

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, "Darwinian" view, 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 both the empirical implausibility and the methodological undesirability of a functionalist study of human concepts in the relational sense of "function", 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, where the notion of a concept is taken in a rather broad and informal sense: in this sense, all species have certain abilities dependent on certain concepts, allowing them to make sense of certain things in their environments rather than others. This capacity, I assume, requires no specific conscious awareness, and is of its nature limited. Thus, while chimpanzees engage in social relations, have a measure of numerical competence, and can concatenate symbols, they exhibit no sensitivity to phrase structures (cf. Fitch and Hauser 2004). Young human children have no such limitation, while presumably having others. Viewed in this way, having some species-typical conceptual capacity, hence having certain concepts rather than others, is just to have another organismic trait. Concepts are structures in the organism, part of its anatomy.

For each concept, now, we can in principle 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 or what it

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.

means). While genesis and function seem conceptually distinct enough, we may well be tempted to explain one in terms of the other. The following two theses capture this 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)? But my opposing thesis 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.

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, 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 *explicates* the meaning of the word, or tells what the meaning *consists* in. 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 a connection between meaning (concepts) and referential function, the latter is not explanatory for the former. 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 analog 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 are not 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 function of any given organismic structure that enters the evolutionary scene may be various. 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 in other functions still. What “function” does it, as a particular organismic structure, subserve intrinsically?² The question seems to have no factual answer. 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 single function for a given structure and no change of it, but 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

² Of course I here implicitly 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: the notion of function plays no explanatory role, it is nothing that is independent from the structure as some given “purpose”, relative to which we can then rationalize the structure (*ibid.*, 234-6).

of “bodily” function, Spencerism is little more than a forgotten chapter in the history of biology, in the study of “mental” function it remains quite 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 into existence, it cannot cause 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.³

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 –

³ Methodological monism in this sense is vigorously defended in Chomsky (2000).

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*.

Thus understood, human language, with whose biological aspects I will be primarily concerned, 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.⁴ To my knowledge, little evidence suggests this. Neither our experience in ontogeny nor that of our ancestors exhibits the needed information to make the inductive step from a merely acoustical description of the external side of language to its more abstract levels of representation, such as phonetics and phonology, let alone syntax and semantics (see sections 4 and 7 for more on this).⁵ At the same time, the case in question shows that more than the mere *necessity* of environmental complexity for the origin of the

⁴ I think that this would be so even if we understand the ECT as a constitutive rather than an explanatory or genealogical claim.

⁵ Language is created by children in the absence of a “language model”, or linguistic structure in the environment. See Goldin-Meadow (2003). See 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 blood around the body”. Still, and again in the case of human language, no such trivialities regarding the externalist’s notion of function may be forthcoming: it is not known what language is “for”, one concrete problem being that its actual uses – say, in communication – do not seem to tell us much about the specific structures that enter into these uses, to the extent that standard generative analyses of these structures are right. Very possibly, current research suggests, crucial 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 what theoretical and practical purpose the identification of the “function of language” serves in the biolinguistic program (Jenkins 2000), where our common sensical *concept* of language and possible general features of it seem basically irrelevant (for a recent discussion see Chomsky 2003b, 312-3). For a doubt on the coherence of the externalism entailed by the ECT see also Sterelny (1997).

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. A structured environment of a certain kind is clearly *necessary* for the language faculty to evolve, both phylogenetically and 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: occurring with no preferred orientation in some adaptive direction, they are “blind”. Natural selection may be said to be blind, too: it does not see into the adaptive future.⁶

In short, if an adaptation arises, this is not because some internal structure was molded on the model of some external one. Nothing is *there* because of what leads it to be selected, or what it is useful for (even though much is not any more there because of what it was not useful for).

⁶ On the so-called “statistical” interpretation of evolutionary theory, 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). Walsh et al. (2002) compare natural selection to a sorting process of biased coins, as a consequence of which the frequency of the heads-to-tails changes over a sequence of trials (the bias of each individual coin corresponds to the fitness of individual organisms). The systematic bias explains the sorting outcome, but the bias is not the property of any individual coin, being a mean. In the same way that sorting is not a *force* acting on an organism; in the same way, natural selection does not cause any individual organism to live or die. Sorting is an ensemble-level phenomenon, depending on a *sequence* of trials. Not applying to individual-level phenomena, natural selection does not offer explanations of them.

Contrary to what Spencer held, organisms do not “adapt themselves” to changing circumstances (although it may be true that it constructs its environment, another matter). They “do” no such thing, do not “react” to an environment. Rather, they *vary*. Supplied with an internal structure, whatever its origin, they find themselves either lucky (selected) or not.⁷

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 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 accounting both for within-generation and across-

⁷ 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, in a controversy with Sober (cf. Sober (1995)), 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), which Walsh et al. (2002) forcefully defend, supports the present view. As for Neander’s suggestions, she appeals crucially 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). It still has no hand in how each single mutation takes place, the mechanisms of which remain the ones they are, 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. On Godfrey-Smith’s (1998) related claims on natural selection’s “creative power” see fn. 11.

