

Volume 3 Number 2
2007

ISSN 1807-9792

abstracta

Linguagem, Mente & Ação

Foundational Evidentialism and the Problem of Scatter
Ted Poston

Hume, Association and Causal Belief
J. P. Monteiro

Dewey's Metaphysics of Mind
Wilson Mendonça

**Uma Reflexão a Respeito da Evolução Humana
e a Natureza da Linguagem**
Alfredo Pereira Jr.

Discussion

**On Schiffer's Arguments against the Fregean Model
of 'that'-clauses:
A Comment on Vignolo**
Paolo Bonardi

ABSTRACTA
Linguagem, Mente e Ação

ISSN 1807-9792

Volume 3 Number 2
2007

Editors

André Abath
Leonardo de Mello Ribeiro
Carlos de Sousa

Executive Editors

Jules Holroyd
Gottfried Vosgerau

Associate Editors

Abílio Azambuja
Miquel Capó
José Edgar González
Vanessa Morlock
Olivier Putois
Erik Rietveld
Giuliano Torrenco

TABLE OF CONTENTS

Foundational Evidentialism and the Problem of Scatter Ted Poston (University of South Alabama)	89
Hume, Association, and Causal Belief J. P. Monteiro (University of São Paulo)	107
Dewey's Metaphysics of Mind Wilson Mendonça (Federal University of Rio de Janeiro)	123
Uma Reflexão a Respeito da Evolução Humana e a Natureza da Linguagem Alfredo Pereira Jr. (São Paulo State University)	138
<i>Discussion</i>	
On Schiffer's Arguments against the Fregean Model of 'that'-clauses: A Comment on Vignolo Paolo Bonardi (University of Geneva)	162
Errata	176

FOUNDATIONAL EVIDENTIALISM AND THE PROBLEM OF SCATTER

Ted Poston

Abstract

This paper addresses the scatter problem for foundational evidentialism. Reflection on the scatter problem uncovers significant epistemological lessons. The scatter problem is evaluated in connection with Ernest Sosa's use of the problem as an argument against foundational evidentialism. Sosa's strategy is to consider a strong intuition in favor of internalism—the new evil demon problem, and then illustrate how a foundational evidentialist account of the new evil demon problem succumbs to the scatter problem. The goal in this paper is to evaluate the force of the scatter problem. The main argument of the paper is that the scatter problem has mixed success. On the one hand, scatter undermines *objectual evidentialism*, an evidentialist theory that formulates principles of basic perceptual justification in terms of the objects (or properties) of perceptual states. On the other hand, the problem of scatter does not undermine *content evidentialism*, an evidentialist view that formulates its epistemic principles in terms of the assertive content of perceptual states. The significance of the scatter problem, especially in concert with the new evil demon problem, is that it provides an argument for content evidentialism.

The problem of scatter is a significant and neglected epistemological problem. Ernest Sosa has done the most in developing the scatter problem. In this paper my central focus shall be Sosa's use of the scatter problem as an argument against foundationalist evidentialism. The scatter problem, however, is larger than just an epistemological problem. In its most general form it is a problem for certain types of non-consequentialist normative theories. The introductory section places the scatter problem within the larger context of types of normative theories and then proceeds to develop the epistemological problem of scatter. The second section develops Sosa's use of this problem. The third section argues that scatter afflicts objectual evidentialism but not content evidentialism. The final section addresses a worry about the truth connection that arises from reflection on epistemological scatter.

I. Introduction

Normative theories fall into two kinds: consequentialist and non-consequentialist theories. Consequentialist theories explain the normative in terms of value and production. A consequentialist theory of right action, for instance, explains right action in terms of the good and ability to produce the good. Mill's utilitarianism is a prime

example of a consequentialist theory. On Mill's view the right act produces the most value. In epistemology Alvin Goldman's early theory of process reliabilism is a good example of a consequentialist epistemological theory.¹ On Goldman's view a belief is justified only if it is caused by a truth-productive process, i.e., a process that tends to produce more true beliefs than false beliefs. In general a consequentialist epistemological view explains justification in terms of truth and ability to produce true belief.²

Non-consequentialist theories explain the normative in some other way. These theories are primarily characterized by a negative thesis: the normative is not properly explicated merely in terms of value and production. W.D. Ross's ethical intuitionism is a non-consequentialist theory of the right.³ The right act, it is claimed, is not properly analyzed in terms of value and production. Rather, according to Ross, it is a matter of an action *fitting* the circumstance. In epistemology evidentialism is a non-consequentialist theory of justification. The evidentialist holds that facts about justification can't be explicated merely in terms of truth and ability to produce true belief. Rather a belief is justified only if it fits one's evidence. For instance, Feldman and Conee propose:

EJ Doxastic attitude *D* towards proposition *p* is epistemically justified for *S* at *t* if and only if having *D* towards *p* fits the evidence *S* has at *t*.⁴

Facts about whether a belief fits the evidence are not simply a matter of whether the belief is produced by a truth-productive process.

Non-consequentialist theories face the scatter problem when their normative principles do not display any deep explanatory unity. Chisholm's epistemological theory is an example of a non-consequentialist theory that suffers the scatter problem because he offers a bevy of basic normative principles—ten to be exact.⁵ Ross's ethical theory faces the scatter problem as well. Ross offers a list of "prima facie duties". These include duties of fidelity, reparation, gratitude, non-injury, etc.⁶ These prima facie

¹ Goldman (1979).

² I assume that gaining truth and avoiding error are the only items of final epistemic value.

³ Ross (1930).

⁴ See Feldman and Conee (2004:83).

⁵ See Chisholm (1989:62-72).

⁶ See Ross (1930:18-36).

ethical principles work together with a general principle of fit—the right act is one that fits the occasion—to explain what is the right act. Yet there is no explanation for the diverse facts about fit. The evidentialist faces a similar problem. Though the evidentialist offers a single normative principle—Feldman and Conee’s principle EJ—there is little explanation regarding why certain beliefs fit the evidence while other beliefs do not fit the evidence.

Ernest Sosa has recently used the scatter problem as an argument against foundationalism evidentialism.⁷ Sosa’s earlier writings appeal to the scatter problem as a reason to reject contemporary forms of internalistic foundationalism. In “Epistemology Today: a perspective in retrospect”⁸ Sosa writes,

The strategy of some contemporary foundationalism (sic) now seems clear: multiply epistemic principles in order to provide for the sources of justification required for our rich knowledge in its various dimensions. But the danger in a free use of such a strategy also seems clear, for a wide scattering of diverse principles does not make for a satisfactory epistemology.⁹

In “Theories of justification: old doctrines newly defended”¹⁰ Sosa presses the scatter problem against evidentialism. Sosa describes evidentialism as the “view that there is a relation among propositions – ‘fitting’ or ‘being supported’ or the like – such that a proposition is justified for one iff it ‘fits’ or ‘is supported by’ the relevant evidence that one ‘has’.”¹¹ Against such a view Sosa complains,

John Stuart Mill would often object to ‘intuitionist’ moral theories by deploring their lists of retail intuitions with no apparent unity, thus charging them with a problem of ‘scatter.’ Suppose evidentialism stops with a scattered set of principles. For example, suppose it includes principles admitting green and blue but ruling out bleen and grue, and supposed it includes principles that allow direct introspection of triangularity but not direct introspection of octagonality or of 23-sidedness. Obviously there would then be a problem of scatter – and that does not yet consider the appropriate experience-introspection ties suitable for extraterrestrials. The problem is that if we hold the abstract level of *fitting* to be *fundamental* we may then lack the resources for greater theoretical unity.¹²

⁷ Sosa in BonJour & Sosa (2003:164-5).

⁸ See Sosa (1991:65-85).

⁹ Ibid:73.

¹⁰ Sosa (1991:108-130).

¹¹ Ibid:127.

¹² Ibid:128.

These quotes bear out that the scatter problem is the problem of theoretical unity. If epistemic principles for perceptual justification are conceived as underwriting transitions from perceptual states to doxastic states the scatter problem arises when a theory licenses different transitions with distinct, fundamental principles. Sosa's mention of the grue problem is illuminating. If an epistemic principle licenses the move from a perceptual state to the belief that something is green then, unless there is something about the perceptual states that indicates green rather than grue, the principle should license the move to the belief that something is grue. The strategy of handling multiple transitions by distinct epistemic principles gives the suspicion that one is merely describing intuition at the level of theory and accordingly not offering a significant explanation about the data.

The problem of scatter arises as well for evidentialist views that try to account knowledge or justification from multiple sources. Consider for instance the following putative sources of knowledge.¹³

1. Perception, by means of the external sense organs;
2. Memory;
3. "Inner consciousness," or the apprehension of our own states of mind – for example, our awareness of our own sensations, of our beliefs and desires, of how we feel, of what we are undertaking to do;
4. Reason, as the source of our a priori knowledge of necessity – our knowledge, for example, of some of the truths of logic and mathematics;
5. Moral consciousness;
6. Intuitive understanding;
7. Religious consciousness;
8. Social understanding, or the apprehension of social facts – for instance, that Jones is uncomfortable around Bill or that Smith has overdrawn his bank account.

It is exceedingly difficult to unify these distinct sources of knowledge within an evidentialist framework. It's quite natural then to see the role for the concept of fit within an evidentialist framework. However, I agree with Sosa that if the facts about fit

¹³ The list is taken from Chisholm (1982:114-5). I have added social understanding to Chisholm's list since this does seem an area of knowledge and it does not fit well with the other sources.

are fundamental then this makes for a less than desirable epistemology. In this paper I shall focus on Sosa's use of the scatter problem in connection with an evidentialist account of perceptual justification. I shall argue that scatter can be avoided by appealing to assertive mental content. Though I shall not stress it I think this solution can handle the problem of scatter that arises for multiple sources of knowledge or justification as well.

As mentioned Sosa's most recent use of the scatter problem focuses on internalist foundationalist accounts of perceptual justification, particularly in connection with the new evil demon problem. He argues that attempts to uphold the internalist intuition contained in the new evil demon problem succumb to the problem of scatter. But, as I shall argue, Sosa's objection has mixed success. On the one hand, scatter undermines *objectual evidentialism*, an evidentialist theory that formulates principles of basic perceptual justification in terms of the objects (or properties) of perceptual states. On the other hand, the problem of scatter does not undermine *content evidentialism*, an evidentialist view that formulates its epistemic principles in terms of the assertive content of perceptual states. The significance of the scatter problem, especially in concert with the new evil demon problem, is that it provides an argument for content evidentialism.

II. Sosa, Scatter, and Demon-world perceptual justification

Sosa's discussion on the scatter problem is situated within a discussion of the new evil demon problem.¹⁴ The new evil demon problem concerns subjects who have the same beliefs and experiences that we do but, because of the presence of a controlling demon, their beliefs are false and their experiences falsidical. Nevertheless there is a strong intuition that these subjects are just as much justified in their beliefs as we are in ours. This intuition supports the key internalist claim that justification is not merely a matter of truth and production. As evidentialists would put it, in the demon world those subjects have good evidence for their beliefs because their beliefs fit their experiences. This intuition is straightforwardly a problem for externalist theories of justification because such theories require beliefs to exhibit alethic production for the kind of justification necessary for knowledge.

¹⁴ Sosa in BonJour & Sosa (2003:162-5).

In Sosa's discussion of this intuition he lays out three internalist attempts to explain the rational appropriateness¹⁵ of the demon-worlders' perceptual beliefs and claims that the most plausible version faces the scatter objection. The three theories Sosa considers are Foley-rationality,¹⁶ a deontological account of justification, and an evidentialist account that Sosa calls "the good reasons" account. Although Sosa does not explicitly acknowledge this the first two internalist theories do not face the scatter objection because they exhibit theoretical unification. Foley's account of rationality offers the single principle that a belief is justified just in case were one to reflect on one's deepest standards one would be satisfied with the belief. This avoids the scatter problem because the same principle is invoked to explain every justified belief. Similarly a deontological account of justification avoids the scatter problem, for it invokes the same principle to explain each justified belief. Nevertheless, Sosa finds these theories suspect. Although Sosa does not put it this way his main reason for rejecting these theories lies in the fact that the theories deny a logical (or modal) connection between justification and truth.¹⁷

Having rejected these accounts he turns to the good reasons view, i.e., evidentialism, to account for the intuition that the beliefs of the subjects in the demon world are rationally appropriate. The good reasons view maintains that the perceptual beliefs of the denizens of a demon world are rationally appropriate because they have good reasons for their beliefs. The good reasons view as such is uninformative because "good reasons" stands as much in need of explanation as "rational appropriateness." It is no improvement to say that their beliefs are rationally appropriate because their beliefs fit their experiences. A more informed attempt will explain when a perceptual experience provides a good reason for (or when it fits) a perceptual belief. Sosa offers the following suggestion on behalf of the good reasons account:

One might suppose that there is a state with an intrinsic mental character whose intrinsic mental character makes it properly characterizable as a state of experiencing thus (as a state of experiencing a white, round item, or the like), and this with logical independence of any modal relation that such a state, with

¹⁵ Sosa describes the problem in terms of "subjective appropriateness" (see pp. 162-3). I prefer "rational appropriateness" because what is at issue is rational belief.

¹⁶ See Richard Foley (1993).

¹⁷ For the reasons see Sosa in BonJour & Sosa (2003:159 & 164).

such an intrinsic character, may bear to the presence of or absence of a white and round item.¹⁸

Sosa objects that the good reasons account faces the problem to “explain how such states could possibly give a reason to believe that there is something white and round before one.”¹⁹ He anticipates that the defender of this view will claim that the state just does provide such a reason. To this Sosa rejoins with the problem of scatter. He writes,

And now one will face the following prospect: the need for a boundless set of principles each with fundamental status, connecting various intrinsically characterized mental states with paired external facts of specific sorts. Even though there is no modal relation between a given intrinsic mental state and its paired external fact type, that state might nevertheless serve as a reason to believe in its paired fact. And this will be so even when, as in the demon world, that sort of intrinsic mental state may in fact – when combined with the modally stable presence of the controlling demon – be a modal *counterindication* of its paired sort of external fact. Each such rational relation uniting the supposed mental/external pairs would be postulated as holding primitively, despite there being no modal relation among the mates, and even when the mental state modally counterindicates the paired external fact (assuming the presence of the demon to be modally quite stable relative to the world).²⁰

We can isolate two issues here. First, Sosa claims that the good reasons account implies a “boundless set of principles each with fundamental status, connecting various intrinsically characterized mental states with paired external facts of specific sorts.” Second, Sosa avers that these rational relations between a mental state with such and such intrinsic character and a belief about an external fact hold independently of any modal relation between the state and the external fact. This is claimed to be especially implausible because in a demon-world the mental state with such and such intrinsic character is a counterindication of its paired external fact.

The second objection is significantly different than the first. The first objection is the scatter problem. The quotes from Sosa’s earlier use of the scatter problem bear out that the essence of the scatter problem is the lack of theoretical unity regarding a theory’s normative principles. The second objection is the problem of the truth connection. These problems are clearly separable. This can be seen from the fact that the problem of the truth connection arises for the good reasons account, Foley-

¹⁸ Ibid:164.

¹⁹ Ibid.

²⁰ Ibid:164-5.

rationality, and a deontological account of justification, whereas the scatter objection arises only for the good reasons account. One common objection to Foley-rationality and a deontological account is that the putative rational relations between the justifiers and the justified hold independently of modal relations between the justifiers and the truth of the justified.

Though the scatter problem and the problem of the truth connection are separable, the problem of scatter raises concern over how a non-consequentialist epistemological theory relates its principles to truth. The scatter problem arises because of a concern about theoretical unity—there seems to be no unifying feature that ties together distinct, fundamental epistemic principles. A consequentialist theory avoids scatter by connecting the epistemic principles to the central epistemic item of value, truth. Thus, the consequentialist maneuver to avoid scatter simultaneously provides an answer to the truth connection. It is understandable, therefore, why Sosa ties these two objections together. Accordingly, an adequate non-consequentialist treatment of the scatter problem should have something to say about the problem of the truth connection. The final section sketches in outline a non-consequentialist account of the truth-connection.

III. Dispersing the Scatter Problem

The scatter problem can be dispersed by recognizing that there are two ways of unpacking the notion of the intrinsic mental character of a perceptual state to yield principles of perceptual justification. One way is to unpack the notion in terms of the objects of the perceptual state. To mesh with the intuition in the new evil demon problem these objects need be logically independent of empirical objects. A sense-datum or a property based view fits the requirement of logical independence. For simplicity, I shall work with a sense-datum view. On this account principles of perceptual justification mention the objects of the perceptual state in the antecedent. For instance, a perceptual state having a round, red sense-datum (or collection of sense-data) provides *prima facie* justification for the belief that there's a red, round object.

A different way of unpacking a state's intrinsic mental character is in terms of mental content, where the content has the kind of independence Sosa refers to. The

content of the perceptual state may be nonconceptual in nature²¹ or it may be conceptual.²² On either view a perceptual state that has the assertive content that p (or as if p)²³ yields prima facie justification for a belief whose content matches the content of the perceptual state.

An evidentialist account that unpacks the notion of intrinsic mental character the first way may be called *objectual evidentialism* and an evidentialist view that takes the second option *content evidentialism*. The scatter problem arises for objectual evidentialism but not content evidentialism. That content evidentialism accounts for the intuition in the new evil demon problem and avoids the scatter problem constitutes a significant reason for it. Moreover, as explained in the final section, content evidentialism can also account for the problem of the truth connection. The following explains how scatter afflicts objectual evidentialism and how content evidentialism avoids the problem.

A. Objectual Evidentialism

It should be fairly obvious by now that objectual evidentialism is afflicted by the scatter problem. Consider its epistemic principles. These principles have the form: when certain sense-data are present one has a good defeasible reason to believe that such and such physical object is present. Different sense-data justify different empirical beliefs. However, not every object (or property) that figures in a perceptual state provides a good reason to believe the appropriate proposition. Plausibly, a perceptual state that includes a large green item provides a good reason to believe that a green item is present but not a good reason to believe that a grue item is present. Moreover, a perceptual state can exhibit more detail than we are plausibly justified in believing is present. Consider the perceptual state that results from looking at an office building. On a sense-datum view the state may include objects that correspond to a specific number of windows, say 23. But one isn't justified in believing that there are 23 of the relevant objects present. One simply can't tell on the basis of a look that an image has some specific, complex property. So the objectual evidentialist is committed to the claim that perceptual states

²¹ See Heck (2000) & (forthcoming); also Peacocke (2001).

²² See McDowell (1994).

²³ The "as if" qualification is needed to bring in nonconceptualist views. For one use of the "as if" formulation see Pollock & Ovid (2005). They defend direct realism, formulated as "(DR) For appropriate P's, if S believes P on the basis of being appeared to as if P, S is defeasibly justified in doing so" (p. 325)

provide good reasons to believe that *certain* objects and properties are present but not a good reason to believe of *every* object and property that it is present. However, on an objectual evidentialist account there is nothing about the perceptual state that explains this difference. So the objectual evidentialist is saddled with the scatter objection.

B. Content Evidentialism

Content evidentialism avoids the scatter problem. This view unpacks the notion of the intrinsic mental character of a perceptual state in terms of mental content. It holds, for instance, that when one is in a mental state with assertive content *p* (or as if *p*) then one has a good reason to believe that *p* is true. Assertive contents are contents that are presented as true.²⁴ In the Müller-Lyer illusion, for instance, one hosts the assertive content that the lines are of unequal length. This content provides a good defeasible reason to believe that the lines are incongruent. It should be noticed that a consequence of content evidentialism is that there exist a potentially boundless number of epistemic principles. But importantly these principles are consequences of the more general and fundamental epistemic principle that a state with the assertive content that *p* provides prima facie justification for *p*.

Content evidentialism can avoid the grue problem and the specificity objection that plagued objectual evidentialism. A plausible development of content evidentialism is that a state with assertive content as if (e.g.) *there's a green object* provides a good reason to believe that *a green object is present* but not a good reason to believe that *a grue object is present*. The problem with objectual evidentialism is that it lacked a principled reason for this asymmetry. The content evidentialist, however, has the reasons to explain this asymmetry. The content of the state is that *a green object is there*, not a grue object. That is, the state has a projectible content that is incompatible with grue-like properties. Most states do not represent grue-like properties. What is needed for the content evidentialist here is to ride piggy-back on solutions to the disjunction problem in the philosophy of mind.²⁵ Presumably content evidentialism can align itself with a view of mental content that avoids the disjunction problem. For

²⁴ Heck (2000) makes similar use, as well, of the notion of assertive contents.

²⁵ See Fodor (1990).

instance, the view may be joined with a teleological account of content.²⁶ Teleological accounts of content avoid the disjunction problem by appealing to the function of the state to represent (e.g.,) green instead of grue.²⁷

The specificity objection is that the objects of perceptual states can be much more specific than one has justification for the appropriate belief. For instance, a perceptual state may contain a 13-sided figure but that state doesn't yield prima facie justification for the belief that the figure has 13 sides. This is a problem for objectual evidentialism because it has no principled reasons for holding a difference in justification between which objects and properties figure in the perceptual state. The content evidentialist can avoid this problem by appealing again to the content of the perceptual state. Perceptual states may not carry contents that match every object or property of the perceptual state. When looking at an office building one may be in a perceptual state that has the content that there are many windows, but not the content that there are 23 windows even though the perceptual state carries information that there are 23 windows. There is much detail that needs to be filled in to this general picture but it is plausible that content evidentialism is much better positioned to handle the specificity objection than objectual evidentialism.

