SORITES

σωρίτης

An International Electronic Quarterly of Analytical Philosophy
Indexed and Abstracted in ***THE PHILOSOPHER'S INDEX***
ISSN 1135-1349
Legal Deposit Registration: M 14867-1995

Regular-Mail Address:

Prof. Lorenzo Peña
CSIC [Spanish Institute for Advanced Studies]
Department of Theoretical Philosophy
Pinar 25
E-28006 Madrid
Spain
Fax +341 564 52 52
Voice Tph +341 411 70 60, ext 18

Prof. Guillermo Hurtado
Instituto de Investigaciones Filosóficas
Circuito Mtro Mario de la Cueva
Ciudad de la Investigación en Humanidades
Coyoacán 04510. México DF
México
FAX (525) 665-4991

INTERNET ACCESS:
<http://www.ifs.csic.es/sorites/>
<http://www.filosoficas.unam.mx/sorites>
Editorial e-mail inbox (esp. for submissions): <sorites@ifs.csic.es>
Inquiries and subscription-requests: <sorites@pinar2.csic.es>

Issue #11 — December 1999

SORITES (ΣΩΡΙΤΗΣ)
ISSN 1135-1349
Issue # 11. December 1999

MAIN INTERNET ACCESS:

http://www.ifs.csic.es/sorites/

<sorites@ifs.csic.es> (Editorial e-mail inbox, esp. for submissions)
<sorites@pinar2.csic.es> (Inquiries and subscription-requests)

Other InterNet access:

(1) By anonymous ftp: open ftp.csic.es, then go to directory /pub/sorites
(2) From the Americas the following URL is more convenient:

http://www.filosoficas.unam.mx/sorites

Please send both your Regular-Mail contributions and books for review to the address of either of the Executive Editors

SORITES

ISSN 1135-1349

SORITES

ISSN 1135-1349

Issue #11. December 1999

TABLE OF CONTENTS

SORITES (ΣΩΡΙΤΗΣ), ISSN 1135-1349
Issue #11. December 1999. Pp. 3-5.
Abstracts of the Papers

ABSTRACTS OF THE PAPERS

REFERENCE CHANGE OF NATURAL KIND TERMS

by Luis Fernández Moreno

Kuhn's thesis of referential incommensurability rests on the thesis of reference change according to which theory change involves reference change. One of Kuhn's disagreements with Putnam's reference theory and in general with the causal theory of reference concerns the question of whether the reference of natural kind terms may change. On examining this disagreement it will be paid attention to the factors which might involve changes of reference and to the doctrines which may lend support to the thesis of reference change. It will be argued that, though the reference of natural kind terms is open to change, the proponents of the thesis of reference change have not conclusively established their thesis.

♣ ♣ ♣ ♣ ♣ ♣ ♣

Was Frege wrong when identifying reference with truth-value?

by Jean-Yves Béziau

We discuss Sengupta's argumentation according to which Frege was wrong identifying reference with truth-value.

After stating various possible interpretations of Frege's principle of substitution, we show that there is no coherent interpretation under which Sengupta's argumentation is valid.

Finally we try to show how Frege's distinction can work in the context of modern mathematics and how modern logic grasps it.

♣ ♣ ♣ ♣ ♣ ♣ ♣

Quasi-Indexical Attitudes

by Tomis Kapitan

Indexical reference reflects indexical consciousness, consciousness from a particular spatio-temporal perspective. In using terms like *this, that, I, you, beyond,* not only do we designate items falling within our experience but also record our conscious orientation to them, and since such orientation embodies a unique perspective, then indexical modes of presentation are essentially subjective. If this is so, then how do we explain the fact that we communicate quite well with indexicals? Moreover, how can we accurately attribute indexical reference to others? While we never exactly duplicate the contents of another's indexical consciousness in our own, we can simulate them in our own thinking by pinpointing the speaker's perspective and referents from our own vantage point and imputing generic indexical modes. We represent our attributions through *quasi-indicators,* the abstract singular terms used to depict another's contents. Consequently, we must be capable of *quasi-indexical* consciousness which, in turn, is the foundation of all communication. Its structure is the topic of this paper.

♣ ♣ ♣ ♣ ♣ ♣ ♣

Are There Mental Entities? Some lessons from Hans Reichenbach

by Jeanne Peijnenburg

The meaning of mental terms and the status of mental entities are core issues in contemporary philosophy of mind. It is argued that the old Reichenbachian distinction between *abstracta* and *illata* might shed new light on these issues. First, it suggests that beliefs, desires and other pro-attitudes that make up the higher mental life are not all equally substantial or real. Second, it conceives the elements of the lower mental life (sensations, impressions) as entities that are inferred from concrete, observable events. As a consequence, it might teach us two lessons: first, to see reliefs in the higher mental map, and second, to acknowledge that qualia are probabilistically inferred rather than directly experienced.

♣ ♣ ♣ ♣ ♣ ♣ ♣

Amounts of Vagueness, Degrees of Truth

by Enrique Romerales

Many theorists think nowadays that vagueness is a widespread phenomenon that affects and infects almost all terms and concepts of our thought and language, and for some philosophers degree of truth theories are the best way to cope with vagueness and sorites susceptible concepts. In this paper I argue that many of the allegedly vague concepts (colour terms, «heap», «town» etc.) are not vague in the last analysis the philosopher or scientist could offer if compelled to, and that much of the vagueness of the properly vague ones (viz. «young», «thin», «far») comes from its contextual dependence alone. I also argue that degree of truth approaches — particularly the infinitist ones — and fuzzy logics do not solve practically any of the puzzles brought about by vagueness and sorites arguments, and conversely they have many additional problems of their own. Concerning recalcitrant cases of vagueness, I would tentatively commend the epistemic theory of vagueness, from an inference to the best explanation (or to the least bad, to speak more properly).

♣ ♣ ♣ ♣ ♣ ♣ ♣

Benardete's Paradox

by Michael B. Burke

Graham Priest has focused attention on an intriguing but neglected paradox posed by José Benardete in 1964. Benardete viewed the paradox as a threat to the intelligibility of the spatial and temporal continua and offered several different versions of it. Priest has selected one of those versions and formalized it. Although Priest has succeeded nicely in sharpening the paradox, the version he chose to formalize has distracting and potentially problematic features that are absent from some of Benardete's other versions. I offer a formalization of a *simpler* version of the paradox, the one that presents most plainly Benardete's challenge to the spatial continuum. Proposed resolutions of Benardete's paradox should address this version of the paradox as well as the one formalized by Priest.

SORITES (ΣΩΡΙΤΗΣ), ISSN 1135-1349
Issue #11. December 1999. Pp. 6-14.
Reference Change of Natural Kind Terms

REFERENCE CHANGE OF NATURAL KIND TERMS

by Luis Fernández Moreno

1. Introduction

Kuhn's incommensurability thesis asserts that the languages in which successive or rival scientific theories are formulated are not mutually translatable. Nevertheless, Kuhn restricted the scope of this thesis in two senses. First, the translation failure between the languages of theories only concerns a sort of terms, namely, the *kind terms*; second, that failure of translation only affects *some* of the kind terms of those languages or rather a small subset of interrelated terms.[1] Thus Kuhn claims that incommensurability has only a *local* character, since most terms common to rival theories are mutually translatable.[2]

Although the incommensurability thesis so conceived, to wit, as a semantic thesis, concerns the concept of translation and hence meaning, this thesis is often justified on the basis of the concept of reference. The *thesis of referential incommensurability* — as it may be called — would assert that kind terms from rival theories have a different reference. This thesis rests on the claim that a change of theory entails changes in the reference of kind terms common to those theories or, in short, that theory change involves reference change. This thesis could be called «thesis of reference change».

The acceptability of the *thesis of reference change* will depend on how the reference of kind terms and especially of natural kind terms is determined. Since according to the causal theory of reference the thesis of reference change should be rejected, it is understandable that Kuhn examined the causal theory of reference. Kuhn [1979] already contained some remarks on this theory, but Kuhn's last judgment on the causal theory is formulated in two later writings, Kuhn [1989] and [1990]. In these two writings Kuhn examined critically the causal theory of reference and especially Putnam's natural kind terms reference theory such as it is formulated in Putnam

---

[1] In the following I shall not generally make explicit this second restriction.

[2] Concerning Kuhn's final, restricted, conception of incommensurability see especially Kuhn [1983], [1991] and [1993].

[1975a]. Though Kuhn asserts that he restricts himself primarily to Putnam's theory,[3] Kuhn claims that his objections against Putnam's theory would also apply to other versions of the causal theory. One of Kuhn's objections, and indeed his main objection, concerns the question of whether the reference of natural kind terms may change. On examining this objection I will pay attention to the factors which might involve changes of reference and to the doctrines which may lend support to the thesis of reference change. I shall argue that though the reference of natural kind terms is open to change, the proponents of the thesis of reference change have not conclusively established their thesis.

2. Kuhn's Objections to Putnam's Theory

It is advisable to begin with a brief characterization of Putnam's natural kind terms reference theory. According to this theory the reference of a natural kind term is determined by two factors, to wit, by some initial act of dubbing samples of the kind — let us call the samples there involved «original samples» — and by the relation sameness-of-kind; this relation is constituted by the so-called essential properties, i.e. by properties concerning the internal structure of samples of the kind, which are discoverable only by scientific research. Thus, in Putnam's words, there is a «contribution of the environment» to the determination of the reference of natural kind terms. Putnam mentions another sort of contribution too, namely, the «contribution of the society»; by this it is meant that there is a linguistic division of labor or, more precisely, there is a subset of members of the linguistic community, the set of «expert» speakers — they will be in the main members of a scientific community — who have a more reliable knowledge than laymen of the reference of natural kind terms; Putnam claims that non-expert speakers rely on the judgment of expert speakers about the reference of these terms.[4] Kuhn agrees with Putnam's assertions on the linguistic division of labor and therefore with Putnam's theory of *reference transmission*, but he does not accept some central claims of Putnam's theory of *reference determination*, which entail the rejection of the thesis of reference change. According to Putnam the original samples of the kind and the relation sameness-of-kind are independent of our theories; thus theory change does not involve reference change and therefore the thesis of reference change has to be rejected.

Kuhn's main objection to Putnam's theory is that this theory rules out and therefore cannot explain *changes of reference*. More precisely, Kuhn questions that the samples determining the reference of natural kind terms and the relation sameness-of-kind have remained *stable* through theory change. Kuhn formulates this objection on analysing Putnam's Twin-Earth thought experiment concerning the term «water». It is noteworthy, however, that Kuhn's remarks will not concern the term «water» as employed in everyday life or by laypeople, but as used «within the community of scientists and philosophers to which Putnam's argument needs to be applied.»[5]

---

3 Kuhn [1989], p. 25.

4 Putnam [1975a], p. 228.

5 Kuhn [1989], p. 26, n. 28 and [1990], p. 318, n. 25.

Kuhn attends specially to the second part of Putnam's thought experiment and to the referential history of the term «water» as used on Earth, more precisely, to the following passage from [1975a]:

> [L]et us roll the time back to about 1750. At that time […] [t]he typical Earthian speaker of English did not know water consisted of hydrogen and oxigen […] Yet the extension of the term 'water' was just as much H_2O on Earth in 1750 as in 1950 […][6]

Kuhn questions that the term «H_20» has the same extension as the term 'water' such as this term was used in 1750. The extension of the term «water» from the perspective of present chemistry or of 1950's chemistry is the set of samples of H_20, regardless of whether these samples are in solid, liquid, or gaseous state. But Kuhn claims that this was not so from the perspective of 1750's chemistry. According to Kuhn, at that time, namely, before the so-called «Chemical Revolution», which took place in the 1780's, it was considered that different chemical substances corresponded to the three states of aggregation, to wit, to the solid, liquid, and gaseous states, as it was regarded that a chemical substance could only exist in *one* of these three states, and water was conceived in 1750's chemistry as having the essential property of *being a liquid*.[7] Hence the term «water», such as this term was used in 1750's chemistry, would not be co-referential with the term «H_20», but with the term «liquid H_2O» (or «close-packed H_2O particles in rapid relative motion»). From here it follows that the term «water», such as it was used in 1750's chemistry, is not co-referential or coextensive with the term «water», such as it is used in present chemistry; from the perspective of 1750's chemistry a piece of ice would not belong to the extension of the term «water», while it does belong to the extension of the term «water» from the perspective of present chemistry or of 1950's chemistry. Thus in this case the relation sameness-of-kind has not remained stable through theory change, at least through changes in our scientific or rather meta-scientific theories about the notion of sameness-of-kind concerning chemical substances.

