

Spencerism and the Causal Theory of Reference¹

Abstract

Spencer's heritage, while almost a forgotten chapter in the history of biology, lives on in psychology and the philosophy of mind. I particularly discuss externalist views of meaning, on which meaning crucially depends on a notion of reference, and ask whether reference should be thought of as cause or effect. Is the meaning of a word explained by what it refers to, or should we say that what we use a word to refer to is explained by what concept it expresses? I argue for the latter view, which I call "Darwinian", and against the former, "Spencerian" one, assuming conceptual structures in humans to be an instance of adaptive structures, and adaptive relations to an environment to be the effect rather than the cause of evolutionary novelties. I conclude with the deficiency – both empirically and methodologically – of a functionalist study of human concepts and the languages they are embedded in, as it would be undertaken in a paradigm that identifies meaning with reference or that gives reference an explanatory role to play for what concepts we have.

1. Introduction: Function and Genesis

I will consider a pair of theses about concepts, locating the notion of a concept in a rather broad and biological context: every species has certain abilities dependent on certain internal cognitive resources, and concepts in particular. Thus, e.g., both human and non-human primates represent numerical content, communicate, and form social relations, all of which depends on certain concepts. But no matter what training effort is made, all and only humans have the full concept of an integer, develop a 'theory of mind' depending on a repertoire of mental state concepts, and exhibit sensitivity to phrase structures as underlying human language (see Carey 2004, Fitch and Hauser 2004).

For each concept, then, that a species can form, we can ask about (i) its *genesis* (where it comes from, how it comes into existence), and (ii) how it *functions* (what it is used to refer to). While genesis and function seem distinct enough conceptually, we may well be tempted to explain the one in terms of the other. The following two theses capture this wide-spread temptation:

- (1) An organism O comes to have a concept, C, of some thing, X, because of functional relations that O takes up in relation to X.
- (2) How O's concept, C, of some thing, X, functions in relation to X, is explained by how O came to have C.

The first thesis explains the genesis of O's internal concepts – where they come from or what brought them into being – by how they function or what they mean in an environment: *function explains genesis*. The second explains how a concept functions – what it means – from its genesis – or where it comes from: *genesis explains function*.

I will deny both theses, but before doing that let me note that both sound quite natural. Take our concept of a doorknob, made famous through Fodor's (1998) discussion of it. What, except for doorknobs, and how we related to them functionally or causally, could possibly explain why we come to have a concept of doorknobs (how that concept came into being)? In fact, that appears to be a basic empiricist intuition. And what if not where a concept comes from or how it arises should explain what *doorknob*, as we use that word, means or refers to (how it functions)? The opposing thesis I explore and defend is this:

- (3) The functional relations to some thing, X, that O takes up by virtue of some concept, C, do not explain how C comes into existence, and are not explained by how C comes into existence.

¹ I am deeply obliged to two anonymous referees from B&P, and particularly the editor for their very useful comments. I am equally highly indebted to Maria E. Kronfeldner for effectively inspiring the whole project.

The first conjunct breaks a causal-explanatory path from function to genesis. While a concept's adaptiveness (functionality) may explain why it is selected, maintained, or proliferated, it does not explain why it came into being, and hence exists. This thesis excludes a causal theory of meaning, if understood in the following terms. Let *reference* be a relation between a word and a thing in the world, understood as something that determines the meaning of the word. Suppose further a *causal* analysis of that notion of reference, on which the meaning of *water*, say, as based on that reference will arise from our causal relations to water – the substance found in our environment on earth. On this view, the latter and how we relate to it will decide what our word *water* means, or what concept it expresses.

Again this view is intuitively plausible, for one should think that were it not for the existence of water, how we deal with it and are causally affected by it, we would not have come to have our concept of water. Hence our standing in these relationships to water, the substance, is surely at least a *necessary* condition for our coming to possess the concept of water; one might even be inclined to go for the stronger view that it is *sufficient* (where *else* are such concepts supposed to come from, one might ask). But the third thesis suggests that this is a *non sequitur*: while there is an important *connection* between meaning (concepts) and referential function, the latter is not explanatory for the former; and how our word *water* functions in respect to water, the substance, may not even be a necessary condition for the origins of that concept, hence our having it (although it is, trivially, a necessary condition for it to *function* as it does). We might come to live in an environment in which there was no water, or even a similar-looking substance. Then there is no more reason that we would not have the concept of water (leaving aside the question of whether we had been alive in this environment), than there is for a cave-dwelling animal that never sees the light not to actually have eyes.

While that analogy points to standard cases of “Darwinian theft”, I will argue that there may be a deeper reason why water, the substance, and relations to it may not be necessary for the origins of our concept of water. There are general arguments from the structure of Darwinian evolutionary theory that not only break the explanatory path from function to organismic structure, but according to which structures may be said to have functions even though these were never enacted in the creature's phylogenetic past. That a concept can be used to talk about water, the substance – could function in that fashion – is, on the same grounds, neither a necessary nor a sufficient condition for the concept to come into existence (see section 2).

The other half of the conjunctive claim (3) states that how a concept came into being need not explain how it functions with respect to an environment and things occurring in it. This viewpoint should be familiar from a Darwinian perspective, since in the latter the functioning of any given organismic structure that enters the evolutionary scene may vary. Which function(s) it will come to subserve in its evolutionary history will depend on accidents of a multifarious and unpredictable kind. Evolution has no foresight and often proves weird. The story of the jawbone is a famous example (and no exceptional case). We see it being used for cracking in reptiles, but also for hearing in mammals, and with other functions still. What function does it, as such or as a particular organismic structure, subserve intrinsically?² The question seems to have no factual answer, and in this sense, there is no intrinsic connection between structure and function. A given structure may have multiple functions, and it has these functions usually not as an isolated structure, but as a part of some functional whole, so that it is unclear to what the function should be ascribed. There *may* be a

² I here understand *function* to mean “selected effect” function. On the understanding of function as “causal role” as conceived by Amundson and Lauder (1994), the relation between a structure and its function is equally not intrinsic, and in a more obvious way: the notion of function plays no explanatory role, it is nothing that is independent from the structure as some given “purpose” of it, relative to which it can then be rationalized (ibid., 234-6).

single function associated with a given structure and no change of it, but even that may be for some accident of history, and the function will still not be something given when the genesis is given, or determined by it. Of course, the same function may also be subserved by anatomically very different structures.

The view denied in (3) is not, I will argue, “Darwinian”, but, as I will say, “Spencerian”. Far from having been superseded by the history of biology, the Darwin-Spencer dichotomy seems strangely alive in current thinking about human meaning and mind. While in the biology of “bodily” function, Spencerism is little more than a forgotten chapter in the history of biology, in the study of “mental” function it remains a success story, even long after the days of behaviorism have passed. A long sequence of externalist views of meaning suggests that meaning is a matter of external function and causal relations: meaning, it is thought, just *cannot* predate its use in referring. It is bootstrapped from reference, consists in the latter. Why this methodological dualism with respect to the two domains, the “mental” and the “physical”? On a unified view, if the environment cannot cause organic adaptive structures to come into existence, it cannot bring concepts into existence. Indeed, for the rest of this paper, I will suppose that “mental” and “physical” will be no more than informal labels picking out empirically different domains for theoretical inquiry, with no implications for either ontology or method.