As I mentioned earlier the scatter problem arises in regard to difference sources of knowledge. My proposal is that content evidentialism can be extended to cover difference sources of knowledge in so far as these sources generate assertive contents. This allows us to view content evidentialism as a principled and theoretically unified account of epistemic justification.

C. A Worry

Let us consider another worry about content evidentialism. Content evidentialism solves the scatter problem by appeal to the assertive content of perceptual states. Yet every mental state with assertive mental content requires justification. So these perceptual states require justification. Hence at some point the content evidentialist will

²⁶ See for example Millikan (2000). For a survey of teleological theories of mental content see Neander (2004).

²⁷ Admittedly this is complicated by the fact that content evidentialism needs to align itself with an internalistic account of mental content. It doesn't seem incompatible, however, to align a broadly teleological account of content with internalism.

require saying how non-contentful states can provide justification. But then the scatter problem will re-arise. So, content evidentialism has, at best, postponed the problem.

This objection has its roots in a problem first noted by Wilfrid Sellars.²⁸ A fully adequate treatment of this problem is beyond the scope of this paper. But it is not implausible that the objection runs awry by requiring that *every* mental state with assertive content requires justification. In ordinary language perceptual states are not candidates for justification. We don't ask what justifies one's perceptual experiences. It seems strange to ask, for instance, what justifies it looking to you as if a blue cup is on the table. Similarly, in the Müller-Lyer case we don't hold you accountable for it looking as if the two lines are incongruent. As long as the content evidentialist can resist this universal demand for justification the objection is unsuccessful.

IV. Content Evidentialism & the Truth Connection

The scatter problem raises concerns about the theoretical unity of an epistemological account: what underlying features, if any, unify the account's epistemic principles? A natural answer to this question is to appeal to the connection to the central item of epistemic value, i.e., truth. This is the consequentialist (i.e., the externalist) option that explicates justification via truth and production. The production concepts appealed to—e.g., *causation, reliability, sensitivity, safety, truth-tracking, truth-aptness*—share the central feature that there is a modal connection between justification and truth. The intuition embodied in the new evil demon problem undermines a modal connection between justification and truth and so raises concern about scatter and the truth connection. Content evidentialism handles the scatter problem by appealing to the notion of assertive mental content. This content, though, is modally independent of its truth. And this leads to worries about the truth connection. How does assertive content that is modally independent from the truth of the presented content give one a good *epistemic* reason to believe that the content is true? The following sketches an account of the truth connection that allows for the modal independence of justification and truth.

Content evidentialism denies a constitutive link between justification and truth and production. This denial is motivated by thought experiments illustrating that a belief

²⁸ See Sellars (1963) "Empiricism and the Philosophy of Mind" for the original statement of the problem and Bonjour (1985) Chapter 4 "The Doctrine of the Empirically Given" for the canonical development of the Sellars' problem.

may be justified even though one's justification is modally independent from the belief's truth. The general feature of such thought experiments is to imagine circumstances in which justified belief and belief caused by a truth-productive process come apart. For ease of reference I shall call a belief that is caused by a truth-productive process a "productive belief". A key argument for externalism is that epistemically justified belief is assessed as to how well the belief fares with respect to the goals of attaining truth and avoiding error.

This intuition about a significant truth-connection can seem to cut against non-consequentialist epistemic theories because those theories deny that justification can be properly explicated merely in terms of productive belief. But I think that the evidentialist has more to say about this than is usually recognized.²⁹ The difficulty for the content evidentialist is to explain how justification is connected to truth in more than the uninformative way that justification implies that one is epistemically justified in thinking that one's belief is true. This is trivial because (e.g.) prudential justification implies that one is prudentially justified in thinking one's belief is true.

The options for the content evidentialist are twofold: claim that there's nothing more to say about the truth connection than the above platitude or explicate the connection with some non-production concept. The first option is not promising because it does not distinguish epistemic justification from other kinds of non-epistemic justification. In the following I explore the second option. The main problem with this approach is that the non-production concept involved is often not distinctly epistemic. For instance, the central non-production concept in Foley-rationality is that of acceptance, specifically acceptance on one's deepest standards after careful reflection. Acceptable belief and productive belief come apart, so it is judged that Foley-rationality is not an adequate account of epistemic justification. A similar problem afflicts deontological accounts as long as responsible belief is analyzed in a way in which it isn't modally connected to productive belief.

The concern that non-consequentialist epistemological accounts do not uphold a strong enough truth connection is largely correct. It shows the need to incorporate productive concepts in one's account of the truth connection. In order to accomplish

²⁹ In Cohen's article (1984) he finds it intuitive that there is some connection between epistemic justification and truth but every account he surveys he finds lacking (see Cohen's conclusion pp. 292-3). For other discussions of the nature of the truth connection see Conee (1992) and Hendricks (2005).

this, however, one needn't appeal to production concepts directly. Rather a plausible account of the truth connection may appeal to production concepts indirectly. The view I propose holds that one's justification implies propositional justification for believing that one's belief is truth-productive. In the abstract this view explicates the truth connection via an epistemically justified belief whose content involves a production concept. On this view the following principle is true: *If S is justified in believing p then S has justification for believing that S's belief that p is caused by a truth-productive process.*

This account of the truth connection is inspired by remarks by Kant and Chisholm. First, Chisholm. In his book *Perceiving* Chisholm writes:

We hope... that our marks of *evidence* will also be marks of *truth*. We hope that, if there is some general mark of evidence, a certain type of state M which is a mark of evidence for a certain type of hypothesis H, then M will be a reliable criterion of truth; we hope that, more often than not, when we believe H while we are in state M, we will believe H truly.³⁰

It is intriguing that Chisholm talks in terms of hope. If hope is simply belief under a different guise then the claim is that persons in fact believe that marks of evidence are marks of truth. This seems right as far as it goes. But the central epistemological issue is whether the belief is rational or not. If the belief that our marks of evidence are marks of truth is not rational then the fact that persons have this belief does not contribute anything to a solution of the problem of the truth connection. If, however, this belief is rational it may provide an interesting solution that fits well with an epistemology that eschews a direct tie between justified belief and productive belief.

An account that the belief that marks of evidence are marks of truth is rational can be made by appropriating the Kantian idea of rational hope. With respect to practical rationality the Kantian hope is that persons are free. This hope is a postulate of practical reason, as such people are rationally entitled to believe that they are free. Although this may not be Kant's rationale the claim that we are free can be justified in the following way. This avoids the embarrassing result that the postulate is an arbitrary assumption. The idea is this. The justification for the postulate lies in the connection between freedom and moral obligation. Practical reason issues moral directives (e.g.,

³⁰ Chisholm (1957:38).

you ought to respect persons). But these moral directives are binding only if persons are free. By acting on moral directives persons exemplify themselves as free beings. That is, by acting on moral reasons persons represent themselves as free beings in a way that carries justification for the postulate. On this conception of Kant's reasoning, moral action implies epistemic justification for the postulate. Kantian hope, in this way, involves epistemically rational belief.

A similar defense can be given for the Chisholmian hope that our marks of evidence are marks of truth. On this line of reasoning perceptual experience issues doxastic directives (e.g., believe that there are tables) but these directives are mandatory only if the perceptual experience is truth-productive. By forming beliefs based on the deliverances of perceptual experience one exemplifies experience as truth-productive. That is, by believing the assertive contents of perception one represents these assertive contents as worthy of belief, as that is truth-productive. In this way the epistemic practice of forming beliefs on the basis of presentational contents implies that one has justification for the proposition that presentational contents are marks of truth.

This analogy between Kantian and Chisholmian hope deserves more consideration than I can presently offer. However, it holds promise for our present difficulties. The problem is that, on the one hand, a consequentialist view about the connection between justified belief and productive belief seems required because of the distinctly epistemic nature of justification and yet, on the other hand, such a view seems false because of the new evil demon thought experiment. One needs a non-trivial account of the truth connection that is consistent with content evidentialism. This Kantian-Chisholmian account accomplishes this. Epistemic justification is connected to the truth not by a logical or modal relation but by an epistemic relation to believe that one's evidence is truth-productive.

In the end the content evidentialist agrees with the consequentialist that production concepts have a central role to play in the solution to the problem of the truth connection. The disagreement is over where to locate those concepts. The content evidentialist avers that reflection on the new evil demon problem provides an argument that production concepts should be located within content of epistemically justified attitudes. This provides a clear connection to the primary item of epistemic value without surrendering the key intuition contained in the new evil demon problem that the

nature of evidential fit is independent on any modal relation between the belief and the fact that makes it true.

Conclusion

Reflection on the scatter problem uncovers significant epistemological lessons. The scatter problem undermines objectual evidentialism. But it leaves unscathed content evidentialism. Content evidentialism provides a principle of justification that displays explanatory unity. Moreover, it holds promise for unifying diverse sources of knowledge within a non-consequentialist epistemology. Also, by appropriating a broadly Kantian view of rational hope the propositionalist may hold that the epistemic practice of forming beliefs on the basis of evidence justifies the proposition that the marks of evidence are reliably connected to truth.³¹

Ted Poston

University of South Alabama

poston@jaguar1.usouthal.edu

³¹ I am grateful to Juan Comesana, Trent Dougherty, Jonathan Kvanvig, Eric Loomis, Kevin Meeker, and Andrew Moon for comments on an earlier draft of this paper.

References

- Bonjour, L. (1985) *The Structure of Empirical Knowledge*, Cambridge, MA: Harvard University Press.
- Bonjour, L. and Sosa, E. (2003) *Epistemic Justification*, Malden, MA: Blackwell.
- Chisholm, R. (1957) *Perceiving*, Binghamton, NY: Cornell University Press.
- _____. (1982) *The Foundations of Knowing*, Minneapolis: University of Minnesota Press.
- _____. (1989) *Theory of Knowledge* 3rd ed, Englewood Cliffs, NJ: Prentice-Hall.
- Cohen, S. (1984) 'Justification and Truth', *Philosophical Studies* 46, 279-95.
- Conee, E. (1992) 'The Truth Connection', *Philosophy and Phenomenological Research* 52, 657-669.
- Conee, E and Feldman, R. (2004) *Evidentialism*, New York: Oxford University Press.
- Fodor, J. (1990) 'Information and Representation', IN P. Hanson (ed.), *Information, Language and Cognition*, Vancouver: University of British Columbia Press.
- Foley, R. (1993) *Working without a Net*, New York: Oxford University Press.
- Goldman, A. (1979) 'What is Justified Belief?', IN G.S.Pappas (ed.), *Justification and Knowledge*, Dordrecht: D. Reidel, pp. 1-23.
- Heck Jr., R. (2000) 'Nonconceptual Content and the "Space of Reasons"', *The Philosophical Review* 109, 483-523.

- _____. (Forthcoming) 'Are There Different Kinds of Content?', IN J. Cohen and B. McLaughlin (eds.), *Contemporary Debates in Philosophy of Mind*, Malden, MA: Blackwell.
- Hendricks, S. (2005) 'Demons and the Isolation Argument', *The Philosophical Quarterly* 55, 403-418.
- McDowell, J. (1994) *Mind and World*, Cambridge, MA: Harvard University Press.
- Millikan, R. (2000) *On Clear and Confused Ideas: An Essay about Substance Concepts*, Cambridge: Cambridge University Press
- Neander, K. (2004) 'Teleological Theories of Mental Content', *The Stanford Encyclopedia of Philosophy (Summer 2004 Edition)*, Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/sum2004/entries/content-teleological/>>.
- Peacocke, C. (2001) 'Does Perception Have a Nonconceptual Content?', *The Journal of Philosophy* 98:5, 239-264.
- Ross, W. D. (1930) *The Right and the Good*, Oxford: Oxford University Press.
- Sellars, W. (1963) 'Empiricism and the Philosophy of Mind', IN *Science, Perception, and Reality*, Atascadero, CA: Ridgeview, pp.127-196.
- Sosa, E. (1991) *Knowledge in Perspective*, New York: Cambridge University Press.

HUME, ASSOCIATION, AND CAUSAL BELIEF *

J. P. Monteiro

Abstract

The associationist interpretation of Hume's account of causal belief is criticized. The origin of this mistaken interpretation is explained. The difference between Hume's views in the *Treatise of Human Nature* and in *An Enquiry concerning Human Understanding* is examined.

Was Hume an associationist about causal belief? Did he really endeavour to explain causal thought by the mechanism of the association of ideas? Some of the most influential of his commentators have presented one form or another of this interpretation — among them Alfred Ayer, John Passmore, William Kneale, Antony Flew, John Mackie and Fred Wilson. And in the Index of the best known edition of the first *Enquiry* it is asserted that according to Hume causal belief “is produced by the principles of association, viz. resemblance, § 41; contiguity, § 42; causation, § 43” (1975:395).

But if we read, in the same work and edition, the last lines of § 40 carefully, we find that Hume, having established that “the sentiment of belief is nothing but a conception more intense and steady” than mere fictions, and is derived from “customary conjunction”, pretends only to supplement this by finding “other operations of the mind *analogous* to it”, in order to discover more general principles of human nature (1975:50; my emphasis).

Hume invokes the principles of association of ideas, presented in Section III of the same work, and asks whether in the relations that occur “when one of the objects is presented to the senses or memory, the mind is not only carried to the conception of the correlative, but reaches a steadier and stronger conception of it than otherwise it would have been able to attain”, exactly as in the case of causal belief (§ 41, 1975:50-

* This paper was presented at the Williamsburg Hume Conference in 2000, and is published here in English for the first time. It was published in Portuguese in *Manuscrito*, Campinas, in 2000, and in *Análise*, Oporto, in 2001. The author hopes that its discussion by the English speaking philosophical public may help to correct its mistakes.

51). And he goes on to present some “experiments”, in order to justify the suggested analogy.

He proposes a comparison between certain mental phenomena where *lively* perceptions, either impressions of sensation or ideas of the memory, are associated with *faint and dull* perceptions, that is, ideas of the imagination (EHU II, 1975:18), and another phenomenon: the formation of causal belief. One example of the first is when the *impression* derived from looking at the portrait of a friend, associated by resemblance with the *idea* of the same friend, communicates some of its vivacity to this idea, which is thus enlivened in the process (§ 41, 1975:51). Another is when we are in the presence of an object, and the impression of that object enlivens the ideas of other objects that are usually *contiguous* to it (§ 42, 1975:52). Other examples are seeing the relics of saints, if we consider them as connected with the latter by “causation”, or seeing the son of an absent friend, the father being also viewed as a “cause” — in these cases also, association by causation tends to give a new vivacity to the otherwise fainter ideas of the “effects” (§ 43, 1975:53).

Now, what exactly do we have here? Hume will conclude on the next page that the same very general principle of “transition from a present object” is responsible for the new liveliness acquired by the formerly faint ideas, in those examples of association, and also for the new liveliness acquired by the idea of the effect, when the impression of its present cause communicates to it its natural vivacity, through the channel established by causal inference (§ 44, 1975:53-4). A new, analogical step is thus taken in Hume’s theory — but it should be clear that this in no way amounts to an explanation of causal belief by the association of ideas.

Causal belief derives, first, from the frequent conjunction mentioned by Hume in § 40, and, of course, from the principle of human nature he identifies as “custom or habit” in part I of the same Section V (1975:43 ff). The new argument adds only that the same belief presents itself as “a more vivid, lively, forcible, firm, steady conception of an object” (1975:49), namely, the idea of the effect we expect to follow from its cause, and that this new vigour is received by this idea from the impression of the present cause, a mental process that is analogous to cases of association between other impressions (or ideas of the memory) and formerly dull ideas that resemble or are contiguous to them, or are taken as their causes or effects.

How could this be mistaken for a supposed “production” of causal beliefs by the mechanism of association itself? There is clearly no suggestion of such production or derivation in Hume’s philosophy — either in the *Enquiry* or, as we shall presently see, in the *Treatise of Human Nature*. We will see that many passages about this matter *are* ambiguous, especially in the *Treatise*. But first one must make up one’s mind about this: is there, at this point, any reasonable doubt that in the passages misinterpreted by, among others, the 1902 editors of the *Enquiry*, and which Nidditch failed to correct in the 1975 revised edition, Hume could never have meant to assert that causal belief derives from the association of ideas?

To persist in that interpretation, one would have to show what else there is to be found in those pages besides an analogy between the formation of causal beliefs by the transmission of vivacity from the present cause to the idea of the effect, which thus approaches the liveliness of an impression — giving credibility to the *reality* of its imminent manifestation — and those cases where an association is established in such a way that a similar transmission takes place. Not, of course, an association between ideas of the imagination — for in that case all ideas involved would be weak and faint, and there simply would be no vivacity to be communicated.

It is impossible that causal belief could ever be produced by the association of ideas, in the original sense this has in Section III, because of the complete senselessness of the idea of transference of liveliness in that kind of association. On the other hand, an association between impressions (or memory ideas) and faint ideas is the kind of association where that transmission may take place, and this transmission is precisely analogous, according to Hume, to the transference of strength from the impression of the cause to the “common” idea of the effect, a transference that transforms the latter in the specially enlivened idea which is called “belief”. The same very general principle is at work here, but the belief, even as an enlivened idea, is not *produced* by any kind of association — only the process of its enlivening is analogous to the enlivening of certain ideas in some cases of association between forceful perceptions and faint ones.

Now, what could lead anyone to a different interpretation? The paragraph where the analogy is suggested ends with the following: “And if the case be the same with the other relations or principles of association, this may be established as a general law, which takes place in all the operations of the mind” (1975:51). That is,

what Hume's examples are intended to show is that the case of the other *relations* besides the causal relation, that is, the three principles of association, is the same as the said causal relation. If we find there the same communication of vivacity between perceptions, then what we have here is a new and very general principle of human nature, the principle of transference of vivacity. But it is impossible to read the phrase "the other relations or principles of association" as meaning that for Hume the causal relation is *also* a principle of association.

We clearly have *four* relations here: three *associative* relations (by resemblance, by contiguity and by causation) and the *epistemological* relation derived from the causal *inferences* whose origin the *Enquiry* explains in Part i of the same Section. No conflation should be allowed here of the causal relation that lies at the root of causal belief and the associative relation "by causation" that may come *afterwards*. Hume's theory of causal belief is not part of his "associationism". Maybe John Passmore (1952) was right when he detected in our philosopher an "associationist project" (1952:116). But in the case of causality, the project never succeeded, or simply never existed. On its success in other fields we shall presently comment. But causal thought and reasoning not only has the status of real *reasoning* in Hume's philosophy, it has no shadow of dependence on any associative mechanism.

In the case of Selby-Bigge himself, perhaps his "associationist" interpretation of Hume's theory of causal belief in the first *Enquiry* derives from his own reading of the *Treatise*. In his edition of that work, the Index (1978:648) quotes p.101, where Hume wrote about the formation of causal belief: "There enters nothing into this operation of the mind but a present impression, a lively idea, and a *relation or association* in the fancy betwixt the impression and idea" (emphasis mine). When Hume himself employs the term "association" in this context, it is probable that many interpreters are unable to resist the temptation to conclude that the *intention* of the philosopher really was to explain causal belief by the association of ideas, or of ideas and impressions.

Does Hume employ that term, in this passage of the *Treatise*, with the same meaning it has when he speaks of the association of ideas? The temptation to accept this should be resisted: that would imply a serious misunderstanding by the philosopher himself of his own philosophy. Granted, he could have failed to see in the

Treatise a problem he clearly understood in its recasting in the *Enquiry*. The possibility exists; but is it plausible that this was really the case? Or was Hume's use of the term "association", in the passage above, merely a case of careless and common use, simply in the general sense of "relation", with no other philosophical intention involved?

It should be noted that this passage appears in approximately the same context as the one in the *Enquiry* on which we have commented. In the *Treatise*, Hume is already presenting the same parallel between enlivened ideas in causal inferences and enlivened ideas in associations with impressions. The text is the same in the two works, §§ 41-43 of the second repeating the text between pp. 99 (5th line in the second paragraph) and p. 101 (13th line) of the *Treatise*, on associations between impressions and ideas by resemblance, contiguity and causation. The passage quoted in the Index comes in the paragraph that follows, comparing those cases of "a relation or transition of the fancy" enlivening an idea with the analogous case in "our reasonings from cause and effect". If Hume meant to say that belief is produced by association, in this context, he would have to assert it explicitly — which he is far from having done. So, the most plausible interpretation consists in taking the phrase "relation or association" in exactly the same sense as the phrase "relation or transition" above, in a common, non-technical sense, that is, with no reference to the association of ideas as a "producer" of causal beliefs.

