caught on.) Several authors have made clear in mentioning "reverse compositionality" that they take it to be a substantial principle. Philip Robbins (2005) or Kent Johnson (2006) even offer paper-length treatments to this effect. Whence this "substantivist" discourse about decompositionality?

Well, crucially, all of these authors are running their discussion over relations between *states of linguistic understanding* (between the state of understanding a compound expression and the states of understanding its constituent expressions), rather than over relations between *states of concept possession* (between the state of possessing a compound concept and the

---

[11]   Fodor and Lepore (2001: 59) uses the term; Fodor (2000: 371, bottom of page) once mentions it in passing. In most of the works in which Fodor adverts to the property, he does not assign any specific label to it at all.

states of possessing its constituents). But there is a big difference between a principle of decompositionality about compound linguistic expressions and a principle of decompositionality about compound concepts. We are considering concepts qua mental representations, i.e., intrinsically contentful states or events. The state of possessing a complex mental representation, by definition, just *is* to possess its constituent mental representations in a given mode of combination. By contrast, the state of understanding a compound expression is *not* by definition conceptually complex. All that the state of understanding a compound expression indeed trivially entails is the formal complexity of the expression itself, not a corresponding decomposability into constituent states of understanding—put differently: not a corresponding complexity of the mental representation which provides its interpretation. Therefore, while the psychological principle of the decompositionality of concepts (CC or CP, no matter which) is trivially true, the linguistic principle of the decompositionality of expression understanding is not. But of course, if a decompositionality principle *were* germane to the discussion of the concept-theoretical notion of LC (which is not the case), then *at least* it would have to be a concept-theoretical decompositionality principle. Mixing the linguistic principle up with the concept-theoretical principle is confused.

Substantivist commentators on decompositionality can hardly be assumed to be making a mistake about the trivial status of concept-theoretical decompositionality itself. Insofar as they intend to connect their remarks to the present discussion about LC and compositional phrasal compounds, the flaw to be lamented is rather that they are not considering concept-theoretical decompositionality at all. They take as their point of departure certain statements of "reverse decompositionality" by Fodor in which he himself mistakenly adverts to linguistic expressions (the reviewed decompositionality dialectic of 1998-II being a case in point), but they ignore many other statements by Fodor in which, more appropriately, he speaks of concepts only (we will encounter such contexts later).[12] Perhaps needless to say in light of the distinction we have just made, the way in which Fodor moves freely back and forth between linguistic expressions and concepts in speaking of decompositionality must be viewed as a further confusion in his discussion.

---

[12] Recanati forms a welcome exception in this respect. He runs his remarks over concepts only, and it is no surprise that he accordingly describes decompositionality as being a triviality in the pivotal paragraph of his note: "[T]hat the … properties that are constitutive of constituent concept are inherited by their hosts … is a trivial matter" (2002: 287).

Independently of this, there is a still more elementary point to be made against potential contemplators of the idea that a non-triviality of linguistic decompositionality might be relevant to the present context of discourse. Decompositionality violation are wholly irrelevant to our concerns already for the reason that they trivially presuppose a shift away from compositional compounds: Cases in which a person understands a compound expression without understanding its constituent expressions are a fortiori not cases in which the compound has a compositional meaning. After all, the compositionality of a compound expression *presupposes* understanding of the constituent meanings from which the compound's meaning is derived. Thus, already for this reason and independently of the distinction between concept possession and expression understanding, decompositionality violations could not possibly be a resort in defending the possibility of confining prototype theory to the lexical level of compositional phrasal compounds.

Let me conclude. Resorting to an alleged coherence of DD in defending LC would be fundamentally confused in three independent respects. The first point has been politely bracketed in the last few paragraphs. **1.** LC is not DD, hence defending the coherence of decompositionality violations would be irrelevant to defending LC anyway. **2.** The conceptual notion of decompositionality that we have been talking about all along is trivially true indeed. A linguistic notion of decompositionality can be distinguished from it; this notion is coherently deniable. But Fodor must have a conceptual one in mind, in bringing decompositionality to bear. **3.** Decompositionality violations violate the presuppositional basis of compositionality violations. They trivially cannot be instantiated by compositional compounds anyway.

5. The decompositionality muddles,
part 2

The fallacious rendition of LC as DD is not a temporary lapse of Fodor's in (1998-I). It is a constant, pervading each of his other relevant works too, in some way or other. These works add quite a number of more red herrings, incoherences and casually dropped bizarrenesses to the thicket through we which have blazed our trail so far. The task in the following is to achieve clarity over these matters, showing that there is no sensible or even coherent argument in Fodor's other aforecited works against LC or against lexical-level prototype theory as such. A number of confusions are cleared away in the course of this.

5.1. How Fodor (1998-I) fails to address LC as stated by Schiffer. Fodor's (1998-I) paper, which already contained his first de re reaction to LC,

as it were, has an afterword appended to it whose exclusive purpose is to re-
turn to LC as proposed by Stephen Schiffer (unpublished). I quote the rele-
vant passage below. Fodor first insists on the fact that CP-decomposi-
tionality must be true. Then he returns to Schiffer, quoting him as he explic-
itly states LC:

> Now look at Steve's proposal, which is that "it is reasonable to
> hold that the possession conditions of complex concepts are de-
> termined by those of their constituent concepts. But for the case
> at hand this simply requires *F&G* to be such that to possess it
> one must be able to recognize good instances of *F* and good in-
> stances of *G*."
>
> As stated, Steve's theory is wrong about PET FISH: It's true,
> of course, that to have the concept PET FISH you have to have
> the concept FISH. But it's certainly *not* true that to have the con-
> cept PET FISH you have to have a recognitional capacity for
> good instances of fish. To have a concept, conjunctive or oth-
> erwise, you have to have the concepts that are its constituents.
> But you *don't* have to have recognitional capacities correspond-
> ing to its constituents; *not even if, by assumption, the complex
> concept is itself recognitional*. So having the constituents of a
> concept can't require having a recognitional capacity in respect
> of their instances. If it did, you could have a complex concept
> *without* having its constituents—which is not an option. (Fodor
> 1998-I: 45, emph. his.)

Consider Fodor's central claim in this passage, that "it's certainly not true
that to have the concept PET FISH you have to have a recognitional capacity
for good instances of fish, not even if, by assumption, the complex concept
is itself recognitional". This involves a general claim and a more specific
one, both of whose presentation by way of an argument in this context is
quite incredible. I address them in reverse order.

Ignoring LC, addressing its opposite. One thing that Fodor is doing is to
hint at the same decompositionality argument which we have already en-
countered. Incredibly, in the specially emphasized part of his negative re-
mark—"not even if, by assumption, the complex concept is itself recogni-
tional"—Fodor is addressing a position that assumes phrasal-level prototype
theory, even though the position to be criticized (even if it is misunderstood
as DD instead of as LC) *had just been described as stating the denial
thereof.*[13] In contrast to the situation with the decompositionality argument
in Fodor (1998-II), where we assumed that Fodor had to be construed, on

---

[13] More than the preceding Schiffer quote can be offered to document this. One need only
take a look at Fodor's prior recounting of Schiffer's reply, a few paragraphs back, on
pp. 44-45.

pain of a blatant petitio principii, as aiming the argument against the positive idea of lexical prototype theory presupposed in LC, rather than against LC itself, in the present case, he really *does* seem to rely on a utilization of the decompositionality argument against LC.

An equally bewildering fact which should not go unmentioned here is Fodor's metatheoretical evaluation of the situation. Two paragraphs before the quoted passage, Fodor assures us with respect to the original compositionality arguments that uniformly assuming prototype theory for both lexical and phrasal concepts "isn't an essential part of my argument" (p. 45). Of course it is indeed. This metatheoretical remark is furthermore bewildering in light of the fact that the uniform lexical-cum-phrasal application of concept theory in question it is precisely what he is focusing on with his specially emphasized "not even if"-remark in the present context.

Denying lexical prototype theory without argument. The other way in which Fodor reasons in the quoted passage consists in the more general negative verdict which precedes his heavily italicized supplementary remark. On one way of understanding this, Fodor is writing here as if he has discovered of allegedly prototype-theoretical lexical concepts *qua constituents* that they do not have the concept-theoretical essences they are claimed to have after all. To repeat, Fodor says that "it's certainly *not* true that to have the concept PET FISH you have to have a recognitional capacity for good instances of fish. To have a concept, conjunctive or otherwise, you […] *don't* have to have recognitional capacities corresponding to its constituents". This is all. Not a hint of an argument against someone who holds lexical prototype theory is given. All that Fodor is doing here is to state the contention that the concept-theoretical claim is not true of lexical concepts, only that he does so in a more convoluted way which takes the perspective of looking down from the compound to the constituents. The rhetoric is vacuous. He might as well replace it with the plain assertion that it is "certainly not true" that PET or FISH satisfy prototype theory.

Denying the strawman of a phrasally driven imposition of lexical prototype theory. Fodor's general negative verdict might also be understood in a slightly different way: as rejecting the utterly outlandish notion that the constituents' concept-theoretical natures are motivated and entailed by their phrasal host's compositional nature all by itself. In fact, this is the natural way of understanding his statement. It accords with the way in which Fodor takes up Schiffer's formulation, according to which a phrasal compound's compositional nature "simply requires" the constituents to have prototype-theoretical natures. "As stated," this is "wrong" Fodor says, going on to em-

phasize what having the compositional compound really only requires. This cannot be sensibly understood as negating anything other than an alleged phrasally driven imposition of the concept-theoretical idea in question. But Schiffer, it goes without saying, is stating LC here from the perspective of somebody who is *already* holding prototype theory for the lexical constituents. He is emphasizing, against this background, the *absence of a requirement* of phrasal prototype theory for someone who is a lexical prototype theorist already and who wants to preserve compositionality. Fodor appears to be perverting this into the *independent imposition of a requirement*. Perhaps this is the most absurd manoeuvre he stages in place of addressing LC.[14]

5.2. How Fodor (2000) fails to address LC as stated by Peacocke. Under some scrutiny, Fodor's (2000) reply to LC as stated by Peacocke can be seen to include the same moves as Fodor's (1998-I) decompositionality dialectic does after the decompositionality argument itself. But there are substantial further elements in this subsequent paper.