2. The Darwin-Spencer Divide

The assessment above regarding Darwin and Spencer may seem surprising in the light of Dennett (1995) calling Spencer “an important clarifier of some of Darwin’s best ideas” (p. 393). These ideas, for Dennett, appear to be Darwin’s theory of evolution as such, for what he criticizes in Spencer’s account are their social-Darwinist applications. “Spencer was a Darwinian – or you could say that Charles Darwin was a Spencerian”, we read, and also: “the modern synthesis is Spencerian to its core” (ibid., 394). As a self-confessed “good Spencerian adaptationist”, Dennett adopts Godfrey-Smith’s “Environmental Complexity Thesis” (ECT), according to which “there is complexity in the organism in virtue of complexity in the environment” (ibid., 395; see Godfrey-Smith 1998).

Spencer presumably defended a strong version of this paradigmatically externalist claim, to the effect that environmental complexity is both necessary and sufficient for organic complexity. Godfrey-Smith opts for its mere necessity – a claim I have already said I will deny – and proposes, more precisely, to understand the ECT as a thesis about the *function* of cognition and its evolutionary *rationale*, in the teleonomic sense of function: cognition evolved for the sake of enabling the agent to deal with environmental complexity. While that appears as an answer to a “why”-question – it aims to explain why cognition exists in the first place, by appeal to a notion of what it is “for” – the thesis can also be read as an empirical/explanatory claim (which it was in Spencer’s version of it): structural complexity *outside* causes/explains structural complexity *inside*.

Note that thus understood, human language, to which I will repeatedly return, would seem a particularly unpromising start for the ECT. If the ECT had its way, externalist structures in the environment should predict or at least stand in some systematic relation to the core structural properties we can empirically attest in the workings of the computational system underlying human language. To my knowledge, little evidence suggests this, as long as one accepts standard textbook accounts of that system (see e.g. Culicover 1997, and see sections 4 and 7).³ At the same time, the case in question shows that more than the mere

³ Language is created by children in the absence of a “language model”, or linguistic structure in the environment. See Goldin-Meadow (2003), and Yang (2002) for a state of the art account of language acquisition. None of the points in the text is to say that the ECT does not have a reading under which it is “obviously true”, as Neander (1997: 567-8) claims, who considers it as “trivial” as the claim that “circulation is for transporting

necessity of environmental complexity for the origin of the internal structures would have to be invoked in defending an externalism as underlying the ECT, since otherwise even the strongest internalist, such as Chomsky (2000), need not disagree: there surely *are* various physical and environmental preconditions that are necessary for the language faculty to evolve phylogenetically and mature ontogenetically.

Irrespective of the particular case of human language, Godfrey-Smith agrees that the “critical feature of a Darwinian mechanism is that the variants produced do not bear a systematic relation to the environmental factors that exert selective pressure on the organism” (Godfrey-Smith 1998: 87). Looking at the matter with Darwinian eyes, indeed, at least the *origins* of organic complexity have not much to do with what the latter is “for”. For Darwin but not Spencer, the environment is silent on matters of design. It does not act on the organism so as to mold it and confer structural complexity to it, or so as to create adaptations. Darwinian environmental selection is like a spectator sport: choosing among forms that present themselves, as a consequence of which these forms get redistributed in following generations. The evolutionary process works with forms that are already there, and have an *internal* cause in random (undirected) mutations: they occur with no preferred orientation in adaptive directions. Natural selection itself is “blind” in a similar way: it does not see into the adaptive future.⁴

In short, nothing is *there* because of what leads it to be selected, or because of what it is useful for (even though much is not any more there because of what it was not useful for).⁵ At least in this sense, then, there is, for the Darwinian, no direct transfer or induction of structure, a direct causal or explanatory path from environmental complexity to internal structure. By crediting the environment with a more “active” role in structuring the organism, Spencer falls on the Lamarckian side in the Darwin-Lamarck-divide, with direct

blood around the body”. Even such “trivialities” regarding the externalist’s notion of what something is “for” may be unavailable for human language, though. All species communicate, but the specific way in which we do does not seem to explain specific linguistic structures. Possibly, mechanisms allowing language processing evolved irrespective of their later use in the human communication system (Hauser et al. 2002, Fitch and Hauser 2004). It remains unclear today what theoretical and practical purpose some identification of the “function of language” serves (see further Chomsky 2003b: 312-3). For a doubt on the coherence of the externalism entailed by the ECT see also Sterelny (1997).

⁴ On the so-called “statistical” interpretation of evolutionary theory (Walsh et al. 2002), moreover, it is a theory about the structure of *populations* only, not a theory of “forces” acting on *individual* organisms (and hence not a theory about any “creative” forces, see the next fn.). Not applying to individual-level phenomena, natural selection does not offer explanations of them.

⁵ The view I am endorsing in this paragraph is not particularly controversial, if indeed “no version of Darwinism holds that natural selection explains how, against a given genetic background, a particular mutation which confers an advantage arises” (Godfrey-Smith 1998: 93). Still, there is of course an ongoing debate on whether natural selection has a creative role to play on the evolutionary scene. Thus, Neander (1995) argues that “natural selection has a creative and not merely distributive role to play” (*ibid.*, p. 586), whereas, on the view above, the creative role is played by random mutations on the individual level alone. The “statistical” as opposed to the “dynamic” interpretation of evolutionary theory (fn. 6), supports the present view. Neander crucially appeals to the *cumulativity* of the adaptive process, in which some earlier preservations and proliferations of some co-adapted sequence of genes changes the probability of what subsequent variations will arise. Still, natural selection cannot but influence what subsequent variations will “*randomly* arise” (Neander 1995: 586; my emphasis), no matter whether they occur with other probabilities on the population level, given a changed gene pool in which previous mutations took place. Neander’s argument, one might say, depends on mixing two “cycles of causation”. In a single-step selection process, as she points out, there is a “causal isolation” between each random/select sequence. Here the causal processes invoked are the organism-internal ones giving rise to a random mutation. In cumulative selection, she further argues, one random/select sequence is *not* causally isolated from the next. But, now, the notion of “causality” has widened and invokes the gene pool in which certain distributional shifts have taken place.

That said, the whole issue about cumulativeness is possibly orthogonal to the concerns of this paper. What explanatory role, if any, natural selection needs to play for the origins of human language is a widely open question (cf. Hauser et al. 2002); and as regards human concepts, if these are *atoms* in the sense of Fodor (1998), hence primitives, it is not clear how they should have arisen gradually by cumulative selection.

repercussions for his paradigmatically associationist psychology. Thus, Spencer took it for a law that

“all psychical relations whatever, from the absolutely indissoluble to the fortuitous, are produced by experiences of the corresponding external relations; and are so brought into harmony with them” (Spencer 1855: 530).

As Godfrey-Smith puts this view: “the environment brings about an organic change exactly in its own image” (1998: 86), a mechanism meant to account both for within-generation and across-generations change. Since variation is acquired for Spencer, by the direct causal flow of environment to organism, for the sake of an increasing optimization in the latter, there is a diminished importance for the process of selection, as what is there to be selected is *already* brought in harmony with environmental pressures.

Spencer remained essentially uncompromising in his “Lamarckian” psychology throughout his life, despite certain concessions on the possibility of trial-and-error-learning in addition to classical conditioning (see Godfrey-Smith 1998: 88-92). Only the former form of learning can be termed “Darwinian”, in that new variants are produced randomly or spontaneously. Spencer’s hyper-externalism radicalizes Lamarckism, in fact, in that the Lamarckian organism’s “creative response to felt needs” – as when the giraffe stretches for a longer and longer neck to reach up to the treetops – is now *itself* thought to be molded by the environment. Lamarck by contrast crucially appealed, internalistically, to *drives* towards adaptive complexity (“inherent tendencies”, “innate proclivities”) as well. But these must come from somewhere, Spencer (and Dennett 1995) would object here, and the environment is meant to turn that trick, too.