This is Kuhn's main objection against Putnam's theory and in general against the causal theory of reference, an objection which concerns the thesis of reference change, but Kuhn formulates in [1989] and [1990] two further objections against the causal theory, which bear on the so-called essential properties. On the one side, Kuhn attributes to the causal theory the thesis that only a single essential property determines the reference of each natural kind term, and then he objects that more than one essential property is required to determine the reference of the term «water» such as it was used in 1750, namely, the properties of being H_2O and of liquidity.[8] However, the causal theory does not seem to be committed to the thesis that Kuhn attributes to the proponents of this theory; there is no reason why Putnam or other causal theorists could not accept that the reference of a term is determined not by a single property but

---

6 Putnam [1975a], p. 224.

7 Kuhn [1989], p. 28 and [1990], p. 311.

8 Kuhn [1989], pp. 26 and 29; [1990], pp. 309 and 312.

by a conjunction of properties. On the other side, Kuhn questions the frequent assimilation by Putnam and in general by causal theorists of theoretical properties to necessary or essential properties and of so-called superficial properties or macroscopical properties to contingent ones. Kuhn claims that the so-called superficial properties are as necessary as the theoretical ones, since if a theory that posits the relevant theoretical properties could not predict the superficial properties or at least some of them, it would not be taken seriously. Thus Kuhn claims that, as theoretical properties have been posited to explain and predict superficial ones, the latter will be as necessary as the former.[9] Kuhn's remark is motivated by a certain underestimate of the so-called superficial properties by Putnam and in general by causal theorists; the point of the causal theory is just that internal structural properties are more determinant of a natural kind term's reference than macroscopical properties, since samples which share the former but not the latter ones should be still considered as samples of the kind, while samples which share the latter but not the former ones should not be regarded as samples of the kind. This contrast between internal structural properties and macroscopical properties could be formulated in the following way: every member of the kind has to have the internal structural properties that determine the kind, but it is not necessary that every member of the kind have all the macroscopical properties usually associated with the kind. Nevertheless, this contrast between internal properties and macroscopical ones is compatible with Kuhn's point that macroscopical properties play an important role in the determination of the reference of natural kind terms; they contribute to specify the relevant relation sameness-of-kind, since the internal properties which constitute this relation will be those that are responsible for such macroscopical properties.

2. Causal Theory and Reference Change

After having taken into account these two further objections let us return to Kuhn's main objection against Putnam's theory and in general against the causal theory of reference. Kuhn argues that the extension of the term «water» has changed between 1750 and 1950, since the relation sameness-of-kind has not remained stable through theory change. Now, if the causal theory of reference does not allow changes of reference, then the causal theory should be rejected.

The claim that theory change may involve changes of reference is very plausible and can be justified on the basis of the thesis that the relation sameness-of-kind does not only depend on the world, but also on our theories or conceptions about the notion of sameness-of-kind; thus changes in these conceptions may entail changes in the relation sameness-of-kind and so may result in changes in the reference of natural kind terms. This conclusion is illustrated and supported by Kuhn's historical example. Concerning this historical example I find advisable to make the following three remarks. First, I do not commit myself to the veracity of Kuhn's historical example; nevertheless, if someone would question the historical accuracy of Kuhn's example, I would ask him to regard it as a thought experiment. Kuhn's historical example, regarded as a thought experiment, illustrates and supports Kuhn's conclusion that the

---

9 Kuhn [1989], pp. 29-30 and [1990], pp. 312-313.

relation sameness-of-kind does not need to remain stable through theory change. Second, the scope of Kuhn's conclusion is, however, not very far-reaching since changes in our scientific theories do not usually entail changes in our conceptions about the notion of sameness-of-kind and so in the relation sameness-of-kind. Third, Kuhn's historical example may also be interpreted as supporting the plausibility of the claim that there have been modifications in the sorts of samples involved in the determination of the reference of natural kind terms.

Anyway, in regard to Kuhn's main objection against Putnam's theory and in general against causal theories of reference it must be noted that, though classical versions of the causal theory seem to make reference change impossible, not every version of the causal theory has to be committed to the immutability of reference. In fact, it may — and should — be allowed that the reference of a term is not only determined by the use of the term in the supposed initial dubbing but also by subsequent uses of the term.[10] Modifications in our uses of terms — due to mistakes or to deliberate choice — may be accompanied by changes of reference, especially if those modifications involve changes in the samples determining the reference of the term or in the notion of sameness-of-kind.

In regard to the first sort of change it is advisable to mention a possible interpretation of Putnam's reference theory according to which the samples determining the reference of a natural kind term are not the original ones, but the so-called «paradigmatic samples», which do not need to be the same as the former ones and which will be determined by experts. In a seldom quoted paper, Putnam [1975b] — written in 1974 -, he makes some claims which could be interpreted in that way; for instance, after asserting that «we may no longer care about the original use of [a] term»,[11] Putnam regards as dubbers and as initiators of a chain of transmissions of a term «the original dubber, or the relevant expert».[12] Thus the reference of natural kind terms would be determined by paradigmatic samples and by the relation sameness-of-kind. Now, if the choice of paradigmatic samples is taken by experts, it is plausible to assume that their choice would depend partly on their theories so that theory change might involve changes in the paradigmatic samples and therefore may result in changes of reference.

Nevertheless, one can accept that there have been changes of reference, while rejecting the thesis of reference change; this thesis is stronger since it amounts to the claim that a change of theory *always* entails some changes of reference.

---

[10] Thus some authors such as M. Devitt assert that the reference of terms is multiply grounded; see Devitt/Sterelny [1987], pp. 62-63 and 71-72.

[11] Putnam [1975b], p. 274.

[12] Putnam [1975b], p. 275.

3. The Justification of the Thesis of Reference Change

Kuhn's justification for the thesis of reference change rests on two sorts of doctrines, one semantic and the other ontological. The ontological doctrine is a sort of antirealism which has been called by some authors, such as R. Nola, «relativist idealism».[13] According to this doctrine a change of theory involves a change of world, or rather of phenomenal world, though not of the world-in-itself. Now, since the concept of reference expresses a relation between language and world, and since the world-in-itself is assumedly inaccessible to us, the reference of terms must be given in the phenomenal world corresponding to a theory. And as a change of theory entails a change of phenomenal world, a change of theory would also involve changes of reference. Relativist idealism makes the thesis of reference change very plausible, but if the former were the only justification for the latter, the thesis of reference change would have no interest by itself, since the acceptability of this thesis would depend on the previous adoption of a certain ontological position.

However there is another way to vindicate the thesis of reference change, namely, to assume a sort of *description theory of reference*. According to the description theory, in its modern or cluster form, the reference of a natural kind term is determined by a cluster of descriptions that speakers associate with the term, where it is allowed that some descriptions are more central than others for the determination of the reference; to the extension of a term will belong the entities which satisfy, in John Searle's words, a «sufficient but […] unspecified number» of such descriptions.[14] This last condition can be modified in order to obtain stronger and weaker versions of the description theory; a strong version would demand the satisfaction of all or, at least, of most the descriptions associated with the term. The description theory may incorporate the distinction between experts and non-experts mentioned by Putnam and claim that the relevant cluster of descriptions consists of descriptions that experts associate with the term or, at least, that they are the most central ones. But since it has to be expected that experts who support rival theories will associate different and even incompatible descriptions with a term, the cluster description theory may make plausible the thesis of reference change. Although Kuhn did not specify precisely how he thinks that the reference of natural kind terms is determined, he made some proposals about how natural kind terms are learned which may also be interpreted as proposals about how the reference of natural kind terms is determined, and which agree with the approach of the description theory. According to these proposals kind terms are learned — and their reference is determined — by recourse to exemplary members of their extensions — which will be ostended to or described — and to symbolic generalizations which contain these terms. Kuhn characterizes symbolic generalizations[15] or, for short, generalizations as sentences of a theory which have the

---

[13] See Nola [1980]. Other authors, such as M. Devitt, call this doctrine «constructivism»; see Devitt [1984].

[14] Searle [1969], p. 169.

[15] Kuhn [1970], pp. 182 f.

form of universal sentences or which can be easily put in that form; to the generalizations of a theory belong specially the laws of the theory. Kuhn's proposals about how the reference of natural kind terms is determined seem to constitute, at least partly, a sort of description theory, where the respective descriptions will be extracted from generalizations containing the terms. Now, insofar as theory change involves a substantial alteration of generalizations of the theory, it is likely to be accompanied by change of reference; thus the thesis of reference change becomes plausible, and especially if a *strong version of the description theory* is endorsed which requires the satisfaction of most the descriptions obtained from generalizations.

Although at present pure description theories do not have many followers, there is a certain agreement that some sorts of descriptions have to play a role in the determination of the reference of natural kind terms; for this reason most of present versions of the causal theory are *descriptive-causal* ones. Among those sorts of descriptions are the following.[16] First, the determination of the reference of a natural kind term requires the association with the term of a sortal or categorial term which contributes to eliminate the indeterminacy of ostension, for instance, the sortal term «metal» in the case of the term «gold». Second, descriptions of certain observable properties of the samples and descriptions which ascribe to the samples certain causal powers are needed to specify the relevant relation sameness-of-kind, since the internal properties which constitute this relation will be those that are responsible for such observable properties[17] and causal powers. Third, in the case of reference to natural kinds which are unobservable, it is necessary to use a description of the causal mechanism through which it is assumed that unobservable entities produce certain observable phenomena.

But if it is conceded that such sorts of descriptions have to play a role in the determination of the reference of natural kind terms, it must be admitted that variations in those descriptions may result in changes of reference. In this regard it may be relevant to take into consideration a plausible claim concerning the reference of names made by Putnam in [1973]:

> [U]nless one has *some* beliefs about the bearer of the name which are true or approximately true, then it is at best idle to consider that the name refers to that bearer in one's idiolect.[18]

Now, this remark concerning the reference of names may be applied to the reference of the rest of the terms, including natural kind terms; thus the referent of a natural kind term must be so as to make true or approximately true at least some of our beliefs or theories about it or rather some of the beliefs or theories of experts about natural kinds.

---

[16] See, e.g., Devitt/Sterelny [1987], Sterelny [1983], Sankey [1997] and other references given in Sankey's paper.

[17] This remark agrees with one of Kuhns's aforementioned objections to the causal theory of reference.

[18] Putnam [1973], p. 203; Putnam's italics.

For this reason the question of what is the referent of a natural kind term, such as it is used by experts, cannot be completely independent of what their beliefs or theories are. Thus it may not be excluded that drastic changes in experts' beliefs or theories be accompanied by changes of reference.

The proponents of the thesis of reference change may regard this conclusion as rather disappointing, since their thesis is stronger, namely, a change of theory *always* entails changes of reference. But in order to justify this thesis, and according to the aforementioned remarks, they would have to argue for an unplausible antirealist position, such as relativist idealism, or for a strong version of the description theory of reference. It is their turn to put forward good arguments for either of these claims.

References

Devitt, M. [1984]: *Realism and Truth*. Oxford: Blackwell; 2nd. ed., 1991.

Devitt, M. y K. Sterelny [1987]: *Language and Reality*. Oxford: Basil Blackwell.

Kuhn, Th.S. [1962]: *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press; 2d. ed., enlarged, 1970.

[1970]: «Postscript-1969». In Kuhn [1962] 2d. ed., pp. 174-210.

[1979]: «Metaphor in Science». In A. Ortony (ed.), *Metaphor and Thought*, Cambridge, Cambridge U.P., pp. 409-419.

[1983]: «Commensurability, Comparability, Communicability». In P.D. Asquith and T. Nickles (eds.), *PSA 1982*, East Lansing, Philosophy of Science Association, vol. 2, pp. 669-688.

[1989]: «Possible Worlds in History of Science». In S. Allén (ed.), *Possible Worlds in Humanities, Arts, and Sciences*, Berlin, de Gruyter, pp. 9-32.

[1990]: «Dubbing and Redubbing: The Vulnerability of Rigid Designation». In C.W. Savage (ed.), *Scientific Theories* (*Minnesota Studies in Philosophy of Science*, vol. 14), Minneapolis, University of Minnesota Press, pp. 298-318.

[1991]: «The Road Since Structure». In A. Fine, M. Forbes, and L. Wessels (eds.), *PSA 1990*, East Lansing, Philosophy of Science Association, vol. 2, pp. 3-13.

[1993]: «Afterwords». In P. Horwich (ed.), *World Changes. Thomas Kuhn and the Nature of Science*, Cambridge, MIT Press, pp. 311-341.

Nola, R. [1980]: «'Paradigms Lost, or the World Regained' — An Excursion into Realism and Idealism in Science». *Synthese*, 45, pp. 317-350.

Putnam, H. [1973]: «Explanation and Reference», en G. Pearce y P. Maynard (eds.): *Conceptual Change*, Dordrecht, Reidel. Reimp. en Putnam [1975], pp. 196-214.

[1975]: *Mind, Language and Reality. Philosophical Papers*, vol. 2, Cambridge, Cambridge University Press.