Oliver Johnson (1995) points out that in Book I of the *Treatise* it is often difficult to tell whether terms have their ordinary meaning or the special meaning Hume gave them as "technical terms of his philosophy" (1995:3). I think that "association" is a case in point here, as a technical term in I, i, 4 and as an ordinary term, meaning the same as "relation" in the passage in question, as well as in several others. It is also plausible to think that Hume noticed this ambiguity, here and on p.93 (where "belief" was already defined as "an idea related to or associated with a present impression"), as one more example of those "negligences in expression", typical of his "juvenile work" that he came to regret in his famous Advertisement (*Enquiries*, 1975: 2) and decided to eliminate in the recasting — which he actually did. In the *Enquiry*, "association" is always employed as a technical term in Johnson's sense.

In his Index to the *Treatise*, Selby-Bigge only quotes Hume's passage, not risking any interpretation, as he unfortunately did in the *Enquiry*. But this derives

from the difference in style between the two Indexes, the first with abundant quotations and the second abstaining from them. Nevertheless, the mistake in the second has not been corrected to the present day, so that it is natural to suppose that Selby-Bigge's and his successors' interpretation of the *Treatise* was the same. That interpretation has been shared by several distinguished scholars and philosophers.

John Passmore has one chapter of his book on Hume dedicated to “the Associationist”, where he writes that for the Scottish philosopher causal connection “would form no part of the Universe for us (at least) were it not for the influence of association” (1952:116). And A.J. Ayer (1980) finds in Hume's philosophy a difference between the two first principles of association and the third, in that resemblance and contiguity “provide tracks for the movements of our attention”, whereas association by causation “is the main source of supply for our factual beliefs” (1980:56). William Kneale (1949) was “shocked” by “Hume's assertion that induction can be no more than association of ideas without rational justification” (1949:55). Also, according to Antony Flew (1961), in Hume's philosophy “the idea of association remains crucial for the whole account of learning from experience” (1961:18). All this is untenable in the face of textual evidence, as I hope is by now completely clear. It is unfortunate that such distinguished authors have contributed to perpetuate a very old muddle about the role of association in Hume's philosophy of knowledge.

Hume certainly was an associationist about the passions, the moral sentiments, and the rules of justice in society, and many other aspects of human life, as different as literature and superstition. There is plenty of evidence of this in Books II and III of the *Treatise*, in the *Dissertation on the Passions* and in the *Enquiry concerning the Principles of Morals*. But the association of ideas has no **cognitive** role in his philosophy, beyond serving as “the cement of complex ideas”. Custom or habit does have such a cognitive role, as is well known. But there is no foundation for the legend of Hume's associationism about “induction”. In sum: according to Hume, causal belief does not **derive** from any kind of association of ideas. If a causal belief is an enlivened idea, like those of the saint and of the absent friend and the corresponding examples for resemblance and contiguity (1975:51-2), this means only that there is a partial analogy between the former and the latter.

So, in the famous passage in the *Abstract* where Hume (if he really was its author) proclaims that the principles of association are “the cement of the universe”, what he may mean is only that those principles “are the only ties of our thoughts” in something other than causal reasoning (1971:86). The *Treatise* comments on the principles of association that, with their “gentle force”, they are “nature in a manner pointing out to every one those simple ideas, which are most proper to be united in a complex one”. This explains, for instance, “why languages so nearly correspond to each other” (1978:10-11; cf. EHU III, 1975:23). If we had no complex ideas, we would obviously have no notion of the universe — which seems enough to justify Hume’s celebrated metaphor.

The late J. L. Mackie wrote a beautiful book with that beautiful title, *The Cement of the Universe*. And in another book he summarises Hume’s central position about causal inference in the following terms: “The truth of the matter is just that when sequences of a certain kind have been observed a number of times an association of ideas is set up, so that on observing the antecedent we expect a successor like those which have commonly followed similar antecedent events” (1985:181). Fred Wilson (1997), notwithstanding his firm position about the rationality of Hume’s theory of knowledge, also believes that the association of ideas lies at the root of his conception of causal inference. He says, for instance, that among the “different kinds of associations” Hume includes “those involved in causal inferences” (1997:34). I believe that Wilson’s defence of Hume’s philosophy would be more convincing if he lost faith in the myth of a Humean “cognitive” associationism. His Hume, like Norton’s (1982) and Flage’s (1990), to mention just two more examples of commentators with a positive evaluation of the Scottish philosopher, does not stand for the destruction of science or common sense. This positive evaluation is incompatible with an acceptance of that myth.

I believe that all those who do accept it are the victims of a special kind of illusion, a “reading illusion” produced by the “negligences” of which Hume was admittedly guilty in the *Treatise*. J. A. Robinson (1968) has long ago commented on “the characteristic contrast between local lucidity and global obscurity which renders Hume’s meaning so often elusive” (1968:129). But in Hume’s first book there is also *local* obscurity: some passages do look, even by themselves, like symptoms of some

kind of cognitive associationism. In fact, they are nothing of the kind, but only explanations, indeed by association, of something other than causal thought or belief.

Hume did write that, besides resemblance, contiguity and causation, there is a further “principle of union among ideas” which seems different from the other three “but will be found at the bottom to depend on the same origin”, and this principle is the well known constant union between two species of objects of experience, when “the appearance of any new individual of either species naturally conveys the thought to its usual attendant” (THN I, iii, 6, 1978:93). But speaking of a common origin is far from saying that repetition, or habit, is at bottom another kind of association, in any technical or philosophical sense.

What our philosopher is presenting here is a special *psycho-linguistic* hypothesis about the use of causal terms: “Because such a particular idea is commonly annexed to such a particular word, nothing is required but the hearing of that word to produce the corresponding idea; and it will scarce be possible for the mind, by its utmost efforts, to prevent that transition. In this case it is not absolutely necessary, that upon hearing such a particular sound, we should reflect on any past experience, and consider what idea has been usually connected with the sound. The imagination of itself supplies the place of this reflection, and is so accustomed to pass from the word to the idea, that it interposes not a moment’s delay betwixt the hearing of the one, and the conception of the other” (*ibid.*). He goes on to say that the linguistic process in this hypothesis is “a true principle of association among ideas”, to which he adds: “I assert it to be the very same with that betwixt the ideas of cause and effect, and to be an essential part in all our reasonings from that relation” (*ibid.*). And here he already employs an expression (“related or associated”; the same in I, ii, 7, 1978:96) similar to the one in I, ii, 8, p. 101, quoted in the Index (“relation or association”).

What this amounts to is, I believe, the following. According to Hume, what happens *after* a causal inference has already been made, with the corresponding belief and expectation, is that every time an *impression*, occasioned by the observation of one of the objects in the conjunction, appears in the mind, the *idea* of the other object also immediately arises. *Only* after this process is “entrenched” is it possible to make the same transition from the *idea* of the first to the idea of the second, for instance when the *name* of the first object is heard by someone already persuaded that this object is always followed by the second.

What we have here is nothing like an identification of the “fourth” principle, repetition or habit, as another case of the association of ideas, or of the association of impression and idea. What we have is simply a Humean theory about what happens when *language* becomes part of causal reasoning — that is, in a late phase of the process. For Hume, causal reasoning and belief is obviously possible without, or before, the use of language — his discussion of “the reason of animals” clearly proves this (THN I, iii, 16; EHU IX). As we have seen above, in a technical sense, there is nothing associative in the making of a causal inference or the formation of a causal belief. Only causal language presents phenomena that seem to justify an explanation by associative mechanisms. This is obviously insufficient even to suggest the existence of any Humean associationism about causal belief. Still, this passage may have been one of those responsible for the interpretation that is being criticised here.

Or, perhaps, the interpretation was inspired by passages like the following: “Reason can never satisfy us that the existence of any one object does ever imply that of another; so that when we pass from the impression of one to the idea or belief of another, we are not determined by reason, but by *custom or a principle of association*” (THN I, iii, 7, 1978:97; emphasis mine). This does *look* like an identification of custom as a principle of association, but that interpretation obviously will not do: the principles of association of ideas are resemblance, contiguity, and causation, and custom is something else — it is an instinct, also called “habit”, that acts in us when we experience repeated conjunctions. What this *must* mean is that causal belief, not being produced by reason, derives *in part* from the sensitivity to repetition we call custom or habit, and *in part* from association in a *non-technical* sense, between an impression and the idea to which it communicates part of its vivacity. Granted, this is a particularly “negligent” passage, more so, perhaps, than any other one in the *Treatise* — but this could never allow us to conclude that Hume *really* meant that habit is one of those principles of association we know from the first chapters of the same work, for this would be a plain absurdity.

On the other hand, Hume explains probability by the “*association of ideas to a present impression*”, and it is from habit that the association is derived (THN I, iii, 12, 1978:130). What does this mean? It means, again, that a relation (or “association” in a non-technical sense) between perceptions is established following their repeated conjunction in experience, by the influence of custom or habit — *not* in the least by

any principle of association of ideas in the technical sense. Resemblance and contiguity, as natural relations, have nothing to do with this, and causation, as a principle of association, has in this its condition of possibility, not the other way around. Nothing here even suggests that probability derives from the association of ideas properly so called — no more than does causal belief in general.

Other passages in the *Treatise* may, when their precise context is ignored, present an associationist appearance. Most of those passages let themselves be interpreted along the lines I have tried to sketch here. But suppose that someone opens the *Treatise* directly on p.112 and simply reads this: “all belief arises from the association of ideas, according to my hypothesis”. Is it not tempting to take this as evidence in favour of the interpretation of Hume as an associationist about causation? The temptation must, I believe, be resisted, until we are able to see this text in perspective, that is, in its proper context.

This context is that of a two-pages-long argument, purporting to present a new “proof” of Hume’s theory that causal belief “is nothing but a lively idea related to a present impression”, a proof consisting in showing that the associative principles of resemblance and contiguity, although their effect in the formation of opinion is much weaker than that of causation, “still have some effect, and augment the conviction of any opinion, and the vivacity of any conception”. He begins with contiguity, presenting the example of *religious* beliefs enlivened by actually visiting holy places, like Mecca and Jerusalem: “The lively idea of the places passes by an easy transition to the facts, which are supposed to have been related to them by contiguity, and increases the belief by increasing the vivacity of the conception. The remembrance of these fields and rivers has the same influence on the vulgar as a new argument; and from the same causes” (1978: 110-1).

The second step in the proof is concerned with resemblance: “Some philosophers have imagined that there is an apparent cause for the communication of motion, and that a reasonable man might immediately infer the motion of one body from the impulse of another, without having recourse to any past observation”. If this opinion were true, it would amount to a demonstration, “and must imply the absolute impossibility of any contrary supposition” — and this is easily “refuted”, simply by clearly conceiving a completely different behaviour in the second body (1978:111; cf. EHU IV, I, 1975:29: “may I not conceive, that a hundred different events might as

well follow from that cause?”). This well-known Humean argument is here the instrument of a refutation by *modus tollens*: if we have the contrary of what was predicted by a certain theory, that theory must be rejected.

But *why* did those philosophers (the Leibnitzians, I presume) make such a terrible mistake? It is here that Hume resorts to association by resemblance, but only to explain that mistaken *philosophical* opinion, just as association by contiguity explained the religious opinions mentioned above. And the phrase we are trying to clarify appears at the end of that explanation — with no room for doubt about its real significance, as will be immediately obvious: “The reason why we imagine the communication of motion to be more consistent and natural (...) than any other natural effect is founded on the relation of *resemblance* between the cause and effect, which is here united to experience, and binds the objects in the closest and most intimate manner to each other, so as to make us imagine them to be absolutely inseparable. Resemblance, then, has the same or a paralell influence with experience; and as the only immediate effect of experience is to associate our ideas together, it follows that *all belief arises from the association of ideas, according to my hypothesis*” (1978:111-2; last emphasis mine).

Maybe my emphasis was superfluous. It is now clear, I am sure, that “all belief” means here only “all those mistaken philosophical opinions”, and not belief in general. How could it be otherwise? Much stranger than the admittedly strange phrasing would be to read it as meaning that Hume’s present hypothesis has to do with anything besides the proposed explanation of a mistaken philosophical belief, by a process of association of ideas by resemblance, following another similar explanation of (supposedly mistaken) religious beliefs by a process of association of ideas by contiguity. Curiously enough, Selby-Bigge’s Index is silent about p.112, although it notes that on p.111 “resemblance and contiguity augment the vivacity of any conception” (1978: 649).

What is here produced by the association of ideas, according to Hume, is only a mistake and a philosophical illusion: some philosophers have imagined or fancied that the resemblance between cause and effect *in this exceptional case* — in general they are quite distinct, as is Hume’s well known doctrine (“the effect is totally different from the cause, and consequently can never be discovered in it”, EHU, V, I, 1975: 29) — may serve as grounds for believing that they really “find the effect in the

cause”. This is simply a Leibnizian illusion, produced by an associative or psychological trick in the fancy. That is all there is to it. As is noted further along in the *Treatise*: “Nothing is more apt to make us mistake one idea for another than any relation between them, which associates them together in the imagination, and makes it pass with facility from one to the other” (I, iv, 2, 1978: 202).

The influence of the interpretation I criticise here is such that even Kant has been its victim, in some of the translations of the *Critique of Pure Reason*. In Part II of the Introduction, he regrets the direction taken by Hume’s theory of causality, mainly because the Scottish philosopher derives the concept of cause only from “frequent concomitancy” (“*öftern Beigesellung*”) (1956:41). This was correctly translated into English by F. Max Müller (1881:401), before the turn of the century, and earlier than that Tissot (1864:35) had chosen “liaison fréquente”, which is quite acceptable. But in the present century many translations have “association” or its equivalent instead of “concomitancy”, or “concomitance”, or “conjunction” as would also be correct¹, thus transforming Kant into one, perhaps the first, of those who read Hume as an associationist about causality — which Kant certainly never did! Those were all interpretative translations. *Traduttori, tradittori?* Maybe all translators are, in some measure (myself included) — but these, I think, went beyond measure. They “corrected” the text of the *Critique*, introducing into it what they thought was the true version of the Humean theory there criticised by Kant. And this may have given strength to that interpretation, due to the prestige of Kant, *at least* as a philosopher who was saved from his “dogmatic slumbers” by reading Hume — and reading him correctly, it is to be presumed. How could Kant be wrong about this? Well, he was not — only his translators in the twentieth century were, as well as many interpreters of the philosophy of David Hume.

Hume’s theory is that effects and their causes are linked by something other than *deduction*, but he never dreamed of replacing “demonstration” by the association of ideas. The relation between effects and their causes may be called a simple “association”, in contrast with the deductive union they cannot have, as was Hume’s most celebrated discovery — but only in a popular sense, not in the Humean technical

¹ Those translators include, at least: J. Barni (revu par P.Archambault) (1976:59); A. Tremesaygues & B. Pacaud (1950:34); Alexandre D. L. Delamarre & François Marti (1980:76); J. Perojo (1967:149);

or philosophical sense of the “gentle force” of the association of ideas (THN I, iii, 4, 1978:10), which may or may not lead the mind from an idea to another that resembles the first, or is “contiguous” to it. This corresponds only to a certain tendency of the human mind to follow, in some cases, certain associative paths.

In complete contrast with this, there is nothing “gentle” about the forces that generate causal belief, whose operation is strong and unavoidable: Let Hume’s philosophy speak for itself: “This belief is the necessary result of placing mankind in such circumstances [repeated experience]. It is an operation of the soul, when we are so situated, as unavoidable as to feel the passion of love, when we receive benefits; or hatred, when we meet with injuries. All these operations are a species of natural instincts, which no reasoning or process of the understanding is able either to produce or to prevent” (EHU V, i, 1975:46-7). I think this well known passage counts against any attempt to derive Hume’s conception of causal belief from any merely associative process. When in 1748, nine years after the *Treatise*, Hume published his definitive philosophy of knowledge in the *Enquiry*, all “negligences” of expression about these matters had been definitely left behind.

J.P. Monteiro

University of São Paulo

jpmonteiro@netcabo.pt

References

- Ayer, A. J. (1980) *Hume*, Oxford: Oxford University Press.
- Flage, D. (1990) *David Hume's Theory of Mind*, London: Routledge.
- Flew, A. (1961) *Hume's Philosophy of Belief*, London: Routledge & Kegan Paul, London.
- Hume, D. (1971) *An Abstract of A Treatise of Human Nature - bilingual edition*, Paris: Aubier.
- Hume, D. (1975) *Enquiry Concerning Human Understanding*, IN *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, edited by L. A. Selby-Bigge, 3rd edition revised by P. H. Nidditch, Oxford: Clarendon Press.
- Hume, D. (1978) *A Treatise of Human Nature*, edited by L. A. Selby-Bigge, 2nd ed. revised by P.H. Nidditch, Oxford: Clarendon Press.
- Johnson, O. (1995) *The Mind of David Hume (A Companion to Book I of A Treatise of Human Nature)*, Urbana and Chicago: University of Illinois Press.
- Kant, I. (1864). *Critique de la Raison Pure*, trans. by J.Tissot, Paris:Lagrange.
- (1881). *Critique of Pure Reason*, trans. by F. Max Muller, London:Macmillan.
- (1934). *Critique of Pure Reason*, trans. by J. M. D. Meiklejohn, London: Dent & Sons.
- (1945). *Critica della Ragion Pura*, trans. by G. Gentile & G. Lombardi-Radice, Bari : Laterza.

----- (1950). *Critique de la Raison Pure*, trans. by A. Tremesaygues & B. Pacaud,
Paris : Presses Universitaires de France.

----- (1956). *Critik der reinen Vernunft*, Hamburg: Felix Meiner.

----- (1967). *Crítica de la Razón Pura*, trans. by J.Perojo, Buenos Aires: Editorial
Losada.

----- (1976). *Critique de la Raison Pure*, trans. by J. Barni (revu par
P.Archambault), Paris : Flammarion.

----- (1980). *Critique de la Raison Pure*, trans. by Alexandre D. L. Delamarre &
François Marti, IN *Oeuvres Philosophiques*, Paris : Gallimard.

----- (1989). *Crítica da Razão Pura*, trans. by M. P. Santos & A. Morujão, Lisboa:
Gulbenkian.

----- (1992). *Critique of Pure Reason*, trans. by N. Kemp Smith,
London:Macmillan.

Kneale, W. (1949) *Probability and Induction*, Oxford: Clarendon Press.

Mackie, J.L. (1985) *Logic and Knowledge*, Oxford: Oxford University Press.

Norton, D.F. (1982) *David Hume, Common-Sense Moralist, Sceptical Metaphysician*,
Princeton: Princeton University Press.

Passmore, J.A. (1952) *Hume's Intentions*, London: Duckworth.

Robinson. J.A. (1968) 'Hume's Two Definitions of 'Cause' ', IN V. C. Chappell (ed.),
Hume, London: Macmillan.

Wilson, F. (1997) *Hume's Defence of Causal Inference*, Toronto: University of Toronto Press.

DEWEY'S METAPHYSICS OF MIND

Wilson Mendonça

Abstract

In *Experience and Nature* Dewey makes “an attempt to contribute to what has come to be called an ‘emergent’ theory of mind”. On a first approach, that doesn’t look very innovative to our contemporary materialist convictions. Indeed, Kim argues persuasively that a central claim of emergentism—concerning the irreducibility of emergent properties—is irremediably at odds with a view of mental causation that follows from some very plausible physicalist assumptions. This is “the problem of downward causation.” I intend to show that Dewey’s brand of emergentism actually allows an adequate reply to the very important worry formulated by Kim.

1. A contribution to an “emergent” theory of mind

There is much in the writings of John Dewey that is now only of historical interest. In many passages, however, Dewey seems surprisingly to be addressing issues that have high priority in our contemporary agenda. This applies specially to the remarks on the metaphysics of mental events as presented in Chapter VII (“Nature, Life and Mind-Body”) of *Experience and Nature*.¹ Dewey makes here “an attempt to contribute to what has come to be called an ‘emergent’ theory of mind” (271). He conceives of these remarks as a rethinking of the premises and assumptions leading to the mind-body problem and the theories offered as “solutions” to it, which “range from the materialism of Hobbes, the apparatus of soul, pineal glands, animal spirits of Descartes, to interactionism, pre-established harmony, occasionalism, parallelism, pan-psychic idealism, epiphenomenalism, and the *élan vital*” (252). On a first approach, that doesn’t look very innovative to our contemporary materialist convictions. Indeed, as Jaegwon Kim points out,² emergentism was a first attempt to formulate the doctrine of nonreductive materialism. It flourished during the first half of the twentieth century and

¹ Dewey (1958). Numbers in parentheses refer to the pages of this book.