There is a caricature of Lamarck, according to which, as Gould puts it, “a giraffe felt a need for a long neck, stretched ever so hard, and then passed the results of these successful efforts directly to its offspring” (Gould 2002: 179). An analogy to this caricature in psychology would be this: a human, feeling the need to name a certain – given, yet nameless – kind by a concept, thought ever so hard, came up with a suitable concept, and passed it on. The Darwinian response to that caricature would be that no need for a new concept will as such bring it into existence.

While Lamarckism in biology is nowadays regarded as *factually* wrong, if not *necessarily* (Dawkins 1983), its analogy in psychology is much alive: in particular, if ontogenetic *learning* is conceived as an “instructive” rather than a “selective” process, it *is* the sort of direct flow of structure or information from the environment to the organism that Darwin denied.⁶ The classical case would be Skinner’s (1969) theory of learning as an accumulation of learned habits or operants. Dawkins (1983: 20-23) however argues that learning, if understood as adaptive *improvement*, must *in principle* be supported by and depend on an independent Darwinian selective mechanism as well. Structure in the organism may be induced, but none of these changes are intrinsically adaptive. As long as explaining progressive evolution is our goal, Dawkins’ argument shows it is not coherent to merely add a Darwinian mechanism of evolution to an independently operative Lamarckian one.

In fact, Lamarckian instructive views of learning create a basic and unresolved paradox: it seems inconceivable that a human could “stretch for a concept it lacks”. The

⁶ As pointed out by Jerne (1967) (see also Piattelli Palmarini 1989 and Gazzaniga 1992), the history of biology provides many examples for a move from instructive models to selective models of some particular phenomenon. In the concrete case of the immune system that Jerne was primarily concerned with, it had been argued that, in the light of the immense numbers of antigens that existed in the environment, an organism could not possibly hold a stock of antibodies perfectly “adapted” to them. Still, it is now thought that an animal

“cannot be stimulated to make specific antibodies, unless it has already made antibodies of this specificity before the antigen arrives.” (Jerne 1967: 201).

What is relevant here is Jerne’s point that a given environmental problem as such does not tell us about its possible solutions. Indeed nothing *is* a *problem*, “objectively”, unless the organism has the conceptual resources to deal with it or grasp it as such, or has a conception of its possible solutions (cf. Piattelli Palmarini 1989: 13).

problem in its Ancient Socratean formulation, is not one about species-specific restrictions on what concepts each individual *can* possess (or how a monkey should come up with the concept of an integer, if by its nature it is incapable of that concept). The problem, as Jerne (1967) saw it, is that if you do not know what you are looking for, lacking the relevant concept, you can neither look for it nor find it. To look for it, you have to know what you are looking for, which by hypothesis you don't, lacking the relevant concept. You also cannot find it, for lacking the relevant concept you would not recognize that it was the solution to your problem.

In the reformulation of this classical problem by Fodor, it runs as follows. Suppose you wish to acquire a concept you lack, say the concept expressed by the word "red", presumably a primitive concept.⁷ Suppose, too, that learning it would require an inductive process of hypothesis-formation and testing. If that is the case, you have, at some stage, to form the hypothesis that some object falls under the concept expressed by "red" by virtue of being *red*. But that is not possible, because by assumption you lack the concept involved in this hypothesis (cf. Fodor 1998: 124). In another formulation, you can *only learn what you can represent* with given conceptual resources, but not what requires more powerful structures than those in place.

The internalist response to the paradox is that "learning concepts" is strictly speaking a misnomer and a child's learning certain concepts at certain times in ontogenetic development will be no more than a mapping given phonetic labels to given representational resources that are in place at those developmental stages, resources that themselves originate on independent and internal grounds. E.g., on Yang's (2002) model, learning grammars consists in *discarding* those grammars not matching environmental data (a form of "learning by forgetting", see Mehler and Dupoux 1994). Similarly, it may be that concepts are basically in place and usable if triggered in relevant critical periods, though discarded if not selected by an environment.⁸ The externalist and Spencerian response is also that the concepts are already there, but now they are supposed to be *there in the environment*; learning is the transfer of this structure to the organism. I return to learning and its paradox in sections 5 and 6.

Spencer, discarding the need for any internalist creative and non-functionally driven element in mental development, committed himself to show that the "mind, supposed passive, is moulded by its experiences of 'outer relations'" (James 1880: 188), arguing that intelligence developed out of cumulative modifications through the direct influence of the environment:

"The cohesion between psychical states is proportionate to the frequency with which the relation between the answering external phenomena has been repeated in experience." (Spencer 1855, cited by James 1880: 184)

That is, what we know is a function of what happened to us (or our ancestors). James objected to what he argued was "an obsolete anachronism, reverting to a pre-darwinian type of thought". Instead,

"new conceptions, emotions, and active tendencies which evolve are originally produced in the shape of random images, fancies, accidental out-births of spontaneous variation in the functional activity of the excessively instable human brain, which the

⁷ That is, a concept not decomposable into others on the basis of which it could be learned on a piece-by-piece basis together with a known mechanism of combination (such as syntax, or logic). Atomism for *conceptual primitives* will be assumed in what follows, a commitment any decompositionalist has to shoulder too, though she would assume a lesser number of primitives. See Fodor (1998), and also (Carey 1982: 47) for empirical and conceptual evidence against lexical decomposition.

⁸ For some striking recent experimental evidence see Hespos and Spelke (2004); see also Carey (2004) for how given representational resources for numerical representation may remain unused. Carey offers a "bootstrapping" mechanism for avoiding the problem that we cannot learn what we cannot represent; she does not, however, show that, ever in the ontogeny of the human integer concept, a stronger representational resource is generated from a strictly weaker one.

outer environment simply confirms or refutes, adopts or rejects, preserves or destroys, - selects, in short (...)." (James 1880: 184)

That picture is true, James sensibly concedes, for the "higher mental departments" only, as the lower ones are clearly not creative (or stimulus-insensitive, or non-input driven) in the sense James emphasizes: while there is (some) sense in which I am under the direct influence of the environment when I feel pain, say, there is no external or internal physical fact or configuration that, as far as we can tell, will force me to say or think anything in particular.

On James' "Darwinian" alternative to Spencer, then, an environment selects a concept for a use, but it does not create it, or shape the organism so as to let it have a concept where it lacked one before. There is evidently no denial of adaptation here, but there is a standard Darwinian denial that adaptation is a causal mechanism for the emergence of novel internal structure in the organism. Adaptedness is the long-term *effect* or *result* – indeed the necessary result – of environmental selection, but not a cause for it, as Spencer held. I will pursue this Jamesian internalist vision of concept possession for the rest of the paper, elaborating first on the meaning of "concept" as a theoretical term.

3. Concepts

On the Jamesian-Darwinian picture, the relation of content holding between an expression (e.g. word) and an object depends on (i) internally given structures the organism finds itself having, and (ii) an environment selecting among these structures, with the result of adaptively complex structures in the long term. The environment does not "produce" the internal structures (concepts); while intentional relations (what our thought and talk ends up being about) *does* depend on history and environmental causation, this needs to tell us nothing about the *origin* of concepts that figure in them. While this is unintuitive, it may be no more so than that the height of treetops and the need for high feeding should not cause giraffes to have long necks.