[1975a]: «The Meaning of 'meaning'», en K. Gunderson (ed.): *Language, Mind and Knowledge* (*Minnesota Studies in the Philosophy of Science*, vol. 7),

Minneapolis, University of Minnesota Press. Reimp. en Putnam 1975, pp. 215-271.

[1975b]: «Language and Reality», en Putnam [1975], pp. 272-290.

Sankey, H. [1997]: «Incommensurability: The Current State of Play». *Theoria*, 12, pp. 425-445.

Searle, J. [1969]: *Speech Acts. An Essay in the Philosophy of Language*. Cambridge: Cambridge U.P.

Sterelny, K. [1983]: «Natural Kind Terms». *Pacific Philosophical Quarterly*, 64, pp. 110-125.

Luis Fernández Moreno

University Complutense of Madrid. Spain

<luisfm@eucmos.sim.ucm.es>

SORITES (ΣΩΡΙΤΗΣ), ISSN 1135-1349

Issue #11. December 1999. Pp. 15-23.

Was Frege Wrong when Identifying Reference with Truth-Value?

WAS FREGE WRONG WHEN IDENTIFYING REFERENCE WITH TRUTH-VALUE?

by Jean-Yves Béziau

0. Introduction

Frege's thesis according to which the reference (*Bedeutung)* of a sentence (*Satz*) is a truth-value (*Wahrheitswert*)[1] is one of the most controversial aspect of his philosophy. Many people think that it is against common sense, according to which the reference of a sentence is a state of affair, a fact.

Some people have even tried to show that Frege was absolutely wrong. This is the case of G.Sengupta in a two pages paper (Sengupta 1983) which seems to have been approved by Dummet (cf. the footnote attached to the last sentence of the paper). Here are the first two sentences of the paper:[2]

> A fundamental assumption in Frege's semantics (henceforth A1) is that the customary reference of a declarative sentence is its truth-value. The purpose of this paper is to prove that A1 is false.

One of the main difficulties in discussing this kind of things is that although Frege's work is the origin of many basic concepts of modern logic, they have been seriously transformed. A typical example is the Fregean stroke (). It is difficult to know exactly what was its exact meaning for Frege (in fact Frege changed several times of idea) but one thing is sure when we interpret «I — *P*» as meaning «*P* is logically true», we are using a conceptual framework which is quite different from Frege's one although Frege's work can be considered as its source.

However self-incoherent interpretations cannot be used against Frege, and it seems that Sengupta's argumentation is based on such an interpretation.

---

[1] We will not discuss here terminological problems and we will stick to these conventional translations. For a discussion about Frege's notion of *Bedeutung*, see e.g. (Angelelli 1982).

[2] Our analysis will lead us to quote most of Sengupta's short paper, so that it is not necessary to read it in order to understand Sengupta's argumentation and our refutation of it.

After stating various possible interpretations of Frege's principle of substitution (section 1), we show that there is no coherent interpretation under which Sengupta's argumentation is valid (section 2). Finally we try to see how Frege's distinction can work in the context of modern mathematics and how modern logic grasps it (section 3).

1. Substitution

Let us first quote a fundamental extract of Frege's *Sinn und Bedeutung* where he justifies in a sense his option of identifying reference with truth-value:

> Wenn unsere Vermutung richtig ist, dass die Bedeutung eines Satzes sein Wahrheitswert ist, so muss dieser unverändert bleiben, wenn ein Satzteil durch einen Ausdruck von derselben Bedeutung, aber anderm Sinne ersetz wird. Und das ist in der Tat der Fall. Leibniz erklärt gradezu: *Eadem sunt, quae sibi mutuo substitui possunt, salva veritate.* Was sonst als der Wahrheitswert könnte auch gefunden werden, das ganz allgemein zu jedem Satze gehört, bei dem überhaupt die Bedeutung der Bestandteile in Betracht kommt, was bei einer Ersetzung der angegebener Art unverändert bliebe ? (Frege 1892, p.35)[3]

According to Frege, we can therefore state the following substitution principle, which he sees himself as an interpreation of Leibniz's principle:

Frege's substitution principle

It two sentences *Q* and *Q'* have the same truth-value, thus the sentence *P* containing *Q* as a subsentence has the same truth-value as the sentence *P'* that we get from *P* substituting *Q'* for *Q*.[4]

Within the framework of present mathematical logic, this principle can be interpreted in two differents ways: on the one hand taking Fregean truth to be simple truth (truth in a model), on the other hand to be logical truth. Accordingly there are two definitions of substitutions which are not equivalent.

---

[3] If our supposition that the reference of a sentence is its truth value is correct, the latter must remain unchanged when a part of the sentence is replaced by an expression having the same reference. And this is in fact the case. Leibniz gives the definition: «*Eadem sunt, quae sibi mutuo substitui, salva veritate*'. What else but the truth value could be found, that belongs quite generally to every sentence if the reference of its components is relevant, and remains unchanged by substitutions of the kind in question ? (Max Black's translation).

[4] This is already an interpretation of Frege, in fact an adaptation to the case where the Satzteil is itself a *Satz*.

Substitution's principle 1 (S1)

It two sentences Q and *Q'* have the same truth-value in a given world (or model, or valuation),[5] thus the sentence *P* containing *Q* as a subsentence has the same truth-value (in this world) as the sentence *P'* that we get from *P* substituting *Q'* for *Q*.

Substitution's principle 2 (S2)

It two sentences *Q* and *Q'* are logically equivalent, i.e. are true in exactly the same worlds, thus the sentence *P* containing *Q* as a subsentence is logically equivalent, i.e. is true in exactly the same worlds, to the sentence *P'* that we get from *P* substituting *Q '* for *Q*.

As we will see in the next section, the mistake of Sengupta is due to the

fact that he interprets Frege's principle as an incoherent mixture of (S1) and (S2).

We will now make a few remarks about these principles to clarify their meanings and in order to give a basis for the analysis presented in our third section.

First let us note that in modern logic the term *substitution* is used in several different ways. In general by the law (rule or theorem) of substitution it is meant something which neither corresponds to (S1) nor (S2), but the fact that if, in a tautology, we substitute a given sentence for all the occurrences of an atomic sentence, it is still a tautology.

(S2) is generally called the *replacement* theorem (e.g. Kleene's terminology) although it is also sometimes presented under the name *substitution* theorem (e.g. Church's terminology).

(S1) itself rarely appears under such a name. This principle is most of the time not stated explictly. It is obviously true in any matrix's semantics. In particular if we say that a logic is *truth-functional* iff it can be characterized by a finite matrix,[6] (S1) holds in every truth-functional semantics. The validity of (S1) in matrix's semantics is due to the fact that in this case the truth-value (under a given valuation) of a compound sentence is a function of the truth-values of its components.

It is possible to prove that (S2) holds in every truth-functional bivalent logic (i.e. logic which can be charcaterized by a two-valued matrix and therefore for which (S1) holds); see e.g. (Béziau 95). It is a consequence of the fact that from the viewpoint of a two-valued matrix, we can replace in (S2) «is true» by «have the same truth-value», i.e. in this case S2 is equivalent to the following principle:

---

[5] We are not precise in order to include the widest range of semantics (sentential, first-order, Kripke, etc.)

[6] We will stick to this definition, which seems to be the implicit one when someone says that modal logics or intuitionistic logic are not truth-functional. For a discussion about this question, see (Béziau 1997).

Substitution's principle 3 (S3)

If two sentences *Q* and *Q'* have the same truth-values in exactly the same worlds, then the sentence *P* containing *Q* has the same truth-values in exactly the same worlds as the sentence *P'* that we get from *P* substituting *Q'* for *Q.*

It is clear that there are some logics in which (S2) holds but not (S1). For example if we consider the current modal logics, from the point of view of Kripke's semantics, (S2) holds but not (S1).[7]

2. Refutation of Sengupta's proof

These definitions being made, let us turn to Sengupta's interpretation and argumentation:

> We shall take for granted the verity of the assumption that the truth-value of a declarative sentence is a function of the references of its parts (henceforth A2). A2 is not only in conformity with Frege's view, but also entailed by Leibniz's principle. A consequence of A1 and A2 is that the truth-value of a declarative sentence containing another as part remains unchanged when the part is replaced by another sentence having the same truth-value, provided that the part as part has only customary reference and expresses a complete thought. Since we have taken the verity of A2 for granted, if the consequence is proved to be false so is A1.

Let us call A3 what Sengupta calls «a consequence» of A1 and A2, deleting the final part which is in fact independent of Sengupta's mistake. Thus we have the following assertions:

> A1. The customary reference of a declarative sentence is its truth-value.
>
> A2. The truth-value of a declarative sentence is a function of the references of its parts.
>
> A3. The truth-value of a declarative sentence containing another as part remains unchanged when the part is replaced by another sentence having the same truth-value.

A3 looks very much like Frege's substitution principle. In particular the question if we must interpret it as (S1) or as (S2) is left open. However we can remark that Sengupta articulates A1, A2 and A3 in a particular way. To take A3 as a consequence of A1 and A2 seems to choose to interpret A3 as (S1). It is not obvious that this articulation corresponds to Frege's one.

---

[7] Łukasiewicz's three-valued logic, in which both (S1) and (S2) hold, was supposed to formalized the notion of possibility, but nowadays nobody considers this logic as a modal logic.

Sengupta gives the following description of the example, according to which he will (allegedly) prove that A3 is false and that therefore, A2 being assumed, Frege cannot claim A1:

> Let us consider the following sentences assuming that Srimati (…) detests long hair:
>
> 1. Two plus two is equal to four
>
> 2. Srimati detests long hair.
>
> 3. It is unfortunate for Ranjan that Srimati detests long hair.
>
> (…)

Sentence 1 is necessarily true and under the assumed circumstances 2 is also true.

We can thus say that Sengupta chooses a world *w* («the assumed cicrumstances») in which 2 is true and 1 also, because according to him the sentence 1 is true in all the worlds («is necessarily true»). Let us note that the sentence 1 is not very well chosen in the sense that the fact that it is a necessary truth is controversial. It would be better to take a tautology like:

1'. If Srimati detests long hair then Srimati detests long hair.

Then Sengupta goes on as follows:

> Now, if Frege were right in assuming that the customary reference of a declarative sentence is its truth-value, then 1 and 2 would be coreferential, and substituting the one for the other in sentence 3 would have no consequence for its truth-value, provided that the embedded sentence in sentence 3 had only customary reference and expressed a complete thought.

After showing that 3 had only customary reference and expresses a complete thought (parts of the argumentation which is of no interest for us here), Sengupta concludes his paper as follows:

> The consequence of substituting 1 for 2 in sentence 3 remains to be seen.The substitution does not necessarily preserve the truth-value of sentence 3. We can easily conceive of possible worlds in which the fact that Srimati detests long hair is unfortunate for Ranjan, but not the fact that two plus two is equal to four. A1 is thus proved to be false.

What Sengupta is saying is that there are some worlds in which 3 is true and the following sentence 3' is false:

3'. It is unfortunate for Ranjan that two plus two is equal to four.

But what can we conclude from that ? All we can say is that 3 and 3' are not logically equivalent. But 1 and 2 are not logically equivalent. Thus this does not contradict (S2).

Imagine now that Sengupta consider that the consequence A3 of A1 and A2 is (S1) and not (S2). How can he say that (S1) is false, and that assuming A2, therefore he has proved that A1 is false ?

In the given world *w*, taking A2 for granted, 3 and 3' should have the same truth-value,[8] since in *w* 1 and 2 have the same truth-value. Thus (S1) is not contradicted.

Therefore the consequence A3 of A1 and A2, should it be (S1) or (S2), is not proved to be false.

In fact it seems that Sengupta in order to refutate Frege is using the following principle of substitution, which is an absurd mixture of (S1) and (S2) that no one would defend:

Sengupta's substitution principle

If two sentences *Q* and *Q'* have the same truth-value in a given world, thus the sentence *P* containing *Q* as a subsentence is true in exactly the same worlds as the sentence *P'* that we get from *P* substituting *Q'* for *Q*.

3. Reference as class of models

We will now try to show how Frege's distinction can be articulated within the framework of mathematics and how modern logic captures it. This account will shed a new light on the relations between truth-functionality, extensionality and intensionality.

Most people identify truth-functionality with extensionality, and therefore, taking intensionality as the opposite of extensionality, they identitfy non-truth-functionality with intensionality. According to these views, current modal logics are intensional because they are not truth-functional.

Our proposal leads us to think that extensionality is expressed by (S2) and that it differs from truth-functionality (only bivalent truth-functionality entails extensionality in the sense of (S2), as remarked in the first section). In particular current modal logics are extensional (because (S2) holds) even if they are not truth-functional.