This time around, at least, it is not after an alleged confrontation of the prototype theorist with the decompositionality argument that Fodor ascribes the decompositionality denial to the LC proponent. A decompositionality argument as characterized above does not occur here at all. Rather, more appropriately, the ascription is made by Fodor subsequently to the original compositionality arguments. In spite of the perverseness of the identification as such, this at least gets the dialectical position of LC right, since it is of course the compositionality arguments in reply to which LC is made. Later renditions by Fodor stick to this picture, more or less (e.g., Fodor 2001, to which we turn in the ensuing subsection).

A new red herring: The notion of merely contributing semantic values. A novel red herring enters the scene when Fodor distorts LC as stated by Peacocke. More precisely, Fodor takes up remarks with which Peacocke elaborates on his own statement of LC. They are as follows:

> "Once one has fixed a semantic rule—something at the level of reference—for a mode of combining constituent concepts, a

---

[14] Fodor's appendix is still not finished here and adds more confusion in imagining a stipulation coming to the rescue; this time, one on which *CP*-decompositionality undergoes an ad hoc denial. The matter is equally puzzling as (and incompatible with) the ad hoc denial of CC-decompositionality and the ad hoc affirmation of CP-decompositionality in the dialectic of (1998-II) criticized earlier. It is only in his (2000) paper that Fodor makes explicit that he takes CP-decompositionality to be coherently deniable indeed, and how he does. We will take a look at this in the next subsection.

rule which takes as input the semantic values of the constitu-
ents, nothing more is required to determine the significance of
that mode of composition. There is no need to go on to associ-
ate additionally with the complex concept any recognitional ca-
pacity, or prototype-and-similarity relation, beyond what has al-
ready been fixed by the rule at the level of semantic value [sic;
probably to be read: "at the level of reference"] and the proper-
ties of its constituents." (Peacocke 2000: 339.)

Now, to be sure, it has to be observed that these remarks by Peacocke are
unhelpful. They come along as a redescription of the fact that lexical-
phrasal uniformity is not needed for the semantic determination accom-
plished by compositionality. But they mix in two irrelevant and misleading
points in describing the weakness of what is required by compositionality.
And as previously with Schiffer, Fodor seizes what was intended as a re-
mark that more or less redescribes LC, and uses it into a gateway for ab-
struse attributions. Only, this time, he does so in a very different way, as we
will see in a moment.

The two respects in which Peacocke's supplementary remark is beside the
point are these. Firstly, Peacocke's description of the weak fact required by
compositionality focuses on compositional relations between semantic val-
ues "at the level of reference". This, however innocuously it may be in-
tended, brings in an issue which has nothing to do with the present one.
Secondly and equally unhelpfully, Peacocke contrasts the notion of the gen-
eration of a new prototype structure with the notion of the generation of a
semantic value by a semantic combinatorial rule. This also misses the point
of LC, which he is trying to redescribe. In fact, Peacocke's contrast is base-
less, since a prototype-theoretical concept's bundle of feature representa-
tions determine the content of their target concept (i.e., they determine what
falls under it and with what degree of typicality) in accord with some rule or
other too, no less than phrasal concepts do. It is just not a rule that is en-
countered in the combinatorial semantics of natural language.

Now, in comes Fodor, jumping at Peacocke's added remark, reapplying the
already familiar attribution of the idea of CC-DD, and attributing to Pea-
cocke the patently incoherent idea that compositional concept formations
require *only the contribution of semantic values,* rather than the contribution
of *concept-constitutive properties too:*

Suppose C is a complex concept, all of whose (non-logico/syn-
tactic) constituents are recognitional. What, now, does composi-
tionality require of C? Not, according to Peacocke, that *it* be
recognitional; only that "the *semantic value* of [C] … be deter-

mined in a given way … from the *semantic values* of its atomic constituents'. [Italics, ellipses, and bracket added by Fodor.] That is, according to Peacocke's story, whatever (e.g. epistemic) individuating properties a primitive concept may have, compositionality requires only that it contribute its semantic value (e.g. its referent) to its hosts. (Fodor 2000: 371, emph. his.)

It is surprising that Fodor still thinks he knows what he is talking about here, given the renewed incoherence imagined. Since concept-constitutive properties and content-constitutive properties are just one and the same thing for Fodor's purposes—rightly so, since no other concept-individuating scheme is relevant in the present context of inquiry—Fodor is ascribing the obviously incoherent idea that a constituent concept enters into a compound under preservation of its content, in spite of having abandoned its content-constitutive properties. Of course Peacocke is not claiming what Fodor says he is claiming; how could he?[15]

A restaging of the DD strawman and the decompositionality dialectic. Paralleling the confused approach of his decompositionality dialectic in (1998-II), Fodor thinks of himself as making critical investigative progress in showing that CC-DD entails CP-DD. It is in the concept-possession formulations that Fodor's attribution of DD in place of LC is most explicit. Passages such as the following document clearly that Fodor's misattribution in (1998-II) is anything but a temporary lapse:

So then, a situation could arise as follows: you have concept AN whose, as it were, "syntactic" constituents are the recognitional concepts A and N, AN is compositional in the sense that its semantic value is determined by the semantic values of its constituents according to the appropriate rule. Notice that, though all this is assumed to be in place, *it doesn't follow, according to Peacocke's story, that you have either the concept A or the concept N.* That's because A and N are recognitional but AN isn't; so, whereas having A requires being able to recognize As [sic] and having N requires being able to recognize Ns [sic], having AN doesn't require being able to recognize anything, *including As or Ns* [sic]. (Loc. cit., emph. Fodor's.)[16]

---

[15] In fact, the passage quoted from Peacocke actually *has* him stating of the compound concept that "the properties of its constituents" are ones that have "already been fixed". Not that Fodor registers this.

[16] The concept name 'AN' in this passage is tacitly understood by Fodor to refer to some unspecified concept corresponding to an adjective-noun combination.

In an equally clear manner, a page later, Fodor introduces the alleged possibility of "fixing things up for Peacocke by brute stipulation", rejecting this with special rhetorical effort as ad hoc. Again, this faithfully reproduces a part of the decompositionality dialectic dissected earlier: maintaining CP-decompositionality by stipulation, while continuing to deny CC-decompositionality, as expressed in the preceding quote. As earlier, the dissociation of the notions remains entirely mysterious. The earlier dismissal reapplies.

Fodor's spurious notion of a *contingently* false CP-DD. Another notable difference which Fodor (2000) exhibits in comparison to earlier treatments is of broader relevance than the one-off attribution to Peacocke concerning CC-DD and a contribution of semantic values. While Fodor (1998-II: 52), recall, had still noted at least with respect to the CP variant of decompositionality that it is "true by definition that mastering a complex [concept] requires mastering its constituents, since, after all, constituents are by definition *parts* of their hosts", Fodor now treats *both* CC-DD and CP-DD as being *merely contingently false*. This treatment is encountered in other papers of his too. But it is in (2000), close on the heels of the passages just quoted, that Fodor expressly assures us at least of the contingency of *one* of the DDs, namely CP-DD:

> Now, it's not, to be sure, any sort of truism that you can't have AN unless you have A: it's perfectly possible, in principle, to suppose that you can have *A as a constituent of a complex concept* even if you don't have A *tout court*. (In effect, you could have only the concept of something that's *A insofar as it is N*.) (Fodor 2000: 371, emph. his.)

Consider for a moment what is going on here. Fodor is trying to tell us *that you can have concept C as a constituent without having it*. But of course, this is just as contradictory as any other variation on the denial of decompositionality. If you do not possess *C*, then, a fortiori, you do not possess *C* as a constituent. Fodor's description of a violation of CP-decompositionality thus cannot be understood literally, if it is to be assigned the non-contradictory content which he intends. But if Fodor is not literally speaking of being able to have a concept as a constituent without having the concept, then he is not really addressing decompositionality at all here—not in the literal sense which has been suggested by him earlier all along. It is *really a spurius form of DD* which he is ascribing with CP-DD.

The shifting meaning of 'having a concept' in Fodor's talk of CP-DD. Which non-contradictory understanding is Fodor relying on in describing CP-DD as contingent? Two possibilities come to mind. The first possibility

is that Fodor is treating the modifying phrase 'as a constituent' as having a suspending, potentially privative meaning. But this would be absurd. Nothing about ascriptions of constituenthood allows for a treatment of the predicate 'is a constituent of a compound' as performing a downgrading of ontological commitment, comparable to what adjectives like 'alleged' or 'potentially', or fully-fledged privative adjectives like 'fake' or 'former' would do. Such adjectives entail the suspension or annulment of a separate applicability of the noun because that is what they mean. The predicate 'is a constituent of a compound', by contrast, has no such meaning. It just means that a representation is a proper part of another one, nothing else.

What Fodor is really doing is to employ the predicate 'having a concept' *with two different meanings inside and outside the locution 'having a concept as a constituent'*. The sense which the expression 'having a concept' is assumed to have within the locution is an altogether weaker one than the sense it is assumed to have outside the locution.

We have, then, seen that Fodor must at least sometimes be intending his talk of CP-DD in a spurious way, and we have seen that it is spurious insofar as the crucial predicate therein undergoes a meaning shift within one and the same sentence. Let me point out that, by contrast, for *CC-DD,* not even a nonliteral understanding is discernible that would make its treatment as contingently false, however unjustified, at least understandable.

A spurious notion of CP-DD as a commitment to intraphrasally confined constituents. Fodor's parenthetical remark indicates which two different notions of concept possession he has in mind in his two uses of 'having a concept': The weaker notion is that of *possessing a concept such that one need not be capable of tokening it outside a specific phrasal concept;* the stronger notion of possessing a concept is one that *is* committed to this capacity. With this, Fodor has thrown yet another incomprehensible red herring into the arena.