Concepts, in my technical usage, then, will be internal structures in organisms that enable these organisms to take up certain functions. In the human case, they will play a role in referential acts, while not having reference intrinsically, which would be like saying that organic structures are intrinsically adaptive. This is not to say that concepts in the present sense are non-semantic, though, a point to which I will return. It is not the case, in particular, that just any of these internal structures could take up any referential function. A concept suited to refer to a person, for example, cannot be used to refer to a tree (as such). The standard *externalist* explanation for this fact would be that persons *are* not trees, and that this factual property of persons enters into our concept of persons, which we then, and *because* of what properties it is caused by, do not apply to *trees*. But here the internalist replies that the response is as plausible as the view that some creature, by being exposed to an environment to which it stands in causal relations, acquires some specific concept of it. Nothing in biology suggests that we acquire internal complexity by standing in causal relations, unless we have suitable internal structures for that to happen. There is no more reason for an Alien to acquire concepts of heat, houses or justice, when placed in an earth-bound environment, than there is for a rat to acquire concepts of beauty, pianos, or mental states.

Concepts as here understood do not determine reference, nor do they originate from reference (i.e., functionally), though they enable a human to take up referential relations in a species-specific way, given a suitable environment. "Concept" *is* a functional notion in the intuitive sense that it enters into the explanation of the organism's functioning. But that does not mean a concept is *picked out* relationally by appeal to what it is used *for*. Amundson and Lauder (1994: 234-5) argue that identifying the function of an organismic form or structural pattern does not, in the factual practice of evolutionary biology, depend on figuring out purposes or selected effects, and in fact not on *any* reference to the environment. Again, the "function" of the form may be what it is *not* used for, and never was: ecological experiments may have to be set up to elicit its *possible* functions, or how it *can* be used. If so, its

ecological function does not precede the structure and is no presupposition for our understanding of it. Concepts in the present sense cannot then be picked out by their actual use or content for the same reasons that organismic structures cannot necessarily be picked out or understood in terms of their actual use. They could be picked out by their *possible* content, or their semantic *potential*, but they are not *explained* by these latter notions either.

In so viewing concepts I am departing from standard philosophical usage, where a concept is usually either (i) the *intentional content* of a mental representation structure, or (ii) a symbolic *mental representation* in the brain that intrinsically *has* such a content. In the first case, a concept is an intension (abstract object) or extension, in the second it is a syntactic object intrinsically *relating* to some external object or set-theoretic construct that it is “true of”.⁹ Concepts in the present sense have neither property. There is a *connection* between concepts in my sense and worldly extensions, in that the former are adaptive structures (they enable humans to make sense of the world). But they need not be caused by anything out there, and the connections in which they stand to external objects that exist and have causal properties are not intrinsic ones. In particular, even when functional, they do not necessarily track properties of the external world, as when we describe material things of the everyday world as solid and impenetrable, but physics tells us that these things consist of atoms that are mostly empty space. We still interpret the world in terms of categories like *fire*, *earth*, *air* and *water*, which are the Ancients’ four elements, even though only the last has turned out to denote a natural kind, and we know this. In the case of ordinary concepts like *mind*, *fate*, *belief*, *gratitude*, *Chinese*, *the heavens*, *perfection*, *war*, *London*, or *city*, we have essentially no idea how to pin down what the referents of these concepts out there are, and certainly it seems we cannot do so in a non-circular fashion, i.e. without re-using the concepts in question.¹⁰ I will henceforth denote concepts in the two *standard* senses above by means of the familiar CAPITALS, and use the word “concept” uncapitalized if I intend my own notion.

Of course, none of the internalism above entails that how a given concept actually functions with respect to an environment can be determined empirically without looking at that environment. The actual functioning in the relational sense of “function” is nothing that is determined “in the head”, but this is trivially so, not substantively. What environment an organism is embedded in is not determinable by looking at internal structures in the organism, just as a Euro-detecting vending machine built in Germany does not as such determine whether it will detect German, Dutch, Italian, or Greek Euros. This will depend on its history, or where the machine is placed and how it is used. This kind of “externalism”, a truism, does nothing to imply the kind of “Spencerian” causal flow from the environment to the organism that I dispute.¹¹

⁹ Note that in the case of the child’s putative concept of a *possible language* (on this see further section 7 below) there literally is nothing in the environment for that concept to be “true of” or to “apply to”: the structural patterns that this concept characterizes – phonetic, phonological and semantic mental representations, including the primes and relations making up these representations – are nowhere visible in the acoustic patterns that describe the outward physical side of language. It is widely argued, in particular, that the most basic syntactic categories – Noun and Verb – have neither a phonetic, nor a cognitive, nor a semantic or pragmatic rationale (Baker 2003 defines them in purely syntactic terms). Carstairs-McCarthy (1999; cf. 90, 161) provides an internalist, phonetic rationale for them.

¹⁰ With respect to *London*, one might have hoped that it is clear what the referent is: a geographical place. But this tells us virtually nothing about how we use that word in referring to things. As Chomsky (2000) points out, *London* isn’t a place, for cities can be destroyed and rebuilt at another spot, while remaining the same; moreover, cities are cultural and moral entities as well, viewed with or without their history; they are environments, as when we say they are polluted; and they can be all these things at the same time, as in the sentence *London is so old, corrupt, and polluted that we should destroy and rebuilt it 100 miles away* (see further Chomsky 2000, and Hinzen forthcoming).

¹¹ The Kripke and Putnam cases for an externalist view of content have equally no force against the present internalist view. Intentional relations depend on environmental causation (in the sense of selection), but they do not explain the concepts that *figure in* these intentional relations. The Twin-Earth story tells us something about

What brings it about that an organism with particular internal structures takes up intentional relations? This happens as a necessary consequence of embedding the organism in an environment to which its internal structures are suited in some ways rather than others, and better or worse so than other such structures. Intentionality does not explain the embedding of an organism in an environment and how it functions within it, for it is simply the embedding itself that will have various (and more or less successful) intentional relations as a result. These relations are enabled but not caused by the organism's internal structures, and certainly they do not cause these internal structures. To say that they do, in the above analogy, would be like saying that our external use of the Euro-detector causes its internal mechanisms to exist, rather than depending on them.

We may say, if we like, of our vending machine, that it has an internal structure (a concept) *referring to Euros*. But what exactly is this referent? As noted, the vending machine will have another "referent" in each country of the monetary union. We might save a theoretically interesting notion of reference by forming an equivalence class of objects in the extension of the internal structure, but this extension will contain not only different kinds of Euros, but a lot of other things than money (intelligently designed toy-Euros, say). There can be no objection to using such equivalence classes as the "semantic values" for internal concepts, but it seems clear that the resulting view of meaning will still be an internalist one, since it will be the internals of the machine or organism that will restrict and explain membership in the extension. We answer the question of which things are in that extension simply by noting which things *do* trigger the internal mechanism. A reference to the mechanism is crucial in delimiting the extension; hence the latter does not explain the former. Moving from vending machines to human infants, the story doesn't change. Thus, we find out which range of possible languages the child's innate knowledge of Universal Grammar (UG) permits, by looking at which languages factually *do* trigger the learning mechanism powered by UG (see further sections 4 and 7).

External uptake of function and reference is thus, at the same time, an arbitrary and a non-arbitrary matter. Arbitrary, because the internal structure as such cannot determine the range of things to which it will actually be applied or with respect to which it will be used during its history; non-arbitrary, because the nature of the organism's internal make-up will allow certain uses, while forbidding others. Again, to understand this restriction, we cannot look at the world. The world is not responsible for what concepts can mean (how they can be used); rather it is (partially) responsible for what they end up meaning (how they are factually used).