The solution of the identity paradox within present mathematical logic and the construction of a real intensional logic seems therefore open problems.[9]

Let us consider the axioms for complemented distibutive lattices, in short CDL. We can say that the reference (*Bedeutung*) of these axioms is the class of their models. That is to say, following Tarski's idea, the class of structures in which they are true. This same class can be given in many other ways, that is to say, with different sets of axioms. For example the axioms IR for idempotent rings.

---

[8] We will not discuss here the question if it is appropriate to think that A2 applied to the sentence 3. This is what Sengupta assumes and assumes that Frege assumed.

[9] For more details on this question see (Béziau 1994).

The fact that CDL and IR refer to the same thing, the class of boolean algebras, is not necessarily evident. This was proved by Marshall Stone after a tedious conceptual work and was a fundamental step for the proof of his famous representation theorem, cf. (MacLane 1981). This result was an important discovery of the same kind as the discovery that Hesperus and Phosphorus refer to the same object.

A boolean algebra can be seen as a complemented distributive lattice or as an idempotent ring, these are two different ways of looking at the same object. CDL and IR are two different manners of having access to one and the same thing. They are two different *meanings* for the same reference, according to Frege saying that the meaning (*Sinn*) is *the way of giving* (*die Art des Gegebenseins*) the reference (*Bedeutung*), cf. (Frege 1892, p.26).

Using the extension/intension terminology, we can say that CDL and IR are two different intensions for the same extension.[10]

Because the replacement theorem is valid in classical first-order logic, formulas (or set of formulas) having the same extension, can be identified (the relation of logical equivalence is a congruence). This is what happens with CDL and IR formalized in the context of classical first-order logic.

In fact classical first-order logic minimizes the rôle of meaning, interpreted along the above lines, and is not able to give an account to it.

From the viewpoint of the mathematician, the difference of meanings between CDL and IR appears relatively clearly: CDL is formulated in the language of order and IR in the language of function, these two languages corresponding to two different basic intuitions. Of course it is a rough distinction and no mathematician has given yet a precise definition which supports such kind of theory of meaning.[11] But it seems that it fits Frege's view according to which the meaning (*Sinn*) of a sentence (*Satz*) is a thought (*Gedanke*). We can say that what the mathematician feels and tries to explain is that CDL and IR are two different ways of thinking (at the same thing).

In first-order logic the difference of these two languages is very tiny. In fact within first-order logic what is emphasized is the possibility of reduction (modulo the replacement theorem): for example, functions can be defined as predicates.[12]

Within the framework of a (classical first-order) modal logic in which the replacement theorem holds (which is the case of current modal logics), the two following sentences are equivalent:

Stone proved that a complemented distributive lattice is a boolean ring.

---

[10] Here we have the following equation: intension = comprehension = axiomatization. The axioms are the comprehensive way of giving the extension, i.e. the class of models.

[11] However Bourbaki's description of mathematics (as it appears in Bourbaki 1950) gives a key to such a theory.

[12] But model theory reaches great achievements through this line, showing that if we succeed to express a class of structures in a particular way, it reveals important properties of it.

Stone proved that a boolean ring is a boolean ring.

Therefore modal logics do not solve the identity paradox. According to them Stone, like George IV, is the son of La Palice. And the reason why is that they are purely extensional and are not able to express the distinction between reference (*Bedeutung*) and meaning (*Sinn*).[13]

Bibliography

Angelelli, Ignacio. 1982. «Frege's notion of 'Bedeutung'». *Logic, Methodology and Philosophy of Science VI*, L.J.Cohen *et al.* (eds), North Holland, Amsterdam, pp.735-753.

Béziau, Jean-Yves. 1994. «Du Pont's paradox and intensional logic». *Logica'93 (Proceedings of the 7th International Symposium)*, P.Kolár and V.Svoboda (eds.), Filosofia, Prague, pp.62-65.

Béziau, Jean-Yves. 1995. *Recherches sur la logique universelle*. PhD. Denis Diderot University (Paris 7).

Béziau, Jean-Yves. 1997. «What is many-valued logic ?» *Proceedings of the 27th Symposium on Multiple-Valued Logics*, IEEE Computer Society, Los Alamitos, pp.117-121.

Bourbaki, Nicolas. 1950. «The architecture of mathematics». *American Mathematical Monthly*, **57**, pp.221-232.

Frege, Gottlob. 1892. «Über Sinn und Bedeutung», *Zeitschrift für Philosophie und philosophische Kritik*, **100**, pp.25-50.

Gabbay, Dov M. and F.Günthner (eds.). 1986. *Handbook of philosophical logic, vol.II*, Reidel, Dordrecht.

MacLane, Saunders. 1981. «History of abstract algebra: origin, rise and decline of a movement». *American mathematical heritage: algebra and applied mathematics*, D.Tarwater *et al.* eds, Texas University, pp.l1-35.

Sengupta, Gautam. 1983. «On identifying reference with truth-value». *Analysis*, **43**, pp.72-74.

Jean-Yves Béziau

Universidade Federal Fluminense. Instituto de Matemática

Niterio, Rio de Janeiro, Brazil

<jybeziau@hotmail.com>

[13] In the above example, we can put instead of «Stone proved»: «Stone believed», «Stone thought», «It is necessary», «It is possible», etc, all these expressions being considered as interpretations of operators of current modal logics or similar logics (epistemic logics, provability logic, etc.). The fact that in the literature these logics are quite commonly presented as intensional logics is typically illustrated by the cover of the second volume of the *Handbook of Philosophical Logic*: it is said that it «surveys the most significant «intensional» extensions (sic) of predicate logic» (Gabbay 1986).

SORITES (ΣΩΡΙΤΗΣ), ISSN 1135-1349
Issue #11. December 1999. Pp. 24-40.
Quasi-Indexical Attitudes

QUASI-INDEXICAL ATTITUDES

by Tomis Kapitan

1. Introduction

Indexicals are inevitably autobiographical, even when we are not talking about ourselves. For example, if you hear me say, «That portrait right there is beautiful,» you can surmise not only that I ascribe beauty to an object of my immediate awareness but also something about my spatial relation to it. Again, if I praise you directly within earshot of others by using the words, «You did that very well!,» my concern need not be to cause them to think the exact thought I have; they might not be in a position to address you as *you* and I might not care what they think of your performance. My purpose is to get them to ascribe to *me* an attitude that I express with a second-person indexical, to convince them that I am an encouraging and supportive person inasmuch as I addressed someone with words of praise. Indexicals are autobiographical not only because they issue from a speaker — all utterances do — but because they reveal something about the speaker's orientation toward and encounter with objects in a way that non-indexical language fails to do.

For this reason, care must be taken in reporting indexically-expressed thoughts. Suppose the Chair of my Department informs me,

(1) I am upset about the Dean's report.

I cannot relate what he said by reiterating his words within indirect discourse, viz.,

(2) The Chair said that I am upset about the Dean's report.

Because 'I' expresses speaker's reference, my assertion of (2) would cause a hearer to misconstrue who is said to be upset.[1] Alternatively, the sentence,

(3) The Chair said that the Chair is upset about the Dean's report.

---

[1] By contrast, reiteration works for attitudes expressed non-indexically. If the Chair had said, «The Provost is upset about the Dean's report,» then my reiterating his words in indirect discourse would be appropriate. The disparity is well known (Kaplan 1989, 553). The task is to explain it?

loses the critical first-person perspective that the Chair meant to convey. If first-person reference is ineliminable, as often argued,[2] then our ascriptions should be sensitive to indexical usage in a way that (3) is not. One might try the *oratio recta*,

(4) The Chair said: I am upset about the Dean's report.

However, this is inapplicable to attitudes an attributee is not disposed to express. More importantly, to have explanatory worth a direct quotation must be supplemented by an interpretation of what the speaker meant, and this is naturally expressed through the indirect format, for instance, «In saying 'I am very upset about the Dean's report' the Chair meant that…» The apparent advantages of direct discourse are illusory.

Although natural languages provide little means for perspicuous ascriptions of indexical thoughts, Castañeda pointed out that certain linguistic types lend themselves to some such use. Consider,

(5) The Chair said that he himself is upset about the Dean's report.

Here, 'he himself' is used as a *quasi-indicator* inasumch as it represents the indexical reference the Chair expressed through 'I', and by employing this reflexive pronoun, the attributor expresses his own *quasi-indexical attitude*. But our quasi-indexical vocabulary is sparse, and there is a problem of explaining how it succeeds in capturing another's thoughts. Indexicals, we are taught, are context-sensitive because their tokens reflect the speaker's *perspective*.[3] My *this's*, *that's*, *you's*, *beyond's*, etc. express what they do partly because they issue from a unique spatio-temporal vantage point that I happen to occupy. From your perspective, my *here* might be your *there*; my *you*, your *she*; and within my own perspective, a *this* differs from a *that*, and one *there* might differ from another *there*. How is it, then, that a distinct listener processing a speaker's indexical utterances can understand *what* that speaker is saying, much less convey this to a third party? How can quasi-indicators accurately depict the indexical references of others? What exactly are quasi-indicators and what is the precise content of quasi-indexical attitudes?

These questions are not mere curiosities within the philosophy of language. They have considerble practical significance. Quasi-indexical attitudes permeate social life; not only do we explain behavior by reference to the indexical thoughts of people, but many of our deepest emotional reactions are responses to our own interpretations of what others think, believe, intend, and feel, attitudes they would most likely express indexically. In criminal courts, for instance, lawyers, judges and jurors try to determine

---

[2] See, for example, Castañeda 1966, 1967, and 1989a, 70-76. Similar claims are made in Perry 1979 and Lewis 1983, chp. 10. Castañeda criticized familiar reductions of indexicals, say, 'I', to 'this person now speaking,' for imputing too much conceptual apparatus to their users, particularly, small children (1989a, 72-5). Nor are the differences in grammatical person reducible to each other, in particular, first-person indexicals are not the most basic (Castañeda 1990a, 736).

[3] See for example, William James, *Essays in Radical Empiricism*, «The Experience of Activity,» note 14 (Dutton), who describes the expressions 'I' 'here' and 'this' as «primarily nouns of position.» See the perspectival approach in Castañeda 1967 and also Forbes, 1989, p. 470, who explains the difference in sense of two tokens of 'that telephone' in terms of «differing viewpoints.»

the precise intentions with which a defendent acted, yet intentions are saturated with indexical references, from the first-person thoughts about what *I* shall do to the demonstrative references used in guiding action, e.g., *I will shoot the guard standing there*.[4] Our respect for a person's moral character might depend upon our judgment that he or she acted from duty, precisely, what he or she took to be *his* or *her* duty — where 'his' and 'her' are used to mark first-person commitments. Our empathic feelings for one who has tried and been unsuccessful, or our resentment over an undeserved triumph, involve not only our awareness of another's situation but also of the sentiments he or she might convey through «I have failed again» or, alternatively, «*Veni, vidi, vici*!» Such recognitions underlie our *reactive attitudes* — respect, sympathy, resentment, blame — states that are vital to our social consciousness and perhaps lacking in beings whose perceptions and communications are otherwise replete with indexicality.[5] Articulating their structure, and that of the quasi-indexical attitudes from which they emerge, is essential to understanding the psychology of social interaction.

2. Indexical Interpretation vs Indexical Production

Quasi-indicators are parasitic on acts of indexical reference. According to the standard token-reflexive or utterance-reflexive accounts championed by David Kaplan, John Perry and others, one interprets an indexical token by applying the *character* (sense) associated with its linguistic type to the *context* of utterance thereby determining the token's *content* (referent).[6] The character can be expressed through a rule of interpretation, for example, that associated with 'I' might be formulated as,

(I) A token of the first-personal indexical 'I' refers to the speaker or writer of the utterance in which it occurs.

whereas that of 'now' is given by,

(N) A token of the temporal indexical 'now' refers to the time of the utterance in which it occurs.

For example, upon hearing you say, 'I am now going to throw the ball over there', my grasping the characters of 'I', 'now' and 'there' and knowing that you uttered the sentence, when you uttered it, and what region you demonstrated, enable me to determine what your referents are.[7]

---

[4] See Castañeda 1975, chapters 2 and 4, for a defense of the view that intentions are first-person practical thought contents.

[5] Presumably the higher animals can discriminate spatial relations which we would normally express with *her, there, near, beyond*, etc. Morover, they seem to react appropriately to our demonstrative pointings, and their own interactions may be replete with indexical communication, e.g., the bee's dance (see Millikan 1990).

[6] See, for example, Kaplan 1989, 505-507, 523-524.