In construing alleged DDs as cases of intraphrasal confinements of lexical concepts, Fodor is adding a second layer of confusion to the one already operative in virtue of his having misidentified LC with DD. The aforequoted passage has Fodor ascribing the notion of an intraphrasally confined lexical concept in place of LC. *But LC not even remotely suggests intraphrasal collocational restrictions on the lexical concepts it is claimed for.* What is the idea that PET and FISH have prototype-theoretical natures, without the binary conjunctive compound PET FISH having a prototype-theoretical nature, supposed to have to do with the bizarre idea that PET or FISH are necessarily re-

stricted to occurrences within the compound PET FISH? Fodor's attribution could not be more gratuitous.

Even apart from the fact that the property of intraphrasal confinement—IPC, henceforth[17]—has nothing to do with LC, it is *positively precluded by the notion of compositional constituency*. Compositionally formed compounds are necessarily derived from recombinable constituents on any customary understanding. To make the point in a more principled and specific way, it is only qua constructedness from autonomously recombinable constituents that compositionality is appealed to in order to explain productivity and systematicity. Fodor's attribution of a commitment to IPC concepts thus acquires an even more bizarre quality. Even if the strictly IPC concepts he is envisaging are a coherent notion, they are ipso facto *not the compositional constituents we are talking about*.

Ironically, instead of making Fodor aware of the obvious mistakenness which already attached to his original misattribution of DD in place of LC, the added misattribution of IPC to which this leads him only prompts him to march further down the blind alley. Instead of becoming suspicious about his attribution of DD in place of LC, Fodor moves on to declare with a sense of refutational triumph that "surely, Peacocke doesn't want to allow" the IPC as which he has just spelled out the spurious CP-DD (op. cit.: 371), adding, "Surely it ought to follow from the theory of compositionality that, for example, having the concept GLOWING THING is a sufficient condition for having the concepts GLOWING and THING." This is a peculiar way of putting things (what theory of compositionality?), but it is clear enough that it expresses the point I have just made about IPC constituents not being compositional, recombinable building blocks.

5.3. More confusion and multiple contradictions in Fodor (2001). Fodor (2001), first of all, continues the failure to recognize LC for what it is. If one were forced to choose exactly one quote only from Fodor's works to exemplify his strawman attribution of DD in place of LC, one might want to choose the following remark about the compositional compound RED SQUARE, which occurs in the text after the compositionality argument has been presented:

---

[17] I introduce the abbreviation not least because it will come in handy also in the next subsection. As we will see, the notion of a constituent's strict intraphrasal confinement reappears in an amazingly different role in Fodor (2001). I will allow myself to use the abbreviation 'IPC' both in the nominal sense just defined and in a corresponding adjectival sense.

> What I haven't shown … is what I promised you that I'd argue
> for; namely that *no* concepts are constituted by epistemic pos-
> sessions [sic] conditions. For, the following possibility remains
> open: Having RED requires being good at recognizing red
> things; and having SQUARE requires being good at recognizing
> square things. But having RED SQUARE doesn't require being
> good at recognizing anything. In effect, according to this ac-
> count, there can be possession conditions on primitive concepts
> that they don't transmit to their hosts. So an epistemic seman-
> tics might work for the former even if compositionality rules it
> out for the latter. (Fodor 2001: 8.)

Presented in the variant of CP-DD from the outset, this confused substitu-
tion of DD for what is in fact LC as the option which "remains open" could
not be clearer.

What follows are two equally long paragraphs which, together with the just
quoted one, fill the space of about a page. The immediately following para-
graph (2001: 8-9) rehashes elements with which we are familiar by now. In
the pattern familiar from the decompositionality dialectic in (1998-II), the
core of this paragraph emphasizes the truth of decompositionality against
the envisaged chimera—this time, Fodor seemingly treats CP-decomposi-
tionality again as the trivial truth that is this.[18] The second paragraph builds
on the reminder of decompositionality, seemingly giving a novel argument
from decompositionality against the lexical-level concept-theoretical claim
itself. This paragraph, particularly perplexing and error-ridden, will be at the
focus of our attention on the next few pages.

An argument *from* IPC against lexical recognitional concepts. Before he
has gotten to the second paragraph, Fodor has introduced no new elements
into the dialectic besides the renewed insistence on decompositionality. But
Fodor's second paragraph tacitly presupposes very substantial shifts with
regard to the target and with regard to the dialectical status of the decompo-
sitionality assumption. I explain in a moment. Here is the core of the second
paragraph:

---

[18] To quote: "But this won't do. […] [C]ompositionality requires not just that having the
constituent concepts is sufficient for having a host concept, but also (and even more ob-
viously) that having the host concept is sufficient for having its constituents. Or, to put
it slightly differently, compositionality requires that host concepts receive their semantic
properties solely from their constituents, *and also that constituent concepts transmit all
of their semantic properties to their hosts*. The argument is straightforward; if constitu-
ent concepts didn't transmit their possession conditions to hosts, there would be nothing
to stop you from having a host concept *without having its constituents*." (Fodor 2001: 8-
9; emph. his.)

> It certainly *looks* as though you could have RED SQUARE with-
> out having a recognitional capacity for red things at large (in
> favourable circumstances or otherwise). For example, you
> might have a concept RED that occurs *only* in host contexts. In
> that case, you'd be able to think *red square* and *red triangle* but
> not *red* tout court; a fortiori, you wouldn't be able to recognize
> red things as such, since that requires being able to think *red*.
> (That kind of mind wouldn't even be very surprising, since
> thoughts about red tout court are more abstract than thoughts
> about red squares.) Well, if this *is* possible, and recognised [sic]
> capacities are content constitutive [sic], then you could have
> RED SQUARE without meeting the (putative) epistemic condi-
> tions on RED. But biconditional compositionality requires that
> satisfying the possession conditions for RED SQUARE *entails* sat-
> isfying the possession conditions for RED. The moral would
> seem to be that there aren't any epistemic conditions on RED. I
> conclude that there aren't any epistemic constraints on any
> primitive [i.e., lexical] concepts. (Fodor 2001: 9, emph. his.)

Before we get to a reconstruction and a more detailed criticism of the argu-
ment put forth here, clarification is needed with regard to what is argued
against by Fodor. Two general points bear mention in advance.

Preliminary clarification 1: As I have indicated, the quoted passage does not
argue anymore against the negative, confining claim of DD/LC itself, but
rather against the positive lexical-level claim which is presupposed by
DD/LC. Registering this fact is not an entirely obvious matter, for, on the
surface, Fodor's manner of exposition suggests the target to be a unitary one
throughout the last-quoted paragraph and the one that immediately preceded
it in his text; he does not flag as the distinct moves which they are the rejec-
tion of DD/LC and the rejection of the concept-theoretical claim presup-
posed by it. But, for one thing, with the concluding sentences quoted above,
it is made explicit that the focus has tacitly shifted away from repudiating an
alleged DD. For another thing, this second paragraph explicitly rests on tak-
ing for granted a *shared recognition of decompositionality* with the imag-
ined opponent.

Preliminary clarification 2: Though I have consciously avoided speaking of
so-called recognitional concepts up to here, often directly transposing mat-
ters into talk of prototype theory, we cannot avoid adopting talk of recogni-
tional concepts now, if we are to have a chance of reconstructing the argu-
ment Fodor offers. The reason is that the argument relies on what Fodor
takes to be an essential feature of recognitional concepts qua recognitional
concepts. His argument does not even implicitly trade on the crypto-
prototype-theoretical element which he usually makes out in the alleged no-

tion of recognitional concepts. His parenthetical remark in the introductory sentence of the passage quoted above accurately anticipates this: Whereas in the usual case, Fodor relies on an allegedly crucial commitment of the alleged notion of concept-constitutive recognitional capacities to favourable circumstances, the parenthesis has him speaking indiscriminately of recognitional capacities "in favourable circumstances or otherwise". My adoption of the notion of a recognitional concept in the present context is only provisional and is not to be understood as constituting any kind of meta-theoretical agreement on this count with Fodor.[19]

Let us first reconstruct the argument which we have just been presented with by Fodor, before saying anything further by way of analysis or critique. It is an argument from precisely the property of IPC which we have already encountered. For purposes of exposition, it will be most convenient to run the argument over the example given by Fodor.[20]

> **The argument as given by Fodor:**
> **1.** Possessing the concept RED consists in being able to recognize red things as such. (*Assumption for reductio*.)
> **2.** One can possess the concept RED while being only able to token it within the compound RED SQUARE, being unable to token it unembeddedly. (*Possibility of IPC*.)
> **3.** Recognizing red things as such requires tokening RED unembeddedly. (*Incompatibility of IPC and recognitional concepthood*.)
> **4.** One can possess the compound RED SQUARE without being able to recognize red things as such. (*2,3*.)
> **5.** One can possess the compound RED SQUARE without possessing RED. (*1,4*.) (= *Violation of CP-decompositionality*.)
> **6.** Possessing the compositional compound RED SQUARE requires possessing RED. (*CP-decompositionality*.)
> $\therefore \perp$ (*5,6*.)

What we have here is an argument from two assumptions, plus CP-decompositionality, against the recognitional nature of RED. One assumption is that of the compatibility of concepthood with IPC; the other assumption is that of the *in*compatibility of IPC and recognitional concepthood. Call the first

---

[19] Again, see chapter 7. I am trying to show here that Fodor's allegedly decompositionality-based argument against recognitional concepts is a shambles no matter whether you believe that the notion of a recognitional concept can be made sense of or not.

[20] Fodor takes the argument to be "perfectly general", as he states subsequently. This is already clear from his formulation of the conclusion drawn.

the *IPC assumption* and the latter the *Incompatibility Assumption*. From these two, it is deduced that lexical recognitional concepts *violate CP-decompositionality*—which is where Fodor consequently locates the refutation of the notion of lexical recognitional concepts, given the truth of CP-decompositionality. Before I move on to critique the argument, two further preliminaries are in order.

Preliminary 3, completing our terminological toolkit: Remember the contingent principle discussed earlier, which Fodor has in mind when he speaks of CP-decompositionality in his deceptive, non-literal way. Let us call this spurious notion *CP-decompositionality**. By contrast, let us call *CP-decompositionality^T* the mereological tautology which the literally understood principle of decompositionality amounts to. I will continue to use the term 'CP-decompositionality' without modification by a superscript when being purposefully noncommittal as to which of the understandings is the operative one.