All that said, the whole issue about cumulativity is possibly orthogonal to the concerns of this paper. What explanatory role natural selection needs to play for the origins of human language is an open question (cf. Hauser et al. 2002). And as regards human concepts, moreover, 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.

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.⁸

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.⁹ Only the former form of learning can be termed “Darwinian”, in that new variants are produced randomly or spontaneously, among which the environment then merely selects. 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 to *internal* drives for adaptive complexity (“inherent tendencies”, “innate proclivities”) as well. But these must come from somewhere, Spencer (and Dennett) 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). According to Gould, Lamarck’s commitment was not quite that, but merely that a “change in environment can, in an almost mechanical way, elicit an organic response in terms of altered habits”. These are what results finally in an alteration of inherited organic form. Darwin fails to endorse such a mechanism as well. Changing circumstances induce new selection pressures, no new variations. Note this analogy to the caricature view of Lamarck in psychology: that a human, feeling the need to name a certain –

⁸ For Darwin, selection *is* crucial, and evolution does *not* refer to some kind of necessary process of increasing perfection, complexity, and specialization, while, according to Spencer, evolution could only end in the establishment of the greatest perfection and the most complete happiness: cf. Spencer 1899, 530).

⁹ Cf. Godfrey-Smith (1998), p. 88-92 for discussion.

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 possessing 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 still 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. The classical case would be Skinner’s (1969) theory of learning as an accumulation of learned habits or operants. Dawkins (1983, 20-23) by contrast argues that learning, if understood as adaptive *improvement*, must *in principle* be supported by and dependent on a Darwinian selective mechanism. Structure in the organism may be induced, but none of these changes are intrinsically adaptive.¹⁰ As long as explaining progressive evolution is our goal, this is to implicitly *deny* the coherence of Spencer’s conceding a Darwinian mechanism of evolution after Darwin’s theory appeared, in addition to but independent of the Lamarckian one he was endorsing.¹¹

¹⁰ Niels Jerne pointed out that factually, whenever biologists postulated an instructive model of learning in some concrete case, a selective model turned out to be more appropriate. In the concrete case of the immune system, biologists felt inclined to argue, in the light of the immense numbers of antigens that existed in the environment, that an organism could not possibly hold a stock of antibodies perfectly “adapted” to them: the only way the antibodies can get their structure is by direct molding from the antigen. But on today’s received view, an animal

“cannot be stimulated to make specific antibodies, unless it has already made antibodies of this specificity before the antigen arrives. It can thus be concluded that antibody formation is a selective process and that instructive theories of antibody formation are wrong” (Jerne 1967, 201).

The design of artificially synthesized antigens confirmed that the organism came fully prepared for them, thus harboring solutions for problems it could not have been caused to solve in evolution. Jerne (1967) sees the immune system merely to provide an illustration for the general point that a given 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. further Piattelli-Palmarini 1989, 13, and Gazzaniga 1992). The pre-theoretic intuition that, given the enormous and unbounded number of circumstances in which a human organism may find itself during ontogeny, it can’t possibly come *conceptually* prepared to deal with all of them, and find appropriate responses to whatever circumstance it is confronted with, may fall in the same way it did in the immune system case. For a Darwin-inspired “selective” view of language learning, see Yang (2002).

¹¹ What makes this Spencerian compatibilist view coherent, Godfrey-Smith suggests, is his commitment to gradualism. “Selection’s creative power, with respect to any new mutation, lies in its contribution to the creation of the conditions under which the mutation is accessible” to a population (Godfrey-Smith 1998, 94). Jamesian views of the kind explained below are “saltationist” for Godfrey-Smith, by contrast. But why call natural selection’s gradual

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 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 a phrase structure, if it lacks that concept). The problem, as Jerne (1967) pointed out, 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.¹² 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). “Learning concepts”, this line of reasoning suggests, is a label for a trick that nothing can turn: if learning denotes a transition from a stage in which you lack a concept to a stage in which you have acquired it, there is no such thing.

restructuring of the genetic composition of a population “creative”? No causal processes of the sort that act on individuals are involved in the restructuring process, and hence we also cannot say that (individual) fitness between organism and environment is “produced” by natural selection, as Spencer put it. It is interesting to see how, from an externalist point of view, Lamarckism must appear as basically right in the end, despite Darwinism, just as it did for Spencer: for although natural selection is no molding mechanism, hence Darwinism supports an internalist view, it ultimately produces the *same* kind of equilibration that the Lamarckian mechanism produces. It just does so more *indirectly*.