[7] See Millikan 1990, 727-728: «to interpret an indexical one must have prior knowledge, one must already know independently and ahead of time, what item bears the indexical's adapting relation to the

Token-reflexive rules like (I) and (N) are fine for *interpreting* indexicals, but for various reasons they explain neither the psychological mechanisms underlying indexical *production* nor the autobiographical dimension of indexical usage. First, one does not have to identify oneself as a speaker, a writer — much less the speaker or writer or producer of a given 'I' token — in order to produce 'I' tokens and think first-person thoughts. Further, any utilization of a rule like (I) presupposes identification of an 'I' token, and this can only happen subsequent to its production. Second, indexical production does not require independent identification of the referent. Demonstration, for example, can occur *autonomously* if I don't know how to classify something that suddenly looms into my visual field, say, other than as *the thing over there* or, simply, as *that*. Third, a token-reflexive rule like (N) reveals nothing about the speaker's involvement or encounter with the referent. It specifies how I, the hearer, can determine an interval when I hear you utter a 'now' token, but it does not inform me how you picked out a time that you referred to. (N) supplies no information about how one is to *apply* the indexical 'now' in the first place. Consequently, rules for the application or production of indexicals must differ from those guiding their interpretation.[8]

Ruth Millikan correctly emphasizes that to interpret an indexical requires an independent means of identifying the referent, but her repudiation of essential indexicals and first-person thoughts ignores the distinction between interpretation and production. Noting that context-sensitive indexical tokens must bear a certain «indexical adapting relation» to their referents — for example, the relation for 'I' is being the producer of the token — she argues that this relation need not be taken into account in action, nor does the indexical signify it:

> …to interpret an indexical one must have prior knowledge, one must already know independently and ahead of time, what item bears the indexical's adapting relation to the indexical token. One must already know both that this referent exists and how it is related to the token, hence to the interpreter. One does not find this out by interpreting the indexical; one needs already to know it in order to interpret the indexical. For example, a token of «I» does not not tell me who the originator of that token is, that it is, say, Alvin. Rather, if I am to understand a token of «I», I must *already know* who the speaker is. (Millikan 1990, 727-728).

Obviously, the interpreter must have an independent means of identifying the speaker to understand a heard 'I' token. But the interpreter is not the producer of that token.

---

indexical token. One must already konw both that this referent exists and how it is related ot the token, hence, to the interpreter.»

[8] See Castañeda 1983, 323 and Recanati 1990, 708-709, and 1993 chp. 4-5. John Perry accounts for cognitive significance in terms of the speaker's understanding that the utterance meets the conditions that the character of the sentence establishes for its truth (1993, 246-7). This requires that the speaker has already identified the utterance and, hence, cannot explain its production. Moreover, Perry requires that a thinker not only conceive of the utterance of the sentence but also that the truth-conditions of the utterance established by the character are satisfied. It saddles the speaker with a higher-order thought about the semantics of utterances that cannot be expected of all speakers.

Millikan's account does not explain how indexical reference origniates, nor does it show that indexical tokens — as applied by the speaker — do not signify an «indexing relation» of token, utterance or producer to the referent. From the *producer's* point of view there must be some such «adapting relation» in order to use an indexical as a referential device. It underlies the mode of presentation correlated with the indexical type (see below), but because it is anchored in the speaker's perspective it is useless to the interpreter in determining the referent without added information about the context.[9]

I conclude that the token-reflexive analysis is appropriate only to the interpretation of indexicals. It is dependent upon the antecedent production of indexical tokens, and very likely cannot even begin without the interpreter's *indexical* identifications of the relevant tokens. To understand quasi-indexical attributions, consequently, we must turn to their source in indexical thinking.

3. Indexicals, Indexical Modes, and Perspective

Indexical terms reflect a direct encounter with items in our experience, whether in perception, imagination, or other types of experiences. They express *thinking references*, that is, acts of consciously picking out some item for the purposes of thinking something about it.[10] All acts of thinking references occur through *modes of presentation*, each of which is a manner of cognizing an item with at least one being an individuating or identifying mode whereby the item is distinguished from everything else. Thinking of the Sears Tower, for example, I consider it as the tallest building in Chicago, or as that monstrous skyscraper over there, or, simply, as *that*. Modes are «guides» for articulating the data of conscious experience, leading many philosophers to view them as ways or manners of apprehension, thus, as properties of thinkers. But this cannot be the whole story; a mode enables one to pick something out only if there is an appropriate fit. I cannot identify something as the object *there* unless it *is* there. However, some caution is needed. If what I identify as the woman across the street is a cross-dresser, then while the mode *being the woman across the street* is not satisfied by that referent it implies a mode that is, say, *being a person across the street*. A satisfied mode corresponds to a property of the referent, and for indexical modes this is always a relational property an item has in virtue of being encountered. Being a *you* or a *this* is a status — an *indexical status* — a thing has only by being experienced in a certain way. Without it, tokens of 'you' and 'this' cannot serve to pick it out.

---

[9] For these reasons, Millikan's repudiation of the essential or irreducible indexical comes to naught. She acknowledges that a person must have some *inner name* for oneself that bears a special relation to dispositions to act, but since one need know nothing about context in order to determine its referent then it is a Millian name whose semantics is exhausted by its referent. In section 3 I argue that the referential use of indexicals make it necessary to invoke indexical «modes of production» inseparable from the producer's perspective.

[10] I follow Castañeda 1989a in this use of 'thinking reference.' Referring *terms* might express an act of thinking reference, as when one thinks out loud, but one's thinking reference must be distinguished from what one intends to commuincate and from what a hearer is caused to think upon perceiving another's token. Each, in turn, is distinct from the denotation, if any, associated with a linguistic type.

If rule (I) does not specify the productive mode that the speaker employs in making first-person reference, what other description is available? The irreducibility arguments block a simple rule of reflexivity like,

(I') A token of the first-personal indexical 'I' is used by the speaker to refer to himself/herself.

Referring to oneself is necessary for a first-person use of 'I' but it is not sufficient. Castañeda (1989c, 42; 1990b, 126) offered this:

(I*) A token of the first-personal indexical 'I' is used by the speaker to refer to himself/herself *qua* self.

An explanation of the '*qua* self' locution is called for, but even as it stands a rule like (I*) reveals something of the mode of production that underlies use of an 'I' token and that is quite distinct from the interpretive mode given by (I).[11]

If an indexical is used referentially, then there must also be *individuating* indexical modes — each a determinate of the character associated with the indexical type — embodying not only a type of encounter but also perspective. My demonstrative in 'this is beautiful' expresses my particular perspective on an item, say, the Hope diamond pictured in a magazine. I might also use 'this' to refer to that very diamond which now appears as a dirty stone before me, subsequently learning, to my own surprise, what I could express by 'this is this' (Burge 1977, 355). The two 'this' tokens reflect a like mode of encounter but each is correlated to a distinct locale within my perspective. Thus, the indexical status a referent has in order to be picked out indexically is as much a matter of *orientation* as it is the thinker's cognitive *encounter* with the referent. The orientation-type associated with 'I' is location at the perspective's point of origin while the encounter-type is one of reflexive awareness *qua* self. Thinking of someone as *you*, on the other hand, is to encounter him as an addressee located in a place distinct from the point of origin yet upon which the subject's utterance can have causal influence. A person with the same orientation may also be the object of a demonstrative encounter expressed through 'he' or 'she'.

Individuating indexical modes are described as follows. Let *i* be the position of an indexical referent X within Y's perspective *p*; the orientation of X is that of *i*-from-the-standpoint-of-*p*, a description with information about the relative distance of X from *p*'s point of origin as well as direction. To accommodate dynamic indexical thoughts like *This is moving fast* where *this* retains its identity though not its spatio-temporal position, *i* can be conceived as an ordering of positions within *p*. Adding to this the

---

[11] Eddy Zemach has questioned whether talk of referring to oneself *qua* oneself is informative (1985, 194); how does the second 'oneself' indicate anything different from the first? Castañeda argued that while the first is a pure reflexive, merely repeating its antecedent, the second conveys one's experiential confrontation with oneself as «a thinker presently involved in the very experience of making the referring [*sic*] in question» (1989a, 170). While he elaborated on this in 1990b, 127-139, he also spoke of «a primitive apprehension of the subject one calls «I,» not mediated by any identification procedure» (129), and of the *I*-properties expressed by first-person pronouns as «indefinable» (1989a, 76). See also my introduction to Castañeda 1998.

encounter-type *k* (whether of the type *I*, *you*, *it*, *he*, *there*, *now*, and so forth) yields this schematic formula for individuating productive modes:

$$\text{Orientation}_{i,p} + \text{Confrontation}_{k} = \text{individuating indexical mode of production}$$

Hence, three factors are involved in analyzing individuating modes: (i) the ordering *i* of positions of the referent within (ii) the agent's perspective *p*, and (iii) the type *k* of encounter. Each is part of the background constituency of an indexical thought and not necessarily a separate referent. The *irreducibility* of indexicals is due to both orientation and encounter-type; their *subjectivity* is due to the uniqueness and privacy of the *p* factor.[12]

4. Indexical Contents

On the «direct reference» view of indexicals, indexical status is not part of what is said and need not be taken into account in specifying the content of indexically-expressed attitudes. This view is soundly motivated when the attitudes in question are more or less stable dispositions; an agent's ways of tracking and reidentifying permanent objects of beliefs and intentions are unlikely to be indexical. However, indexical status is relevant to the contents of *conscious* states of thinking and reasoning (Castañeda 1989b. 126-131). Suppose I believe,

(6) I am obliged, all things considered, to give the annual Medal of Efficiency to Henry at 10 am on May 15.

Having this belief together with the intention to comply with my self-avowed obligation is not enough to explain my subsequent action of giving Henry the medal. When 10 am on May 15 comes around, I must *also* pick out Henry, the medal, and the time and link them to the appropriate elements in my commitment. How is this achieved? Indexically, of course. I see a medal on the table and think that *This medalllion is the Medal of Efficiency*; my attention is directed to the candidates seated in a row of chairs and I realize that *That one is Henry*; I glance at my watch and conclude that *Now is 10 am* or *Now is the time to act*. In each case, I accept observational statements of the form: *i is the same as a*, where *i* is an indexical and *a* is a non-indexical. By their means I infer from (6),

(7) I am obliged, all things considered, to give this medal to that man now.

and from this, the intention,

(8) I shall now give this medal to that man.

My action is explained by my acceptance of (8) and this, in turn, by my acceptance of (7) (Castañeda 1975, chp. 5). The inference from (6) to (7) could not be made if the

---

[12] See note 2 on irreducibility. The subjectivity thesis was articulated in Castañeda 1981, 1989a, 1989b and 1990a. Analogous reasoning appears in Frege 1967, 25-6 with respect to first-person reference; «everyone is presented to himself in a particular and primitive way, in which he is presented to no-one else.» See also Searle 1983, 220-230; McGinn 1983, 17; and Nagel 1986, chps. 2-4. Boer and Lycan 1980 provides a contrasting view of indexical reference, as do Perry 1979 and 1983, Kaplan 1989, and Millikan 1990, but see Castañeda's responses to each of these positions in Castañeda 1984, 249-256; 1983, 313-328; 1989b; and 1990a.

sameness propositions I accept are of the form: *a is the same as a*. Were (7) the very same proposition as (6) then (6) alone should be sufficient for my inferring (8) and explaining my action. Since it is *not* sufficient, (7) must differ from (6), and this difference can only be in the modes associated with the referring expressions. Consequently, indexical modes are relevant to the implicational behavior of indexically expressed propositions.

This conclusion is not based solely on the role of indexicals in action. There are other intuitively acceptable inferences that can also be sanctioned. For example, if it is true *that today is March 26* then it follows *that tomorrow is March 27*, but the latter is not implied by *that George's birthday is March 26* even though George's birthday is today. One can make the same point in conditional form: the counterfactual *if today were March 26 then tomorrow would be March 27* is true, but *if George's birthday were March 26 then tomorrow would be March 27* is not. Again the truth-conditions of the proposition, *I am presently in DeKalb County* — as thought by me — require not merely that a certain organism identical with myself is in DeKalb County, but that this organism *qua* self-reflective is in DeKalb County.