Preliminary 4: The IPC assumption is in itself surprising enough and of dubious coherence. For the time being, however, I will not complain about it, similarly as I go along here with Fodor's demonstrably ill-conceived notion of a recognitional concept for the sake of argument. I briefly return to this towards the end of this subsection. But Fodor's argument is so seriously flawed in other respects that we will not spend much space on discussing the intrinsic plausibility of the IPC assumption itself.

Fodor's self-contradiction: A CP-decompositionality* violation both as a positive test case and as a reductio. Remember that Fodor (2000) first brought up the alleged possibility of IPC concepts in the course of saying what, "in effect", a CP-decompositionality* violation would amount to. Indeed, for all that Fodor hints at, it appears that the non-trivial notion of CP-decompositionality* which he has in mind just *is* the notion of a compound's being decomposable into constituents that are autonomously employable. In any case, what this means is that Fodor's argument above *explicitly rests on considering a CP-decompositionality* violation as a positive test case which must be accommodated,* and on simultaneously *rejecting* a theory that fails to accomodate this test case *on the very grounds that it lead to a CP-decompositionality violation*. This is surprising, since, after all, Fodor (2001) describes the decompositionality violation that constitutes the present reductio in the same way in which Fodor (2000) described the decompositionality violation which, effectively equating it with IPC, he claimed to be possible.

On pain of Fodor obviously contradicting himself, then, the notion of CP-decompositionality which is operative in lines 5 and 6 of Fodor's present argument *must* be that of *decompositionality^T* rather than that of decompositionality*. Notice, however, that the present text clearly suggests it to be a decompositionality* violation which Fodor takes to constitute the reductio (just as Fodor practically invariably aims at decompositionality* anyway, even though one would not know it from his immediate descriptions). Fodor is speaking of the specific principle of what "biconditional compositionality requires" (2001: 9), rather than of a perfectly general mereological tautology. More decisively, the compositional constituency which Fodor is insisting on in speaking of "biconditional compositionality" *presupposes decompositionality**. Thus, is truly seems to be the case that Fodor is performing the most elementary self-contradiction here. His argument simultaneously assumes that decompositionality* violations form a positive test case requiring accomodation, and that decompositionality* violations constitute a reductio ad absurdum.

Let us resolve to patch matters for Fodor by assuming him to be haplessly *oscillating* in his present, varyingly explicit employments of the notion of decompositionality. The situation given such an interpretation would, in overview, have to be as follows. Fodor is imposing as a positive requirement the ability to accomodate alleged IPC cases, which, as noted, are nothing but *violations of decompositionality**. What he then *derives* is a *violation of decompositionality^T* for lexical recognitional concepts. Then, confusedly *swinging back again,* he claims the derived result to form a reductio ad absurdum in virtue of violating the potentially non-trivial principle of *decompositionality*,* when he really should be lamenting the violation of the tautology of decompositionality^T.[21] Let us assume this interpretation and leave the present matter at that.

Fodor is confused about his own argument in a more fundamental respect still. It is not only that his alleged reductio *cannot* be relying on decompositionality*. As I point out now, his alleged reductio clearly actually *does not*

---

[21] Fodor must also be criticized for a related inconsistency here which I avoid mentioning in the text only for fear of complicating matters further. Once Fodor is focusing on the alleged possibility of IPC constituents, he has not only ipso facto shifted his focus to constituents that violate decompositionality*; he has also ipso facto abandoned consideration of compositional constituents, in the explanatorily interesting sense of the term. Yet, oblivious to this, Fodor still assumes—as his remark on what "biconditional compositionality" requires makes clear—that he is dealing with potentially compositional constituents. This sort of mistake matches Fodor's oversight, criticized in the previous subsection, of IPC's trivial preclusion of compositional constituency (in the explanatorily interesting sense which Fodor is focused on).

depend on any idea of decompositionality at all; a fortiori, it does not depend on the idea of deriving a violation of decompositionality^T.

The irrelevance of decompositionality to Fodor's own reductio. The part of Fodor's argument above which focuses on RED's hosting compound RED SQUARE (premises 4 to 6), and on its alleged violation of decompositionality, is entirely irrelevant to his attempted reductio. It is a red herring of Fodor's, revealing a confusion of his as to what argument he is truly depending on. For his attempted refutation of recognitional lexical concepts *already directly follows from the IPC assumption plus the Incompatibility Assumption.* This, accordingly, is how the argument should really be put forth by Fodor:

> **Fodor's cleared up argument:**
> **1.** Possessing the concept RED consists in being able to recognize red things as such. (*Assumption for reductio*.)
> **2.** One can possess the concept RED while being only able to token it within the compound RED SQUARE, being unable to token it unembeddedly. (*Possibility of IPC*.)
> **3.** Recognizing red things as such requires tokening RED unembeddedly. (*Incompatibility of IPC and recognitional concepthood*.)
> **4*.** One can possess the concept RED without being able to recognize red things as such. (*2*,3.*)
> ∴ ⊥ (*1,4*.*)

Best not to wonder about how Fodor came to add the twist of a decompositionality violation and came to present it as the crux of his attempted reductio argument. What counts is that we now have a clear view of what Fodor's argument really reduces to. The agreeable side effect of this clarificatory progress is that, in moving on to evaluate Fodor's argument against lexical recognitional concepts, we do not have to worry about the additional complication of Fodor's confused dialectical handling of a CP-decompositionality* violation.

The defectiveness of Fodor's core argument. On the other hand, just look where we are now: Not only have we been dealing with the chimera of the notion of a recognitional concept all along, we also are not even dealing with anything that remotely resembles an argument from compositionality against it. Still, can at least any cogent idea be detected in Fodor's cleared up argument? Not at all. Even all by itself, the argument is a non-starter. It is so even in more respects than one. We can be relatively quick about this. To begin with, the idea of IPC, of being able to have a first-order predica-

tive concept[22] which cannot possibly be tokened outside combinations with other predicative concepts is extremely dubious already. We are told nothing about how deep the mechanisms are supposed to run that underlie the nomological impossibility of this confinement of the concept in question to a special kind of conceptual construction. Surely it would have to be some quite shallow form of nomological impossibility, and it is hard to see how it could warrant an IPC assumption of interest.

But even disregarding the highly dubious nature of the general idea of IPC: In the case of the alleged posits of recognitional concepts, Fodor can *certainly not* take the possibility of IPC for granted. After all, for recognitional concepts, he has already laid down that an impossibility of IPC reflects a constitutive fact about recognitional concepthood. So it would require an extremely substantial and non-obvious argument, from the standpoint of the imaginary endorser of "recognitional concepts", to show that IPC must be conceded to be possible for recognitional concepts. Instead, however, Fodor only gives a quick, parenthetical remark on RED thoughts, hinting at a kind of argument which is in fact quite ridiculous. Fodor's "existence proof" to the effect that there can be an IPC concept of redness consists in suggesting that simple thoughts involving the perception-based concept RED are likely to be ontologically subsequent to thoughts involving the complex concepts RED TRIANGLE or RED SQUARE. But they are not; this idea is trivially false. Fodor is mixing it up with the fact that simple thoughts involving the perception-based concept RED (and complex ones too, for that matter) are subsequent to preconceptual, *sensory representations* of red squares, from which the perception-based concept RED would have to be abstracted. Hence, the possibility of a concept RED which is confined to the complex concept RED SQUARE is left just as implausible as it was.

This does not exhaust the confusion of Fodor's core argument against "recognitional concepts". To mention another fatal flaw, there is the self-defeatingness of Fodor's rejection of "recognitional concepts" on the basis of their alleged incapability to accommodate the alleged possibility of IPC concepts. In pointing this out, we come full circle to the ill-foundedness of Fodor's target notion of recognitional concepts which was laid bare in chapter 7 and which, for the sake of argument, I had bracketed in the present context up to here. If there is *any* type of theory of concepts which immediately entails a commitment to the autonomous applicability of concepts, and which consequently precludes a possibility of IPC for its target concepts, it is that of *co-*

---

[22] This specification had better be included because the existence of predicative concepts that are, e.g., categorially restricted to higher-order predications is a much more debatable matter.

variationist theories of concepts, such as Fodor's own. It is precisely these theories which are committed to concepts being constituted by a form of selective causal sensitivity. In fact, for all that Fodor alludes to in the present context, it is completely mysterious how he could take himself to be addressing anything other than a covariationist theory of concepts, or at least any other aspect about the targeted notion than a covariationist one.

6. On UA itself: Illusory side steps and a pseudo-argument in its defence

Let us now return to LC itself and let us focus on its opposite assumption, which the compositionality arguments against prototype theory take for granted: that prototype theory does not apply to the lexical level short of taking lexical-phrasal scope tout court. Judging from the way that this scope assumption concerning prototype theory is taken for granted by Fodor, it would seem that he is relying here on a *general assumption of uniformity* concerning the content-constitutive properties of lexical concepts and compositional phrasal concepts. I will call the assumption that compositional phrasal concepts have the same content-constitutive property as the lexical constituents which they contain the "Uniformity Assumption", or UA, for short.

Profoundly inscrutable self-descriptions by Fodor about UA are encountered in two of the earlier examined papers. In passing, Fodor twice (1998-I: 45 and 1998-II: p. 61, n. 7) portrays himself as being officially neutral on the truth of UA, *even as defined over the specific concept theory that is targeted*.[23] Fodor declares official neutrality concerning UA in the context of the decompositionality argument, and in that of the ensuing decompositionality dialectic.[24] But of course Fodor is indeed relying at the very least on a UA that is specific to the concept theory at issue (in particular, prototype theory). Nothing needs to be added to what we have observed already earlier

---

[23] UA is called the "agglomerative principle" on these occasions; it is specifically defined over recognitional concepts and conjunctive compounds formed from them in (1998-I: 45, top of page).