4. Semantics from teleology

Suppose that intentional relations to the world reduce to (i) structures in the organism, syntactically individuated, (ii) external entities, and (iii) causal relations between (i) and (ii). Call this the *causal-externalist* view of meaning. Concepts in the present sense, hence viewed as internal to the organism and as enabling it to take up certain referential acts rather than others, are not a part of this picture. In the absence of concepts-in-the-head playing this role, it must be symbols and their relations to the world alone that bootstrap meaning, are necessary and sufficient for it. Clearly, however, many things in the world, over and above the referent as such, may cause the occurrence of a mental representation having that referent as its content. Even though, for example, cow-skins are necessarily co-instantiated with cows

the possible uses of our concepts, nothing about why they exist. When being transferred to Twin-Earth without realizing it, our concepts will end up being used to refer to (what is, in one sense of the term "water") non-water. That is as interesting as history gets. Equally, Kripke's argument that Gödel remains Gödel even if we discover that he did not discover the Incompleteness Theorems, receives an internalist answer: it gives us an indirect piece of evidence about the nature of our concept of a person. The insight is that persons for us are not individuated by their scientific achievements. This does not explain why that is the concept it is (see further Hinzen forthcoming, a).

whenever they occur, humans of course distinguish the concepts of a cow and of a cow-skin. If human meaning is to remain a matter of causal and functional relations, one has to sieve out somehow *those* relations that do *not* intuitively constitute the meaning of a mental representation. The notion of a *proper* function has been thought to accomplish just this: the functions that are thought to explain and properly individuate meaning must be biological functions that the organism has been *selected for*.

Specifically, then, our concept of a cow means *cow* rather than *horse* or *donkey* because indicating the vicinity of any of the latter is not its proper function: it is selected for being produced in the presence of all and only cows; in this case and this case only it functions in accordance with its purpose and design. But again we may notice that what is *selected* in a Darwinian context is a *functional relationship* that *depends* on concepts in the present sense being there, rather than causing these concepts. In the absence of the relevant concept in the organism, the presence of cows alone will not necessarily produce it; and if it does trigger it in a particular human at a particular time, we cannot predict or explain this from an *independent* or *objective* identification of the external stimulus, because we only know *which* stimulus we are talking about from the *response* itself that makes use of a certain concept: we might have responded by thinking “cow-skin”, and then we would have concluded that the stimulus supposedly controlling or producing this response was the skin (cf. Chomsky 1959). *Most* human concepts do not seem to be associated with particular adaptive functions in a way that they could be understood as reflexive response-mechanisms to independently given and identified environmental variables that cause or control them.¹²

In the possibly more hopeful case of non-human concepts, say a frog’s neural representation of a fly, again *many* things will trigger the same response mechanism mediated by this representation, not just flies (inedible fly-imitations will do). But we may now take the wider ecological context into account and then point to what made the frog and its ancestors flourish as opposed to what did not, thereby finding reasons to distinguish among stimuli triggering what is in one clear sense the same response. Still, from this very conclusion we see that if we *limit* our attention to the internal structures of the organism and the frog brain’s representational powers, exactly the *same* mental representations or concepts are involved in responding to the flies and the fly-imitations (just as the internal mechanism of the Euro-detecting machine does not become different when it, through some historical accident, comes to detect toy-Euros). No matter what wider ecological context and history we include in our picture of the frog, it won’t distinguish between different concepts in two such cases, and not explain the origin of the concept (in the present sense) that is involved. It *does* distinguish, though, quite trivially, among the different (human) CONCEPTS associated with these concepts. Which CONCEPTS a given concept comes to support depends of course on history, also on the present account.

Generally speaking, we may regard adaptive functioning as setting a problem for biological explanation, rather than providing a solution. At times, describing adaptive functions *may* already be part of the answer: these are adaptationist explanations, which would then not only explain why some internal mechanisms is *selected*, but why it comes to *exist* in the first place. But their range of applicability is an empirical question and should not be rated as having the status of a default assumption, particularly not if we regard adaptation – with Amundson and Lauder (1994: 254), or G. C. Williams – as an “onerous” concept, appeal

¹² As Hauser et al. remark, “[U]nlike the best animal examples of putatively referential signals [such as vervet monkey alarm calls, W.H.], most of the words of human language are not associated with specific functions (e.g. warning cries, food announcements) but can be linked to virtually any concept that humans can entertain. Such usages are often intricate and detached from the here and now. Even for the simplest words, there is typically no straightforward word-thing relationship, if ‘thing’ is to be understood in mind-independent terms.” (Hauser et al. 2002: 1576) Cf. also fn. 15 above: it seems impossible to individuate e.g. a city in “objective” terms, i.e. in ways that do not depend on our concept of a city.

to which is a kind of last resort. That aside, it is not clear how these explanations would work in the case at hand, the case of *concepts*, given that the frog's fly concept is precisely the same, no matter whether it has adaptive or maladaptive consequences (reacts to flies proper, or to fly-imitations whose digestion kills the frog).

Papineau (2001) asks more generally what teleosemantics adds to our understanding of the nature of contentful states, and answers: "it tells us about the underlying nature of cognitive design, and thereby directs us to the past selectionist processes which fixed the real purposes of our cognitive parts" (p. 288). But as I want to argue in the rest of this section, there is no general reason why being so directed should help to understand the nature of cognitive design, or reveal anything about its "underlying nature". It is an open and entirely empirical question what explanatory role natural selection, if any, has played in the emergence of human cognition and design.

To illustrate this with our human knowledge of language, no one doubts that this particular competence evolved. Neither does anyone doubt that it "evolved by natural selection", if this option leaves open, as it should, which evolutionary factors precisely have carried which explanatory weights in its evolutionary origins. Human language function is here again part of a *problem* for biological explanation, and a part of the current *answer* is Universal Grammar (UG). We may view UG as a *structural type* characterizing core knowledge of language present in humans and underlying their use of particular languages, all of which are *variants* of this type. We should note that speculations on evolutionary history have factually played no role in unearthing the generative principles characterizing UG. An adaptive rationale for UG has been stipulated long after it existed (Pinker 1994), and not indeed *because* the specific structures of UG posited on empirical grounds would empirically *suggest* an adaptive rationale (which indeed they do not, by and large; see Uriagereka 1998 and Carstairs-McCarthy 1999 for discussion).

UG as viewed in the biolinguistic tradition (Jenkins 2000, Anderson and Lightfoot 2002) may thus rather be viewed as similar in nature to the "unities of type" of the rational morphologists of pre-Darwinian times, where they were taken to *constrain* the evolutionary process as independently explanatory generative principles, rather than being the *results* of it (on this interpretation of UG see further, Hinzen, forthcoming). By explaining commonalities of structure across widely divergent functionalities by appeal to a common ancestor of the several species that a given type was meant to unify, Darwin effectively subordinated the rational morphologists' explanatory principle – the "unity of type" – to the functionalist's – the "conditions of existence". Still, Darwin found the unities of types well in place before declaring them common ancestors; and declaring them common ancestors will no doubt not explain *these* (Amundson 1998). More generally, there appears to be little reason to believe that the basic theoretical concept of the formalists, the concept of *homology*, only makes sense in the light of the crucial theoretical concept of the functionalists, the concept of *adaptation*.¹³

By the same token, there is no reason, as of now, to follow the teleosemanticist's advice, as presented by Papineau, and to *presumptively* look for "selectionist processes" as something that "fixed the real purposes of our cognitive parts". We would be begging interesting empirical questions on what our design and its origin and rationale really are.

5. Spencerian semantics at work

Dretske (1988) develops the conflation of function and genesis into a whole explanatory framework. He centrally defends the claim that what our words or mental representations mean is a function of how they came to be, namely through "learning". His theoretical goal is

¹³ See again Amundson and Lauder (1994: 243) against the hegemony of functional explanations; and see Amundson (1998) and Gilbert et al. (1996) on the return of the "unity of type", a movement that has continued since.