If indexical modes are relevant to truth-conditions and implications, then they are internal to propositional content. If a sentence '*i is F*' contains a term '*i*' that refers to an item *qua* some indexical mode *M*, it does not follow that *M* is a separable component *about* which one thinks in thinking the proposition *i is F*. The indexical '*i*' expresses or connotes *M*, but *M* is itself neither a subject nor a predicated item in *i is F*. It is the unconceptualized manner through which one conceptualizes the referent of '*i*' and by which '*i*' packs the inferential potential it does. Modes are internal to propositional content because they are constitutive of propositional components.[13]

Explaining what immediate indexical referents are is another matter. But whatever theoretical approach is followed, it must accommodate the fact that we not only make indexical references but identify indexical referents with each other and with other thinking referents. When I come to believe that *Henry is that man* I not accepting a trivial identity statement of the form *a=a* governed by Leibniz's Law, otherwise my identification would amount to my accepting nothing more than *Henry is Henry* or *That man is that man*. More is involved in preparing myself for action since I can accept these latter without being prompted to do anything. So, *Henry* and *that man* are distinct in my immediate thought, but I am affirming that they are in some sense the «same thing». Statements to the effect *that a is the same as b* — henceforth abbreviated as, *a≈b* — are informative precisely because they are not statements of identity, rather, of

---

[13] Modes are not, as such, referred to, nor in need of identifying modes of presentation. This allows us to circumvent the regress argument offered in Schiffer 1990, 255. For a different account of the way in which modes of presentation or senses are employed in accounting for indexicals, see Perry 1977, 1983, the concerns raised by Wettstein 1986, and a reply in Perry 1988. For additional discussion of indexical modes of presentation see Evans 1982, chp. 6 and 1985, chp 10; Searle 1983, 220-230; Peacocke 1983, chps. 5-6; Forbes 1987, 14-25; Smith 1989, 71-79; Recanati 1990, 706-715 and 1993, chps. 4-5; and Bezuidenhout 1996. One benefit of taking indexical modes as internal to propositional content is given in my 1993 where I resolve the problems raised by Richard's context-hopping «Phone-Booth» argument.

an equivalence or congruence relation that falls short of identity. As with immediate indexical referents, a theoretical account of congruence awaits a deeper investigation of thinking reference.[14]

5. Attributing Indexical Reference

Because of the perspectivity of individuating modes, a person's complete indexical content is subjective. Yet, to some extent, we can understand both *what* other people refer to indexically and *how* they refer. For one interested in attributing indexical content, the former requires an independent route to the referent, while the latter is achieved by access to generic indexical modes. Let me show how this twofold interpretive strategy can be used to clarify quasi-indexical attributions.

Besides emphasizing their use in depicting indexical references, Castañeda noted that quasi-indicators (a) occur only within the scope of psychological verbs to attribute indexical references; (b) are anaphoric pronouns which are referentially and syntactically dependent upon antecedents occuring outside the scope of those verbs; (c) are not replaceable by these antecedents *salva veritate* let alone *salva propositione*; (d) express what is interpersonal and repeatable; (e) are not themselves indexicals; and (f) express, *in part*, what their antecedents express. As anaphors bound by operators outside attitudinal scope they are more akin to variables than to singular referring terms.[15] In this respect they are like other anaphors embedded within attitudinal scope. Suppose Henry hears his colleague Robert describe another colleague, Alexander, as a fool. If, unlike Robert, Henry knows that Alexander is going to be appointed the next Provost of the university, it would be unfair of him to report,

(9) Robert thinks that the next Provost is a fool.

Not being privy to Henry's information, Robert does not think of Alexander *as* the next Provost. Let us say that 'the next Provost' *occurs externally* in (9) with respect to the property of *being the next Provost* inasmuch as the attributor, Henry, does not assume it to express one of Robert's referential modes. A regimentation using the familiar *de re* format is preferable:

(10) The next Provost is such that Robert thinks that he is a fool.

What does 'he' signify in this ascription? It does not appear to function as a singular term that Henry uses to pick out Robert's thinking referent. It is certainly not a pronoun of laziness if (10) differs from (9), and it is not a demonstrative designation of a thinking referent different from what Henry picks out with 'The next Provost.' Instead, the relation between the antecedent and the anaphor in (10) suggests that 'he' is a

---

[14] See, for example, the Guise-theoretical approach advanced by Castañeda 1989a, chps. 13-14.

[15] Castañeda voiced this view in speaking of first-person self-reference as being reference to oneself *qua* oneself: «…the locution '*oneself* (like its substituends) in the contexts 'ONE refers to ONEself as *oneself*' depicts, is a proxy for, a first-person reference attributed to *one* (or the entity denoted by its substituends). Here '*oneself*' is what I have called a *quasi-indicator*, more precisely, a quasi-indexical variable» (Castañeda 1989c, 38).

variable ranging over thinking referents and bound by a quantifier falling outside attitudinal scope, perhaps on the order of,

(11) $(\exists x)$(x = the next Provost and Robert thinks that x is a fool).

But this is also deficient. For one thing, the employment of the unrestricted variable 'x' obliterates the fact that with 'he' Henry meant to convey that Robert thought of the referent *as* a male. In that case 'he' *occurs internally* with respect to the property of *being a male* because the attributor takes it be one of Robert's modes.[16] For another thing, if the very person Robert takes to be a fool is *identical* to the referent of 'the next Provost' then (11) implies that Robert's full content is *the next Provost is a fool*. But then we have not advanced beyond (9).

We can circumvent these problems by two maneuvers. First, insisting that the variable ranges over thinking referents, (10) does not say that what Robert thinks to be a fool is identical to the next Provost but, rather, that it is congruent to what Henry refers to with 'the next Provost'. Second, to capture Robert's gendered reference let 'x^M' be a complex expression composed of a variable and an internally occurring modifier expressing the attributed mode of *being a male*. (11) can be replaced by,

(12) $(\exists x)$(x $\approx$ the next Provost and Robert thinks that x^M is a fool).

Even greater economy can be achieved by letting 'x_p' range over all only thinking referents that are congruent to what the speaker refers to with 'the next Provost', a maneuver that avoids the conjuctive construction not apparent in (10), yielding,

(13) $(\exists x_p)$(Robert thinks that x_p^M is a fool).

Stipulating that superscripts occur internally while subscripts occur externally, then (13) shows us how 'he' in (10) occurs internally relative to the mode *being a male* as well as externally relative to the property *being congruent to the next Provost.*

The external/internal contrast is vital to understanding quasi-indicators. They are not merely external given their natural habitat within attitudinal scope. Thus, by using 'he himself' in,

(5) The Chair said that he himself is upset about the Dean's report.

I intend to convey how the Chair referred to himself, namely, in a first person way, implying that 'he himself' occurs internally relative to the *self* mode. At the same time, I am reporting the Chair's reference to *him*self, not to *my*self, *your*self, or some other self, and the third-person character of 'him' expresses *my* modes, not the Chair's.

Accordingly, while 'he himself' is internal relative to the generic productive mode associated with the type 'I', it is external with respect to the modes that I, the attributor, express with 'the Chair' and 'he'.

---

[16] The terminology of 'external occurrence' and 'internal occurrence' is employed by Castañeda in 1980, 780-783 and 1989a, chp. 5, and also by Clark 1980, and Forbes 1987. The contrast between external (*de re*) and internal (*de dicto*) properly belongs to occurrences of individual terms and noun phrases, an interpretation upheld in Zalta 1988, 171; Castañeda 1989a, 93-97; Richard 1990, 128; and Kapitan 1993, 1994.

For these reasons, neither of the following is an accurate paraphrase of (5):

(14) (∃x)(x = the Chair and the Chair said that x is upset about the Dean's report).

(15) (∃x)(x ≈ the Chair and the Chair said that x is upset about the Dean's report).

Nor do familiar analyses in terms of senses work. For example, letting bracketed expressions represent senses, e.g., '[Self]' represents the generic self-mode, and '^' express the manner by which senses combine to form complex senses or thoughts, the following won't do:

(16) The Chair said that [[Self] ^ [is upset about the Dean's report]].

since it could not distinguish between what the Chair is said to think from what is attributed to the Provost by,

(17) The Provost said that he himself is upset about the Dean's report.

Proper regimentations must be sensitive to the distinct perspectives expressed by quasi-indicators. Suppose '$[Self]_{\text{the Chair}}$' is a description satisfied by the particularized first-person mode through which the Chair refers to himself *qua* self (at the time in question). Does

(18) The Chair said that $[[Self]_{\text{the Chair}}$ ^ [is upset about the Dean's report]].

accurately paraphrase (5)? I think not. By the subjectivity thesis, the Chair's identifying mode cannot be the individuating mode expressed by my use of 'he himself' (contrary to suggestions in Peacocke 1981, 191 and Forbes 1987, 21). Alternatively, if '$[Self]_{\text{the Chair}}$' is read non-referentially, then the question concerns scope. A small scope Russellian analysis yields something equivalent to,

(19) The Chair said that (∃s)(s and s alone is the Chair's first-person identifying mode and s ^ [is upset about the Dean's report]).

fails to provide a necessary condition of (5) if the Chair does not think of himself *qua* the modes which I, the speaker, express through 'the Chair' or have conceptualized his first-person identifying mode as *predicative*. This is avoided on the large-scope reading,

(20) (∃s)(s and s alone is a the Chair's first-person identifying mode and the Chair said [s ^ [is upset about the Dean's report]]).

which is similar to a proposal in Perry 1983, 25. Yet this Fregean analysis fails to specify a thinking referent. With his use of 'I' in (1), the Chair thinkingly referred to *himself*, a particular, not to referential modes, and he predicated something *of* this particular, namely, a certain emotional state. Such information can be and should be captured in an accurate attribution.

With the treatment of 'he' in (10) as a precedent, a quasi-indicator is best viewed as a complex term expressing (i) an attributee's first-person mode of production, and (ii) the speaker's reference via an antecedent. The subscript-superscript format again achieves the right blend of external and internal content. First, using

indexical types to specify generic indexical modes, the closest we come to comprehending what the Chair expressed with (1) is to attribute to him an attitude toward a proposition of the type, x_c^I *is upset about the Dean's report* where 'x_c^I' depicts what is the same as the Chair and referred to by the Chair *qua* Self — or, in other words, to what has the self property within the Chair's perspective. (5) gives way to,

(21) $(\exists x_c)$(the Chair said that x_c^I is upset about the Dean's report).

while the correlated analysis of (17) is,

(22) $(\exists x_p)$(the Provost said that x_p^I is upset about the Dean's report).

If the Provost refers to the Chair via the demonstrative 'he' and notes his anger over the Dean's report, we could report,

(23) $(\exists x_c)$(the Provost said that x_c^{HE} is upset about the Dean's report).

where '$x_c^{H\ E}$' depicts what is the same as the Chair and referred to through the demonstrative *he* mode. Had the Provost addressed the Chair with this observation we might report,

(24) $(\exists x_c)$(the Provost said that x_c^{YOU} are upset about the Dean's report).

And so it is with all attributions of indexical reference. Each quasi-indicator is a referentially composite term which conveys reference to what its antecedent refers while expressing the referential modes used by the subject in making a reference. These modes, whether being a *this*, a *you*, a *there*, a *beyond*, etc., are, at best, generic indexical properties whose determinates are accessible only to the occupants of particular non-repeatable perspectives.

Tokens of standard indexical types can also be used quasi-indexically, for example,

(25) I now feel that I am in danger.

Here there is risk of ambiguity. If I use the second 'I' indexically with no intention of revealing how I think of myself then it is not a quasi-indicator. But if I wish to emphasize my possession of a mechanism for making first-person references then I intend (25) to be read as,

(26) I now feel that I myself am in danger.

with 'I myself' as a quasi-indicator used to attribute to myself first-person awareness (Forbes 1987, 18). If so, it is not an indexical though it conveys indexical reference in just the way that 'he himself' in (5) conveys a third-person reference. How is (26) to be accommodated? Compare it with a past-tensed,

(27) Yesterday, I felt that I myself was in danger.

The first 'I' is indexical, but 'I myself' is used to report what I thought yesterday, namely, that I am the same as something which I then felt *qua* Self to be in danger. That past self is not my present thinking referent. Instead, I am now attributing to myself possession yesterday of a first-person referential mechanism;

(28) $(\exists x_{i,y})$(Yesterday I felt that $x^{I}_{i,y}$ was in danger).

where the variable ranges over that which is the same as me yesterday. Moroever, taking 'was' as quasi-indexical we get,

(29) $(\exists t_y)(\exists x_{i,y})$(Yesterday I felt that $x^{I}_{i,y}$ is at t_y^{NOW} in danger).

where 't_y' ranges over intervals the same as yesterday. But (28) and (29) are not the only readings of (27). If 'I myself' and 'was' are genuine indicators, (27) can be taken at face value. The same might be true of (26), though giving it the suggested quasi-indexical reading yields,

(30) $(\exists x_{i,n})$(I now feel that $x^{I}_{i,n}$ is (am) in danger).

where the subscripted 'n' abbreviates the indexical 'now'. This is as appropriate for (26) as (28) or (29) is for (27).

One difficulty common to all accounts of ascriptions concerns iterated attitudes. For instance, the indexicals in,

(31) Isabella knows that Maria believes that I am happy.

are best understood as expressing speaker's reference only and given an external construal. One reading (Castañeda 1989a, 105) is as follows:

(32) $(\exists x)$(I ≈ x and Isabella knows that $(\exists y)$(x ≈ y and Maria believes that y is happy)).

The most controversial aspect of this analysis is the appearance of theoretical notions within attitudinal scope, a problem common to most attempts to deal with iterated belief (Forbes 1993). But if we remember that ascriptions are the attributor's interpretations of what the attributee thinks and that insight into the form, composition, and entailments of the attributed content is a matter of theoretical investigation, then interpretations like (32) cannot be ruled out.