¹² That is, a concept not decomposable into others on the basis of which it could be learned on a piece-by-piece basis. Atomism for conceptual primitives will be assumed in what follows (a commitment also for decompositional views, though for the decompositionalist there are fewer primitives than Fodor assumes). It is further useful to reserve the label of “concept” for primitives, so that structured meaning entities come out as phrases or propositions.

The internalist or Platonic response to the paradox is that “learning concepts” is a misnomer and what we call concept acquisition is no more than a process of fitting phonetic labels to concepts we already have, originating on independent and internal grounds. Just as, on Yang’s (2002) language model, learning grammars consists in discarding those grammars not matching environmental data (a form of “learning by forgetting”), available concepts not selected by an environment will remain unused.¹³ The externalist or Spencerian response is also that the concepts are already there, but now they are supposed to be there in the environment, and learning is the transfer of this structure to – or its induction in – the organism.¹⁴

We return to learning in sections 5 and 6. For now we note that for Darwin (and partially even for Lamarck) there is a “creative” element in the genesis of new variants that is lacking in Spencer: the creativity in question in the case of Darwin is nature’s creativity in generating new forms spontaneously, in no adaptive direction and independent of functional needs. Articulating this view with respect to psychology, William James criticized Spencer precisely for the complete lack of this creative element in his theory of mental development in his *Principles of Psychology* (1855), which committed itself to show that the “mind, supposed passive, is moulded by its experiences of ‘outer relations’” (James 1880, 188). Spencer argued that intelligence developed out of cumulative modifications through the direct influence of the environment, according to the following behaviorist principle:

“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)

¹³ For striking recent experimental support for this idea see Hespos and Spelke (2004).

¹⁴ For a neglected third (developmental) possibility to resolve the learning paradox – which does not change the basic internality story I will be telling – see Molenaar (1986).

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” (*ibid.*, 189):

“I can easily show that throughout the whole extent of those mental departments which are highest, which are most characteristically human, Spencer’s law is violated at every step; and that as a matter of fact the 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 outer environment simply confirms or refutes, adopts or rejects, preserves or destroys, - selects, in short, just as it selects morphological and social variations due to molecular accidents of an analogous sort.” (James 1880, 184)

To put this in slightly different terms, the mind is like a generative engine randomly producing structures that, while maybe conforming to certain constraints, are genuinely new on the scene. The environment is a mere arbiter over their fate. 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, or see a red object,¹⁵ 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.¹⁶

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

¹⁵ Also that, we should note, will only be the case under the assumption that I am an organism having the relevant internal structures in place (a child born with Hereditary Sensory and Autonomic Neuropathy Type-5, in particular, feels no pain).

¹⁶ Just that point – an insight of common sense – startled Descartes (1637, pt. V, sec. 7 and 9): for any possible situation, S, and any utterance U that I make in S, I need not have made U in S. But it is much harder to refuse to see red, or feel pain, if there is a red object, and there is pain.

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” (Spencer’s word), 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 these intentional relations. While this is unintuitive, it is also unintuitive that the height of treetops and the need for high feeding do 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, in particular, 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 severally return. It is not the case, for example, 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 (unless we endow trees with personal properties, as in fairy tales). The standard *externalist* explanation for this fact would be that persons *are* not trees, and that this factual property of persons causes our concept of persons,

which we then, and *because* of what external object we got it from, 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.¹⁷ 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.

Crucially, what a concept refers to in an environment cannot *individuate* a concept in the present sense. 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. If the environment is not suitable, it may be that the internal structure that a new conceptual variant is remains unused.¹⁸ “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 need not, in the factual practice of evolutionary biology, depend on figuring out purposes or selected effects, and in fact it need not depend on *any* reference to the environment. The “function” of the form may in fact also 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 and is no presupposition for our understanding of it. Concepts in the present sense of given organismic structures cannot then be picked out by their actual use or content for the same reasons that organismic structures cannot

¹⁷ In which sense the so-called “innateness hypothesis” is a platitude. From the present point of view, innatism essentially means that we adopt a selective rather than instructive view of environmental causation. In the case of concepts this means that the environment does not causally bring particular concepts into existence, but selects them.

¹⁸ Or it may be used, while not referring to anything, in which case we have a fictional concept.

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 it.