“to install the informational processes underlying concept learning as the source of meaning and content” (Dretske 1994: 260), a goal that in turn subserves another, to show the intentional contents of the mind to be causally relevant in the explanation of behavior. Concepts, according to Dretske, flow naturally (without the mediation of intentional propositional attitudes) from our encounters with external environments. Behavior is strictly *conditioned*. It is not even a Lamarckian “creative” response to recurring situations, and even more clearly not a Darwinian selection of an independently originating internal structure.

That the acquisition of a concept could not depend on anything other than learning, hence could not have an *internal* cause, is inconceivable for Dretske, as only its having an external cause in the sense of being appropriately conditioned, makes it acquire causal relevance in the explanation of behaviour:

“(…) the content of beliefs and desires, of fears and intentions, the representational content that is featured in every explanation of behaviour by reasons, must derive from the development in learning of those circuits that constrain and structure voluntary action.” (Dretske 1994: 262)

To get intentional contents with causal relevance, one must “look to systems whose control structures are actually *shaped* by the kind of dependency relations that exist between internal and external conditions” (Dretske 1988: 95, my emphasis). Meaning is causally relevant to behaviour *only* in situations in which it is “associatively learned” (ibid., 96), by way of a “mechanical” response (ibid., 85):

“By the timely reinforcement of certain output – by rewarding this output *when*, and generally *only when*, it occurs in certain conditions – internal indicators of these conditions are recruited as causes of this output.” (ibid., 98)

But in a Darwinian frame of mind, Dretske’s insistence that internal structures subserving intentional behavior “must” derive from learning is the fallacy of conflating genesis and function: Darwin showed us how even the most perfect adaptedness of an organism to an environment does *not* require us to invoke Lamarckian shaping or environmental instruction. There is no inherent relation between adaptation and genesis, function and causal history, how something works and how it came to be. And of course there are standard empirical objections to Dretske’s need for learning, too: children know more than they can have learned from the kinds and amounts of inputs they receive, and what they know does not necessarily reflect what they hear (for a review of poverty of the stimulus arguments see Mehler and Dupoux 1994, Yang 2002, Anderson and Lightfoot 2002; for doubts about the significance of “associative learning” at large see also Gallistel and Gibbon 2001).

The notion of “development in learning” as such, informally understood, may in principle involve any amount of creativity on the side of the organism. But this is clearly not allowed in the technical notion of learning employed by Dretske. The internal control structures that are acquired are supposed to crucially not depend on the organism’s active contribution and the internal resources it brings to bear on understanding, but instead on what they *do* mean or indicate about the external circumstances to which the organism’s behavior is appropriated during learning (cf. Dretske 1988: 88). As Dretske (1994: 260) points out, “[W]hat converts a physical state – some condition of the brain – into a belief is outside the head”. This latter assertion has a reading on which it is trivially true, for the embedding of an organism and its concomitant functioning in the environment in which it is placed is nothing that the organism causes. But this is not the reading Dretske intends, where the conceptual structures subserving intentionality are mechanically induced, as in empiricist learning theories more generally. My point is again that it is an *empirical* question, when we look at particular processes that we intuitively describe as “learning”, whether the process we look at is no more than the *selection* of a given neural structure, whatever its origin. Wherever this would be arguably true, we could still say that:

instances of a mental representation, R, carry information about instances of an external property, P, if “Ps and only Ps cause Rs” is a law, which is essentially how Fodor (1990: 57) summarizes Dretske’s approach. But these co-variance relations would then merely indicate a *result* of an evolutionary process of natural selection and variation. As a result of this two-step evolutionary process, we do have reference, but the concepts that enable the latter would have a different cause.

By and large, the *behaviourist* view of meaning as deriving from associative learning or conditioning is as clear an expression of Spencerism as one can get. Dennett (1995) endorses this heritage unabashed (both the Spencerism and the behaviourism). Even Fodor (1990) is fully explicit that he regards Dretske’s as well as his own causal co-variance-approach to meaning as continuous with that of Skinner. Skinner’s account of language acquisition reduces language learning to social reinforcements mediating alterations in the strength of verbal operants: an operant response comes under the *control* of a type of discriminative stimulus as a function of the frequency with which the response elicits reinforcement when produced in the presence of stimuli of that type. While Chomsky’s critique of this view, Fodor argues, has effectively demolished his learning theory, it does nothing against the *semantics* that went with it. This seems immediately problematic for a view like Dretske’s, where learning precisely is what grounds semantics, and the difference between a learning theory and a theory of meaning becomes moot. Fodor, by contrast, gives up on the learning theory, and focuses on what he calls a “Skinnerian” semantics, which he regards as independent of a learning theory. It is to this view that I now turn.

6. Doing without learning

On Fodor’s (1990) view, instances of the mental representation underlying the word *dog* denote dogs because the former are under the causal control of instances of dogs in the environment. This is still a theory of meaning that bootstraps meaning from reference, or explains the origin of concepts functionally, through causal relations. Fodor renounces behaviourism, of course, but for a wrong reason, from the present point of view. The Fodorian reason is that his Skinnerian theory of how meaning comes about is reformulated as a theory of the semantics of *thoughts* (“propositional attitudes”), the existence of which Skinner denied (Fodor 1990: 55). The advantage that Fodor claims for this reformulation is that while speech acts are *actions*, and it is plain that what we *say* is not a function of the situation we are placed in, what we *think* is no action in this sense. However, what I *think* in a situation seems no more a function of what is true in it than what I *say* in it, not least because what can happen in it from my point of view is a function of what thoughts I, as opposed to a monkey, say, can think. For a concern with Spencerism, what matters is not the view we take on the metaphysics of propositional attitudes, but the view we take on *semantics*. What the Darwinian is opposed to is the idea that *concepts* – rather than functional relations to an environment as *based* on such concepts – are causally or mechanically generated by an environment or the things within it.

If the dependency between concepts and their external function is as I have stated it, extensions cannot be constitutive for concepts, as they remain also in Fodor’s later views. *Which* extension is supposed to determine the content of a concept can only be identified by looking at (understanding) the concept, hence is no cause of the concept. The present objection to Fodor’s account of concepts is thus that it provides a relational individuation of concepts. In fact, Fodor’s externalism is hard to square with his nativism. In Fodor (2001), Fodor considers a theory according to which the relational properties of “proto-concepts” (innate concepts, considered prior to being externally triggered), relations that he supposes to be constitutive for them, supervene on the proto-concepts’ (possibly unactualized) *dispositions* to enter into causal world-to-mind relations: “Maybe what makes a mental

representation a token of the proto-concept type CAT is its disposition to be triggered by cats” (Fodor 2001: 137). But when precisely am I disposed to react to cats by saying or thinking something involving the concept of a cat? I am only so disposed if I *have* the concept of a cat, so that there *is* something to be triggered in the first place, and indeed some appropriate thing (the concept of a mouse would not do).¹⁴ Hence, however concepts are individuated, they must be individuated independently of their dispositional relations to particular external objects.