[24] Fodor (1998-I: 45) claims non-reliance on UA before launching the objection to Schiffer, which is already thoroughly confused without it and which I have criticized in subsection 5.1: "Now, it does seem to me that somebody who holds that there are primitive recognitional concepts should also hold the agglomerative principle […]. But my thinking this isn't an essential part of my argument." Fodor (1998-II: p. 61, n. 7) claims non-reliance of UA after having insisted on CP-decompositionality against the imaginary denial of CC-decompositionality in the unfolding dialectic which I have dissected in section 3: "Notice that the argument now unfolding is neutral on the parenthetical caveat, i.e., it is *not* committed either way with respect to what Schiffer calls the 'Agglomerative Principle'. See [Fodor 1998-I]."

on in this chapter, quite apart from the fact that the compositionality arguments themselves obviously rely on a prototype-theoretical UA. Fodor's metatheoretical remarks about his own moves are simply not comprehensible and add a further element to the confusion he generates.

Let me recapitulate matters up to here. Not only does Fodor usually fail to recognize LC as the idea that it is, misidentifying it with DD instead. He more than once even suggests that he is not relying on the preclusion of LC (/on a prototype-theoretical UA) at all, which of course he is. Apart from this, we have also addressed and repudiated, in section 2, a lone exception in which Fodor (1998-I: 41) directly argued against LC, in his first reaction to Schiffer. I now complete our survey of loci in which Fodor responds to the notion of LC as such. Let us thus turn to yet another, novel pair of rejoinders which are made in Fodor and Lepore (2001) in reaction to Paul Horwich. Horwich (1998: 161, n. 7) rejects the prototype-theoretical UA behind the compositionality arguments, stating that the pet fish argument "is flawed, because it unjustifiably presumes that if the meanings of the primitive terms are given by stereotypes, then so must the meanings of the complexes in which those terms appear". He goes on to express skepticism about a general UA (p. 161 and p. 162, n. 8). This is where Fodor and Lepore set in.

Fodor and Lepore's (2001: 58-60) rejoinders are made in the context of addressing the general UA principle. They consist in two arguments to very different effects. They are no better than the ones which we have already scrutinized. The first rejoinder has Fodor claiming that he does not really need UA, shifting to an entirely new argument against lexical prototype theory. The second rejoinder has Fodor arguing that a general UA is indeed true, even though he claims not to depend on it.

6.1. The side step to an argument from context-dependence in Fodor and Lepore (2001). The first argument, as noted, has Fodor claiming that he does not need UA to argue against lexical prototype theory, and presents a wholly new reason why this is supposed to be the case. This new argument has nothing to do with the compositionality or even decompositionality arguments anymore and can only be understood as amounting to a concession of the point of LC against the compositionality arguments.[25] The new argu-

---

[25] Fodor explicitly speaks on p. 59 of the passage in question (middle of page) as if the argument to follow shows that the compositionality arguments do not presuppose UA. His description is false and utterly misleading, for of course the compositionality arguments themselves do indeed rely on a UA. As for the property of context-independence which the following argument relies on, whether it is imposed by compositionality or not, as Fodor stresses, it does not yield any argument against prototype theory, as I now show.

ment rests on an alleged context-dependence of prototypes. I now argue that it is a non-starter.

Fodor's new argument against lexical prototype theory is one from the context-independence of lexical meanings and an alleged context-dependence of prototypes, and it is offered within the space of a sentence (Fodor and Lepore 2001: 58-59). In the familiar quick and casual manner, Fodor helps himself without argument to the assumption that "stereotypes aren't context-independent (for example, the stereotype of people-swimming is much different from the stereotype of flounder-swimming since the latter, but not the former, adverts to the exercise of fins); so lexical meanings can't be stereotypes" (p. 59).

Now, if you take the proper concept-theoretical perspective on this, there is not even the beginning of an argument here. Once a prototype theorist has postulated a given prototype-centered structure to constitute a given concept, then, once she is postulating *that concept* to be contained in different compound concepts, she is *thereby* positing those different compounds to contain that very prototype-centered structure as a part. She cannot be postulating different prototype-centered structures in different compounds short of assuming different concepts to be involved in them in the first place. Thus, it is prima facie unclear how a conflict of the sort that Fodor is presuming could even arise. Fodor is unaware of this, as far as his complaint goes. Put differently, his complaint would require the prototype theorist to acknowledge a situation in which she *simultaneously postulates involvement of one and the same concept* across different compounds, *yet feels pressed to assign different prototype-centered structures* to the different tokens of the concept. But assuming that she would acknowledge such a situation is entirely question-begging. Fodor's argument from context-independence against lexical prototype theory must therefore be dismissed.

There is another way of construing Fodor's objection from context-dependence, namely as claiming that there is context-dependent variation as to *which prototype-centered structures are associated with a given linguistic expression in different contexts*. In effect, on this understanding, he is claiming that a prototype theorist would have to assume that understanding 'swimming people' involves connecting a different prototype to 'swimming' than the prototype connected to 'swimming' in understanding 'swimming flounders'. But this is not the concept-theoretical claim he would need against prototype theory. If it is indeed a point about context-dependent associations of prototypes with a given expression that he is

complaining about—and it seems that it is—, then there are two objections to make.

Firstly, *even if* we granted his claim of context-dependence, it would take an extremely non-trivial further assumption to obtain the reductio which he believes he can derive. The reductio would only be obtained on the assumption that a given linguistic expression must express one and the same concept on all occasions of its use, short of fully-fledged homonymies. This general methodological assumption (which, it happens, forms a dominant presupposition throughout Fodor's work) is in fact not only extremely non-trivial. It is also implausible. One could easily write a book about how detrimental this assumption has proven in various contexts of investigation (and such a book might well be needed). But we cannot begin to go into these matters here, neither do we need to. Secondly, the claim of a context-dependence of the association of prototype-theoretical concepts with a given linguistic expression, even if it *did* impugn prototype theory, is in itself a highly substantial assumption to make. Fodor cannot just help himself to it.[26]

6.2. The argument from "meanings" and mereology for a general UA in Fodor and Lepore (2001). The second argument in Fodor and Lepore's aforementioned passage is another entirely new one, and the only one that Fodor ever offers for UA as a general principle. Here, he argues that a general UA is indeed true, even though he alleges not need it.[27]

This argument is at least as bizarre as any of the previously reviewed ones. It is an argument for UA, from the joint fact of linguistic compositionality and decompositionality, plus a *mereological explanation* thereof. Having insisted on the idea of "reverse compositionality" with respect to the understanding of phrasal expressions again, Fodor and Lepore go on to argue:

> [A]s far as anybody knows, compositionality and reverse compositionality *always go together*. [...] It would be nice to have

---

[26] This last point is also one that Horwich makes on behalf of the prototype theorist in a footnote rejoinder (2001: 381, n. 15) to the context-independence objection. Horwich correctly points out that "any alert defender of the stereotype theory of meaning will simply deny [the context-dependence claim] and will maintain that his theory, properly stated, is that the stereotype associated with the isolated word 'swim', whatever it may be, is the meaning of that word wherever it occurs". What he says on behalf of the prototype theorist against the context-dependence objection, however, reduces to this; the more principled objections which I have just made.

[27] Fodor and Lepore this time explicitly speak of UA as the "uniformity thesis", characterizing it on p. 59 as the idea "that if the meanings of the primitives are stereotypes (or uses, or prototypes, or inferential roles, or whatever), then the meanings of the complexes are *also* stereotypes (uses, prototypes, inferential roles, etc.)".

an explanation of why the meanings of complex expressions supervene on the meanings of their parts; and of why the meanings of parts supervene on the meanings of their complex hosts. And it would still be nicer if the explanation of these two superveniences also explained why they always turn up together. In fact, the explanation is obvious; the meaning of 'dogs bark' supervenes on the meanings of 'dogs' and 'bark' because the meanings of 'dogs' and 'bark' are parts of the meaning of 'dogs bark'; and *the meanings of 'dogs' and 'bark' supervene on the meaning of 'dogs bark' for exactly the same reason*. But the idea that the complex meanings (don't just supervene on, but) actually contain the constituent meanings is the 'uniformity thesis' in a very strong form. So it looks like the 'uniformity thesis' must be true. (Fodor and Lepore 2001: 59-60, emph. theirs.)

Fodor is claiming here that linguistic compositionality is explained by a *mereological relation* between "the meanings" of a phrasal expression and "the meanings" of its constituent expressions. The same sort of mereological relation is likewise claimed to explain linguistic *de*compositionality. This idea of a part-whole relation between phrasal and constituents "meanings" is then assumed *to entail UA*.

I find it baffling how Fodor can take himself to have provided an argument here, let alone a cogent or even potentially pertinent one. Most fundamental objections first: Fodor's ascription of a mereological relation here is entirely vacuous in the only remotely relevant way that he can be construed. Let us be sure to have a clear conception of what Fodor is speaking about. The manner in which Fodor speaks of "meanings" above and of mereological relations between "meanings" is not only quite nebulous; it is inappropriate. What is at issue with a prototype theory of concepts and with UA is the question of the uniformity of the *content-constitutive properties of the mental representations,* i.e., concepts, that are expressed by lexical expressions on the one hand and by compositional phrasal expressions on the other. What Fodor's talk of *part-whole relations* between "meanings" must be understood to aim at here, thus, are part-whole relations between content-constitutive properties of mental representational states. But what can be possibly meant by part-whole relations between properties of states? It would seem that, properly speaking, Fodor can only be meaning part-whole relations between *states with* those content-constitutive properties. But the lexical and phrasal concepts which Fodor's claim is supposed to apply to *already* stand in a part-whole relation to each other *by assumption*. Thus, Fodor's claim of a part-whole relation between content-constitutive properties ("meanings") of a compound concept and its constituents simply redescribes the subject matter. It does not succeed in stating anything non-pleonastic *of*

that subject matter. A fortiori, it does not state anything of the subject matter of lexical and phrasal concepts that would warrant UA.

Now, other authors have rejected Fodor and Lepore's argument already. In particular, Horwich (2001: 380-81) and Robbins (2005: 264-67) have. But their rejections are situated at a different level than the basic rebuttal just provided. Both of them presume that there must be some initial sense to the consideration that Fodor is putting forth here, and they wonder *what kinds of meaning entities* Fodor might have in mind when he claims a mereological relation to obtain between "meanings" of words and phrasal expressions.