My notion of a concept is thus a departure from standard philosophical usage, where a concept is usually either (i) the *intentional content* of a mental representation structure, or (ii) an internal physical structure in the brain that intrinsically has or determines such a content or a worldly *extension*. In the first case, a concept is an abstract object (an intension), in the second it intrinsically relates to some external object or set-theoretic construct that it is “true of”.¹⁹ Concepts in the present sense have neither property. There is a connection between concepts in my sense and worldly extensions, to be sure, 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 stand in no *intrinsic* connections to external objects that exist and have causal properties. Even once functioning, they do not necessarily track properties of the external world. 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; or that all there is is strings vibrating in eleven dimensions. The disparity between ordinary and scientific vocabularies and understanding is so vast that no sense even of contradiction arises when the two are used in parallel. 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 categories, nobody *knows* whether these categories in fact “track” or “pick out” anything in particular, and it does not seem to matter for ordinary understanding. In

¹⁹ Note that in the case of the child’s putative concept of a possible language (see further section 7 below) there literally is nothing in the environment for that concept to be “true of” or to “apply to”: for the structural patterns characterizing this concept – 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 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, semantic or pragmatic rationale; Baker (2003) defines them in purely syntactic terms. A recent account of language evolution that, on the basis of acknowledging these problems, provides a rather unexpected rationale for their existence – namely a phonetic one) is Carstairs-McCarthy (1999; cf. pp. 90, 161).

the case of *mind, fate, belief, gratitude, Chinese, the heavens, perfection, war, 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.²⁰ I will denote concepts in the two *standard* senses above by means of the familiar CAPITALS, and use the word “concepts” uncapitalized if I intend my own notion.

Note on the other hand that 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, which in turn depends on human intentions. This kind of “externalism” I consider a truism, and it does nothing to imply the kind of “Spencerian” causal flow from the environment to the organism that I centrally dispute.²¹

²⁰ What exactly are we talking *about* if we talk about fate, which external object *exactly* is Chinese, what *precisely* are the heavens? Even asking students uncontaminated by the philosophy of language what actually the referent of *Socrates* is yields the weirdest results. With respect to *London*, or *city*, one might have hoped that it is clear what the referent is: a geographical place, and a collection of geographical places, respectively, which might be a non-circular way of describing the referents in question. But then as Chomsky (2000) points out, *London* isn’t a place, for cities can be destroyed and rebuild at another spot, while remaining the same; moreover, cities are cultural and moral entities as well, which can be viewed together and without their history; and they are environments, as when we say they are polluted; or 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).

²¹ The Kripke and Putnam cases for an externalist view of content have thus equally no force against the present internalist view. Intentional relations depend on environmental causation (selection), but they do not explain the concepts that *figure in* these intentional relations. The Twin-Earth story tells us something about 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, but no more. Equally, Kripke’s argument that Goedel remains Goedel 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, but it does not explain why that is the concept it is (see further Hinzen forthcoming, a).

What brings it about that an organism with particular internal structures takes up intentional relations? This happens *unavoidably* by 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 structures. But their explanation is a matter of internal mechanisms only.

We may say, if we like, of our vending machine, that it has an internal structure (concepts) *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, when the machine happens to be placed in a children’s room). 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, as the *internals* of the machine or organism will determine, restrict, and explain membership in the extension. The structure of its internal mechanism determines the semantic values, not the other way around.

On the picture given, external uptake of function and reference is, 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 may be applied or with respect to which it can be used.

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 our internal concepts can mean. It is (partially) responsible for what they end up meaning.

In the next section, I will argue that the addition of natural teleology (teleonomy) to an externalist theory of meaning does not make it more plausible to bootstrap meaning from reference, or concepts from their external functions.

4. Causal-Teleological Views of Meaning

On one version of the causal view of reference, the meaning of a linguistic expression *is* the referent it is used to talk about. The meaning of a name, it used to be said, *is* its bearer. Sometimes the meaning is also said to derive by causal *chains* from it. At the end of the causal chain, though, some physical object *is* apparently a necessary and sufficient reason for a concept (say the one expressed by the name *Aristotle*) to come into existence (given, say, the *perception* of that object), and to mean what it does. Crucial to this view is the lack of any *mediation* of this referential relation by a semantically relevant internal structure in the human organism. Not what's in your head, but the actual, causal mind-world relations are supposed to determine the internal structure's functioning or meaning. If there is thought to be more to meaning than reference on this externalist picture, it is something *determining* the reference, hence understood in terms of reference. Standard observations about the *rigidity* of nominal reference then intervene in this move, implying that whatever these internal determinants are, reference fixation is actually independent of them. Hence the conclusion on that view is that meaning is, too, since meaning is tied externalistically to reference and truth conditions.