We face the same difficulty of combining a biological nativism with an externalist individuation of content in Fodor’s discussion of the learning paradox described in section 2. Fodor’s own (1998) response to the paradox is interestingly rather more in line with Spencer’s than Plato’s. Fodor starts by considering giving up on *cognitivism*, the view that concept acquisition involves a process of hypothesis-formation and testing, and stipulating instead that acquiring a concept is *nomologically locking* onto the *property expressed* by the concept. Internal operations of the mind – forming beliefs, testing of hypotheses, concepts – drop out of the process, and since these internal operations caused the problem in the first place, the learning paradox disappears. This moves us into an externalist direction of explanation, but the internalist response is predictable. Suppose indeed that having a concept is “resonating to” the property that the concept expresses. Then two questions arise: (i) what if not having the concept itself *explains* our resonating to *the particular* property expressed by it, rather than some other property (in which case the concept would not be explained by our resonating to the property)?; (ii) why the dualism of “concept”, on the one hand, and “property” expressed by it, on the other (the latter being a metaphysical notion with possibly no naturalistic status at all)? The concept cannot be explained by its relation to the property, since if we ask what property that is, we have to refer to the concept.

While finding the above externalist and anti-cognitivist move attractive, however, Fodor himself raises the problem that even locking on to a property does ultimately not seem to be a process that is unconditioned internally. It clearly seems to depend on exposures to the *right kind* of things given in one’s experience: in particular, these typically being things *falling under* the concept to be learned. The concept of a doorknob is learned from exposures to (typical instances of) *doorknobs*, not giraffes or oysters, say. For Fodor this means that hypothesis-testing must be a part of the picture after all, for the learner has to use experiences with *doorknobs* to test and confirm hypotheses about what property the word “doorknob” denotes. Indeed, suppose the process of concept acquisition was *not* so cognitively mediated, or “brutely causal”. Then we would not predict that the concept of a doorknob is typically acquired from exposure to doorknobs and doorknobs and them only. But it is. Therefore the process is not brutally causal. In short, the relation between a concept and the experience from which it is learned or which it is “true of” *is* a special case of the evidential relation between a generalization and its confirming instances (Fodor 1998: 127, 132).

But then again, ask just *when* an experience would be a confirming instance of a generalization formed over what the word “doorknob” expresses. Clearly, only if it was a doorknob. But this thought the learner cannot think, lacking the concept that figures in it. Thus hypothesis-testing is not involved, the internalist would conclude. The process in which a concept that originated on independent grounds is triggered *must* be “brutely causal”, and the doorknob/DOORKNOB problem raised is a *reductio* of the hypothesis-testing model.

But why then *is* DOORKNOB acquired from doorknobs and not giraffes? The appearance may be as unsurprising as the one that organic structures not fitted to particular uses are not selected for or selectively retained. The sought relation between the concept of doorknobs and doorknobs is the relation of selection among given variants. It is non-cognitive

¹⁴ It seems similarly unclear how, if part of what’s innate about a concept is a specification of its “proprietary trigger” (ibid.: 138), we would know *what* trigger is “proprietary” if we do not know the concept it is proprietary for.

(and in this sense: brutally causal, though still not arbitrary), and it is non-instructive. This is a Darwinian-internalist response to the problem of concept acquisition, and a clear tension emerges in combining the view that the primary evolutionary mechanism is a selective one with the view that internal structures such as concepts are externally caused.

A concept may be like a key that opens a lock, but as in the case of other adaptive structures, the lock need not be referred to in explaining how the concept comes into existence.

7. Possible Languages

I have argued that concepts do not arise through content, though they can be individuated in terms of their *possible* content (or *possible use*), which varying environments then actualize. Put differently, while reference does not determine concepts, the environment provides external referents for concepts within a range of ways permitted by these concepts. This claim is testable, since the modalities used in this theoretical commitment imply that certain intentional relations would be such that we could *not* take them up, and that there would be CONCEPTS we could not learn, because we did not have, and could not, by the mere power of our will, develop the concepts needed for these CONCEPTS.

To exemplify a test of this prediction, we may return to the theory of UG (see section 4) as capturing the child's concept of a *possible language* (itself involving concepts of a *possible phoneme, word, concept, phrase, etc.*), a concept as such not depending on any actual language spoken, though any actual instance of a possible language depends, by assumption, on it. As a structural type, UG commits us to the human *impossibility* of certain languages not obeying the constraints of that type. These latter languages would be *logically* possible (and constructible in the laboratory, so to speak), but the commitment is that they would not be natively learnable by children in the way that actual human languages are. Smith and Tsimpli's (1995) experiments with Christopher, a prodigious language learner with severe cognitive deficits, who could *not* learn languages in which the experiments included some violation of a principle of UG, exemplifies just this.¹⁵

While, of course, in some intuitive sense, children "learn" languages, it is an empirical question whether human languages are learnable in the sense required for externalist and empiricist learning theories. There is no constraint in human biology for languages to be learnable in the latter sense, just as there is no constraint in human biology for immune responses to be so learnable. On the view that language acquisition is rather enabled by UG – that is, UG is viewed as a "language acquisition device" or LAD – UG, as characterizing the child's mind, provides "channels" within which experience is bound to flow. We know this acquisition process to be surprisingly independent of a large amount of feedback, corrections, and environmental control parameters, hence of hypothesis-testing in the Fodorian sense. Also in the case of the lexicon, the minor role that feedback and correction play in the child's acquisition of a vocabulary, its effortlessness, speed, and essential uniformity, despite a great variance in scanty evidence and cultural background, and the little specificity of the sensory input needed, all conspire to suggest that this is the right conclusion, too. The inductive gap between the system of knowledge attained and the evidence on which this happens seems equally vast.¹⁶

¹⁵ Examples of "impossible words" in the sense of Hale and Keyser (2002) are another case in point. These are words derived in violation of a syntactic constraint, and hence not natively learnable (e.g., the invented transitive verb *book* as in **He booked on the shelf*, with the meaning *He put the books on the shelf*. Still another case may be concepts that are impossible, not on syntactic, but on categorization grounds, such as the concept of a thing comprised of the four legs of a cow, which, as opposed to things like a flock or a hand that also consist of various parts, is not *one* unified thing for a creature with a conceptual system like ours.

¹⁶ See further Bloom (2000), and Gleitman and Landau (1994). Chomsky (2000: 120) points to "the rate of lexical acquisition (...), with lexical items typically acquired on a single exposure, in highly ambiguous

One can describe the process of learning a language without bringing *any* representational notion of propositional attitude into play, of the sort that centrally figures in a Fodorian “Computational-Representational Theory of Mind”. The child (as opposed to the linguist) needs no more beliefs or a “theory” of which structures develop in its head and get selected than it does in the case of other biological organs. Nor does it need to represent the contents of such a theory. The relational notion of representation drops out of the conceptual framework of the theory of UG, as it is not clear how understanding of the language faculty and its use is enhanced by letting structures in the organism “represent” or “correspond” to constructs in the environment of which they are “true” (see also fn. 14). In short, the theory of the LAD as investigated in the biolinguistic framework gives relational notions of “content” or “propositional attitude” no explanatory role to play (see further Chomsky 2003a).

Generative grammar in this way remains instructive for its lack of support for the hypothesis-testing model and the externalist idea of learning as based on representationally understood propositional attitudes. But it also teaches a philosophical moral regarding the “semanticality” of concepts in my sense: the range of conceptual structures determining a (humanly) possible language variant, while not being defined or generated relationally, are also not semantically *inert*. As I described it, they delineate a possible content. The concepts involved will not determine the contents of the child’s *actual* experiences – since it is a historical matter of what linguistic environments it will happen to be placed in – but it *will* determine what French, Bantu and German have in common. This is also a *content* (though a more abstract one), and not merely a “form”. It is in this sense that concepts in the sense of this paper, though they are part of the organism’s intrinsic structure and thus independent of an environment that the organism is placed in, *have* a “content”. They do not “reach out” to the world in the sense of determining whether French or German is spoken, but they have semanticality all the same.¹⁷

8. Conclusions

I have argued for the good use of a notion of ‘concept’ on which concepts enter into the explanation of meaningful reference and language use, while not being either intensions or worldly extensions. They do not come into existence through their functioning in an environment, nor do they cause their functioning. The origination of concepts and their functioning must be kept apart. Applying a Darwinian rather than Spencerian conception of progressive evolution to the case of concept and language learning thus suggests abandoning prevailing tendencies to regard the environment as a (direct or indirect) cause of our concepts. While externalism appears to be a general tendency in our intuitive thought, the empirical study of the human language faculty does not suggest, as of now, to give external function an explanatory role to play in the genesis of that part of our cognitive structure and design. A serious doubt concerning the *explanatory significance* of notions such as content, representation, and reference, appears warranted.