² In Kim (1998: 226).

gave away to those forms of nonreductive materialism that substitute supervenience or realization relations for the more or less metaphorical “emergence of higher-level properties.” Moreover, Kim argues persuasively that a central claim of emergentism—concerning the irreducibility of emergent properties—is irremediably at odds with a view of mental causation that follows from some very plausible physicalist assumptions. This is “the problem of downward causation.” As I intend to show that Dewey’s brand of emergentism actually allows an adequate reply to the very important worry formulated by Kim, I will start with a brief characterization of the problem of downward causation.³

2. The problem of downward causation

Consider the case, where the instantiation of an emergent property *M* causes the instantiation of another emergent property *M**. The emergentist assumes that the appearance of emergent properties depends on the presence of appropriate basal conditions. Also, the emergentist is a physicalist in the sense that physical conditions ultimately determine the instantiation of all the properties there are. So we have for the emergent property *M** a determining physical property *P**. The counterfactual implication of *M*’s claim as a cause of *M** says that *M** could not have been instantiated, if *M* were not present on this occasion. The determination relation between *P** and *M**, on the other hand, implies that unless *P** were present on this occasion, *M** could not have been instantiated. The plausibly coherent description of the situation seems to be: the instantiation of *M* causes the instantiation of *M** by causing the instantiation of *P** in the first place; the later instantiation determines then the instantiation of *M**. The first part of this description is of course downward causation. To it the emergentist is committed. For, as Kim argues, if emergent properties are really new, then the causal powers associated with them are irreducibly distinct from the causal powers of the properties defining the conditions out of which they emerge. This means that the causal role of *M* in the process by which *P** is brought about cannot be “preempted” by any physical property. However, as we have a good reason to assume that the instantiation of *P** has as its cause an instantiation of another physical property—this reason being derived from the assumption that the physical world is

³ I will draw here on Kim (1998: 229-233).

causally closed—, the purported distinctness of emergent causal powers results in the weird supposition that physical phenomena underlying emergent phenomena are systematically overdetermined. They are overdetermined in the sense that they have two independent causes, a physical cause and an emergent one. The weirdness of the supposition comes from the fact that the joint operation of two causes, each one being sufficient to bring about the effect, should manifest itself not occasionally, but *whenever there is causation by emergent properties*. Moreover, if we decide to apply to the relation between the emergent property *M* and its emergence-base *P* the same reasoning applied to the relation between *M** and *P**, we arrive at the conclusion that it is ultimately *in virtue of* some necessarily co-instantiated physical property *P* that the instantiation of *M* causes the instantiation of *P** (and also the instantiation of *M**). While this would solve the problem of massive overdetermination of physical phenomena, it would also make higher-level causal relations ultimately dependent on, derivative from the causal processes at the physical level: all irreducible causal powers would turn out to be physical properties. Kim is right when he points out that the emergentist, committed as she is to downward causation, could hardly accept this view of higher-level causation. But the only alternative seems to be the abandonment of the physical causal closure, which is not really open for the emergentist, insofar as she keeps committed to physicalism.

3. A multilayered model of the world: “physical,” “psycho-physical” and “mental”

Dewey's brand of emergentism agrees with Kim's emergentist in giving the physical a fundamental role in a multilayered model of the world. Actually, the model proposed by Dewey distinguishes three such layers, three fields of interaction among events *that are ontologically homogeneous*. The layers are termed by Dewey “physical,” “psycho-physical,” and “mental;” and the distinction between them is “one of levels of increasing complexity and intimacy of interaction *among natural events*” (261, my emphasis). Accordingly, “the idea that matter, life and mind represent separate kinds of Being” (261) is in Dewey's view a “philosophic error.” For the layers proposed by Dewey reflect only differences in the way natural, ultimately physical events are connected: while there is in his ontology no isolated occurrence in nature, “interaction and connection are not wholesale and homogeneous” (271).

The first layer, “the scene of narrower and more external interactions,” is “physical.” Dewey associates with this layer the properties “of the mathematical-mechanical system discovered by physics and which define matter as a general character” (272). The second layer is “psycho-physical.” It doesn't mean “an abrogation of the physico-chemical” or “a peculiar mixture of something physical and something psychical” (255). All that is implied in the idea of psycho-physical is a certain degree of organization as an empirical trait of some events. The peculiar organization justifies the choice of “life” as a general character of the events on this level. In this sense “there is no problem of the relation of physical *and* psychic” (255). Living activity is essentially characterized “by needs, by efforts which are active demands to satisfy needs, and satisfactions” (252). Dewey immediately explains the intended sense of these words in terms of completely physical conditions: “tensional distribution of energies,” “states of uneasy or unstable equilibrium,” “movements which modify environing bodies,” “recovery of equilibrium pattern” (253). That is why his theory makes no ontological difference between life processes and the activity of inanimate bodies.

The difference between the animate plant and the inanimate iron molecule is not that the former has something in addition to physico-chemical energy; it lies in the *way* in which physico-chemical energies are interconnected and operate, whence different *consequences* mark inanimate and animate activity respectively. (253f.)

Finally, the third layer in Dewey's model of the world is that of mind. In this somehow peculiar sense, mind is “a further process in life” (281), “a new scheme of affairs to which both organic and environmental relations contribute, and in which they both partake” (283). The important point is that the mental, as well as the psycho-physical, is ontologically homogeneous with the physical. “The external or environmental affairs, primarily implicated in living processes and later implicated in discourse ... are as ‘physical’ as ever they were” (285). The relevant difference between mind and the physical is not to be understood in terms of being, but again in terms of the organization and the degree of connection among natural events. Association, communication and participation represent for Dewey a form of integration of organic-environmental connections, which is vastly superior to those of animals without language, and which endows the events so connected—and, by extension, the organisms involved—with new properties properly termed “mental.” This is summed up in the definition of mind as “an

added property assumed by a feeling creature, when it reaches that organized interaction with other creatures which is language, communication” (258) or as “what actually takes place when a living body is implicated in situations of discourse, communication and participation” (285).

Ontological continuity between physical and mental is further guaranteed by empirical facts such

... that animals are connected with each other in inclusive schemes of behavior by means of signaling acts, in consequence of which certain acts are deferred until a joint action made possible by the signaling occurs. In the human being, this function becomes language, communication, discourse, in virtue of which the consequences of the experience of one form of life are integrated in the behavior of others. (280)

As there is, in Dewey's view, no philosophical problem of the relation of organic life *and* the physical, there is likewise no problem of the relation between mind *and* body. Organic acts, which are only peculiar ways of operation of physico-chemical energies, are “a kind of fore-action of mind” (282). They turn to full-fledged mental acts when “organisms get more complex and human” and their ability to procure support of needs from surrounding media “involves more extensive and more enduring changes in the environmental order” (283). All relevant distinctions here are differences in degree. Vital acts of utilization, of biological adaptation, form for Dewey the “immediate material of thought when social communication and discourse supervene” (270). Through the organism's participation in communicative interactions, biological acts acquire sense and meaning, but they also *persist* and “supply mind with its footing and connection in nature” (290). “Our physical names for mental acts like seeing, grasping, searching, affirming, acquiescing, spurning, comprehending, affection, emotion are not just ‘metaphors’” (290). As Dewey frequently points out, meanings, the mental items *par excellence*, are meanings *of*: they “intrinsically have reference to natural events” (288). Even mental images and “ideas,” as the conscious, *qualitative* aspects of meanings, are in Dewey's view only further properties “of *partial* organic behaviors, which are their ‘stuff’” (291).

For Dewey, it is only “natural” that, as the functions of mind developed out of organized patterns of physiological and vital affairs, “the presence and operation of meanings” (290) constituting mind just *is* the use of “structures which are biological adaptations of organism and environment,” these structures being mind's “own and only

organs” (277). “If thinking is naturally serial with biological functions ... it will have as the material of thought, even of its erratic imaginings, the events and connections of this environment” (279). Accordingly, a mind essentially involves “a world or nature temporally and spatially ‘external’ to itself but ‘internal’ to its functions” (278).

4. A contextualist account of qualitative differences: having and knowing

Although some influent contemporary philosophers of mind—to mention just a few: Thomas Nagel, Ned Block, David Chalmers—could perhaps accept Dewey's naturalist and *externalist, contextualist* account of the mind's *functions*, they would insist that the more *qualitative* aspects of mental states cannot be captured by it. At least some mental states, these philosophers would argue, have “subjective characters” (Perry), of which we are immediately conscious. These are the “qualia,” the “what-it-is-like” properties of conscious events, which are supposed to be *intrinsic* to conscious events and which would presumably resist any “reductionist” attempt as put forward by Dewey's contextualist theory of mind. Here is how John Perry defines them:

It seems clear ... that the subjective character of a mental state is not an historical or contextual property of it. It is a property of it that is determined by current inner events. The phenomenal event will typically have external causes and effects, and it may have many current properties that are determined by such external factors. But the subjective character of the event will not be one of these properties. The subjective character is a matter of what it's like to be in the state, not its typical causes, nor its causes on a given occasion.⁴

Now, one of the most remarkable features of Dewey's emergentism is its explicit recognition of qualitative differences in mental events. In his view qualities are for real. Not by chance, Dewey's preferred word for emergent features just is “qualities.” While it is initially and frequently used by him in the sense of “objective” properties of natural events, which are accessible from the “third person” point of view, it is also interestingly connected with the conscious, “first person” aspects of experience aimed at by the above mentioned philosophers. A remarkable instance of this double sense is the assertion that “in feeling a quality exists as quality” (266).

This view requires comment. A good starting point is the claim that the “basis of sensitivity,” as Dewey puts it, is given when the activities of the parts of an organism

⁴ Perry (2001: 35).

are so organized as to tend to perpetuate the whole patterned activity of which they are parts. “This pervasive operative presence of the whole in the part and of the part in the whole constitutes susceptibility—the capacity of feeling” (256). Psycho-physical interactions involving even lower forms of life already show this capacity, irrespective of the fact that it can remain unrealized. For the constituent parts of even a plant or a lower animal tend to act *selectively* in the environment so as to maintain the organized body to which they belong. Locomotive organs and distance-receptors as found in higher animals bring with them a substantial enhancement of this selective power involved in the maintenance of vitally relevant patterns of energy-organization. Organically connected with the remote as well as with the nearby, mobile animals endowed with organs for distal events are able to act “with reference to a spread-out environment as a single situation” (279). Psycho-physical activities are then organized into a comprehensive unity such that the present phase embodies cumulatively what has occurred and, at same time, anticipates encounters with the realities of the environment having to do with needs and their satisfaction. “Each immediate preparatory response is suffused with the consummatory tone of sex or food or security to which it contributes” (257). Dewey states that in this case sensitivity is “realized as feeling, even though only as vague and massive uneasiness, comfort, vigor and exhaustion” (256).

The operative assumption behind this statement seems to be that the states of a higher animal's body, connected as they are to events in the environment, put the organism in condition simply to have sensations or “feelings” that reflect qualitatively the total organic disposition of the body and, by extension, the inclusive happenings “outside.” (Dewey's is a *representational* theory of the conscious mind, as it has been recently—and exemplarily—developed by Michel Tye.⁵) This is not tantamount to the assumption of a mysterious mode of cognitive access to the world. And the main reason why the later assumption is not part of Dewey's theory, is explicit in Dewey's denial that the organism is always aware of its “feelings” in the sense of their being *epistemically accessible* to it. Normally, the organism is not aware of the distal events in the “external” world represented in its “feelings.” Neither must the organism be aware of the “surrogate events” in the nervous system—that is, of the proximal aspects immediately reflected in qualitative differences presented “in the having:”

⁵ Tye (1995).

Complex and active animals *have*, therefore, feelings which vary abundantly in quality corresponding to distinctive directions and phases—initiating, mediating, fulfilling or frustrating—of activities, bound up in distinctive connections with environmental affairs. They have them, but they do not know they have them. (258)

By *having* sensations the organism can be put in a condition conducive to its *knowing* them—if *other* conditions are satisfied.⁶ The latter include for Dewey the *discrimination* of the submerged unidentified qualities or general tones of whole situations—the vague feelings characteristic of psycho-physical activities in animals—through their use “as common and shared means to common ends” (260). To become specifically known (and therefore “mental”), differences in immediate sentience must be “employed as indications of acts performed and to be performed and as signs of their consequences” (258). Dewey is at pains to make it plausible that language and social intercourse are necessary means for the “objectification” of immediate qualitative differences as the cognitive contents of our epistemic states. *Discriminated* qualities, that is, qualities referred to “external” aspects, are “meanings” in Dewey’s sense of the word. Being incorporated in a system of signs pointing actively to vital aspects of the relationship of the organism and the natural and social environment, immediate qualities acquire the “sense” of the consequences they have in living and become meanings. They turn then to “traits of things.”

To term a quality “hunger,” to name it, is to refer to an object, to food, to that which will satisfy it, towards which the active situation moves. Similarly, to name another quality “red,” is to direct an interaction between an organism and a thing to some object which fulfills the demand or need of the situation. ... organically conditioned qualities ... are discriminated only as they are employed to designate objects; red, for instance, as the property of a dress or toy. ... The child has to learn through social intercourse that certain qualities of action mean greediness or anger or fear or rudeness; the case is not otherwise with those qualities which are identified as red, musical tone, a foul odor. (259f.)

There are thus two mechanisms connecting immediate qualities to environing features. By the operation of the first mechanism, properties of organic interactions are realized in the organism as qualitative differences in sentience. On this level, qualities are simply felt, that is, merely *had*, not *known*. The second mechanism endows felt qualities with

⁶ “Sentiency in itself is anoetic; it exists as any immediate quality exists, but nevertheless it is an indispensable means of any noetic function.” (259)

sense, by referring them back to their causes and active consequences. Immediate qualities become on this level meanings, qualities simultaneously had *and* known. The second mechanism presupposes of course the first. Both mechanisms may be so complex as to defy a detailed theoretical description. But they are not problematic from the philosophical point of view. The first mechanism can be made equivalent to the nervous system, “the mechanism of the connection or integration of acts” (293), if we are prepared to see “the nervous system in organism,” “the organism *in* nature,” where the *in*-relation, as Dewey insists, has beyond the spatial dimension also a temporal one: “when thus seen they will be seen to be *in*, not as marbles in a box but as events are in history, in a moving, growing never finished process” (295).⁷

The second mechanism is much more complex, involving as it does communication and social intercourse. Irrespective, however, of how exactly we come to have knowledge of our qualitative states, the mechanism doesn't seem to depend on philosophically mysterious principles concerning essentially private, non-natural entities. After all, as psycho-physical properties, “the qualities never were ‘in’ the organism; they always were qualities of interactions in which both extra-organic things and organisms partake” (259). Dewey is aware that a lot more must be said about the process by which immediate qualities are “objectified” as traits of things by being referred back to their contextual origins. He is only urging that if an organism is able to use qualitative differences as successful indications of consequences of acts past and future, this operation will have the character of knowledge.

It is worth noticing how Dewey's contextualist account of immediate qualities differs from fashionable theories that identify them with what-it-is-like properties of experiential states. As we saw, these are defined by Perry as intrinsic properties only contingently connected to external causal factors. Accordingly, knowledge of these properties is supposed to be a matter of attending to the experience and recognizing *in it*, in abstraction from all relations to contextual features, the kind of experience that it

⁷ Immediately felt qualities, as conceived by Dewey, are therefore akin to the “nonconceptual contents” of representational events in the brain, which are at the very center of Tye's theory of phenomenal consciousness. As such, they are the output side of proprio-ceptors and extero-ceptors “mechanically” attuned to contextual happenings in the natural environment as well as in the body itself. Thus, representational events in the nervous system explain how properties of integrated events previously occurring on the physical level (qualities in the wider sense) are realized as immediately felt qualitative differences (qualities in the narrow sense). This is the content of the cryptic remark that “in feeling a quality exists as quality.”

is. The expression of this knowledge is, for instance, “This is what it is like to see red,” where the “this” is conceived as an inner demonstrative referring to the type of the experience, that is, to the intrinsic property instantiated in the experience itself. As Perry puts it, “This is what it is like to see red” says “that the normal experience of seeing red is of that type, has that character.” And the person uttering this sentence “is using her new experience as an exemplar of the type, in order to refer to the type.”⁸ Dewey, by contrast, individuates qualitative aspects of experience *relationally*, by reference to “extrinsic” affairs. Qualities in this sense are always qualities *of*: “of inclusive situations” (265), “of events in a peculiar condition” (258), “of action” (260), “of cosmic events” (267), “of the active relationship of organism and environment” (259), “of the things engaged as [much as] of the organism” (259). Even when qualities exist only as a general tone of the situation, they are “potentially and proleptically” (258) significant of objective differences in external things. Dewey would also take radically intrinsic properties *à la* Perry to be epistemically inaccessible for us. The very idea of knowledge of what-it-is-like would appear to him as based on “the notion that sensory affections discriminate and identify themselves, apart from discourse, as being colors and sounds, etc.” That sensory affections could “thus *ipso facto* constitute certain elementary modes of knowledge, even though it be only knowledge of their own existence,” is, however, in Dewey’s eyes, only a “preconception about mind and knowledge” (259).

5. The specific causal efficacy of qualities

Thus, qualities are *the* emergent properties in Dewey’s emergentism. With reference to them we can discuss all the issues around the problem of downward causation, which, as we saw, is Kim’s main instrument to reveal the failure of emergentism as a suitable form of materialism. Let us ask, to begin with, how the causal powers of qualities relate to *physical* causal powers. Although Dewey states that “qualities become specifically effective ... in psycho-physical situations” (268) and even associates with qualities “efficacies not displayed by the inanimate” (255), he leaves no doubts that these “new efficacies” do not go *beyond* physical causal powers in a sense contrary to the main tenets of ontological materialism. There is no incoherence here. Notice, first, that

⁸ Perry (2001: 78).

qualities have physical conditions for their emergence. Dewey only extends the dependence relation between the occurrence of qualitative properties, on the one hand, and physical properties, on the other hand, to the very *causal relations specifically involving qualities*: “it is in virtue of the character of events termed matter that psycho-physical and intellectual affairs can be differentially determined” (263). Causal facts involving physical processes determine, then, the facts about the causation of events by qualitative properties, so that “what is known about the earlier ‘physical’ series is applied to interpret and direct vital phenomena” (284). Life and mind, as Dewey puts it, have a “mechanism,” which means “an addition to our resources” and without which “education, deliberate modification, rectification, prevention and constructive control would be impossible” (263). Given this “mechanistic” dependence of qualities on physical aspects of the world, it would be wrong to assign direct, non-derivative efficacy to qualities. Indeed, this is for Dewey the error of Greek science, which, as he sees it, tried to describe and explain the world in terms of the efficacy of “qualities like wet and dry, hot and cold, heavy and light and ... such qualitative differences in movement as up and down, to and fro, around and around” (265). Dewey’s ontology aligns here with the “mechanistic” approach of modern science in its characteristic “denial of causal status (and hence of significance for science) of these and all other direct qualities” (265) and the corresponding replacement of qualitative differences “by non-qualitative indices of number and form” (266f.). “Fruitful science of nature began when inquirers neglected immediate qualities ... in behalf of ‘primary,’ namely, signifying, qualities, and when they treated the latter, although called qualities, not as such but as relations” (263).

However, while the immediate individuality of qualities “is impertinent for science, concerned as the latter is with relationships” (266), there is for Dewey a justifiable way of vindicating Greek science’s “underlying assumption that qualities count for something highly important” (268). “When knowing inanimate things, qualities as such may be safely disregarded” (266). At this level, “the career of an event can indeed be fully described without any reference to its having red as a quality” (268). (Dewey’s is incidentally a metaphysics of mind that recognizes the principle of the physical causal closure, which is so congenial to contemporary materialism.) But *in life and mind*, qualities can be reinstated in an active, although derivative role. Dewey is very careful

in locating adequately the “new efficacy” of qualities. Thus, he warn us against attributing to them, *apart from organic action*, “efficiencies which qualities possess only through the medium of an organized activity of life and mind” (265). The key element in this account of “new” causal powers is organization of *physical affairs* in comprehensive wholes constituting psycho-physical and mental phenomena. For if life and mind, as we saw, are only “characters” of complex and extensive interactions of events, they do not represent the surreptitiously introduction of mysterious forces or powers beyond what can be found at the purely physical level. The specific efficacy of qualities, as Dewey writes, is not external to the events connected by it: “it is all one with the organization that permeates them, and which in permeating them, converts prior limitations of intensity and direction of energy into actual and intrinsic qualities, or sentient differences” (266).

Hence the “new” causal powers assigned to qualities are really just the “old” physical ones, organized into manageable units. This fact alone is sufficient to block Kim's general argument against emergentism. For it makes it clear that Dewey's emergentism, as opposed to Kim's, is *not* committed to downward causation as an ontologically irreducible form of causation. To be sure, Kim would then immediately deny independent reality to purportedly new, emergent properties that do not have irreducible causal powers of their own. Dewey would counter with an observation to the effect that “the most adequate definition of the basic traits of natural existence can be had only when its properties are most fully displayed” (262)—which they are, Dewey would urge further, when the complexity of interacting events attain the levels of life and mind.