Intentional relations to the world thus reduce to (i) physical structures in the organism (intrinsically carrying a certain content), (ii) external things making up this content, (iii) causal

relations between (i) and (ii). The causal-*teleological* view, now, adds a fourth element to this list: (iv) biological functions between (i) and (ii), in the teleological, relational, or historical sense of “function”. Defenders of the teleological theory spell out the teleological element by appeal to natural selection operating on ancestors, learning in ontogeny, or both.²² The teleological theory is often regarded to be a necessary addition to the causal one, for it quickly transpires internal to the causal-referential view itself that causal relations alone will not ground meaning sufficiently. 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.²³ Every causal theory will have to bring out that even though cow-skins are necessarily co-instantiated with cows whenever they occur, meaning cows and meaning cow-skins, when using the word *cow*, is a different matter. If human meaning is to remain a matter of causal relations, one has to sieve out the causal relations that do not intuitively constitute the meaning of a mental representation somehow. For this, however, it seems that one has to *refer* to the meaning, which thus cannot be derived from some relation to external objects in a non-circular way (see further Loewer 1999). How teleologically understood functions solve this problem is unclear, in fact even in the case of the most likely candidate, automatic response mechanisms: in general *many* things will trigger the same response mechanism, as in the case of the frog’s concept of a fly, which leads it to snap for objects it considers flies. The internal representation cannot intuitively “mean” or “represent” *whatever* it is used for, as these will be too many clearly distinct things (in the latter case little black food pellets, for example, designed to fool the frogs; see Dennett 1995, ch. 14).

²² Dretske (1988) in particular employs natural selection for *perception* (which clearly is not “learned”), and learning for *thought*: thought appears intuitively more like a candidate for what can be acquired by ontogenetic learning.

²³ Again, a cautionary note as regards “causing a mental representation”: by and large no external thing or event appears ever to be a cause (or sufficient reason) for any human being to think or say anything in particular (as differences with mere reflexes and pathological cases like phobias, where such causes do exist, or with the functioning of alarm calls in monkeys, make manifest).

Given stimulus-insensitivity, most human concepts are *not* response-mechanisms of this kind. That how words function and refer is relevantly homologous to how vervet monkey call for predators or frogs snap seems rather doubtful.²⁴ That they can be individuated by appeal to their function in the relational sense, I have denied above, and that within evolutionary biology we can identify traits by assigning them a “proper function” or a “purpose” in the relational sense (Millikan 1984), has been generally denied by Amundson and Lauder (1994). And as noted, even in the case of the stimulus-sensitive frog, it is not clear how teleology should help the causal theory to establish some intrinsic relation between an internal structure (“mental representation”) in the frog, on the one hand, and flies-as-*opposed-to*-pellets, on the other. The same concept can be used for both flies and imitations of them, and the mechanism is the same in both instances. Of course, we could not say that the same CONCEPT is used to both purposes, since a CONCEPT is intrinsically related to an extension, things it is “true of”. But while some technical definition of function distinguishing the two CONCEPTS can presumably be given, it does not seem very intuitive. Why should we want to distinguish the “representation” of a fly from that of a pellet, and thus ignore the fact that the frog’s causal-role functioning does *not* distinguish these two cases?

While an appeal to functions for which a trait was selected may surely add to our understanding of organismic functioning in particular cases, there appears to be no reason to believe that it must so add in general, and hence that the causal-teleological can hold the claim of a *general* theory of what it is to have content. Papineau (2001) in particular asks what teleosemantics adds to our folk understanding of contentful states, and answers: “it tells us about

²⁴ Thus, as Hauser et al. remark, “[U]nlike the best animal examples of putatively referential signals, 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).

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 I see no reason why being so directed should help in general to understand the nature of human cognitive design, or reveal anything about its “underlying nature”.

Consider a state such as “knowing a language”, presumably a state characterizing our psychology. Suppose we assume, with Papineau, that psychological states of their essence are “selectional” ones, and claim that a properly historical and functional perspective on this state will add to our understanding of it. But the functionality that our knowledge of language factually has sets a problem for explanation, while not providing one. Having identified a function to be explained, we aim for a mechanism that explains it. On the “biolinguistic” approach to Universal Grammar (UG) in the generative tradition (Jenkins 2000, Baker 2001, Anderson and Lightfoot 2002), part of the explanation is that the human brain represents a computational procedure implementing a recursive mechanism generating a potentially infinite output of any human language. This procedure is thought to operate in a step-by-step fashion, in a way that each step can be traced to the interaction of a number of universal principles of language with a number of parameters whose values are contextually fixed. There is an empirical question, then, which of infinitely many possible procedures generating the same infinite output is the *correct* one (there surely *is* one such procedure, even if it will forever remain unknown). The one that is correct, is correct even if Aliens actually brought all animals to earth only recently. It would simply reflect the way in which we are structured, which no amount of history can change. Our ancestry and evolution may *help* in finding which characterization of the generative procedure is correct. But there is no reason to believe that it *necessarily* helps, and it has *factually* not done so, it seems. An evolutionary and adaptationist rationale (as in Pinker 1994) was given to the structures posited in the generative tradition *after* they were posited, and not because the structures

particularly *suggested* an externalist or functional rationale, which indeed they do not (see Uriagereka 1998 for discussion).