Thus, Horwich assumes that "one might naturally construe [the assumption of a mereological relation] as implying that the meanings of complexes are ordered *sets* whose members are the meanings of words" (p. 381). He contrasts this with the fact that "word meanings—which are presumably *not* sets—would be very different kinds of entity from the meanings of complexes". Which in turn moves Horwich to assume that we should precisely *predict* (and here, his text shifts back to the *meaning-constitutive properties* it was talking about) that the UA with respect to the meaning-constitutive properties of words and their compounds is false. Horwich thus rejects Fodor's step from the mereological assumption to UA as a non sequitur that inverts the evidential situation. Robbins, who goes into the matter more meticulously, likewise considers what the mereological claim about lexical and phrasal "meanings" would require. Differently from Horwich, he argues that "the mereological hypothesis rules out the identification of denotations with sets, pluralities, or functions" (p. 264) and that it naturally invites the identification of "meanings" with *properties* (p. 266). What he objects to is not the step from the mereological claim about "meanings" to UA, but rather the prior assumption (which Horwich grants for the sake of argument) that the mereological assumption explains linguistic decompositionality.

The problem about these approaches to Fodor and Lepore's argument is that Horwich and Robbins go too far by way of an anticipatory interpretational speculation about Fodor's claim of a mereological relation between "meanings," and, more specifically, that they go down a wrong alley in doing so. They engage with an irrelevant issue in interpreting the mereological claim. What their conjectures are based on is the assumption that Fodor and Lepore's argument for UA *implicitly intends to commit itself to a certain kind of semantic value* to the exclusion of other kinds of semantic value that are on offer at the intersection of metaphysics and the philosophy of language. But questions about the ontologically categorial nature of semantic values are independent of the question of what makes a mental representa-

tion (or, for that matter, a linguistic expression) have the semantic value that it does.

I emphasize this because it is crucial to a general assessment of Robbins' approach to Fodor's decompositionality-related rejection of LC, and to Fodor and Lepore's isolated defence of UA just encountered. Within section 4, I have already criticized Robbins (2005) as engaging in a wrong-headed investigation, insofar as he focuses on the status of linguistic decompositionality instead of conceptual decompositionality. I claimed this to form one of three fundamental respects in which the investigation of linguistic decompositionality is beside the point, when pursued with an eye on the evaluation of LC (which has indeed been the only motivation for investigating decompositionality). But Robbins' conception of what is at issue with LC and UA is fundamentally confused also in the respect that I am presently complaining about. In fact, this is the most perspicuous error in the strategic framework presupposed by the whole discussion of Robbins' noteworthy (2005) paper, and it is important to correct it.

Here is a representative statement by Robbins, which documents what he understands prototype-theoretical LC to amount to. Having himself quoted one of the strange loci in Fodor's work where CP-DD is attributed in place of LC, Robbins voices no protest and directly moves on to explain what he takes Fodor to be attributing, offering "a specific example of the sort of theory Fodor has in mind", namely prototype theory (the only relevant target, in fact). Robbins construes prototype-theoretical LC, then, as follows:[28]

> Say we split the difference between denotation-based and prototype-based semantics, as follows: the meaning of a word is given by a pair consisting of its denotation and the associated prototype, and the meaning of a phrase by its denotation alone. [...] This hybrid theory complies with PC by (a) incorporating denotations at both the lexical and the phrasal level and (b) dispensing with phrasal prototypes. [...] Notice, however, that the theory fails the test of reverse compositionality, since it blocks the possibility of deriving lexical meanings solely from phrasal meaning plus syntax. By bringing PRC into play, neo-Russellians seek to force hybrids like these out of the running. (Robbins 2005: 253.)

I find this similarly puzzling as many of Fodor's own attributions and arguments through which we have worked our way up to here. To be fair, Robbins avoids attributing the contradiction which Fodor is attributing in ascrib-

---

[28] The abbreviations employed by Robbins refer to the principles of compositionality and reverse compositionality, respectively.

ing DD in place of LC. In other words, there does not even seem to be an implicit commitment here to the incoherent idea of concepts abandoning their essences once they enter into a compositional compound. Nonetheless, the picture given in this quote, which expresses the strategic assumptions on which Robbins predicates his whole investigation, is confused in a different, novel way.

What Robbins is assuming here is that prototype theory is a *rival* to "a purely denotational (/truth-conditional) approach to semantics", as Robbins briefly characterizes on p. 2 the "denotation-based semantics" which he refers to in the quote. He is *assuming prototypes to form additional kinds of semantic values besides denotations*. But this is obviously not the case. Prototypes are *not* additional kinds of semantic value besides denotations. Prototype theory, like concept definitionism, or like covariationist concept atomism, is a theory of *what property of a mental state makes it have* the semantic value it does. Put differently, it is a theory of a concept's content-constitutive property. Any theory of concepts has to posit some such kind of property or other. As for the semantic value in question, psychological theories of concepts, such as prototype theory, practically only address concepts which can be naturally and non-technically be assigned a denotation in the first place (in effect, concepts expressed by adjectives or nouns). If anything, actual theories of concepts invariably *presuppose* denotations as semantic values. Prototype-theoretical structures posited for concepts are posited as structures that determine which entities *fall under* the concept. Saying that prototype-theoretical structures determine a denotation or reference is little else than a stylistic variation on this.

Robbins is mixing up different levels of theorizing—and correspondingly, different disciplines—in his understanding of prototype-theoretical LC. The theory family which really would have to play the role which he assigns to a "purely denotational semantics" (or "denotationalism", or "neo-Russellianism", as he often calls it) is the theory family of *concept atomism*. The comparison which he should be having in mind is one between theories which explain the fact that conceptual states represent what they do by positing certain kinds of content-determining mental representational structure, and theories that explain this without resorting to representational structures. Concept atomism and this sort of contrast, however, are left entirely unmentioned by him.

Let me relate my comments more directly to Robbins' quoted passage. The core idea of prototype-theoretical LC is described by him as the idea that "the meaning of a word is given by a pair consisting of its denotation and

the associated prototype, and the meaning of a phrase by its denotation alone". But one only need properly consider the notion of a prototype for a moment in order to discern, without much general metatheoretical reflection, that the statement cannot be made sense of. There are two different ways in which the term 'prototype' can be sensibly used in the context of prototype-style theories of concepts. On one way of using the term (the one which I have been adhering to), a prototype is the best instance of a concept. Hence, as a member of the concept's extension, the prototype is an entity at the level of denotation. On the other frequently encountered way of using the term 'prototype', a prototype is the probabilistically structured mental representation itself which constitutes the concept *of* a class of entities centering on a prototype. It is not a semantic value at all. Either way, it does not make any sense to equate the "meaning" of a concept—its semantic value(s)—with "a pair consisting of its denotation and the associated prototype". For the notion of a prototype is itself either located at the level of denotation, or it is the mental representation of that denotation. But is it not, as noted, a kind of semantic value besides denotation. As noted already, then, the notion of a prototype as such brings no new semantic value into play.[29] A fortiori, it does not form a semantic value that could be abandoned once the level of compositional phrasal concepts is reached.

Robbins' confused perspective on LC in his (2005) is of a piece with the uncashed talk of "meanings" and of a "theory of meaning" which he adheres to throughout. It appears that, under the umbrella of his talk of "meanings", he runs together the notion of a type of semantic value with what are in fact meaning-*constitutive* properties of representations, be they linguistic or mental.

The conflation is certainly one that is encouraged by Fodor's work. We have had ample opportunity to observe that loose talk is the order of the day here, and Fodor's way of speaking of "meanings" in the context of the quote given earlier from Fodor and Lepore (2001)—which Robbins reacts to as well—is no exception. Here, Fodor uses the term 'meaning' numerous times in a way that is really intended to refer to meaning-constitutive properties.[30] This is at least highly misleading. In fact, this is just one example of many which could be provided, and the employment of talk of "meanings" in a

---

[29] The situation is no better if you shift back to words, which Robbins speaks of. I do not want to complicate matters further here, and assume the point is clear enough.

[30] Thus, on the same page of Fodor and Lepore (2001: 59), the UA is reported as stating "that if the meanings of the primitives are stereotypes (or uses, or prototypes, or inferential roles, or whatever), then the meanings of the complexes are *also* stereotypes (uses, prototypes, inferential roles, etc.)".

manner that cuts across distinct theoretical categories occurs very often in Fodor's work on concepts.

7. Dismissing UA

In this final section, I will show that the UA together with the assumption of the existence of phrasal concepts analogous to compositionally understood phrasal expression directly leads to a reductio ad absurdum. Following that, I will say something about the utter baselessness of UA on general grounds.

Recall the point made in chapter 6. The compositional phrasal compounds of actual languages, and the compound concepts posited in analogy with them, have non-probabilistic content-determining structures. They trivially exhibit "definitional", "classical", structure. Let me refer to this trivial truth as the *definitionality of compositional phrasal compounds,* or DCPC, for short.

Before we move on, we should pause to remind ourselves what DCPC does not claim: It does not claim that phrasal expressions as such cannot express prototype-theoretical concepts. Of course they can, and maybe they very often do have such an interpretation. DCPC only points to the fact that actual phrasal expressions do not express prototype-theoretical concepts on their compositional interpretations. The inevitable availability of compositional phrasal interpretations does not preclude the possibility of assigning a correct, distinct interpretation of the prototype-theoretical sort. Conversely, an assignability of a prototype-theoretical interpretation to a phrasal expression does not remove the availability of assigning a correct, distinct interpretation of a compositional sort. This was emphasized in a separate section of chapter 6 already.