6. Property emergence and emergent causation

In Dewey's theory, facts about causation by higher-level properties emerge from, are determined by, causal facts about physical processes. Therefore, causation of events by emergent properties, that is in our case: causation in virtue of qualities, is itself *emergent causation*. As opposed to downward causation, which is causation *across levels*, it is a case of “same-level” causation. It is entirely deployed on the level of emergent phenomena and connects facts constituted by aspects of a temporally and spatially spread-out environment organized into unity. This shows itself on both sides of

the causal relation. The living organism, to which the causal powers “not displayed by the inanimate” are assigned, “is not just a structure; it is a characteristic way of interactivity which is not simultaneous, all at once but serial” (292). It “acts with reference to a time-spread, a serial order of events, as a unit, just as it does in reference to a unified spatial variety” (279). As to the effects generated by higher-level properties, the behavioral responses of the organism, Dewey argues against their identification with locally individuated tokens of bodily movements. “The remote and the past are ‘in’ behavior making it what it is. The action called ‘organic’ is not just that of internal structures; it is an integration of organic-environmental connections” (279).⁹ Qualities become then specifically productive of effects by giving rise to a whole “mode of action,” to “a certain pattern of energy-organization” (268), whereby “the former terms of a historic process are retained and integrated in this present phase” (281)¹⁰—that is: *not* by giving rise to “local” movements.

As remarked earlier, facts about higher-level causation have an extensional counterpart on the physical level, where the mechanism of life and mind is located. The point of separating hierarchically the level of life and mind from the physical level is based on the assumption that there are fields of interacting events, which, although extensionally identical with the physical, cannot be *interestingly captured* by the terms of the latter alone. Indeed, Dewey's general argument for emergentism states that, whereas the application of what is known about physical affairs to psycho-physical and mental phenomena is possible and even desirable, “this application does not exhaust their character nor suffice wholly for their description” (284). In agreement with the emphasis put on the incapacity to describe—and this means: to describe *interestingly*—Dewey asserts that each one of the levels in his metaphysical picture of the world “has its own categories,” which are fundamentally “categories of description, conceptions required to state the fact in question” (272f.). When dealing with the facts about higher-level causation, we are, therefore, dealing with *non-extensional* relations between

⁹ The possibilities of this integration are widened in the case of an organism endowed with the functions of language and communication: “Not merely its own distant world of space-time is involved in its conduct but the world of its fellows.” (280)

¹⁰ The most general fact, to which Dewey is calling our attention in these passages, is the one “so much ignored and virtually denied by traditional theories:” “The thing essential to bear in mind is that living as an empirical affair is not something which goes below the skin-surface of an organism: it is always an inclusive affair involving connection, interaction of what is within that organic body and what lies outside in space and time, and with higher organisms far outside.” (282)

category-dependent facts, whose individuation reflects, moreover, the explanatory interest of putting the phenomena “in better order, because in a wider context” (284). In particular, Dewey leaves no doubt that his emergentism answers to the interest of countering the “pre-occupation with what is specific, particular, disconnected,” in behalf of a holistic “sense of the intimate, delicate and subtle interdependence of all organic structures and processes with one another” (295). Anyway, the irreducibility of higher-level causation to physical causation, that is: of *emergent causation* to causation *tout court*, turns out to be an explanatory issue.

7. An instrumentalist theory of mind?

Is Dewey's emergentism, then, at bottom only a form of “instrumentalism,” as the latter is conceived—and generally rejected—by contemporary philosophers of mind? *Not* in the sense in which instrumentalism immediately implies anti-realism. There is, after all, in Dewey's account of mental causation a clear sense in which the widely scattered aspects of the world, organized into unities, *are really there*, “waiting” to be discovered and systematized by our theories. But since the “hard” facts of higher-level causation are made dependent, for their individuation, on our categories and explanatory strategies, we may still be inclined to call Dewey's approach instrumentalist. So be it! It seems to me, anyway, that it is a *defensible* form of instrumentalism. What is more to the point: it seems to me that it is a form of instrumentalism that can be put to work in the contemporary debate on the possibilities of nonreductive materialism.

Wilson Mendonça

Federal University of Rio de Janeiro

mendonca@ifcs.ufrj.br

References

Dewey J. (1958) *Experience and Nature*, New York: Dover.

Kim, J. (1998) *Philosophy of Mind*, Boulder: Westview Press.

Perry, J. (2001) *Knowledge, Possibility and Consciousness*, Cambridge, Mass.: MIT Press.

Tye, M. (1995) *Ten Problems of Consciousness*, Cambridge, Mass.: MIT Press.

UMA REFLEXÃO A RESPEITO DA EVOLUÇÃO HUMANA E A NATUREZA DA LINGUAGEM ¹

Alfredo Pereira Jr.

Abstract

In this paper I discuss scientific and philosophical issues related to the interdisciplinary area of Neurolinguistics, focusing on the origin and nature of human language. What kind of evolutionary process leads to the emergence of language? I propose that this process began with the adoption of bipedalism, triggering the co-evolution of new brain structures and forms of social organization. This process created a context that favoured the emergence of articulated vocalization using a combinatorial linguistic system. The identification of sufficient conditions for the emergence of human language in this context depends on the paradigm adopted for the analysis. I look for a synthesis of nativist and constructivist paradigms, showing that both contribute to an understanding of the origin and nature of human language.

1. Introdução

Que processos cerebrais apóiam a expressão vocal e atribuição de significado para palavras e sentenças? Que processos suportam o uso de regras gramaticais? Estas operações requerem mecanismos especiais, ou são feitas pelas mesmas estruturas responsáveis por outros processos cognitivos? A neurociência cognitiva da linguagem (*neurolinguística*) procura dar respostas satisfatórias, ou minimamente consensuais, a estas perguntas.

Duas escolas de pensamento dominaram as recentes discussões sobre a base neurobiológica de linguagem. Baseada em estudos de lingüística e psicologia do desenvolvimento, a escola *nativista* concluiu pela existência de mecanismos inatos específicos para a linguagem (Fodor 1976; Pinker 1994). Tais mecanismos envolveriam um grupo de genes exclusivo para a espécie humana, que se expressariam em um módulo cerebral especializado. Por outro lado, a escola *construtivista* nega a existência de genes ou módulos exclusivos para a linguagem, e tenta explicar dados lingüísticos e psicológicos em termos de um processo de aprendizagem, facilitado pela co-evolução da espécie humana e da prática lingüística. A competência lingüística se derivaria de

¹ Agradecimentos aos pareceristas da Revista Abstracta, à FAPESP e CNPQ, por apoio financeiro à minha pesquisa, e a Michael B. Wrigley (in memoriam), com quem aprendi Filosofia da Linguagem.

princípios gerais de plasticidade neuronal e de propriedades cognitivas do córtex pré-frontal dos primatas (Deacon 1997; Quartz & Sejnowski 1997).

Um aspecto central da corrente nativista contemporânea é o funcionalismo computacional, que se expressa na tese metodológica de que a cognição deve ser estudada "de cima para baixo" (abordagem "top-down"), o que significa que primeiramente deve-se identificar as funções da linguagem, no plano de uma abordagem psicolinguística, para depois se procurar as estruturas cerebrais que suportam tais funções. O modelo de cérebro adotado por estes autores - o qual, diga-se de passagem, é adequado à metodologia que adotam - tem sido o modelo de processamento "serial" de informação, por meio de módulos macroscópicos, concebidos como "caixas pretas", conectados através de canais de entrada e saída de informação. É assumido que cada módulo implementaria uma função lingüística, por exemplo, a função de representação lexical, das regras gramaticais, etc.

Será que o cérebro trabalha de modo semelhante a tal modelo? O apoio para esta proposta viria de resultados de escaneamentos (PET-scanning) e ressonância magnética (fMRI), mostrando um grau de especialização regional no cérebro, por exemplo para diferentes classes de palavras. Porém, há outras evidências que sugerem que o paradigma de computação modular não seria adequado para se entender o funcionamento do cérebro.

Medidas de atividade eletromagnética feitas por EEG (através de uma malha de eletrodos na superfície do escalpo) têm duas vantagens, relativamente às técnicas acima: melhor resolução temporal (no alcance de milissegundos, para EEG, e de segundos, para PET e fMRI); e a possibilidade de distinguir entre influxo (picos positivos) e efluxo de energia (picos negativos), relativamente a uma determinada região em que é feita a medida. Estudos de atividade lingüística monitorados por EEGs (Roy John et al. 1997; Llinas 1992) têm revelado um fenômeno que é contraditório com o paradigma de modularidade: a existência de mudanças de fase rostro-caudal rápidas. Este fenômeno implica que cada região, considerada pelos modularistas como estando continuamente ativada durante uma tarefa cognitiva (por exemplo, "guardando" uma palavra na memória), estaria sendo sucessivamente ativada e inativada cerca de 25 vezes por segundo. Este fenômeno parece ser uma condição necessária para o processamento

lingüístico, uma vez que alterações do padrão oscilatório parecem estar relacionadas a disfunções como a dislexia.

Portanto, a aparente localização das funções lingüísticas em regiões específicas do cérebro, sugerida pelas medidas em PET-scanners e fMRI, pode ser um artefato dessas metodologias, as quais, devido à sua baixa resolução temporal, trabalham com *médias* da atividade cerebral na escala de segundos. Quando a atividade cerebral é medida na escala de milissegundos, fica evidente que qualquer tarefa lingüística envolve uma ampla rede neuronal, incluindo o córtex pré-frontal, áreas parietais e temporais do córtex, e estruturas subcorticais como o cerebelo. Também a dissociação entre processos sintáticos e semânticos, um pressuposto bastante comum nas abordagens modularistas, pode ser mais uma questão de ativação diferencial de núcleos de uma mesma rede neuronal, que uma dissociação entre módulos diferentes.

Se o cérebro não tem módulos macroscópicos para executar as funções definidas em teorias lingüísticas, então como este processamento é realizado? O esboço de uma visão alternativa foi apresentado no livro de T. Deacon, *The Symbolic Species* (Deacon 1997), possivelmente a crítica mais contundente contra o paradigma de modularidade feita por um neurocientista cognitivo. O processamento lingüístico é realizado, de acordo com Deacon, através de amplas redes neuronais no cérebro humano, envolvendo as áreas classicamente estudadas (áreas de Broca e Wernicke), e outras áreas relacionadas à memória de trabalho e processos executivos. O construtivismo apresenta a vantagem, frente aos nativistas, de não necessitar a postulação de genes ou macro-módulos exclusivos para a linguagem. Por outro lado, relativamente às evidências psicolingüísticas, a visão nativista parece estar melhor armada para explicar a facilidade com que as crianças aprendem a linguagem.

Para se conciliar ambas as propostas, consideramos que a base neurobiológica de competência lingüística possa ser resumida em duas condições suficientes:

- (a) a existência e operação de um sistema combinatorial determinado geneticamente, que não seria exatamente a "linguagem do pensamento" proposta por Fodor (1976) mas um sistema bioquímico e multi-funcional, nos termos propostos por Black (1991) e;

- (b) certas mudanças na organização do cérebro (em especial, o controle do aparelho fonador pelo córtex motor), que teriam acontecido durante a evolução da espécie humana, sendo selecionadas positivamente e herdadas pelos atuais membros da espécie.

Nativistas e construtivistas concordam que a linguagem tem as características de um sistema combinatorial. Por exemplo, Jackendoff (1994) convincentemente discutiu, de um ponto de vista chomskyano, que a linguagem seria basicamente um "sistema generativo". Deacon (1997) propôs que a linguagem envolveria uma dimensão de *referência simbólica*, ou seja, os símbolos lingüísticos formariam um sistema combinatorial fechado onde os componentes se articulam entre si, sendo que apenas uma *totalidade* de símbolos se referiria a objetos e processos no ambiente. Tal sistema de comunicação difere de outras formas, não-combinatoriais, de comunicação animal, como as de tipo icônico, em que os sinais individuais não se articulam entre si, e se referem diretamente a objetos presentes no ambiente.

O aspecto combinatorial da linguagem pode se apoiar em um sistema de sinais eletroquímicos, operantes na sinapse e nos canais de transdução de sinais, que não é exclusivo aos humanos. Recentes estudos têm revelado que a complexidade algorítmica das cadeias de sinais codificadas em macromoléculas biológicas é comparável à complexidade da linguagem humana (Popov, Segal & Trifonov 1996). Rocha (1997) mostrou, através da utilização de lógica nebulosa, a possibilidade de derivar do sistema combinatorial de sinais celulares uma linguagem formal com a mesma complexidade das linguagens naturais. Podemos ainda constatar que as operações envolvidas no processamento fonético, fonológico e sintático são semelhantes à “estratégia combinatorial” (Black 1991) usada em processos sinápticos e de transdução de sinais no domínio intra-neuronal. Para a expressão desta atividade combinatorial em termos do comportamento lingüístico observado nos seres humanos, sem dúvida existem vários estágios de processamento que ainda não foram elucidados pela neurociência. Podemos imaginar que seria preciso um tipo adequado de arquitetura das redes neuronais, permitindo que a dinâmica combinatorial bioquímica, em diversas populações de neurônios dedicadas aos diversos aspectos do processamento lingüístico, possa ter influência em outras redes que diretamente controlam o comportamento.

A estrutura em múltiplas camadas do neocórtex e do cerebelo, assim como a dinâmica neocortical, controlada pela formação reticular e tálamo, parecem satisfazer às condições para a expressão das operações combinatoriais que ocorrem no nível bioquímico. Nas espécies não humanas, tais operações combinatoriais dariam suporte a diversos tipos de comportamento inteligente, inclusive, em chimpanzés, a processamentos genuinamente lingüísticos por meio de sinais manuais. Tendo em vista a generalidade biológica dos sistemas combinatoriais, podemos inferir que a especificidade humana frente à linguagem não decorreria do primeiro, mas sim do segundo aspecto acima apontado, a saber, de mudanças estruturais que tornaram possível a organização hierárquica e o controle voluntário da fonação e a expressão vocal de cadeias articuladas de sinais, indo além das restrições da comunicação vocal icônica, à qual as outras espécies biológicas estão restritas.

O aparecimento do comportamento lingüístico em humanos, a partir da vocalização articulada, poderia então ser explicado em termos de mudanças na organização de cérebro, que permitiram *ao sistema combinatorial preexistente* controlar a produção da vocalização. Nesta perspectiva, a natureza da linguagem humana residiria no sistema combinatorial bioquímico do cérebro enquanto combinado com uma série de mecanismos cerebrais típicos da evolução humana, que possibilitaram a organização hierárquica e a expressão vocal das seqüências geradas por tal sistema.

2. Uma ordem plausível da evolução da linguagem

A hipótese de que a linguagem estaria baseada em um mecanismo combinatorial geral conduz ao problema de se explicar por que outras espécies não desenvolveram a comunicação lingüística ao grau em que nós o fizemos. Os tecidos corticais, em outras espécies de mamíferos, têm uma arquitetura semelhante à encontrada na espécie humana, e além disso certas espécies, inclusive não-mamíferos (como o papagaio) têm poderoso aparato fonético. A explicação de singularidade humana relativamente à linguagem então requer referência a mudanças organizacionais que aconteceram durante a evolução humana, mas não em outras espécies. Tais mudanças permitiram a hierarquização e expressão externa de sucessões de sinais produzidas pelo sistema combinatorial interno, e, deste modo, o seu uso para a comunicação com outros indivíduos.

Com base em estudos sobre a evolução humana, podemos conjecturar que as mudanças cruciais começariam com o controle cortical de vocalização, que possivelmente decorreu de eventos deflagrados pela adoção de bipedalismo. Sabe-se que os primatas não humanos têm um controle sub-cortical de suas vocalizações (Deacon 1997), centrada na área periaqueductal grisada. Estudos neste sentido têm mostrado que o substrato da vocalização não é uma área discreta e pequena, mas um sistema extenso que alcança desde a medula, passando por áreas subcorticais, até o sistema límbico (Jurgens 1998). Conseqüentemente, a comunicação vocal nestes animais está relacionada a respostas "instintivas" a estímulos, e expressão de estados emocionais. O aparecimento de conexões eferentes, do córtex motor (responsável pelo desencadeamento das ações voluntárias) para o aparato fonético, permitiu aos humanos estender a vocalização para outras funções comunicativas, como a expressão de raciocínios e o planejamento de ações coletivas.

Simultaneamente ao controle cortical das vocalizações, ocorreram modificações no aparato fonético, especialmente a descida da laringe (veja-se Deacon 1997), dotando o aparato fonético de maior grau de liberdade, e conseqüentemente o tornando capaz da produção de seqüências articuladas de gestos vocais. Ao longo da evolução da linguagem humana, também se atingiu a organização *hierárquica* das seqüências de gestos vocais, caracterizando a gramática. Não devemos colocar a gramática no rol dos eventos primitivos geradores da linguagem, uma vez que vários desenvolvimentos na comunicação vocal de nossos antepassados podem ter ocorrido antes da sintaxe ter se cristalizado na forma encontrada nas línguas atuais.

Finalmente, devemos também considerar o aparecimento da capacidade de monitoramento da fala em tempo real, pelo cerebelo. Tal monitoramento é necessário para que o falante possa corrigir cada comando enviado pelo córtex motor, relativamente ao estado prévio do aparato fonético. Tal necessidade deriva do alto grau de liberdade do aparato, possivelmente o maior entre todos os sistemas musculares dos animais. Esta função nova do cerebelo pode ter sido induzida pela adoção da postura vertical, acompanhando a mudança para a forma bípede de locomoção, envolvendo também o sistema vestibular e os gânglios de base.

Todas essas mudanças organizacionais seriam geneticamente transmissíveis, tendo como conseqüência que as novas gerações teriam sido cada vez mais aptas à

comunicação por meio da linguagem. À medida em que esse padrão de comunicação se tornou um padrão seletivo na sociedade humana, várias outras especializações cerebrais foram induzidas, em especial as especializações no córtex auditivo, que nos permitem distinguir a fala em meio a outros tipos de sons, e a reconhecer formas complexas de articulação lingüística.

Entender a origem e evolução da linguagem é uma tarefa complexa, que requer contribuições de diversas áreas científicas. O diálogo transdisciplinar, que pode conduzir ao progresso do entendimento, é frequentemente difícil, em parte por causa de uma discordância sobre o que deveria ser explicado. Nesta situação, uma estratégia efetiva pode ser a de se restringir a explicação a um aspecto da linguagem, considerado como o mais primitivo, que teria deflagrado o processo evolutivo que conduziu a todos os demais. Seguindo este método, me concentrarei na discussão de possíveis fatores evolutivos que tornaram possível a *produção da fala*, que teria sido o passo crucial na origem da linguagem.

Discutindo as origens de linguagem, devemos fazer uma distinção entre *capacidades mentais e produção de fala*. Podemos assumir que o surgimento da fala foi possivelmente precedido pela capacidade de formação de abstrações, as quais poderiam eventualmente ser comunicadas por meio de ícones, como pinturas rupestres. É concebível que nossos antepassados, como também outras espécies que têm um córtex pré-frontal desenvolvido, possuam capacidades de abstração; porém, estes indivíduos não *falam*. Nos raciocínios que se seguem, faço portanto a *suposição* de que a fala, e não simplesmente a formação de conteúdos mentais abstratos, seria o fator crucial que teria disparado o processo evolutivo no qual a espécie humana veio a se distinguir de nossos antecessores primatas.

Assumo que antes que nossos ancestrais começassem a falar eles poderiam ter formado mensagens abstratas em seus cérebros/mentes, sem contudo possuir um meio adequado de comunicá-las (exceto ícones, os quais, apesar de atingirem um certo grau de abstração, seriam insuficientes para especificar tais mensagens). Esta suposição pode, sem dúvida, ser contestada por aqueles que defendem a anterioridade da linguagem relativamente ao pensamento; para estes, a formação de determinados tipos de conteúdo mental abstrato só seria possível após o surgimento da linguagem falada. Contudo, não é minha intenção polemizar com esta posição teórico-filosófica. O

objetivo, mais modesto, é apenas de se encontrar um fio condutor para a discussão da emergência evolutiva da linguagem em sua forma propriamente humana.

Em tal situação - existência de pensamento abstrato sem a possibilidade de sua comunicação pela fala - o desenvolvimento dos sistemas cerebrais especializados em decodificar seqüências estruturadas de sinais, e lhes atribuir um significado, era simplesmente impossível, pois tais estímulos não existiam no ambiente. Só depois que seqüências organizadas de gestos vocais começaram a ser produzidas, as especializações para uma correspondente atribuição semântica puderam se desenvolver. Deste modo, podemos conceber que o desenvolvimento da fala teria sido anterior ao desenvolvimento dos mecanismos de decodificação semântica da fala, pois a produção de seqüências articuladas de sons vocais já era possível mesmo quando os ouvintes ainda não estavam preparados atribuir um significado adequado a elas, ao passo que a atribuição de tais significados não poderia ocorrer na ausência de tais seqüências.

A precedência da produção da fala pode ser ilustrada pela observação de um fenômeno que ocorre em nível ontogenético, a saber, o período em que as crianças 'balbuciam' (do inglês 'babbling'). Este fenômeno revela que os humanos podem produzir seqüências (relativamente) longas e estruturadas de gestos vocais, sem atribuir um significado lingüístico adequado a elas (veja-se o estudo de Greenfield 1991). Esta observação sugere que os humanos poderiam ter iniciado sua atividade lingüística falando sem (muito) sentido, abrindo o caminho para que mensagem complexas pudessem vir a ser transmitidas e adequadamente interpretadas. A conseqüência lógica desse argumento é que uma explicação satisfatória das origens de linguagem deveria começar pela discussão de como a produção de fala se tornou possível, pois só depois que alguém começou a falar os ouvintes passaram a dispor de estímulo para desenvolver suas capacidades lingüísticas de ordem semântica.