Rather than as a search for ultimate causes or rationales for existence, the theory of UG is an empirical theory of a *structural type*, according to which all human languages happen to pattern (Hinzen forthcoming).²⁵ As such it has nothing to do with “psychological states”, whatever these may be, or their evolutionary purposes. The above notion of type is that of the pre-Darwinian typological tradition of the “rational morphologists”, as described by Amundson (1998). By explaining homologies among given species by appeal to a common ancestor, Darwin aimed to effectively subordinate the rational morphologists’ explanatory principle – the “unity of type” – to the functionalist’s explanatory principle – “conditions of existence”. But the finding and the study of structural types effectively preceded their incorporation into evolutionary theory, hence was as such manifestly independent of any search for an ultimate rationale of the kind that the evolutionary mechanism of natural selection was supposed to provide. Moreover, declaring the common type an ancestor does not explain why it exists in the first place. There appears to be little reason to believe that the crucial theoretical concept of the formalists, the concept of homology, only makes sense in the light of the crucial theoretical concept of the functionalist, the concept of adaptation.²⁶

Assuming the explanatory independence of both concepts, we may view the theory of UG as exemplifying the autonomy of the study of form (structure) over the study of external function. That any structural type *must* ultimately show its “underlying nature” as being the result of the shaping of natural selection is a form of “presumptive adaptationism” that Georges Cuvier put

²⁵ While the generative approach to UG remains controversial, I agree with Nowak’s et al. (2002) conclusion that the “debate around an innate UG should not be whether there is one, but what form it takes” (p. 615). The theory of UG reflects an as such uncontentious search for a “unity of type” as underlying human languages. Such a theory can be empty or uninteresting, but not false or illegitimate in principle. It clearly cannot be contested on the grounds that no functionalist rationale has been offered for the universal structures posited (cf. Dennett 1995, ch. 13)..

²⁶ As Amundson and Lauder (1994, 243) put it, the claim that biological categories are only definable in functional terms is “almost completely groundless”.

against his formalist adversaries in the 19th century, and that today's functionalists, such as Dennett (1995, ch. 13) revive against their formalist counterparts. As regards the "underlying nature" of UG, if it has any, it seems we should first look for "proximal" mechanisms, and have natural selection enter the scene of evolutionary explanation only as a last resort, or as the "onerous concept" that Amundson and Lauder (1994, 254) – with G. C. Williams – claim it to be (see also Hauser et al. 2002).

5. Spencerian Semantics Today

Within teleosemantics, Dretske's (1988) emphasis on the need for "learning", and his way of conceiving this notion, makes him a good example for the conflation of function with genesis. Dretske contends that what our words or mental representations mean is a function of how they came to be, namely through "learning". His theoretical goal is "to install the informational processes underlying concept learning as the source of meaning and content" (Dretske 1994, 260), hence in my terms to conflate function and genesis, 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 a *conditioned*, not even a Lamarckian "creative" response to recurring situations, and even clearer not a Darwinian selection of an independently originating internal structure.

That the acquisition of concepts 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 explanatory significance of “associative learning” see 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 that it is placed in 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.²⁷ My point is 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.²⁸ 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 behaviorism). Also Fodor (1990) is fully

²⁷ The kind of learning that connectionist neural networks model reflects most closely the kind of learning that Dretske intends his theory of content to be based on. See Dretske (1988, 98).

²⁸ Though considerations of a general theoretical nature may rule out instructive learning, too. See again Dawkins (1983).

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, as (Fodor 1990, 53) views it, 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 its learning part, he points out that 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 from a learning theory. It is to this view that I now turn.

6. Doing without learning (Fodor)

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 can think. Faced with inputs that humans perceive as organized phrase-structurally, tamarin monkeys do not represent any phrase structures. For a concern with Spencerism, what matters is not what view we take on the metaphysics of propositional attitudes, but what view we take on *semantics*. What the Darwinian as we understand him here 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 told 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. Fodor's externalism does not square with a form of biological 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), which are supposed 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). If I don't have the appropriate concept, no amount of cats or their features will trigger it, on Fodor's own

view.²⁹ Hence, however concepts are individuated, they must be individuated independently of their dispositional relations to particular external objects.³⁰

The same problem of making a biological nativism coherent with an externalist individuation of content transpires 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 to give up on *cognitivism*, the view that concept acquisition involves a process of hypothesis-formation and testing, and to stipulate instead that acquiring a concept is to get *nomologically locked* to the *property expressed* by the concept. Internals of the mind – formation of beliefs, testing of hypotheses, concepts – drop out of the process, and since these internals caused the problem in the first place, the learning paradox disappears. It is clear that this moves us into an externalist direction of explanation. But the internalist response to the move 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* property expressed by it, rather than another property (in which case, the concept would not be explained by our resonating to the property)?³¹ (ii) why the dualism of “concept” and “property” expressed by it (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.

²⁹ It seems similarly unclear how, if part of what's innate about a triggered 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.