Multiple operators with quasi-indicators introduce special ambiguities. Contrast (31) with,

(33) Isabella knows that Maria believes that she herself is happy.

If the speaker intends to represent Isabella's self-reference without claiming that she attributes any particular mode of reference to Maria, then (33) is,

(34) $(\exists x_i)$(Isabella knows that $(\exists y)$(y ≈ x^{I}_i and Maria believes that y is happy)).

where 'x_i' ranges over thinking referents congruent to what the speaker refers to with 'Isabella'. On the other hand, (33) can also be taken as reporting Isabella's attribution of self-reference to Maria:

(35) Isabella knows that $(\exists x_m)$(Maria believes that x^{I}_m is happy).

Other ambiguities lurking in (33) are due to the differences in the interpretation of 'Maria'. (35) works as an analysis of (33) if the occurrence of 'Maria' is intended to reveal Isabella's reference, but if it is the speaker's mechanism only, the following paraphrase might be more appropriate:

(36) $(\exists x)(x \approx$ Maria and Isabella knows that $(\exists y_x)(x$ believes that y_x^I is happy).

Even more complicated analyses are in order if we wish to capture the temporal parameters implicit in (33).

6. Conclusion

The foregoing offers an account of the role of quasi-indicators in attributing indexical thoughts and references to others. The attributions are themselves attitudes, quasi-indexical attitudes, that make possible communication with indexicals. What emerges from this analysis is that these attitudes, so vital in our reactions to each other, require the intellectual feat of abstraction since interpreting indexicals can only yield the *type* of content a person grasps. This may come as a surprise to those who believe that the reactive attitudes, including the feelings of love, sympathy and respect for particular persons, are among the more concrete and least mathematical emotions we have. Yet if the foregoing account is correct, quasi-indexical attitudes involve precisely that feature of human intelligence which permits us to discern form and pattern in the mass of information that impinges upon us daily, namely, the abstractive power represented by our use of general terms, anaphoric pronouns and variables. It follows that these very important emotions have an intimate relation to the workings of human reason.

References

Boer, S., and Lycan, W., 1980. «Who Me?» *The Philosophical Review* 89, 427-466.

Burge, T., 1977. «Belief *De Re*,» *The Journal of Philosophy* 74, 338-362.

Castañeda, H-N., 1981. «The Semiotic Profile of Indexical (Experiential) Reference,» *Synthese* 49, 275-316. Reprinted as «Indicators: The Semiotics of Experience,» in Jacobi and Pape, eds., *Thinking and the Structure of the World* (de Gruyter), 57-93.

Castañeda, H-N., 1983. «Reply to John Perry: Meaning, Belief, and Reference,» in J. Tomberlin, ed., *Agent, Language, and the Structure of the World* (Hackett), 313-328.

Castañeda, H-N., 1984. «Philosophical Refutations,» in J. Fetzer, ed., *Principles of Philosophical Reasoning*, 227-258.

Castañeda, H-N., 1986. «Self-Profile,» in J. Tomberlin ed., *Hector-Neri Castaneda*, (Reidel).

Castañeda, H-N., 1987. «Self-Consciousness, Demonstative Reference, and the Self Ascription View of Believing,» in J. Tomberlin, *Philosophical Perspectives 1* (Ridgeview Publishing Company), 405-454.

Castañeda, H-N., 1989a. *Thinking, Language & Experience* (Minneapolis: University of Minnesota Press).

Castañeda, H-N., 1989b. «Direct Reference, the Semantics of Thinking, and Guise Theory» in Almog *et al*, eds., 105-144.

Castañeda, H-N., 1989c. «The Reflexivity of Self Consciousness: Sameness/Identity, Data for Artificial Intelligence.» *Philosophical Topics* XVII, 27-58.

Castañeda, H-N., 1989d. «Persons, Egos, and I's,» *Psychopathology and Philosophy*, in M. Spitzer *et al*, eds. (Berlin: Springer-Verlag), 210-234.

Castañeda, H-N., 1990a. «Indexicality: The Transparent Subjective Mechanism for Encountering a World,» *Nous* 24, 735-750.

Castañeda, H-N., 1990b. «Self-Consciousness, *I*-Structures, and Physiology,» in *Philosophy and Psychopathology*, M. Spitzer and B. Maher, eds. (Berlin: Springer- Verlag), 118-145.

Castañeda, H-N., 1998. *The Phenomeno-Logic of the I: Essays on Self-Consciousness*. J. Hart and T. Kapitan, eds. (Bloomington: Indiana University Press).

Chisholm, R. 1981. *The First Person* (Minnesota).

Forbes, G., 1987. «Indexicals and Intensionality: A Fregean Perspective,» *The Philosophical Review* 96, 3-31.

Forbes, G., 1989. «Indexicals» in *Handbook of Philosophical Logic*, ed., D. Gabbay, vol IV (D. Reidel), 463-490.

Forbes, G., 1993. «Solving the Iteration Problem,» *Linguistics and Philosophy* 16, 311-330.

Frank, M. 1995a. «The Subject v. Language,» *Common Knowledge* (Oct. 1995), 30-50.

Frank, M. 1995b. «Is Subjectivity a Non-Thing, an Absurdity [*Unding*]?» in K. Ameriks and D. Sturma (eds), *The Modern Subject* (SUNY), 177-198.

Frege, G., 1967. «The Thought: A Logical Inquiry,» in P. Strawson, ed., *Philosophical Logic* (Oxford), 17-38.

Jacobi, K. and H. Pape, eds. 1990. *Thinking and the Structure of the World* (Berlin: De Gruyter).

Kapitan, T., 1993b. «Keeping a Happy Face on Exportation,» *Philosophical Studies*.

Kapitan, T., 1994. «Exports and Imports: Anaphora in Attitude Ascriptions,» in J. Tomberlin, ed., *Philosophical Perspectives 7/8: Philosophy of Logic and Language* (Ridgeview Press).

Kaplan, D., 1989a. «Demonstratives, An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals,» in Almog *et al*, eds., 481-564.

Lewis, D. 1979. «Attitudes *De Dicto* and *De Se*,» *The Philosophical Review* 88, 513-543.

McGinn, C., 1983. *The Subjective View* (Oxford).

Millikan, R., 1990. «The Myth of the Essential Indexical,» *Nous* 24, 723-734.

Nagel, T., 1986. *The View From Nowhere* (Oxford).

Neale, S., 1990. *Descriptions*. Cambridge: MIT Press.

Pape, H., 1988. «Was sind Quasi-Indikatoren? Eine Einfuhrung in Castañedas Theorie der Subjectivitat,» in M. Spitzer, F. A. Uehlein, and G. Oepen (eds.) *Psychopathology and Philosophy* (Springer-Verlag), 198-209.

Perry, J., 1977. «Frege on Demonstratives,» *The Philosophical Review* 86, 474-497.

Perry, J., 1979. «The Problem of the Essential Indexical,» *Nous* 13, 3-22.

Perry, J., 1983. «Castaneda on He and I,» in Tomberlin 1983, 15-42.

Perry, J., 1990. «Individuals in Informational and Intentional Content,» in E. Villanueva, ed., *Information, Semantics and Epistemology* (Blackwell), 172-189.

Reboul, A. 1992. «How Much am *I* I and how much is *she* I?» *Lingua* 87, 169-202.

Richard, M., 1990. *Propositional Attitudes: An Essay on Thoughts and How We Ascribe Them* (Cambridge).

Saarinen, E., 1986. «Castaneda's Philosophy of Language,» in Tomberlin 1986, 187-214.

Searle, J., 1983. *Intentionality* (Cambridge).

Tomberlin, J. ed., 1983. *Agent, Language and the Structure of the World* (Hackett).

Tomberlin, J. ed., 1986. *Hector-Neri Castaneda* (Reidel).

Whitehead, A. N., 1929. *Process and Reality* (Macmillan).

Zalta, E., 1988. *Intensional Logic and the Metaphysics of Intentionality* (MIT).

Tomis Kapitan

Northern Illinois University

<Plotinus@netscape.net>

SORITES (ΣΩΡΙΤΗΣ), ISSN 1135-1349
Issue #11. December 1999. Pp. 41-65.
Amounts of Vagueness, Degrees of Truth

AMOUNTS OF VAGUENESS, DEGREES OF TRUTH

by Enrique Romerales

The view that vagueness is a omnipresent phenomenon has, in recent times, become a sort of philosophical dogma. The sugestion is that there are not only vague words, sentences and concepts, but also properties, states of affairs and objects. Moreover, sometimes it is even claimed that *every* object and *every* state of affairs is inescapably vague[1].

Objects, properties, relations, and states of affairs all belong to what is usually known as ontology (or metaphysics), and they are interrelated in such a way that the vagueness of one of them will necessarily have an impact on all the rest. For instance, if an object O is a vague one, then this will be due to the vagueness of its properties (either due to it definitely possessing a vague property, or for it being indeterminate whether it possesses a perfectly precise property). In the first case, vagueness relies upon the property (that is, upon the predicate), in the second upon the object (that is, upon the grammatical subject); in both cases the result is a vague state of affairs: the state of affairs in which it is unclear whether the object O possesses or lacks the property P (or a relation R with another object).

Let us leave for another ocasion the question whether there exist metaphysically vague objects, because of the metaphysical problems involved in the very conception of an object. In this paper I want only to discuss the semantical aspect, so I will take «object» to stand for the referent of the grammatical subject of a sentence. The question now is: are all objects vague? Let us suppose we use the following criterion for vagueness:

1) **an object is vague iff it is possible to predicate some vague term of it**.

---

[1] Lorenzo Peña cites and approves Unger's and Heller's claim according to which all terms, with the exception of physical predicates (and Peña disaproves even that exception), bring about problems of borderline application and are sorites susceptible (Peña, 1996; 123). Peña explicitly also claims that this fuzziness must be not merely semantical, but ontological, because our language reflects the way the world is. M.Sainsbury (1995) has brought out various problems of intelligibility – and of plausibility – that this last view raises (M. Dummet, in a famous sentence, asserted: «the notion that things might actually *be* vague, as well as being vaguely described, is not properly intelligible» (Dummett, 1975; in Kefee, 1996; p. 111)).

Let us also suppose that we accept the standard definition of what is to be a vague term: that of which there are, or could be, borderline (i.e. doubtful) cases of application. This criterion, which seems to be assumed by many philosophers, is extremely liberal, and to my mind accords neither with normal uses of the predicate «vague» by native speakers of the English language, nor with any of our intuitions, semantical and ontological alike. With such a criterion, every object is vague indeed, because of any imaginable object whatsoever (with the possible exception of mathematics) not only is it conceivable that *some* vague term applies to it, but the contrary seems indeed inconceivable. For example, focus your attention on John, whose head is entirely covered with hairs. Nevertheless, surely there is some possible world in which John begins to lose his hair (may be due to a disastrous diet, with plenty of fat), upto the point at which it is proper to say of him – for instance from his thirties on– that *in that possible world* John is bald. Not «rather bald», or «dubiouisly bald», but plainly bald. But «bald» is a typically vague predicate, because we do not know how many hairs one has to lose to be counted as definitely bald. There are innumerable *actual* cases in which it is doubtful whether the predicate «bald» applies or does not apply (in fact, dermatologists have classified alopecia in six degrees, but, quite obviously, there are many cases in which it is doubtful to which degree some scalp belongs). So, our John (the one who inhabits the actual world) is to be counted as a vague object merely because it is logically possible to predicate of him a vague term like «bald». Surely this is absurd; John is a perfectly precise individual (let us suppose at the moment), and clearly not bald, so if John is to be counted among the vague objects, he will be so counted for other reasons yet to be spelled out.

The trouble here is the modal form of the first criterion. There are too many logically possible worlds, and every object from within the actual world surely also exists in another possible world at which some vague predicate or other properly applies to it. In order that vagueness becomes not a trivial phenomenon (tautologous) but a substantial fact, let us restrict our attention just to the actual world. We can try a more restricted criterion:

2) **An object is vague iff in fact some vague term is predicated of it**.

Now, this criterion continues to be excessively liberal. For, let's take for instance an orange, paradigmatic in all its properties. It has an orange shape, smells of orange and tastes of orange, and, most importantly, looks a splendid orange colour, not in some other possible world, but here, in the actual world. But «orange» (refering to the colour) is supposed to be a vague term, because there actually are cases in which the application of that word is doubtful (objects that are in between orange and red, for instance). Therefore, the orange just refered to is a vague object according to this second criterion because a term applies to it –although *with total precision*– which *in fact* (that is, in some other actual cases) works vaguely. But again, quite obviously, the orange refered to is not a vague object at all, nor are any of *its* token properties, colour included: it is a perfect and unequivocal token of an orange. So something continues to fail in the second criterion. Let's formulate a yet more restrictive version.