Within the present chapter, in the course of rejecting the argument from context-dependence, I already bemoaned the tacit presumption that exactly one type of concept must be expressed by any given expression, short of fully-fledged homonymies. There, the platitude set against this tacit presumption was relevant with regard to *words*. Right now, it bears mention with regard to *phrases*. In particular, we must make a general, systematic distinction between potential prototype-theoretical understandings of phrasal expressions and the minimal, compositional understandings of them. The reason why this bears emphasis is that, when UA is addressed, it indeed

tends to be addressed against the backdrop of the tacit dogma that a given expression can express exactly one concept only.[31]

With this preemptive clarification in place, we can turn to a principled rejection of UA. It seems to me that, once DCPC is recognized by somebody who has hitherto endorsed or seriously considered UA, her perspective rationally cannot but undergo a gestalt shift. Given DCPC, the concept-theoretical nature of compositional phrasal compound concepts is trivially fixed. But this reduces UA to absurdity. For together with DCPC, UA would yield the following argument:

1 DCPC (*Trivial truth*)
2. UA (*Assumption for reductio*)
∴ All lexical concepts have definitional structure.

The bottom line of this argument obviously gives us a reductio ad absurdum of UA. For of course, the trivially true assumption that concepts which are posited as compositional phrasal compounds are defined by their overtly posited constituent structures cannot hold the highly substantial discovery that whatever lexical constituents they might have are definitionally structured—effectively, that any lexical concepts tout court are definitionally structured. We need not even recourse to such facts as that definitionism about lexical concepts (let alone all of them) is regarded as false by virtually all psychologists and most philosophers, or that Fodor himself is staunchly opposed to it. It is sufficient to note that lexical concept definitionism is highly non-trivial. That is, it is sufficient to note that the UA would let a trivality about the embedding structures of lexical concepts decide about a highly non-trivial question which it cannot conceivably have *anything to legislate* about.

---

[31] For example, Robbins (2001: 330-32) does so too, which is the only explicit defence of the general principle of UA that I am aware of, apart from the brief remarks we have examined from Fodor and Lepore. In contrast to what is implied by his already criticized treatment in (2005), Robbins (2001) appropriately formulates UA over properties *constituting* semantic values indeed. The considerations which his earlier paper advances in support of UA, however, are merely tentative methodological considerations which in no way count against the *possibility* of, say, forming a compound PET FISH whose meaning is exhausted by its being formed by means of a conjunctive combinatorial rule from a prototype-centered structure PET and a prototype-centered structure FISH. I take it that, over and above the principled considerations I advance in the text, there is little of interest to be specifically argued against the relevant remarks in Robbins (2001). Generally speaking, I strongly doubt that Robbins would pursue his considerations in favour of UA if, contrary to the one-expression-one-concept picture, he allowed a potential systematic ambiguity in phrasal expressions between prototype-theoretical interpretations and minimal, compositional interpretations.

Moreover, UA would lead to an *infinite regress,* in requiring all representational constituents of a definitionally structured concept to be themselves definitionally structured, ad infinitum. UA cannot but be confused. The confusion is part and parcel of overlooking the trivial truth of DCPC.

In fact, a reductio argument analogous to the one displayed above could already be given independently of the recognition of DCPC, at a more generic level. Compositional phrasal concepts, whether or not they are noticed to have the definitional structures they trivially do have, are by definition concepts whose contents are determined by their constituent structure. This highly generic type of content-constitutive property contrasts with atomistic content-constitutive properties, which, by definition, do not involve representational structures. Since any concept theory, on pain of circularity or regress, must assume there to be a level of representational primitives at *some* point, it is clear enough that *any permissible concept theory will already violate UA with respect to content-constitutive properties at the interface between primitive and complex mental representations anyway*—wherever one happens to want to locate that interface.[32] (For the concept atomist, the interface at which UA would be violated lies between lexical and phrasal concepts. For the opposing, prevalent type of view, it lies at some non-obvious structural level below that of lexical concepts.)

Especially in light of this latter, more generic observation, Fodor, being the UA-endorsing concept atomist which he is, must be endorsing the position that UA does not really aim at a uniformity of content-constitutive properties *tout court* at all, but only at a uniformity claim that is restricted to structure-based content-constitutive properties. In other words, he must be thinking that the duality between types of content-constitutive properties that are atomistic and ones that involve representational structure would be claimed to be permissible, while a further duality *within* the class of content-constitutive properties involving representational structure would be impermissible. But for all we know, this would be *entirely ad hoc*. What is more, the reductio of UA from the top-down-generalizability of definitionism *remains* a reductio even relative to *structured* concepts. The UA-endorsing concept atomist, by performing the ad hoc restriction of UA's scope to structural content-constitutive properties, could save his atomism itself from an immediate refutation by the absurd top-down-generalizability of defini-

---

[32]   I deliberately avoid saying that the interface between primitive and complex *concepts* is the place where any permissible concept theory will violate UA. What is inevitable is not that some concepts are eventually representational primitives, only that some mental representations are: The representational primitives can also be nonconceptual representations. The distinction is orthogonal to the present issue.

tionism. But this would not change anything about the fact that UA, conjoined with the triviality of DCPC, would entail far too much. *The core absurdity derived for UA lies in the fact that it would miraculously exclude any probabilistic theories of lexical concepts merely on the basis of lexical concepts' potential constituency in compounds with classical structure.*

What all of this reflects is simply the utter baselessness of UA. Let me conclude with a few more general words on UA, before we dismiss UA for good, finally completing the task of rehabilitating prototype-theoretical LC. What UA would amount to, to borrow a term from Nelson Goodman, is that having a prototype-theoretical structure is, in some sense, an *expansive* property—a property which is had by everything that has a part which has it. As Goodman notes by way of examples, "the predicates 'is large' and 'is populated' are expansive".[33] It is obviously not the case that having an internal structure of a given kind is an expansive property, neither in some metaphysically necessary sense, nor in some nomologically necessary sense. If matters were otherwise, there would be no micro-, meso-, or macroscopic entities in the universe apart from entities recursively repeating a primitive inventory of the most basic structures of microphysics. Contrary to what UA seems to claim, then, the presence of a certain type of internal structure at a given level of aggregation cannot constrain what type of structure can be encountered at a higher level of aggregation. The only thing that *can* impose such a nomological constraint is *for the same kind of combinatorial mechanism to underlie* generation of the lower-level and the higher-level structure in a given case. But the involvement of different kinds of combinatorial mechanisms underlying lexical concepts and concepts providing the purely compositional interpretations of phrasal expressions is, in effect, precisely what LC commits itself to in the first place.

It seems that these matters could only come to be overseen by taking for granted, entirely without justification, that human cognitive resources hold exactly one combinatorial mechanism, namely a classical, definitionist, languagelike one to the exclusion of any other.[34] The present uniqueness pre-

---

[33] The quote is from Goodman (1977: 38), as is the definition, which I have transposed from the formal mode into property talk. Thanks to Oliver Scholz for help with locating the reference.

[34] A notable exception to this in the context of the compositionality arguments is Jesse Prinz (2002: ch. 11), though he does not mention the duality in the course of addressing a UA (he does not address LC or UA as such at all, surprisingly). Prinz does a very important job in directing attention to the modal fact that the mechanism for forming concepts corresponding to purely compositional interpretations of phrasal expressions is one that we have to be *able* to employ, not one that must be *actually operative* all or even most of the time ("a fallback system, not a mandatory mode of operation," as he puts it on p. 293). Still, I dissent with several basic presuppositions of Prinz's which his

sumption is intimately related to the one-expression-one-concept picture critiqued earlier, since with only one cognitive combinatorial mechanism taken into consideration as underlying the concept assigned to a phrasal expression, only one concept will be considered to be expressable by a phrasal expression (short of fully-fledged homonymies). This is precisely how Fodor runs his discussions, with many theorists following suit.

Just as in all other domains of empirical inquiry, we have to postulate whatever cognitive combinatorial mechanisms we feel required to postulate in their own right. Postulations of probabilistic concept structures on the one hand, and of "definitional"/"classical" concept structures on the other, are to be judged *on their own merit* and in *case-to-case consideration*. The postulation of each of these kinds of mechanisms has its own motivation and each of these mechanisms would appear to be operative relatively often in human beings. (In any case, I do assure the reader that I *can,* e.g., form a purely conjunctive concept from my plausibly probabilistically structured concepts PET and FISH. Who cannot?) Considering the reasons for such independent postulations has not been the business of this paper, however. The business of this paper was to expose the radically defective and confused nature of the alleged arguments against LC and against lexical prototype theory as such. There is no such thing as a true UA, an assumption that would magically exclude one of the two sorts of concept-construction mechanisms that we have ample independent motivation to postulate. The UA by itself is comparably baseless as the notion that the human biologist's dual existence assumption concerning type I, slow-twitch muscle fibres and type II, fast-twitch muscle fibres violates a uniformity constraint in human biology.

---

descriptions of the situation reflect. I cannot begin to raise these criticisms here, but Prinz's lucid treatment would make doing so worth a separate paper.

Let me note too that James Hampton (2000: 303), a cognitive psychologist noted for his work on concept combination, remarks in a comment which he makes rather deceptively in passing that "no [p]rototype theorist denies that we are *able* to think in terms of sets and interpret the logical form of sentences in terms of intersections, disjunctions or complements". I suspect that this remark reflects the influence by "the work of Jesse Prinz" which Hampton acknowledges in a general footnote to his book review of Fodor's *Concepts* (though he does not give any bibliographical reference or specific mention of Prinz in the rest of his review).

References

All published papers are referenced according to their year of original publication (except for a posthumously published manuscript). References in the dissertation follow the pagination of whichever source is adduced in the bibliography. Unless indicated otherwise, these sources are original places of publication.

Anderson, J.R. (1996). *Kognitive Psychologie,* 2nd German edition. Translation of *Cognitive Psychology and Its Implications,* 4th edition (1995). Heidelberg: Spektrum Akademischer Verlag.

Armstrong, D.M. (1968). *A Materialist Theory of the Mind*. London: Routledge.

Block, N. (1995). "On a confusion about a function of consciousness". *Behavioural and Brain Sciences* **18**: 227-247.

Boden, M. (ed.) (1990). *The Philosophy of Artifical Intelligence*. New York: Oxford UP.

Boghossian, P.A. (1991). "Naturalizing content". In B. Loewer and G. Rey (eds.): *Meaning in Mind: Fodor and His Critics*. Cambridge, MA: Blackwell. 65-86.

Crane, T. (1995). *The Mechanical Mind: A Philosophical Introduction to Minds, Machines, and Mental Representations*. First edition. Harmondsworth: Penguin.

Chalmers, D.J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford UP.

Colman, A.M. (2001). *Oxford Dictionary of Psychology*. Oxford: Oxford UP.

Cummins, R. (1989). *Meaning and Mental Representation*. Cambridge, MA: MIT Press.