³⁰ I am assuming they would have to be individuated independently of their conceptual roles (dispositional internalism), too, as I share Fodor's stance against lexical decomposition.

³¹ As Fodor agrees, for the resonating to happen, you need to have the right kind of mind, a “mind like ours” – made up by whatever conceptual structures it is made up of – which thereby is a presupposition for the story, not what gets explicated in it. Put differently, if “what has to be innately given to get us locked to *doorknobhood* is whatever mechanisms are required for doorknobs to come to strike us as such” (Fodor 1998, 142), I'd say that the concept doorknob has to be. But that does not mean that DOORKNOB – a relational notion – is innate.

While finding his externalist move attractive, however, Fodor himself raises the problem that locking on to a property does ultimately not seem to be a process that is unconditioned internally. For it depends on exposures to the *right kind* of things given in one's experience: these will typically be things *falling under* the concept to be learned.³² The concept of a doorknob is learned from exposures to (typical instances of) *doorknobs*, not giraffes, say. For Fodor this means that hypothesis-testing as an acquisition method is back in the picture: the learner has to use experiences with *doorknobs* to test and confirm hypotheses about what property the word "doorknob" denotes. For 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 rather than giraffes. But it is. Therefore the process is not brutally causal. 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 hypothesis the learner cannot form, lacking the concept that figures in the hypothesis. 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 is a *reductio* of the hypothesis-testing model.³³

³² If so, the internalist would wonder, how could reference to such things *explain* what it is to have this concept?

³³ Though not of innateness, if this is merely the commitment that the environment does not cause any concepts into existence. This internalist take on the problem makes good sense of an independent observation in the case of the human language faculty: here there is, as I noted, nothing the knowledge of language acquired is "true of". While the child does of course acquire the language spoken in its environment and not another language, this still does not mean that there was this language there in the environment to start with, and then it wandered into the mind, whose internal structures are then "true of" it. Linguistic structures (described in linguistic rather than acoustic terms) *are* never there in the environment for a creature that lacks linguistic concepts. Again, the environment merely selects from the range of possible languages the child's mind affords the child with. The doorknob/DOORKNOB problem,

Still, why *does* it appear that DOORKNOB is 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 (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.³⁴ 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.³⁵

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 actualize. While reference does not determine concepts, the environment provides external referents for concepts within a range of ways permitted by these concepts. The modalities used in this theoretical commitment implies that certain intentional relations would be such that we could *not* take them

at least in this instance, is not really formulable. Contrary to what Spencer thought, the environment does *not* bring about an organic change exactly in its own image (cf. Godfrey-Smith 1998, 86).

³⁴ Fodor discusses a Darwinian response of this type unfavorably. Quite plausibly, he remarks: “just as you can’t satisfy the conditions for having the concept *F* just in virtue of having interacted with *Fs*, so too you can’t satisfy the conditions for having the concept *F* just in virtue of your grandmother’s having interacted with *Fs*. ” (Fodor 1998, 133) This, in my terms, is the claim that functional relations of a causal or whatever type do not give you genesis. Fodor goes on to demand that something must have gone on in your head in consequence of the interactions – namely, inductive learning. But that consequence is not explained by the interaction either (if it takes place), as it depends, on my view, on the concept’s being in place. Mother nature, if it wants to select for the doorknob/DOORKNOB relation between concepts and their experiential causes, does NOT, if I am right, have to select a “*mechanism that produces* that relation between one’s concepts and their causes” (*ibid.*, 129), unless that mechanism is understood as the mechanism of natural selection itself, as something that sieves out unsuitable correlations.

³⁵ Note that nothing in the above account turns on concepts “being in the genes”. Everything rather turns on the dissociation of genesis and function, and the fact that the process in which a given structure is selected is no process in which an external structure is transferred into the organism. The account in the text is thus consistent with the idea, for example, that concepts are memes whose replication and selective retention is independent of the genes.

up, and that there would be CONCEPTS we could not learn, because we did not have, and could not, by the power of our will, develop the concepts needed for these CONCEPTS.

To test this prediction, we may return to the theory of UG as capturing the child's concept of a *possible language*, a concept as such not depending on any actual language spoken, though any actual instance of a possible language depends by assumption on it. Cultural history selects among the possible instances of this single structural type. As such, the latter allows for an infinite set of language variants, but it commits us to the impossibility of certain languages not obeying the constraints of that type. These latter languages would be *logically* possible (and constructable 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.³⁶

While, of course, in some intuitive sense, children "learn" languages as part of their development, it is an empirical question whether human languages are learnable in the technical sense required for externalist and empiricist learning theories. There is no constraint in human biology as such 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 enabled by UG – 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, or within which data fall into place and become subject to evaluation through the child's mind. We

³⁶ Examples of "impossible words" in the sense of Hale and Keyser (2002) may be another case in point. These are words derived in violation of a syntactic constraint, and hence not natively learnable. Still another case may be concepts impossible on other (though still internal) 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 simply not *one* unified thing for a creature with a conceptual system like ours.

know this process to be surprisingly independent of a large amount of feedback, corrections, and environmental control parameters, hence of hypothesis-testing in the Fodorian sense. Crucially, moreover, one can describe this process without bringing any notion of “propositional attitude” into play, of the sort that figures in a Fodorian “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. Whatever “content” or “intentional directedness” precisely means, the theory of UG as a theory of the acquisition process proceeds without giving these relational notions an explanatory role to play (see further Chomsky 2003a).