The picture I have defended endorses nativism in the form of what amounts to a truism in a biological context: the organism’s response to the environment depends on its internal resources, and in a Darwinian frame of mind, internal cognitive structures will not generally arise from reference or function, or from standing in causal relations, alone. Nativism thus understood – a form of biological *internalism* as inaugurated in the 19th century (see Amundson 1998 and Gould 2002, Ch. 4) – gives rise to a basic conflict with current

circumstances, but understood in delicate and extraordinary complexity that goes vastly beyond what is recorded in the most comprehensive dictionary”.

¹⁷ The older notion of a *synthetic a priori* judgement in the Kantian sense is what is needed here: the child’s judgements on what a possible human language is are *a priori* – they do not derive from experience – but they are not *analytic*, for they have a *content* in the sense just described: the structure of UG does provide substantive *information*, which follows from nothing in logic.

externalism, if the latter is to be more than the non-contentious claim that what our internal concepts end up being used to refer to is a matter of history, contingency, and chance.

References

- Anderson, S. R. and D. W. Lightfoot (2002), *The Language Organ. Linguistics as Cognitive Physiology*, Cambridge University Press.
- Antony, L. and N. Hornstein (eds., 2003), *Chomsky and His Critics*, Blackwell.
- Amundson, R. (1998), "Typology Reconsidered: Two Doctrines on the History of Evolutionary Biology", *Biology and Philosophy* 13, 153-177.
- Amundson, R. and Lauder, G. (1994), "Function without Purpose", in D. Hull and M. Ruse (1998), *The Philosophy of Biology*, Oxford University Press, 227-257.
- Baker, M. (2003), *Lexical Categories*, Cambridge University Press.
- Bloom, P. (2002), *How Children Learn the Meanings of Words*, Cambridge: MIT Press.
- Carey, S. (1982), "Semantic Development: The State of the Art", in Gleitman, L., E. Wanner (eds.), *Language Acquisition: State of the Art*, New York: Cambridge University Press.
- Carey, S. (2004), "Bootstrapping and the origin of concepts", *Daedalus*.
- Carstairs-McCarthy, A. (1999), *The Origins of Complex Language*, Oxford University Press.
- Chomsky, N. (1959), A review of B. F. Skinner's *Verbal Behaviour*, *Language* 35,1, 26-58.
- Chomsky, N. (2000), *New Horizons in the Study of Language and Mind*, Cambridge UP.
- Chomsky, N. (2003a), "Reply to Rey", in Antony and Hornstein (eds.), 274-286.
- Chomsky, N. (2003b), "Reply to Millikan", in Antony and Hornstein (eds.), 308-315.
- Dawkins, R. (1983), "Universal Darwinism", in D. Hull and M. Ruse (1998), *The Philosophy of Biology*, Oxford University Press, 15-37.
- Dennett, D. (1995), *Darwin's Dangerous Idea. Evolution and the Meanings of Life*. New York.
- Dretske, F. (1988), *Explaining Behavior*, Cambridge, MA: MIT Press.
- Dretske, F. (1994), "Fred Dretske", in Guttenplan, S. (ed.), *Companion to the Philosophy of Mind*, 259-265.
- Fitch, W. T. and M. C. Hauser (2004), "Computational Constraints on Syntactic Processing in a Nonhuman Primate", *Science* vol. 303, 377-380.
- Fodor, J. (1990), *A Theory of Content and other Essays*, MIT Press.
- Fodor, J. (1998), *Concepts. Where Cognitive Science Went Wrong*, Oxford: Clarendon.
- Gallistel, C. R., & Gibbon, J. (2001), "Computational versus Associative Models of Simple Conditioning", *Current Directions in Psychological Science*, 10, 146-150.
- Gazzaniga, M. (1992), *Nature's Mind*, New York: Basic Books.
- Gazzaniga, M. (1997), *Conversations in the Cognitive Neurosciences*, MIT Press.
- Gilbert, S. F., Opitz, J. M., and R. A. Raff (1996), "Resynthesizing evolutionary and developmental biology", *Developmental Biology* 173, 357-372.
- Gleitman, L. and B. Landau (eds., 1994), *The Acquisition of the Lexicon*, MIT Press.
- Godfrey-Smith, P. (1998), *Complexity and the Function of Mind in Nature*, Cambridge University Press.
- Goldin-Meadow, S. (2003), *The Resilience of Language: what gesture creation in deaf children can tell us about how all children learn language*, New York.
- Gould, S. J. (2002), *The Structure of Evolutionary Theory*, Harvard University Press.
- Hale, K. and J. Keyser (2002), *Prolegomena to a Theory of Argument Structure*, MIT Press.
- Hauser, M., Chomsky, N., and W. T. Fitch (2002), "The Faculty of Language: What is it, who has it, and how did it evolve?", *Science* 298, 1569-1579.
- Hespos, S. J. & Spelke, E. S. (2004), "Conceptual precursors to spatial language", *Nature*, 430, 453-456.
- Hinzen, W. (forthcoming), *Minimal Mind Design*, two volumes, Oxford University Press.

- James, W. (1880), "Great Men and Their Environment". In *The Works of William James*. Cambridge, MA, 1975, vol. 6, 163-189.
- Jenkins, L. (2000), *Biolinguistics*, MIT Press.
- Jerne, N. K. (1967), "Antibodies and learning: Selection versus instruction" In G. C. Quarton, T. Melnechuk, & F. O. Schmitt (Eds.), *The neurosciences: A study program*, New York: Rockefeller University Press, 200-205.
- Mehler, J. and E. Dupoux (1994), *What Infants Know*, Oxford: Blackwell.
- Millikan, R. (1984), *Language, Thought and Other Biological Categories*, MIT Press.
- Neander, K. (1995), "Explaining complex adaptations", *British J. Phil. Science* 46, 583-587.
- Neander, K. (1997), "The Function of Cognition", *Biology and Philosophy* 12, 567-580.
- Papineau, D. (2001), "The Status of Teleosemantics, or how to Stop Worrying about Swampman", *Australasian Journal of Philosophy* 79, 2, 279-289.
- Rey, G. (2003), "Chomsky, Intentionality and a CRTT", in Antony and Hornstein (eds.), 105-139.
- Smith, N. and Tsimpli, I.-M. (1995), *The Mind of a Savant*, London: Blackwell.
- Spencer, H. (1855), *The Principles of Psychology*, Westmead: Gregg, 1970.
- Spencer, H. (1899), *First Principles*, New York: Appleton.
- Skinner, B. F. (1969), *Contingencies of Reinforcement*, Appleton-Century-Croft.
- Sterelny, K. (1997), "Where does Thinking Come From?", *Biology & Philosophy* 12, 551-566.
- Walsh, D.M., T. Lewens, and A. Ariew (2002), "The Trials of Life: Natural Selection and Random Drift", *Philosophy of Science* 69, 452-473.
- Yang, C. (2002), *Knowledge and learning in natural language*, Oxford University Press.