Cussins, A. (1990). "The connectionist construction of concepts". First published in Boden (ed.) (1990): 368-440.

Dennett, D.C. (1969). *Content and Consciousness*. London: Routledge.

Dennett, D.C. (1984). "Cognitive wheels: The frame problem of AI". Reprinted in Boden (ed.) (1990): 147-170.

Dennett, D.C. (1991). *Consciousness Explained*. Boston: Little, Brown, and Co.

Devitt, M. (1996). *Coming to Our Senses: A Naturalistic Program for Semantic Localism*. Cambridge: Cambridge UP.

Dretske, F. (1981). *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press.

Dretske, F. (1988). *Explaining Behaviour: Reasons in a World of Causes*. Cambridge, MA: MIT Press.

Dretske, F. (1993). "Conscious experience". Reprinted in his *Perception, Knowledge and Belief: Selected Essays* (2000): 113-137. Cambridge: Cambridge UP.

Dretske, F. (1995). *Naturalizing the Mind*. Cambridge, MA: MIT Press.

Field, H. (1978). "Mental representation". Reprinted with a postscript in his *Truth and the Absence of Fact* (2001): 30-67. New York: Oxford UP.

Field, H. (2001). "Postscript to 'Mental representation'". In his *Truth and the Absence of Fact*: 68-82. New York: Oxford UP.

Fodor, J.A. (1981). "The present status of the innateness controversy". In his *Representations: Philosophical Essays on the Foundations of Cognitive Science* (1981). Cambridge, MA: MIT Press. 257-316.

Fodor, J.A. (1983). *The Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, MA: MIT Press.

Fodor, J.A. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: MIT Press.

Fodor, J.A. (1990). *A Theory of Content, and Other Essays*. Cambridge, MA: MIT Press.

Fodor, J.A. (1994). "Concepts: A potboiler". *Cognition* **50**: 95-113.

Fodor, J.A. (1996). "Connectionism and the problem of systematicity (continued): Why Smolensky's solution *still* doesn't work". Reprinted in his (1998-b): 113-125.

Fodor, J.A. (1998-a). *Concepts: Where Cognitive Science Went Wrong*. New York: Oxford UP.

Fodor, J.A. (1998-b). *In Critical Condition: Polemical Essays on Cognitive Science and the Philosophy of Mind*. Cambridge, MA: MIT Press.

Fodor, J.A. (1998-I). "There are no recognitional concepts, not even RED". Reprinted with an afterword in his (1998-b): 35-47.

Fodor, J.A. (1998-II). "There are no recognitional concepts—not even RED, Part 2: The plot thickens". First published in his (1998-b): 49-62.

Fodor, J.A. (2000). "Replies to critics". *Mind and Language* **15**: 350-374.

Fodor, J.A. (2001). "Language, thought and compositionality". *Mind and Language* **16**: 1-15.

Fodor, J.A. (2004). "Having concepts: A brief refutation of the 20th century". *Mind and Language* **19**: 29-47.

Fodor, J.A., Garrett, M.F., Walker, E.C., and Parkes, C.H. (1980). "Against definitions". *Cognition* **8**: 263-367.

Fodor, J.A. and Lepore, E. (1992). *Holism: A Shopper's Guide*. Oxford: Blackwell.

Fodor, J.A. and Lepore, E. (1996). "The red herring and the pet fish: Why concepts still can't be prototypes". Reprinted in Fodor and Lepore (2002): 27-41.

Fodor, J.A. and Lepore, E. (2001). "Why compositionality won't go away: Reflections on Horwich's 'deflationary' theory". Reprinted in Fodor and Lepore (2002): 43-62.

Fodor, J.A and Lepore, E. (2002). *The Compositionality Papers*. New York: Oxford UP.

Fodor, J.A. and McLaughlin, B. (1990). "Connectionism and the problem of systematicity: Why Smolensky's solution doesn't work". Reprinted in Fodor (1998-b): 91-111.

Ford, K.M. and Pylyshyn, Z.W. (eds.) (1996). *The Robot's Dilemma Revisited*. Norwood, NJ: Ablex.

Frege, G. (1897) (posthumous). "Logik". In G. Frege, *Schriften zur Logik und Sprachphilosophie: Aus dem Nachlaß*, ed. G. Gabriel (⁴2001): 35-73. Hamburg: Meiner.

Frege, G. (1918). "Der Gedanke". Reprinted in G. Frege, *Logische Untersuchungen*, ed. G. Patzig (⁴1993): 30-53. Göttingen: Vandenhoeck & Ruprecht.

Goodman, N. (1976). *Languages of Art: An Approach to a Theory of Symbols,* 2nd edition. Indianapolis: Hackett.

Goodman, N. (1977). *The Structure of Appearance,* 3rd edition. Dordrecht: Reidel.

Hampton, J.A. (2000). "Concepts and prototypes". Review of Fodor (1998-a). *Mind and Language* **15**: 299-307.

Horwich, P. (1998). *Meaning*. New York: Oxford UP.

Horwich, P. (2001). "Deflating compositionality". *Ratio* **14**: 369-385.

Johnson, K. (2006). "On the nature of reverse compositionality". *Erkenntnis* **64**: 37-60.

Keil, F.C. and Wilson, R.A. (2000). "The concept concept: The wayward path of cognitive science". Review of Fodor (1998-a). *Mind and Language* **15**: 308-318.

Laurence, S. and Margolis, E. (1999). "Concepts and cognitive science". Introduction to Margolis & Laurence (eds.) (1999): 3-81.

Machery, E. (2005). "Concepts are not a natural kind". *Philosophy of Science* **72**: 444-467.

Margolis, E. (1994). "A reassessment of the shift from the classical theory of concepts to prototype theory". *Cognition* **51**: 73-89.

Margolis, E. (1998). "How to acquire a concept". *Mind and Language* **13**: 347-369.

Margolis, E. and Laurence, S. (eds.) (1999). *Concepts: Core Readings*. Cambridge, MA: MIT Press.

McCarthy, J. & Hayes, P. (1969). "Some philosophical problems from the standpoint of artificial intelligence". In B. Meltzer and D. Michie (eds.), *Machine Intelligence 4:* 463-502. Edinburgh: Edinburgh UP.

Medin, D.L., Lynch, E.B., and Solomon, K.O. (2000). "Are there kinds of concepts?". *Annual Review of Psychology* **51**: 121-147.

Moore, G.E. (1899). "The nature of judgment". Reprinted in G.E. Moore, *Selected Writings*, ed. T. Baldwin (1993): 1-19. London: Routledge.

Murphy, G.L. (2002). *The Big Book of Concepts*. Cambridge, MA: MIT Press.

Osherson, D.N. and Smith, E.E. (1981). "On the adequacy of prototype theory as a theory of concepts". *Cognition* **9**: 35-58.

Peacocke, C. (1992). *A Study of Concepts*. Cambridge, MA: MIT Press.

Peacocke, C. (2000). "Fodor on concepts: Philosophical aspects". Review of Fodor (1998-a). *Mind and Language* **15**: 327-340.

Perry, J. (1986). "Perception, action, and the structure of believing". Reprinted in his *The Problem of the Essential Indexical and Other Essays,* expanded edition (2000): 101-124. Stanford, CA: CSLI Publications.

Pitcher, G. (1971). *A Theory of Perception*. Princeton, NJ: Princeton UP.

Prinz, J.J. (2002). *Furnishing the Mind: Concepts and Their Perceptual Basis.* Cambridge, MA: MIT Press.

Pylyshyn, Z.W. (1980). "Computation and cognition: Issues in the foundations of cognitive science". *Behavioural and Brain Sciences* **3**: 111-132.

Pylyshyn, Z.W. (1984). *Computation and Cognition: Toward a Foundation for Cognitive Science*. Cambridge, MA: MIT Press.

Pylyshyn, Z.W. (ed.) (1987). *The Robot's Dilemma: The Frame Problem in Artificial Intelligence*. Norwood, NJ: Ablex.

Quine, W.V. (1951). "Two dogmas of empiricism". Reprinted in his *From a Logical Point of View* (1953): 20-46. Cambridge, MA: Harvard UP.

Raffman, D. (1995). "On the persistence of phenomenology". In *Conscious Experience*, ed. T. Metzinger: 293-308. Paderborn: Schöningh.

Recanati, F. (2002). "The Fodorian fallacy". *Analysis* **62**: 285-89.

Rey, G. (1983). "Concepts and stereotypes". *Cognition* **15**: 237-262.

Rey, G. (1998). "Concepts". Entry in *Routledge Encyclopedia of Philosophy,* Volume II: 505-517. London: Routledge.

Robbins, P. (2001). "What compositionality still can do". *Philosophical Quarterly* **51**: 328-336.

Robbins, P. (2005). "The myth of reverse compositionality". *Philosophical Studies* **125**: 251-275.

Russell, B. (1903). *The Principles of Mathematics*. New York: Norton.

Russell, B. (1911). "On the relations of universals and particulars". Reprinted in B. Russell, *Logic and Knowledge*, ed. R.C. Marsh (1956): 105-124. London: Routledge.

Schiffer, S. (unpublished). Comment on Fodor (1998-I), read at the 1997 Central Division Meeting of the American Philosophical Association.

Smith, E.E. (1995). "Concepts and categorization". In *An Invitation to Cognitive Science*, 2nd edition, Volume 3, ed. E.E. Smith and D.N. Osherson: 3-33. Cambridge, MA: MIT Press.

Smith, E.E. and Medin, D.L. (1981). *Categories and Concepts*. Cambridge, MA: Harvard UP.

Stich, S.P. (1990). *The Fragmentation of Reason: Preface to a Pragmatic Theory of Cognitive Evaluation*. Cambridge, MA: MIT Press.

Stalnaker, R. (1994). "Stalnaker, Robert". Philosophical self-description in *A Companion to the Philosophy of Mind*, ed. S. Guttenplan: 561-568. Oxford: Blackwell.

Weiskopf, D. (unpublished). "The plurality of concepts". Manuscript available online.

Zimbardo, P.G. (1995). *Psychologie,* 6th German edition. Translation of *Psychology and Life,* 12th edition (1988). Berlin: Springer.