While this “biolinguistic” (Jenkins 2000) stance is widely accepted for the computational system of language, it may be appropriate for the acquisition of the lexicon as well. The very minor role that feedback and correction play in the child’s acquisition of a vocabulary, its effortlessness, speed, and essential uniformity despite great variance in scanty evidence and cultural diversity, and the little specificity of the sensory input needed, all conspire to suggest that this is the right conclusion also in this case.³⁷ The gap between the system of knowledge acquired and the evidence on which this happens seems equally vast.³⁸

The theory of UG viewed as above, finally, teaches a philosophical moral regarding the “semanticality” of concepts in my sense: there are a range of concepts, which, while not being defined or generated relationally, are also not semantically *inert*. They determine (delineate) a

³⁷ 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 circumstances, but understood in delicate and extraordinary complexity that goes vastly beyond what is recorded in the most comprehensive dictionary”.

³⁸ The most plausible idea for *how* concepts are learned – namely by means of learning their parts, and composing the whole – seems to have little empirical confirmation, developmental psychologists having found “no good evidence that a word’s meaning is composed, component by component, in the course of its acquisition” (Carey 1982, 47). Fodor’s more general arguments against theories of lexical decomposition of various kinds remain a forceful challenge (Fodor 1998, 2001). In any case, as noted, any decompositional view will leave *some* concepts as selectively learned, since decomposition must bottom out somewhere, with some primitives, which can’t be learned on the basis of their constituents.

possible content: a (humanly) possible language variant. These concepts will not determine the contents of the child's actual experiences – since it is a historical matter of what environments it will happen to be placed in whether it will eventually speak French rather than Bantu or German – 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”. We thus see that concepts in the sense of this paper, though they are part of the organism's own structure and thus independent of an environment it 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 some semanticality all the same.³⁹

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 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, at least the empirical study of language in the biolinguistic program in linguistics does not suggest giving external function an explanatory role to play in the genesis of that part of our cognitive structure and design.

I have suggested a different picture, which endorses a basic “nativism”, although it reduces the latter to a virtual truism, assuming psychology, in the spirit of a methodological

³⁹ 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.

monism, to be continuous with biology: the organism's response to the environment depends on its internal resources, and no internal cognitive structures 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 (Amundson 1998, Gould 2002) and as underlying the biolinguistic program today (Jenkins 2000, Hinzen, forthcoming) – gives rise to a basic conflict with externalism, I have argued, if the latter is to be more than the truism that what our internal concepts end up being used to refer to – to the extent that it depends on more than the concepts themselves – 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, Ron (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. (2001), *The Atoms of Language*, New York: Basic Books.
- 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.
- Carstairs-McCarthy, A. (1999), *The Origins of Complex Language*, Oxford University Press.
- 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.
- Descartes, R. (1637), *Discours de la Méthode*, Paris: Vrin 1984.
- 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*, Cambridge, MA: 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*, Cambridge, MA: MIT Press.
- Gleitman, L. and B. Landau (eds., 1994), *The Acquisition of the Lexicon*, Cambridge, MA: 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*, Cambridge, MA: 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.
- Hinzen, W. (forthcoming, a), “Explaining Rigidity”, Ms., Universiteit van Amsterdam.
- 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*, Cambridge, MA: 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.
- Loewer, B. (1999), “A Guide to Naturalizing Semantics”, in Hale, Wright (eds.), *A Companion to the Philosophy of Language*, Blackwell.
- Mehler, J. and E. Dupoux (1994), *What Infants Know*, Oxford: Blackwell.
- Millikan, R. (1984), *Language, Thought and Other Biological Categories*, MIT Press.
- Molenaar, P. (1986), “On the Impossibility of Acquiring More Powerful Structures: A Neglected Alternative”, *Human Development* 29, 245-251.
- 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.
- Nowak, M.A., N. L. Komarova, and P. Niyogi (2002), “Computational and evolutionary aspects of language”, *Nature* vol. 417, 6 June 2002, 611-617.

- 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.
- Sober, E. (1995), "Natural Selection and Distributive Explanation: A Reply to Neander", *British Journal for the Philosophy of Science* 46, 384-397.
- 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.