Portanto, faço a suposição de que só se poderia caracterizar o comportamento humano propriamente lingüístico a partir da emergência da fala, esteja esta precedida ou não de pensamento abstrato (como argumentei acima, acredito que a primeira opção seja mais aceitável). Nas próximas seções, minha atenção se volta para a identificação de possíveis eventos evolutivos que propiciaram tal emergência.

3. Há uma causa comum para as mudanças organizacionais que geraram o aparecimento de fala?

Como assinali anteriormente, duas mudanças organizacionais parecem estar envolvidas na origem da fala: o controle cortical de vocalização, e a descida da laringe. Como estes dois eventos, ambos necessários para a fala em sua forma humana, teriam acontecidos juntos? Uma explicação satisfatória, que não dependa de coincidências evolutivas, requer o recurso a uma causa comum.

A causa comum seria possivelmente a adoção de bipedalismo, 5 milhões de anos atrás (o papel de bipedalismo na evolução para a linguagem foi discutido por Corballis 1991). Com o advento do bipedalismo, estruturas corticais (no córtex premotor e motor) especializadas para a coordenação da locomoção em árvores por meio de movimentos de braços e mãos, perderam sua função original, podendo ter sido redirecionadas a uma função nova. A área do córtex motor correspondente à atual área de Broca, como também os mecanismos de sequenciamento hierárquico do córtex pré-frontal, provavelmente tinham se desenvolvido para dar suporte ao comportamento de escalar árvores e se mover de um ramo a outro. Com a adoção de bipedalismo, parte desta rede ficou inutilizada, e possivelmente foi reutilizada em uma função nova, com uma natureza semelhante: a especificação de seqüências organizadas de movimentos, não mais dos braços e mãos, mas *dos músculos fonadores*.

O aparecimento do trato piramidal descendente do córtex motor para os músculos fonadores seria, nesta hipótese, o movimento crucial que permitiu o controle voluntário de fala. Um "deslocamento" (Deacon 1997), ou reestruturação plástica das conexões axonais, pode ter possibilitado o controle cortical da vocalização ao redor 2 milhões de anos atrás. Como esta mudança organizacional foi geneticamente transmissível, o processo teria sido do tipo neodarwiniano, i.e., inicialmente gerado por mutações nucleares ou recombinações genéticas devido a trocas sexuais em uma população, e posteriormente selecionado positivamente de acordo com o caráter adaptativo que o novo traço conferiu a seus portadores.

O deslocamento anatômico de axônios teria estabelecido a base para uma função nova, a vocalização voluntária, sendo que posteriormente tal função poderia ter passado por um novo processo evolutivo. O controle de movimentos de braços e mãos obedece a uma dinâmica temporal que é mais imprecisa que a dinâmica temporal da coordenação

dos músculos vocais. É possível que a participação de estruturas sub-corticais desenvolvidas no suporte à locomoção bipedal (cerebelo, ganglios de base e sistema vestibular) tenham permitido a evolução do mecanismo de sequenciamento de movimentos de locomoção para o sequenciamento de gestos vocais, permitindo um controle mais fino (isto é, com definições temporais mais precisas) da vocalização.

Nesta visão, o aparecimento da área de Broca, como um subsistema especializado que envia axônios ao aparato fonético, se explicaria pela consideração da estrutura prévia do córtex motor, e dos eventos evolutivos que poderiam ter induzido uma mudança na sua função. A área motora que controla movimentos do braço e mão direitos, no cérebro humano contemporâneo, está próxima à área de Broca, no hemisfério cerebral esquerdo. Greenfield (1991, 1998) discutiu convincentemente que essas áreas são homólogas (i.e., derivam da mesma estrutura) nos níveis ontogenético e filogenético. Uma evidência da conexão íntima da área de coordenação da mão com a área de Broca é a facilidade de se rediregir as funções lingüísticas da área de Broca para a mão, na aprendizagem da linguagem escrita ou da linguagem de sinais. Recente pesquisa (Rizzolatti & Arbib 1998) sobre os chamados "neurônios especulares" (do inglês *mirror neurons*), localizados no córtex pré-motor ventral do macaco, apóia a hipótese de que a área F5, que controla o movimento da mão desses animais e também responde quando o animal observa movimentos semelhantes feitos por terceiros, seria o homólogo evolutivo da área de Broca.

Outra evidência de que parte da área de controle do movimento de braço e mão, em nossos ancestrais, foi redirecionada ao controle da musculatura vocal, nos humanos, é o processo de especialização hemisférica. O processo de especialização lateralizada do cérebro humano, tendo afetado tanto a área de Broca quanto as áreas de controle do braço e mão, sugere que tais processos não teriam ocorrido de forma independente. Esta idéia é apoiada por evidências de simultaneidade da destreza manual dos fabricantes de ferramentas de pedra, ao redor de 2 milhões de anos atrás, e vestígios do surgimento da área de Broca em um crânio do mesmo período. Analisando um fóssil de aproximadamente 2 milhões de anos de idade, nomeado "ER1470", Dean Falk (1992) identificou rastros da área de Broca no hemisfério esquerdo. Como este fóssil é presumivelmente contemporâneo dos fabricantes destros das ferramentas de pedra na África Oriental, ela notou que "como a área de Broca e a parte do cérebro que controla a

mão direita estão bem próximas uma da outra, não é surpreendente que sinais de habilidades de linguagem e destreza apareçam simultaneamente no registro fóssil" (Falk 1992: 170). Outra evidência é provida por Reynolds (1993), que comparou construções humanas intituladas (no inglês) "polyoliths" (estruturas que têm características combinatoriais) com arranjos semelhantes de símbolos na linguagem.

Contudo, o aparecimento do controle cortical da vocalização não seria suficiente para a produção da fala, porque o aparato fonético de nossos antecessores não poderia implementar os comandos para a produção de seqüências mais longas e complexas de gestos. Como poderíamos explicar que as mudanças do aparato fonético tenham acontecido ao mesmo tempo que o controle cortical, tornando a produção da fala possível? Assumindo que ambos os eventos têm o bipedalismo como a causa comum distante, como então a adoção do bipedalismo teria induzido mudanças organizacionais no aparato fonético, mais especificamente a descida da laringe?

Uma primeira hipótese seria que as mesmas mudanças genéticas que geraram o deslocamento de conexões axonais do córtex motor para o aparato fonético teriam produzido mudanças estruturais no aparato. Porém, se é o mesmo grupo de genes que controla ambos os fenômenos, o aparecimento de conexões do córtex motor para os músculos vocais, em crianças recém-nascidas, deveria ser sincronizado à descida da laringe. Como o aparecimento de tais conexões precede o nascimento (i.e., o bebê recém-nascido já possui o controle voluntário de suas vocalizações) mas a descida da laringe acontece depois de nascimento, os dois fenômenos seriam possivelmente controlados através de grupos de genes diferentes.

Uma segunda hipótese seria que a descida da laringe seria uma conseqüência anatômica de outras mudanças evolutivas, independentes do controle cortical de vocalização. Este tipo de hipótese foi apresentado por Wills (1993) e Deacon (1997). Wills propõe que "esta liberdade de movimento [da laringe - APJ] foi...um dos muitos resultados de nossa aquisição da postura vertical, que colocou a cabeça mais para trás relativamente à coluna espinal. A parte frontal da garganta se prolongou, permitindo à laringe crescer mais e operar com maior mobilidade" (1993: 156-7). Estudos de Deacon sugerem que "a redução relativa da boca e face, e expansão relativa do crânio na evolução humana, resultaram em uma descida da laringe...e amplificação da faringe, e aumento no papel da língua modulando a forma das cavidades orais" (1997: 356).

O problema com estas hipóteses é que elas não dão uma explicação satisfatória de como uma mudança de postura ou no tamanho da boca, face e crânio poderiam ser suficientes determinar a posição da laringe. O tamanho e posição relativa de cada estrutura anatômica são controlados por fatores genéticos, e não pelo espaço disponível para crescimento ou movimento, o qual só pode ter uma influência limitada durante o desenvolvimento embriológico. As hipóteses apresentadas por Wills e Deacon parecem assumir que o tamanho e posição de algumas estruturas anatômicas pudessem ser determinadas apenas pelas mudanças em outras estruturas, como se o processo embriológico fosse um jogo mecânico de encaixe de peças.

Uma terceira hipótese é que a descida da laringe seria determinada por um grupo específico de genes, independentemente dos genes que determinam o controle cortical de músculos vocais. Esta hipótese pode ser combinada com a proposta geral de Deacon, de que a evolução da linguagem envolveria uma seleção de tipo *baldwiniana*. Segundo esta proposta, seria plausível que a existência de *uma prática cultural repetida* ao longo de gerações, em uma população biológica, afete o genótipo progressivamente, e selecione positivamente os genes que facilitam o desempenho de tal prática.

A adoção do bipedalismo implicou na adoção de uma postura vertical para a locomoção, e a liberação correspondente de braços e mãos, facilitando o uso de ferramentas e novos tipos de interação social. Neste contexto, habilidades comunicativas se tornaram um fator positivo de seleção, criando um parâmetro seletivo novo. Os indivíduos que apresentavam maiores habilidades articulatórias fonéticas (devido a um genoma que produzia uma laringe na nova posição) teriam sido selecionados positivamente, conduzido a população para genomas geradores de aparatos fonéticos progressivamente mais adequados. Nesta visão, o bipedalismo pode ser considerado a causa distante que favorece um novo tipo de interação social, vindo a selecionar os indivíduos geneticamente capazes de produzir uma faixa mais ampla de vocalizações articuladas. Tais mudanças organizacionais teriam sido provavelmente acompanhadas de outras mudanças genéticas, sugeridas pelo fato de que o sistema vestibular, os ganglios de base e o cerebelo foram envolvidos tanto em funções motoras quanto no monitoramento da vocalização.

4. Aspectos da abordagem neurocognitiva das linguagens naturais

A expressão "linguagem natural" se refere ao domínio do comportamento lingüístico dos humanos. Tal comportamento é constituído por ações chamadas de "utterances" (ou "proferimentos", em português). Estas podem ser analisadas em três aspectos inter-relacionados: *sintaxe*, *semântica* e *pragmática*.

A relação entre sintaxe e mecanismos neurobiológicos foi discutida no contexto das teorias nativistas. Chomsky e seus seguidores discutiram convincentemente que a capacidade de gerar sentenças novas e gramaticalmente corretas, exibida por crianças, implicaria na existência de um mecanismo biológico inato. Que tipo de mecanismo biológico seria este? Os proponentes do paradigma nativista, trabalhando no contexto das ciências humanas, não puderam identificar a base neurobiológica de linguagem, se limitando a referências vagas à "biologia" e ao "instinto" (como em Pinker 1994).

Poderíamos considerar as capacidades lingüísticas, inclusive as características da "gramática universal" chomskiana, como resultantes do 'modus operandi' do sistema combinatorial eletroquímico do cérebro, adicionado de um grupo de características organizacionais geneticamente transmissíveis, típicas da espécie humana? A existência de um sistema combinatorial interno, e de mecanismos que permitem sua manifestação externa, em termos de processos de comunicação com outros indivíduos, pode favorecer a hipótese anti-chomskyana da existência de aprendizagem implícita da linguagem por crianças pré-escolares. Deacon (1997) argumentou que as crianças podem aprender a linguagem sem supervisão ou treinamento especial, porque o cérebro humano e os sistemas de suporte da linguagem natural co-evoluíram juntos, ao ponto de princípios da gramática refletirem o próprio modo de funcionamento do cérebro.

Uma possível derivação da hipótese de Deacon seria que a gramática não precisaria ser inata (no sentido de uma estrutura gerada por um grupo específico de genes), mas poderia simplesmente ser isomórfica ao modo de trabalho do sistema combinatorial eletroquímico do cérebro. Tal hipótese é semelhante a uma sugestão feita por Piaget (em Piatelli-Palmarini 1980), com relação à existência de um 'bias' para a aprendizagem de física por crianças; ele sustentou que há uma tendência favorável porque as leis físicas afetam nossos próprios corpos. Analogamente, há um 'bias' que favorece a aprendizagem da linguagem por crianças, pois tal linguagem trabalharia da mesma forma que o sistema eletroquímico combinatorial do cérebro.

O isomorfismo entre combinatória eletroquímica e gramática das linguagens naturais constitui uma possível área de estudo. Ligações entre palavras têm uma estrutura semelhante à ligação entre proteínas, e a ordem das palavras nas sentenças tem estrutura semelhante às seqüências de reações metabólicas. Proteínas têm sítios que permitem sua ligação com efetores e substratos: sua estereoespecificidade pode ser comparada ao papel dos verbos nas orações. Verbos se ligam com sujeitos e predicados, e também ligam duas ou mais proposições, produzindo sentenças mais longas.

Uma visão nativista do significado dos termos lingüísticos foi proposta por Fodor (1980), que considerou a *linguagem do pensamento* como sendo responsável tanto pela estrutura gramatical quanto pela semântica. Uma crítica desta visão foi apresentada por Jackendoff (1994), reivindicando que a existência de um sistema combinatorial geneticamente determinado na espécie humana, incluindo estruturas sintáticas (ou instruções para o desenvolvimento de tal estrutura), não implicaria na existência de *significados* ou *conceitos inatos*. Uma possível interpretação dessa crítica seria que o sistema geneticamente baseado define *um universo praticamente infinito de possíveis combinações de sinais*, mas não os *significados atuais* destes sinais. A fixação de significados por um indivíduo emergiria como resultado das suas experiências e ações. Tal papel da experiência é assumido também na filosofia da lingüística por Lakoff (1987) e Johnson (1987).

A questão principal aqui não diz respeito à capacidade geral de avaliação semântica, que realmente parece ser prevista geneticamente, mas o critério para se atribuir significados específicos, ou melhor, o critério de *estabilização* do significado. Este pode ser entendido da seguinte maneira: embora uma criança esteja geneticamente preparada para ouvir e vocalizar sons articulados, e também para atribuir significado aos mesmos, *só quando ela puder estimar o significado que outras pessoas atribuem aos sons, ela poderá estabilizar os significados de suas próprias representações lingüísticas*. Uma semântica estável, segundo este critério, seria um produto conjunto da atividade de cérebros semelhantes em corpos semelhantes.

Portanto, embora os indivíduos possam *gerar* significados de modo privado, seria a *interação social* que *estabilizaria os significados* atribuídos individualmente. Desse modo, os aspectos privado e público da linguagem se harmonizariam e complementariam. Um exemplo disso seria o uso de dicionários; a constituição destes

compêndios se funda no uso social das palavras, porém o uso destas regularidades sociais não impede a emergência de novos termos e/ou significados gerados por indivíduos criativos. Por outro lado, significados privados são voláteis e só se estabilizam quando adquirem dimensão social (como no caso dos neologismos e gírias).

A estabilidade de significado das sentenças também requer interação social, pois o sistema combinatorial dos cérebros individuais permite inúmeras interpretações diferentes de regras gramaticais comuns. Assim sendo, a estabilidade observada no uso de determinadas regras, em um grupo social, seria o fruto de hábitos culturais que estabilizaram as interpretações. Portanto, sistemas lingüísticos privados seriam possíveis, mas permaneceriam idiossincráticos e altamente instáveis enquanto não atingem a dimensão da comunicação social.

A abordagem semântica de Deacon (1997) se baseia em um conceito particular de *símbolo*. Além do uso comum do termo, relacionado à relação convencional entre um sinal e seu significado, ele também se refere à distinção peirceana entre ícone, índice e símbolo. Na sua interpretação, a distinção entre estes conceitos diz respeito a tipos diferentes de relação entre sinais e o seus referentes. O conceito chave é o de *referência simbólica*, considerada como o tipo de operação onde um sinal se refere a outro sinal que pertence a um sistema de sinais. Portanto, o símbolo é concebido como um *sinal de segunda ordem*, que se refere a um sistema de sinais, ao invés de se referir diretamente a um objeto ou processo físico externos. O cérebro humano, nesta perspectiva, passaria por um processo de desenvolvimento adequado para a aprendizagem de tais operações simbólicas, o que explicaria a facilidade do aprendizado da linguagem pelas crianças, sem o concurso de qualquer estrutura lingüística inata.

Não desconsiderando a validade da visão de Deacon, gostaria de notar que o caráter simbólico da linguagem natural também se deve a um outro aspecto, não tematizado por ele. Se fôssemos seguir apenas o seu critério, as representações eletroquímicas do cérebro também deveriam ser consideradas simbólicas, uma vez que sinais em uma parte do cérebro (p. ex., córtex pré-frontal) podem se referir a sistemas de sinais em outra parte do cérebro (p. ex., córtex parietal). Uma característica fundamental das redes neuronais é a de funcionar como um sistema, onde a atividade de cada parte é relativa à atividade das outras partes, mas isto não implica que tais redes

operem em nível simbólico. Qual seria, então, o fator determinante para se atribuir a um sistema de sinais o caráter de simbólico?

De acordo com o critério da estabilização do significado apresentado anteriormente, o caráter de simbólico se aplicaria aos sistemas de sinais que se situam no domínio da *comunicação social*. Sinais internos ao cérebro são restritos à economia deste cérebro, ou seja, à comunicação entre partes deste cérebro, e não à comunicação entre pessoas. A relação convencional entre sinais e significados, em um linguagem natural, implicaria portanto a existência de interação social entre indivíduos. Em uma sociedade, a dinâmica de cada cérebro afeta a dinâmica de outros cérebros, por meio da comunicação, e a linguagem natural é o modo mais poderoso de fazer isto (alguns outros são discutidos por Freeman 1995).

Como o aparecimento da linguagem criou um contexto social novo, que teria operado ao longo da evolução como um parâmetro restritivo no processo seletivo de tipo baldwiniano (Deacon 1997), um dos resultados deste processo é que o contexto comunicacional se tornou determinante da semântica das linguagens naturais. Conseqüentemente, não é possível estabelecer uma correspondência unívoca entre regras e significados lingüísticos, por um lado, e eventos no cérebro de um único indivíduo, de outro. A razão é a mesma por que não é possível existir uma partida de futebol jogada por só uma pessoa; os jogos de linguagem, como já havia sido apontado por Wittgenstein, requerem um contexto público. De acordo com a análise aqui desenvolvida, isto se deve ao fato de que apenas na interação entre diversos indivíduos é possível alcançar a estabilidade semântica e gramatical observada nas linguagens naturais.

A existência de uma dimensão pragmática da linguagem pode prover uma explicação da impossibilidade de uma linguagem privada estável. A influência das intenções dos falantes e ouvintes no significado lingüístico foi proposta, originalmente, no contexto de filosofia de linguagem, por Grice (1957). Entendo aqui estas intenções como *sub-metas* individuais, ou seja, se o falante e ouvinte tiverem êxito na interpretação das intenções recíprocas, estabelece-se uma comunidade de significação como uma sub-meta realizada que pode ser um passo para o alcance das metas principais de ambos.

Em contextos sociais, onde as linguagens naturais existem, a realização das metas de um indivíduo interfere com a realização das metas de outros indivíduos. Quando um indivíduo fala com outro, suas intenções relativas ao ouvinte correspondem a sub-metas que podem contribuir para a realização de sua meta, isto é, se o ouvinte interpretar sua fala do modo intencionado, este fato aumenta sua probabilidade de obter a cooperação do ouvinte para alcançar a meta almejada. O ouvinte tentará atribuir à fala o que ele acredita ser o significado intencional do falante, caso ele também tenha intenções (submetas) com relação ao locutor. Quando a maioria dos falantes e ouvintes age deste modo, um "pacto" é estabelecido, o qual estabiliza modos de atribuir significado à expressão vocal dos pensamentos, em uma coletividade humana.

Evidentemente tais interações sociais, correspondentes à dimensão pragmática da linguagem, têm influência sobre o processamento cerebral dos indivíduos envolvidos no processo de comunicação. Enquanto se realiza uma operação na memória de trabalho (p. ex., uma pessoa pensa silenciosamente: "as melhores coisas da vida são grátis") o envolvimento cerebral pode não ir além da ativação do mecanismo de "fala interna" (*inner speech*). Porém, quando se profere uma sentença, isto é, quando se fala algo para alguém, em um determinado contexto e com uma determinada intenção (por exemplo, uma cantora que enuncia a frase acima, que é o título de uma canção, para outros músicos no palco, com a intenção de cantar a canção) além da ativação de comandos comportamentais dos sistemas premotor e motor, ocorre também um aumento de atividade no sistema executivo, para a inferência dos *efeitos* que o proferimento pode ter, com relação ao contexto onde é feito, e às metas do falante. Esta implicação sugere a oportunidade de um diálogo entre estudos lingüísticos na área de pragmática, e estudos neurocognitivos de funções executivas.

5. A superação do "binarismo semântico", e a aprendizagem da linguagem

Na abordagem aqui adotada, assumi que as regras usadas para a construção gramatical e atribuição semântica tenham suas características atuais derivadas dos eventos evolutivos que conduziram ao aparecimento de fala. Mecanismos subjacentes à fala humana parecem ser evolutivamente relacionados com circuitos do sistema motor especializados para o sequenciamento de movimentos binários. Assim, uma pergunta interessante se apresenta: seria possível superar, de modo consistente, os esquemas binários que

presumivelmente constituem a base de nossas capacidades lingüísticas? Mais precisamente, se os esquemas subjacentes às nossas capacidades lingüísticas derivaram de esquemas motores binários, teria o pensamento humano superado os limites do binarismo?

O *binarismo semântico* - entendido como um esquema que considera haver disjunção exclusiva entre todos os pares de termos lingüísticos opostos - não parece ser essencial para o pensamento humano; pelo contrário, em alguns casos parece ser um hábito que impõe limitações desnecessárias. A história de filosofia revela uma longa luta contra os limites do binarismo semântico. Heráclito e Aristóteles tentaram expressar o movimento e as transformações que ocorrem na natureza, para tal desenvolvendo formas de pensamento - a teoria da potência e do ato - que superassem o binarismo entre ser/não ser de Parmênides; Kant teve que desafiar o binarismo para encontrar uma solução para as antinomias da razão pura; Hegel desenvolveu uma epistemologia dialética baseada em processos de superação de oposições binárias, e Marx usou a concepção dialética de Hegel para expressar processos históricos da sociedade humana.

Mais recentemente, lógicas chamadas *não-clássicas* têm sido desenvolvidas - como a lógica paraconsistente e a polivalente - para se enfrentar uma diversidade de situações nas quais um pensamento binarista parece não ser adequado, como as dificuldades de interpretação da teoria quântica e problemas psicológicos como o auto-engano, em que um sujeito apresenta idéias auto-contraditórias formando um todo coerente. Estes esforços de superação do binarismo podem ser entendidos como uma auto-organização do pensamento humano, procurando ir além das limitações cognitivas derivadas da origem evolutiva de suas capacidades lingüísticas.

A suposição de que o pensamento humano teria superado a origem binária, evoluindo para formas dialéticas de operação, pode ajudar a resolver algumas dificuldades do construtivismo com relação à aprendizagem da linguagem por crianças. Embora de uma perspectiva neurocognitivista os construtivistas tenham apresentado muitas vantagens explicativas frente ao nativismo, há alguns aspectos de suas teorias da linguagem que parecem ser vulneráveis a um contra-ataque. Proponentes de construtivismo (Deacon 1997; Quartz & Sejnowski 1997) tentam explicar capacidades lingüísticas humanas em termos de processos de aprendizagem que se beneficiam de

certas predisposições, mas não está claro como tais predisposições não seriam derivadas de mecanismos genéticos.

Como os fatores genéticos e ambientais estão inextricavelmente emaranhados no desenvolvimento de um organismo, teorias que assumem sua independência (seja em nível epistemológico e/ou ontológico) se apóiam em bases fracas. Uma solução para este problema seria, em vez de distinguir entre genético (inato) e aprendido, fazer uma outra distinção, entre *componentes elementares* dos comportamentos, que são de origem genética, e as *combinações complexas* de tais elementos básicos, que sempre se formam em qualquer comportamento humano. Todos os aspectos do comportamento lingüístico, nesta visão, são um produto comum de elementos básicos de origem genética, e combinações de tais elementos básicos induzidas pela aprendizagem. Características do comportamento lingüístico que só dependam dos elementos básicos (p. ex., a capacidade fisiológica de articular um determinado fonema) são insuficientes para explicar qualquer evento lingüístico.

Uma fraqueza do construtivismo - com a possível exceção da proposta de Deacon - tem sido sua dificuldade em explicar a universalidade da estrutura gramatical ("universalidade" no sentido de ser comum a todas as linguagens naturais; não no sentido mais fraco, de ser comum a todos os indivíduos de uma espécie biológica). Contudo, construtivistas podem assumir a existência de uma herança genética comum que suporta uma gramática universal, sem assumir que tal herança seria suficiente para se dar conta do comportamento lingüístico. Mais precisamente, tal mecanismo genético seria importante para a explicação das predisposições favoráveis ao aprendizado da linguagem, mas não suficiente para explicar qualquer comportamento lingüístico em particular.

No quadro conceitual proposto por Chomsky, há uma separação estanque entre competência e performance lingüísticas, que é semelhante à distinção entre as regras estruturais internas ao cérebro (linguagem-I) e as regras da linguagem enquanto fato social (linguagem-E). Proponho, aqui, uma perspectiva teórica mais integradora, em que tais dimensões da linguagem não seriam estanques, mas interagiriam dinamicamente entre si. Uma solução de compatibilidade pode ser alcançada por intermédio de uma concepção dialética do processo de aprendizagem. Se a competência lingüística herdada for descrita formalmente por um sistema paraconsistente (ou seja, um sistema que

permita a existência de contradições, dentro de certos limites, sem que isso redunde em sua trivialização) é possível gerar (isto é, deduzir), a partir do mesmo, tanto sentenças gramaticalmente corretas quanto incorretas, sem que ocorra uma inconsistência geral que conduzisse tal processo lógico a uma situação de trivialidade.

Para entender como a competência lingüística de origem genética se encadeia com capacidades aprendidas, é preciso entender como a operação de se gerar seqüências de letras e palavras espontaneamente se encadeia com a operação pela qual sua correção semântica ou gramatical é determinada. A primeira capacidade segue uma predisposição genética para se gerar novas construções, sejam elas consistentes ou inconsistentes. Esta capacidade seria possivelmente melhor descrita como um sistema que obedece a uma lógica paraconsistente ou nebulosa, pois em termos da lógica clássica - implícita na maioria dos atuais modelos da competência lingüística - se gerariam apenas as construções corretas, o que implicaria na existência de um conhecimento inato dos padrões culturais de uma língua, o que seria inaceitável.

Em crianças, tal processo aconteceria durante o período do balbuciar (*babbling*), quando expressões vocais gramaticalmente e/ou semanticamente incorretas são freqüentemente observadas. Proponho, portanto, diferentemente do que foi defendido pela escola chomskiana, que a competência lingüística possibilitaria gerar *tanto as construções corretas quanto as incorretas*. Esta seria a razão por que é "universal". A segunda capacidade, de avaliação da correção gramatical e semântica, seria relativa à linguagem natural específica do ambiente da criança, sendo portanto *aprendida*.

Quanto aos critérios para se considerar uma construção lingüística correta ou incorreta, é preciso levar em consideração tanto as regras internas quanto as externas, uma vez que ambas, na prática, se ajustam reciprocamente, por meio de interações complexas que se estabelecem entre os indivíduos biológicos e a cultura de sua sociedade. Portanto, a competência lingüística universal seria uma entidade logicamente paraconsistente ou nebulosa, para que possa se ajustar, em suas expressões, a diferentes culturas, que muitas vezes utilizam regras incompatíveis entre si.

6. Conclusão

Neste ensaio, procurei combinar uma discussão sobre a natureza da linguagem, a partir dos paradigmas nativista e construtivista, com uma discussão sobre a evolução humana,

concluindo pela possibilidade de se fazer uma síntese de contribuições de ambos os paradigmas. Propõe-se, portanto, que a linguagem teria uma natureza bio-psico-social, envolvendo desde os mecanismos bioquímicos do cérebro, a atividade gerativa da mente individual, e as interações estabilizadoras dependentes do contexto social.

Tal concepção bio-psico-social não é difícil de se formular, mas apresenta dificuldades consideráveis em aplicações práticas, pois requer novas ferramentas analíticas para se estudar as interações entre determinações biológicas, psicológicas e culturais da linguagem. Estas ferramentas dizem respeito à metodologia de estudo de sistemas complexos organizados em hierarquias fracas, na quais há tanto determinação ascendente (“de baixo para cima”) quanto descendente (“de cima para baixo”). Talvez a deficiência da escola Chomskyiana tenha sido se basear em um pressuposto metodológico que postula a separação estanque entre competência e performance. Este pressuposto vem a ser implicitamente questionado pela abordagem construtivista, quando sugere que a aprendizagem - que se situa ao nível da performance - possa contribuir para a formação da competência lingüística. Procurei, aqui, organizar esta discussão, e apontar uma direção para o encaminhamento desta pesquisa, sem ter a pretensão de ter atingido uma conclusão significativa, mas esperando despertar no leitor o interesse para com uma investigação mais aprofundada na área.

Alfredo Pereira Jr.

Universidade Estadual Paulista

apj@ibb.unesp.br

Bibliografia

- Black, I.B. (1991) *Information in the Brain*, Cambridge: MIT Press.
- Corballis, M.C. (1991) *The Lopsided Ape: Evolution of the Generative Mind*, Oxford: Oxford University Press.
- Deacon, T.W. (1997) *The Symbolic Species: The Co-Evolution of Language and the Brain*, New York: W.W. Norton and Co.
- Falk, D. (1992) *Braindance*, New York; Henry Holt.
- Fodor, J. (1976) *The Language of Thought*, Cambridge: MIT Press.
- Freeman, W.J. (1995) *Societies of Brains*, Hillsdale: Lawrence Erlbaum.
- Grice, P. (1957) 'Meaning', *Philosophical Review* **66**, pp. 377-388.
- Greenfield, P. (1991) 'Language, Tools, and Brain: The Ontogeny and Phylogeny of Hierarchically Organized Sequential Behavior', *Behavioral and Brain Sciences* **14**, pp. 531-51.
- Greenfield, P. (1998) 'Language, Tools, and Brain Revisited', *Behavioral and Brain Sciences* **21**, p.159-163.
- Jackendoff, R. (1992) *Languages of the Mind*, Cambridge: MIT Press.
- Johnson, M. (1987) *The Body in the Mind*, Chicago: The University of Chicago Press.
- Jurgens, U. (1998) 'Primate Communication: Signaling, Vocalization' IN Adelman, G. & Smith, B. (Eds.) *Elsevier's Encyclopedia of Neuroscience*, Amsterdam: Elsevier Science.

- Lakoff, G. (1987) *Women, Fire and Dangerous Things: What Categories Reveal About the Mind*, Chicago: University of Chicago Press.
- Llinas, R.R. & Ribary, U. (1992) 'Rostrocaudal Scan in Human Brain: A Global Characteristic of the 40-Hz Response During Sensory Input' IN Basar, E. and Bullock, T. (Eds.) *Induced Rhythms in the Brain*, Boston: Birkhouser.
- Piatelli-Palmarini, M., Org. (1983) *Teorias da Linguagem/Teorias da Aprendizagem*, São Paulo: Cultrix.
- Pinker, S. (1994) *The Language Instinct: How the Mind Creates Language*, New York: William Morrow.
- Popov, O., Segal, D.M. & Trifonov, E.N. (1996) 'Linguistic Complexity of Protein Sequences as Compared to Texts of Human Language', *ByoSystems* 38: 65-74.
- Quartz, S.R. & Sejnowski, T.J. (1997) 'The Neural Basis of Cognitive Development: A Constructivist Manifesto', *Behavioral and Brain Sciences* 20: 537-596.
- Reynolds, P. M. (1993) 'Tools, Language, and Cognition in Human Evolution', IN K.R. Gibson & T. Ingold (Eds.) Cambridge: Cambridge University Press.
- Rizzolatti, G. & Arbib, M.A. (1998) 'Language Within Our Grasp', *Trends in Neuroscience* **21**, p. 188-194.
- Rocha, A.F. (1997) 'The Brain as a Symbol-Processing Machine', *Progress in Neurobiology* 53: 121-198.
- Roy John, E., Easton, P. & Isenhardt, P. (1997) 'Consciousness and Cognition May Be Mediated by Multiple Independent Coherent Ensembles', *Consciousness and Cognition* **6**, p. 3-39.

Wills, C. (1993) *The Runaway Brain: the evolution of human uniqueness*, London: Harper Collins.

DISCUSSION

**ON SCHIFFER'S ARGUMENTS AGAINST THE FREGEAN MODEL
OF 'THAT'-CLAUSES: A COMMENT ON VIGNOLO**

Paolo Bonardi

Abstract

In "Propositions: What They Could and What They Could Not Be", Massimiliano Vignolo counters the arguments put forward by Stephen Schiffer ("The Things We Mean") against the so-called *Fregean model of 'that'-clauses*. My purpose here is to show that some of Vignolo's objections to Schiffer's arguments do not hit the mark. I shall also present a new argument against the Fregean model, which takes its cue from two of Schiffer's arguments.

Call a *Fregean model of 'that'-clauses* the conjunction of the following theses:

- (A) 'That'-clauses are singular terms standing for propositions.
- (B) Propositions are structured entities determined by the referents of the expressions forming the 'that'-clauses and by their syntactic structure.
- (C) Expressions occurring in 'that'-clauses have concepts [i.e. Fregean senses] as referents.¹

In his interesting article "Propositions: What They Could and What They Could Not Be" Massimiliano Vignolo argues that a group of arguments put forward by Stephen Schiffer (2003) against the Fregean model of 'that'-clauses are unsound.

The purpose of my paper is to discuss some of Vignolo's objections to those arguments. The *first section* will consider one of the main arguments by Schiffer. According to Vignolo, this argument can be read in two different ways and, on either reading, it is unsound. Regarding one of these readings, I shall claim that Vignolo is right, though further considerations will be added in order to make his objection entirely convincing. Regarding, on the other hand, Vignolo's objection to the other possible reading of the argument, I shall argue that it is problematic. Taking my cue from the argument by Schiffer considered in the first section, I shall construct in the *second section* a new argument against the Fregean model. I shall maintain that, no matter how

¹ Vignolo (2006: 129).

such a new argument is read, it is sound. Finally, the *third section* will consider two further arguments by Schiffer, showing that Vignolo's objections to them do not hit the mark.

I

One of the main arguments put forward by Schiffer against the Fregean model of 'that'-clauses goes as follows:

i) If the Fregean model is correct, then (a) 'Fido' occurs in "Ralph believes that Fido is a dog" as a singular term whose referent is a concept of Fido.

ii) If (a), then the following inference (Inf) is valid:

(Inf)
Ralph believes that Fido is a dog.
 $\therefore \exists x(x \text{ is a concept \& Ralph believes that } x \text{ is a dog}).$

iii) But the inference is not valid; given the truth of the premise, the conclusion is also true only in the unlikely event that Ralph mistakes a concept for a dog.

iv) \therefore The Fregean model is not true.²

According to Vignolo, this argument (call it *Fido argument*) is unsound for the following reason:

I take the premise of (Inf) to be ambiguous. It allows for a *de dicto* reading and for a *de re* reading. [...] If we construe the premise of the inference (Inf) as a *de dicto* belief, then step iii) is false. If we construe the same premise as a *de re* belief, then step i) is false.³

Let us examine first the case in which the premise is read *de dicto*.

If we construe the premise as a *de dicto* belief, then "Ralph believes that Fido is a dog" is true if and only if Ralph stands in the believing relation [(B)] to the proposition *that Fido is a dog*. According to the Fregean model, this proposition is formed by the concept of Fido [C_{Fido}] and the concept of [the property of] *being a dog* [$C_{\text{being a dog}}$]. The logical form of the premise of (Inf) is:

$B(\text{Ralph}, \langle C_{\text{Fido}}, C_{\text{being a dog}} \rangle).$

The logical form of the conclusion of (Inf) is:

² Ibid., p. 130. This argument originates with Adam Pautz.

³ Ibid., pp. 130-1.

$\exists x(x \text{ is a concept} \ \& \ B(\text{Ralph}, \langle x, C_{\text{being a dog}} \rangle))$.

The conclusion is true if and only if there is a concept that, together with the concept of *being a dog*, constitutes the proposition that Ralph believes. In quantifying in 'that'-clauses of *de dicto* beliefs, variables range over concepts. Therefore, according to the Fregean model, if it is true that Ralph believes *de dicto* that Fido is a dog, it is true that there is a concept that, together with the concept of *being a dog*, constitutes the proposition that Ralph believes.⁴

I think that Vignolo is right in suggesting that the inference (Inf) is valid if its premise and conclusion are *de dicto* interpreted. Nonetheless, it is worth examining more carefully the justification provided by Schiffer in step iii) of the Fido argument for his claim that (Inf) is not valid (i.e.: "given the truth of the premise [of (Inf)], the conclusion is also true only in the unlikely event that Ralph mistakes a concept for a dog"). What exactly goes wrong in such a justification? Vignolo's explanation on this point is not entirely clear:

To achieve his goal, Schiffer should require that in quantifying in 'that'-clauses of *de dicto* beliefs variables range over the ordinary referents of expressions. Only in this case the conclusion of (Inf) would be false unless Ralph mistakes a concept for a dog.⁵

I will try to suggest a different explanation. If, as Schiffer claims, the conclusion of (Inf) is true only if Ralph mistakes a concept for a dog, then the *de dicto* logical form of the conclusion should be:

(1) $\exists x(x \text{ is a concept} \ \& \ B(\text{Ralph}, \langle x, \text{being a dog} \rangle))$.

The linguistic construction (1) states that Ralph believes a proposition constituted of a concept and the property of *being a dog*, i.e. he believes that a concept exemplifies such a property. But this is incompatible with the Fregean model, assumed in step i), according to which only concepts can enter into propositions as their constituents. The *de dicto* logical form of the conclusion of (Inf) is indeed (2), which differs from (1) in that the property of *being a dog* is replaced by its concept.

⁴ Ibid., p. 131.

⁵ Ibid., p. 132.

(2) $\exists x(x \text{ is a concept} \ \& \ B(\text{Ralph}, \langle x, C_{\text{being a dog}} \rangle))$

So, this is, I think, the reason why the justification provided by Schiffer in favour of the presumed invalidity of (Inf) is incorrect when the premise of this inference is read *de dicto*. Let us now examine the case in which the premise of (Inf) is read *de re*.

Construed as a *de re* belief, the premise Ralph believes that Fido is a dog turns into:

(3) Ralph believes of Fido that it is a dog.

[...] In (3) the proper name 'Fido' stands for Fido and not for the concept of Fido. In (4) the variable 'x' must range over things that are ordinary referents of singular terms.

(4) $\exists x(\text{Ralph believes of } x \text{ that it is a dog})$.

[...] The advocate of the Fregean semantics is not committed to denying that in (3) 'Fido' stands for Fido. He has the resources to [formulate the truth-conditions of (3) within the Fregean model]: Ralph believes of Fido that it is a dog if and only if the proposition $\langle C_{\text{Fido}}, C_{\text{being a dog}} \rangle$ is a mode of presentation (MP) of the state of affairs $\langle \text{Fido, being a dog} \rangle$ and Ralph believes [(B)] such proposition. The logical form of (3) is:

(5) $\text{MP}(\langle C_{\text{Fido}}, C_{\text{being a dog}} \rangle, \langle \text{Fido, being a dog} \rangle) \ \& \ B(\text{Ralph}, \langle C_{\text{Fido}}, C_{\text{being a dog}} \rangle)$.

The logical form of (4) is:

(6) $\exists y \exists z (\text{MP}(\langle z, C_{\text{being a dog}} \rangle, \langle y, \text{being a dog} \rangle) \ \& \ B(\text{Ralph}, \langle z, C_{\text{being a dog}} \rangle))$.⁶

So, in the Fido argument, according to Vignolo, step i) is false if the premise of (Inf) is read as (3) and the logical form of (3) is (5). Furthermore, if the conclusion of (Inf) is read as (4) – omitting in the latter the expression 'x is a concept &' contained in the former – and the logical form of (4) is (6), then (Inf) will be valid, contrary to what step iii) states.

But can (5) and (6) really be logical forms of (3) and (4), as Vignolo claims? This is actually disputable, as I am going to show now. (I shall confine myself to showing that (6) can hardly be the logical form of (4). Using a similar strategy, the reader could show that (5) can hardly be the logical form of (3).)

Reconsider the construction (4), in which the occurrence of the pronoun 'it' clearly works as an anaphor linked to the variable 'x'. Taking into account that, by

⁶ Ibid., pp. 132-3. The logical constructions (5) and (6) originate with Kaplan (1969).

definition, anaphora keep the same referent as the expressions to which they are linked, the occurrence of 'it' in (4) could be replaced by an occurrence of 'x', yielding:

(4*) $\exists x(\text{Ralph believes of } x \text{ that } x \text{ is a dog})$.

Consequently, if, as Vignolo claims, (6) is the logical form of (4), then it would also be the logical form of (4*). But, if so, the following question should be answered: what variable in (6) plays the role of 'x' in (4*): 'y' or 'z'?

(4) $\exists x(\text{Ralph believes of } x \text{ that it is a dog})$

(6) $\exists y \exists z(\text{MP}(\langle z, C_{\text{being a dog}} \rangle, \langle y, \text{being a dog} \rangle) \ \& \ \text{B}(\text{Ralph}, \langle z, C_{\text{being a dog}} \rangle))$

No answer to this question is actually made available. In fact, 'y' cannot play the role of 'x' because, according to (4*), x occurs within the proposition believed by Ralph – since the variable 'x' occurs within the 'that'-clause in (4*) – whereas this does not apply to y , as (6) reveals. On the other hand, even 'z' cannot play the role of 'x', because x has an occurrence outside the proposition believed by Ralph, while z only occurs within such a proposition. Hence, the conclusion that (6) cannot be the logical form of (4) follows.⁷

II

Consider another argument by Schiffer against the Fregean model (call it *brother-in-love argument*):

There are cases where it seems that some singular terms occurring in 'that'-clauses cannot but refer to their ordinary referents. Schiffer gives [this] example: your husband's brother says to you: "I believe I am falling in love with you". Schiffer holds that it is *obvious* that [both utterances of 'I'] refer to your husband's brother.⁸

⁷ A possible reply to my criticism could be that although (6) is not the logical form of (4), the former presents the meaning of the latter (and the same would also apply to (5) and (3)); this could be sufficient for Vignolo to reach the conclusion that in the Fido argument, steps i) and iii) are false if the premise of (Inf) is read *de re*. But I think such a reply does not work, for the following reason: generally speaking, if a sentence expresses a structured proposition, then such a proposition is identified by the logical form of the sentence; if so, it is impossible that (6) presents the meaning of (4) without being its logical form, because, according to the Fregean model, the meaning of a sentence is a structured proposition.

⁸ Vignolo (2006: 129).

Although – I think – Schiffer is right in claiming that both of these occurrences designate their ordinary referent, such a claim is not immediately obvious. In this section I would like to present a new argument against the Fregean model (call it *new brother-in-love argument*) which, in a sense, develops Schiffer's intuition contained in the original brother-in-love argument. My new argument will take pattern by the Fido argument examined in the first section, but, unlike the latter, it will be sound under any reading/interpretation.

i*) If the Fregean model is correct then (b) the second occurrence of 'I' in the sentence "I believe that I am falling in love with you", asserted by your husband's brother and addressed to you, refers to the concept of your husband's brother.⁹

ii*) The following inference (Inf*) is intuitively valid:

(Inf*)

(P*) I believe that I am falling in love with you.

∴ (C*) There is someone who believes that he (himself) is falling in love with you.

iii*) Taking into account that 'he' in (C*) is an anaphor linked to the pronoun 'someone' and anaphora, by definition, keep the same referent as the term they are linked to, if (C*) is true, then there is an individual satisfying both the following conditions:

(d) He/It has the belief described in (C*);

(e) He/It contributes to the proposition designed by the 'that'-clause in (C*);¹⁰

iv*) Suppose that (P*) is true. Since (Inf*) is intuitively valid, then (C*) will be true as well. If so, the following question should be answered: who is the individual satisfying both of the aforesaid conditions (d) and (e)? Actually, no plausible answer to this question is possible if claim (b) in step i*) is true.

v*) ∴ The Fregean model is not true.

In an attempt to refute this argument, the advocate of the Fregean model could look for a plausible answer to the question put forward in step iv*). Possibly, some aid towards

⁹ Notice that the claim (b) is true even if the sentence at issue is read *de re*, i.e. as "I believe of myself that I am falling in love with you".

¹⁰ Notice that the conditions (d) and (e) are satisfied even if (P*) is read *de re* (i.e. as "I believe of myself that I am falling in love with you") and (C*) is consequently read as: "There is someone who believes of himself that he is falling in love with you".

that end could come from the alleged *de dicto* and the *de re* interpretations of (C*), i.e. respectively (C*_{de dicto}) and (C*_{de re}), where *a* is you.¹¹ The linguistic construction (C*_{de dicto}) says that there is an *x* such that, together with the relational concept of *falling in love with*, i.e. C_{falling in love with}, and the concept of you, i.e. C_a, constitutes the proposition that *x* believes. On the other hand, the linguistic construction (C*_{de re}) says that there is a person *y* and there is a concept *z* such that the proposition $\langle\langle z, C_a \rangle, C_{\text{falling in love with}} \rangle$ is a mode of presentation (MP) of the state of affairs $\langle\langle y, a \rangle, \text{falling in love with} \rangle$ and *y* believes (B) such a proposition.

$$\begin{aligned} & (C^*_{de dicto}) \exists x B(x, \langle\langle x, C_a \rangle, C_{\text{falling in love with}} \rangle) \\ & (C^*_{de re}) \exists y \exists z (\text{MP}(\langle\langle z, C_a \rangle, C_{\text{falling in love with}} \rangle, \langle\langle y, a \rangle, \text{falling in love with} \rangle) \& \\ & \quad B(y, \langle\langle z, C_a \rangle, C_{\text{falling in love with}} \rangle)) \end{aligned}$$

Now, reconsider the question “Who is the individual satisfying both conditions (d) and (e)?” put forward in step iv*) and suppose first that (C*) is *de dicto* interpreted. Under such an interpretation, the obvious answer to the aforesaid question will be “the value of the bound variable ‘*x*’ in (C*_{de dicto})”. But who/what could such a value be? Taking into account that (C*) follows from (P*), asserted by your husband’s brother, just two answers are *prima facie* available: either *x* is your husband’s brother or it is a concept of him. Unfortunately for the advocate of the Fregean model, either of these answers is inadmissible. In fact, if the value of ‘*x*’ were your husband’s brother, then, on the basis of (C*_{de dicto}), an occurrence of him would be contained in the believed proposition. But this would be incompatible with claim (b) in step i*), according to which the referent of the second occurrence of ‘I’ in (P*) is a concept and not a person. On the other hand, if the value of ‘*x*’ were a concept of your husband’s brother, then, on the basis of (C*_{de dicto}), such a concept would also occur as first *relatum* of the belief relation B, i.e. it would have the belief described by (C*). But claiming that a concept has a belief is highly implausible or even senseless.

(P*) I believe that I am falling in love with you

(C*) There is someone who believes that he himself is falling in love with you

¹¹ I have used here the adjective ‘alleged’ since, as I argued on pp. 165-6, it is disputable that a construction like (C*_{de re}) can be a logical form of (C*).

So, the question formulated in step iv*) has no admissible answer if (C*) is *de dicto* interpreted. Let us now see what happens if (C*) is interpreted as (C*_{de re}). In this case, who could be the individual satisfying both conditions (d) and (e)? Again, just two answers are *prima facie* available: either it is the value of 'y', i.e. your husband's brother; or it is the value of 'z', i.e. a concept of him. But neither of these answers is actually acceptable. In fact, y is not a good candidate for the wanted individual because, according to (C*_{de re}), y does not occur within the believed proposition, and this forbids the fulfilment of condition (e). Even z is not a good candidate for that role since, according to (C*_{de re}), z does not occur as first *relatum* of B, i.e. z does not have the belief described by (C*), differently from what condition (d) would require. So, if (C*) is *de re* interpreted, the crucial question put forward in step iv*) also has no answer.

(C*) There is someone who believes that he himself is falling in love with you
 (C*_{de re}) $\exists y \exists z (\text{MP}(\langle\langle z, C_a \rangle, C_{\text{falling in love with}} \rangle, \langle\langle y, a \rangle, \text{falling in love with} \rangle) \& \text{B}(y, \langle\langle z, C_a \rangle, C_{\text{falling in love with}} \rangle))$

Condition (d): He/It has the belief described in (C*)

Condition (e): He/It contributes to the proposition designed by the 'that'-clause in (C*)

As an extreme move, the advocate of the Fregean model could introduce, besides (C*_{de dicto}) and (C*_{de re}), a third interpretation of (C*), i.e. (7), which differs from (C*_{de re}) solely in that at its beginning there is only one quantifier, whose domain contains ordered pairs constituted of a person, y, and a concept, z.

(7) $\exists \langle y, z \rangle (\text{MP}(\langle\langle z, C_a \rangle, C_{\text{falling in love with}} \rangle, \langle\langle y, a \rangle, \text{falling in love with} \rangle) \& \text{B}(y, \langle\langle z, C_a \rangle, C_{\text{falling in love with}} \rangle))$

In light of (7), one could claim that the individual satisfying the conditions (d) and (e) is an ordered pair including your husband's brother and a concept of him. Actually, such a pair would be a very bad candidate for the wanted individual, because it neither has the belief described by (C*) – in fact, $\langle y, z \rangle$ does not occur as first *relatum* of B – nor does it occur within the believed proposition, which conditions (d) and (e) respectively require. So, none of the three considered interpretations of (C*) is able to refute the new brother-in-love argument, QED.

III

Another argument put forward by Schiffer against the Fregean model is the following:

If expressions occurring in 'that'-clauses have concepts as referents, then in (f) "John believes that Fido barks" the name 'Fido' refers to the concept of Fido. Hence, if Mary asserts (f), she refers to the concept of Fido, which is the concept by which John thinks of Fido. The Fregean model must explain by which concept Mary thinks of the concept by which John thinks of Fido. [Furthermore, given the possibility of iterating the structure of propositional attitude sentences,] we admit a hierarchy of concepts: concepts, concepts of concepts, concepts of concepts of concepts and so on. The Fregean model must explain what such concepts are and what it is to grasp all of them.¹²

Vignolo highlights that this argument (call it *Mary argument*) presupposes an endorsement, on the part of the Fregean model, of the principle (P):

(P) Whenever we refer to something, we do it by grasping a concept under which that thing falls.

My contention is that the problem arises because of a sort of ambiguity of 'refer'. There are at least two readings of 'refer' that should be distinguished:

- (i) To contribute to truth-conditions.
- (ii) To think/speak of.

When Schiffer says that the speaker refers to the concept of Fido in asserting (f), he might mean two different things:

- (i*) The truth-condition of (f) involves the concept of Fido.
- (ii*) The speaker speaks of the concept of Fido.

If Mary asserts (f), certainly the truth-condition of her assertion involves the concept of Fido. Her assertion is true if and only if John stands in the believing relation to the proposition $\langle C_{\text{Fido}}, C_{\text{barking}} \rangle$. So, in the sense of (i), it is true that Mary refers to the concept of Fido. Yet, [...] she speaks of Fido, not of the concept of Fido. In conclusion, Principle (P) should be constrained: if 'refer' is taken to mean thinking/speaking of, then principle (P) holds. If 'refer' is taken to mean contributing to truth-conditions, principle (P) does not hold.

[...] We do not need any concept of concept in order to specify the truth-condition of (f). But we do not need any concept of concept to construct the proposition expressed by (f) either. The proposition expressed by (f) is made of the concept of John, the concept of the believing relation and the proposition $\langle C_{\text{Fido}}, C_{\text{barking}} \rangle$. We can represent such proposition as $\langle \langle C_{\text{John}}, \langle C_{\text{Fido}}, C_{\text{barking}} \rangle \rangle, C_B \rangle$. [...] We can form propositional attitude sentences more and more complicated without being forced to generate any hierarchy of concepts.¹³

¹² Vignolo (2006: 130).

¹³ Ibid., pp. 135-6. (Incidentally, this quotation has been slightly modified: I have exchanged the example presented in Vignolo's article involving George Eliot for an example involving Fido barking.) It should be noted that Vignolo's use of the word 'refer' in (i) is rather unusual. It should also be noted that, unlike

So, taking for granted Vignolo's distinction between (i) and (ii) and his suggestion of confining the validity of (P) to the case in which 'refer' means (ii), Mary's competent assertion of (f) only requires her grasping of the first level concepts C_{John} , C_{Fido} , C_{barking} , C_{B} .

If, on one hand, Vignolo's proposal seems able to block the infinite regress described in Schiffer's Mary argument, on the other hand, another objection could be put forward to such a proposal. According to Vignolo, Mary has to master C_{Fido} , i.e. the concept by which John thinks of Fido, in order to assert competently (f). But this is absurd, because, intuitively, Mary should be able to make such an assertion even if she has no idea about what C_{Fido} is. She could indeed have a concept of Fido which completely differs from John's.

As a reply, the advocate of the Fregean model could maintain that the logical form of (f) is not (8) but (9), the latter involving, instead of C_{Fido} , a quantification over it. In this way, Mary would not need to master C_{Fido} in order to assert competently (f).

- (f) John believes that Fido barks
- (8) $B(\text{John}, \langle C_{\text{Fido}}, C_{\text{barking}} \rangle)$
- (9) $\exists x \exists y (x \text{ is the concept of Fido for John} \ \& \ y \text{ is the concept of the property of barking for John} \ \& \ B(\text{John}, \langle x, y \rangle))$

It is worth noticing that this sort of reply also (apparently) works for another of Schiffer's arguments against the Fregean model, which goes as follows:

Consider the [...] sentence: (g) "Everyone who visits New York believes that New York is noisy". The second occurrence of 'New York' refers to the concept of New York. But it should be a concept shared by all people who visit New York and it is very unlikely that all those people share the same concept of it.¹⁴

The reply to this argument (call it *NY argument*) could be that the logical form of (g) is not the problematic construction (10) but (11), where the single concept of New York (C_{NY}) leaves room for a quantification over many concepts of this city.

Vignolo, for Frege the proposition expressed by (f) is not $\langle \langle C_{\text{John}}, \langle C_{\text{Fido}}, C_{\text{barking}} \rangle \rangle, C_{\text{B}} \rangle$ but rather $\langle \langle C_{\text{John}}, \langle C_{\text{Fido}}^*, C_{\text{barking}}^* \rangle \rangle, C_{\text{B}} \rangle$, where C_{Fido}^* and C_{barking}^* are respectively *second level* concepts of Fido and the property of barking.

¹⁴ Vignolo (2006: 130).

- (g) Everyone who visits New York believes that New York is noisy
 (10) $\forall x (x \text{ visits NY} \rightarrow B(x, \langle C_{\text{NY}}, C_{\text{being noisy}} \rangle))$
 (11) $\forall x (x \text{ visits NY} \rightarrow \exists y \exists z (y \text{ is the concept of NY for } x \ \& \ z \text{ is the concept of the property of } \textit{being noisy} \text{ for } x \ \& \ B(x, \langle y, z \rangle)))$

Unluckily for the advocate of the Fregean model, these replies actually present some difficulties. Particularly, in Mary's case, the logical form of the sentence (f) – asserted, supposedly sincerely, by Mary – cannot even be (9). Otherwise, by asserting sincerely (f), Mary would assert sincerely that there *exists* a concept x of Fido for John and there *exists* a concept y of the property of barking for John such that John believes a proposition constituted of these concepts. As a result, Mary would commit herself to the existence of the concepts x and y . But this is absurd because, intuitively, a speaker should be able to assert sincerely (f) even if she refuses the existence of concepts.

- (f) John believes that Fido barks
 (8) $B(\text{John}, \langle C_{\text{Fido}}, C_{\text{barking}} \rangle)$
 (9) $\exists x \exists y (x \text{ is the concept of Fido for John} \ \& \ y \text{ is the concept of the property of barking for John} \ \& \ B(\text{John}, \langle x, y \rangle))$

This sort of difficulty also affects the reply to the NY argument, which, in addition, presents another problem. Since, according to such a reply, different visitors of New York can have different concepts of this city, in a sentence like ' a believes that New York is noisy', a being any visitor of New York, the referent and consequently the meaning of the name 'New York' could change according to whom a is. If so, 'New York' would have not just one meaning (or two: one for the direct linguistic contexts and another for the indirect contexts) but many. This plainly contrasts with the principle that the meanings of a word should not be multiplied unnecessarily.

How could these further difficulties be overcome? Vignolo puts forward a proposal, according to which the logical form of a sentence like (g) is really (10) and not (11), so that whoever visits New York grasps the very same and only concept of New York, C_{NY} . (Similarly, in Mary's case, the logical form of (f) is (8) rather than (9), so that Mary masters the very same concept by which John thinks of Fido, C_{Fido} .)

But how can such a proposal work if the visitors of New York have different views of this city, i.e. they associate different pieces of information with it? Vignolo's

answer to this question starts from a characterisation of the notion of concept, drawn from the so-called *use conception of meaning*.¹⁵

The central idea of the use conception of meaning is to individuate concepts through regularities of referential and inferential uses of linguistic expressions. The claim is that the regularities of the use of a certain expression are constitutive of the property of expressing a certain concept.

[...] Individuating identity criteria for concepts: the concept W is identical to the concept Y if and only if [the locution] w [expressing the concept W] and [the locution] y [expressing the concept Y] have the same constitutive uses.

[...] In the light of a use theory of meaning, we can resist Schiffer's charge that there exists no single concept shared by all people who visit New York. Even if visitors will have different views of New York, this does not imply that there is no constitutive use of the proper name "New York". We can imagine a constitutive use with different levels of expertise. Although it is not necessary that all speakers master the constitutive use completely, that use constitutes the property of expressing the concept of New York. The division of the linguistic labour and the deference to experts enable us to attribute propositional attitudes towards propositions to speakers even though those propositions are made of concepts that they do not master completely.¹⁶

The characterisation of the notion of concept suggested here by Vignolo is noteworthy, but – I think – does not succeed in overcoming Schiffer's NY argument. To see why, let $\{NY_i\}$ and $\{T_j\}$ be respectively the sets of the constitutive uses of the names 'New York' and 'Toronto'. Suppose that some constitutive uses of 'New York' are identical to some constitutive uses of 'Toronto' and let A be the set of such uses. In other words, $A = \{NY_i\} \cap \{T_j\}$.¹⁷ For purposes of application, consider the following case. Tom is a visitor of New York who believes that New York is noisy (he has, in fact, the disposition to assert sincerely and competently the sentence "New York is noisy", referring to New York). Suppose that his only uses of 'New York' are those included in A . Since A is a *proper* sub-set of $\{NY_i\}$, Tom's mastery of the constitutive uses of 'New

¹⁵ Among the advocates of this conception, Vignolo mentions P. Horwich (1998) *Meaning*, Oxford: Clarendon Press.

¹⁶ Vignolo (2006: 142-3). Vignolo mentions C. Peacocke [(1992) *A Study of Concepts*, MIT Press, Cambridge Mass.] among the advocates of the thesis according to which it is possible to attribute attitudes towards propositions to thinkers who master partially the constitutive use of a proper name and defer in their use of it to the expert members of their community.

¹⁷ What does it mean here to say that the use of a word is identical to the use of another (i.e. that two words share the same use)? In light of the fact that words are used within linguistic contexts, we could determine a given use of a word through a given linguistic context in which it is used. So, a use of the name 'New York' could be individuated by the linguistic context '... is noisy', within which the name is used. On the basis of the fact that the name 'Toronto' can also be used within the very same linguistic context, we could claim that 'New York' and 'Toronto' share a use.

York' will be partial. Nonetheless, according to Vignolo's proposal, such a partial mastery suffices for Tom's grasping of the concept of New York (C_{NY}). But we have supposed that the elements of A are constitutive uses of the name 'Toronto' as well, since by definition $A = \{NY_i\} \cap \{T_j\}$. So, how can we be sure that, by his uses of 'New York', Tom grasps C_{NY} instead of the concept of Toronto (C_T)? Tom could indeed have grasped C_T instead of C_{NY} , mistaking the former concept for the latter. In this case, the name 'New York' within the sentence "Tom believes that New York is noisy" would refer to C_T instead of C_{NY} , so that the logical form of (g) could not be (10), contrary to what Vignolo claims.

Notice that this sort of difficulty becomes even more serious if we modify the aforesaid case in such a way that Tom's uses of 'New York', because of his ignorance, coincide with the constitutive uses of 'Toronto'. In this modified case, Tom's mistaking of C_T for C_{NY} would be highly probable, so that the name 'New York' within the sentence "Tom believes that New York is noisy" would end up referring to C_T and the logical form of (g) could not consequently be (10). In this way, Vignolo's attempt of rescuing the Fregean model of 'that'-clauses from the arguments by Schiffer considered in this section would not hit the mark.¹⁸

Paolo Bonardi

University of Geneva

pbonardi@libero.it

¹⁸ Many thanks to Kevin Mulligan, Marco Santambrogio and Jennifer Saul for their help and encouragement.

References

Kaplan, D. (1969) 'Quantifying In' IN Davidson, D. & Hintikka, J. (eds.), *Words and Objections*, Reidel, Dordrecht.

Schiffer, S. (2003) *The Things We Mean*, Clarendon Press, Oxford.

Vignolo, M. (2006) 'Propositions: What They Could and What They Could Not Be', *Abstracta* 2, 2: pp. 129-147.

abstracta

Linguagem, Mente & Ação

ERRATA

VOLUME 2, NUMBER 2, 2006

In “A Pill against Epiphenomenalism” (by Patrick Spät):

- Page 176, line 20 from top, the sentence that begins with “To the contrary...” should be indented from the margin so as to be clearly characterized as a new paragraph.
- Page 177, line 8 from top, the sentence that begins with “How could it be possible that the PE...” should be indented from the margin so as to be clearly characterized as a new paragraph.
- Page 177, line 26 from top, for *over-think* read *reassess*.
- Page 178, a reference is missing:
Huxley, T.H. (1874) ‘On the Hypothesis that Animals are Automata, and its History’,
Fortnightly Review, 22, 555-580.