3) **An object is vague iff there is some term whose application to it is doubtful**.

This time let's take a ripe grapefruit, somewhere between orange and yellow in colour. It seems that this time we do have something straigthforwardly and actually vague. In this case, even if «orange» and «yellow» were precise colour terms, our grapefruit is a semantically vague object, because we do not know, or it is dubious, whether the color term «orange» rather than «yellow» applies to it.

But, once more, the grapefruit in question is not a vague *object* at all. At most what is vague is its colour, which is a property of the grapefruit, but not the grapefruit itself. Ripe or not, it is definitely a grapefruit. There are no problems of individuation (how many grapefruits are in front of us?), nor of identity (what kind of fruit is it?), nor of identification (what spatio-temporal item are we refering to?) concerning *this* fruit, in spite of *its colour* being doubtful. What stands in front of us is not a vague object, but at most a doubtful (vague) color shade of a well defined object. According to this criterion, there may well be vague properties, but this is insufficient to show that there exist semantically vague *objects*.

So, let's try another yet more restrictive criterion:

4) **An object is vague iff it is doubtful which sortal predicate in fact applies to it**.

Here it seems that we have hit the nail on the head at last. Let's get a different example, and imagine there is in front of us an object similar to a chair but slightly wider, and slightly shorter than normal, and with only one arm. Then, we don't know whether it is a chair, an armchair or a new kind of object. We don't know what to call it: it is a semantically vague object. Of course, this vagueness does not stem from our lack of adequately fine-grained perceptual discrimination: we perceive the object perfectly well, we see its color with clarity (let's suppose it to be perfectly white), its size, its shape; we know it is made out of oak and so on. The root of the problem lies in our lack of conceptual discrimination. We have only two concepts under which this object could roughly be included: chair and armchair. But neither concept does it fit exactly. Maybe it is just an armchair of which one arm is missing; maybe it is a new kind of object created by a designer with a purpose we don't know of. Surely in that case the designer will give it a name, and once the function and purpose of the object is grasped we shall have a new concept. If its function is socially useful, objects of the same kind will be reproduced, the concept will become common, and the word –say «onearmchair»– will be added to the English language.

But we are in no need of strange and artificial examples: the world of nature provide continuously lots of them. A stream has become permanently so full of water in that condition that now we don't know whether it is a stream or a river. A mountain is so eroded by wind and rain that we don't know whether it is a hill. All objects of this kind are semantically vague because it is doubtful which sortal predicate (if any of the ones we actually have) applies to them. But note, it is vague only the mountain so little that is liable to seem rather a hill; the remaining mountains are not vague objects at all. And the same applies to rivers: the cases dubious between river and stream are the exception (if they were the rule, we would have a concept and a term for them).

Now, someone may reply as follows: there is a trick in the latter move. If only mountains that *in fact* are in between a mountain and a hill are to be counted among the vague objects, we are not talking about semantical objects, but about physical objects, about beings. But the point at issue is whether a *term* like «mountain» is or is not vague. And it is indeed vague when there are doubtful cases of application *of the term* –it does not matter whether many or few– that is, when there exists a physical object of which we don't know, or we are not able to determine whether the term «mountain» applies to it with truth in order to say «that is a mountain». So in the former case we can blame the physical object for vagueness (for example the little mountain of King Arthur's seat, in Edinburgh) for being in between a mountain and a hill, or, what amounts to the same, we can blame the predicates «mountain» and «hill» for lacking sharp boundaries and posing many cases of dubious application. And if we are indeed concerned with the semantical question, then the vagueness relevant will be the one concerning which substantive terms (nouns) are vague. That is to say, since we agree that vagueness applies to a large extent to predicate terms, let us inspect whether it applies to subject terms to the same extent. Now it seems we are finally in touch with the semantical question. Let's formulate a criterion in terms of substantive nouns, which are genuinely linguistic entities, rather than in terms of objects:

A) **A substantive term is vague iff there are possible borderline or dubiuos cases of application**.

Again, because modalized this criterion is too liberal: for any substantive term we can conceive of, there is a possible world in which there are objects of which it is doubtful whether the term applies or does not apply to it. For example, «tiger» would be vague, because there is some possible world in which there are some mammals similar to our tigers, but also similar to our leopards (let's dub them «tigepards»), so that it is dubious whether they are a subspecies of tigers, or of leopards, or if they form an altogether different species (to get things yet more dubious and complicated, let's suppose that in general tigepards are able to reproduce only among themselves, but that occasionally, they can reproduce with both tigers and leopards alike, some of their offspring being fertile, some not). Well, whith this criterion surely all or almost all substantive terms are vague, because we can conceive of dubious cases of application. But, is this an adequate definition for the vagueness of a substantive term? I don't think so. Albeit «vague» is vague, we do have a concept of vagueness, and when we have a concept and its corresponding term *usually* there are cases in which it applies and cases in which it does not apply, or at least cases in which the differences in application are very remarkable.That is, if some terms are radically vague, some others must be precise, or at least vague in a much smaller degree. And it seems totally unfair to regard the term «tiger» vague simply because there could have been animals of which it would be difficult to decide whether they were tigers or not.

If someone is not yet convinced by the tiger example, we can provide another. Let's have take a gold ingot. Is the term «gold» vague? Of course, for the ordinary speaker of English there will be cases in which he is in doubt whether to apply or not apply the term to some object, because he is liable to mistake some other metal for

gold, provided they look[2] alike. Nonetheless, there is a well established scientific criterion to determine whether a given ingot (or a single atom, if you like) is or is not gold. And in a case like this it is even controversial whether there are possible worlds at which there is gold with a different number of protons and electrons from our gold. So, I will demodalize the criterion once more:

B) **A substantive term is vague iff there are actual cases of borderline or doubtful application**.

With this criterion it seems clear that not all terms are vague: «ballpen», «gold», «tiger», «quartz», «star» etc. are all cases of non dubious application. True, for the lay person it can be very doubtful whether a watch is or is not made of gold (particularly if its origin is dubious), but the physicist can answer the question without trouble. By the same token, a zoologist can tell whether a certain mammal is a tiger, a geologist whether a piece of mineral is quartz, and the astronomer whether a point of light in the nigth sky is a star (instead of a planet, a comet or a distant galaxy).

What I am trying to say is that most sortal terms have well defined criteria of application, so that troublesome cases for ordinary people can be definitively solved by the expert. Nevertheless, this, unfortunately for the philosopher –and fortunately for ordinary language– does not happen with every sortal. «Mountain» and «hill» are typically vague sortal terms, as also are «city» and «town». Certainly, the geographer could stipulate the border between hill and mountain as being for instance 500 m high (either over the sea level, or more plausibly over the surrounding ground's level), and the political geographer could stipulate the border between town and city as 100.000 inhabitants. Then we would have absolutely precise terms and concepts, although arbitrarily precise. Arbitrariness does not need to be a shortcoming: we are the authors of houses, villages, towns and cities; so we are entitled to define (i.e. to delimit with so much precision as desired) our concepts in that field. The real problem is that in fact we don't do that (or don't always)[3].

Things are different when sortals correspond to atoms, molecules, minerals or biological species. There, nature has established definitive differences which prevent the choice of sortals from being arbitrary. If I remove a single hydrogen atom from a benzene molecule, it is no longer benzene at all. «Benzene» is, then, one among many substantive terms with no borderline cases of application, and which, as a result, is not to be counted among the vague terms. I would suggest that when a sortal admits of borderline cases of application, that is, when we would have to stipulate arbitrarily a

---

[2] Kit Fine termed «extensionally vague» the predicate which has actual cases of borderline application, and «intensionally vague» the predicate which has possible but non-actual cases of borderline application («Vagueness, Truth and Logic», in Kefee, 1987; p. 120). My intention here is to claim that only the first case is crucial and properly relevant for vagueness.

[3] Crispin Wright has objected to this move that to try to make precise terms that are in fact vague is contrary to semantics and would prevent its use (Wright, 1976; in Evans & McDowell, 1977; p. 230). I think the second is partially true, but the first seems a matter of priorities: it might be *contrary to semantics* but in accordance with logic. But, of course, Wright chooses the witggensteinian alternative: if some use of language is contrary to logic,so much the worse for logic.

sharp cut off as the sortal limit –like between a town and a city, or between a hill and a mountain– we haven't got a true sortal, but instead a quasi-sortal. This seems to me more proper than saying that there really are vague sortals. Since «sortal» is a specifically philosophical term, I would use «substantive» for «vague sortals» and non-vague alike, and keep «sortal» just for non-vague sortals.

A barrier should not be set up between natural and artificial sortals. Mountains are natural beings, but the sortal «mountain» gives rise to problems of application in borderline cases that the artificial sortal «table» never (or almost never) gives rise to. Natural sortals are less problematic only when their referents have a clear individuation in the *scala naturae*. Either by being little complex entities (as an atom, a molecule, maybe a prion and certain viruses) or by being highly differentiated. For example, of a mammal, rare, silly, angry, big or whatever as it can be, to predicate «is a cat» is absolutely true (or false). And here there seems to be no place for things like the infinitely perfect cat (the cat more similar to the ideal cat), nor, conversely, for the least cattish entity possible.

Now, the opponent can reply as follows: concerning the last criterion we are confronted with a dilemma: either it is inapplicable in many cases, or its application renders many more cases as ones of vagueness that it seems at first sight. For instance, let's go back to tigers. Probably *now* there are no species alive similar enough to tigers in the relevant aspects and traits, consequently the use of the term «tiger» at present gives rise to no problems. But it is very likely that in the history of the evolutionary process there have been intermediate species, nowadays extinct, of which it was in fact doubtful whether they were really tigers or not. In that case, although for ordinary people «tiger» is not a vague term, it is so for the zoologist –the expert, and consequently the one who is entitled to decide in the end whether the application of a term is correct or not– and for the paleontologist[4]. Moreover, let's suppose that we find remains of species morphologically and anatomically very similar to tigers, and that they seem to be their close ancestors in the evolutionary line. Let's dub them «?tigers». Now, suppose that the sound criterion of individuation for biological species is this: two exemplars (obviously of different sex) belong to the same species iff they are capable of reproducing among them, and also are their descendants. In that case, since ?tigers are extinct, we presumably are not able to know whether they were capable of reproducing with tigers or not. But ?tigers do have existed; so, applying our criterion, it is doubtful (we cannot know) whether the term «tiger» is vague. Now, there is nothing special about the case of the tigers. What happens in the case of so an evolved and differentiated a species like tigers happens in all other cases too, because evolution has been gradual and highly branched, and is full of nowdays extinct species (compare the problem of fixing which one of the hominids was the first human being).

Well, we could reply: possibly paleontology can establish some cases of species now extinct as doubtful concerning their assignment to a species already identified and classified, or their forming a new species. Maybe even in zoology there are some exemplars alive which are only doubtfully classified into a certain species. But it is not

---

[4] This line of argument has been developed in Williamson 1990, 114-15.

so in most cases. Cases of dubious classification are the exception, not the rule. And, since we have decided to regard as vague only the terms whose application is in fact sometimes doubiuos, even if there are vague terms of biological species, they are minimal.

But the critic may reply as follows. Well, paleontology is still progressing, and surely there are many fossil remains of now extinct species yet to be found or maybe unfortunately lost for ever. At any rate, these species have existed, so species terms are in all probability all vague (or almost all). Or, at best, we never can be assured that an apparently precise term won't turn out to be vague in the end. And an argument along the same lines can be offered for artefacts. From the fact that we have never heard of any object in between a chair and an armchair, you cannot infer that never and nowhere has any carpenter made up something intermediate. To put the argument in general terms: for any well defined artefact kind A, it is possible that some civilization or other has at some time made up some artefact token *a*» sufficiently similar to *a* tokens as to be doubtful whether *a*» is to be counted as an A or not. Then we never can be assured that apparently non vague terms like «ballpen», «chair» or «spoon» are not vague after all.

I think it is proper to reply to that line of argument in this way: a term must be counted as vague-innocent until proven guilty. This means that while you have not found a single dubious actual case of application of the term, the fact than in the distant past there could have existed doubtful cases of application is as irrelevant as the fact that there could be cases in the future (from the fact, say, that in the future someone will make something intermediate in between a car and a motorbike, it does not follow that those terms are vague *now*).

We might even claim that if the remains of ?tigers are discovered, this does not turn immediately the term «tiger» into a vague one, if we accept as a criterion for existence of species the fact that there be exemplars alive. Nevertheless, I concede this may sound paradoxical. Because, as we have above admited, it is the expert, the scientist, who is allowed to decide whether the application of a term to a given object (gold, water, quartz etc.) is right or wrong, and as a result it is up to him to define with precision what would otherwise be vague objects. But in this case it is the other way round. Because a term apparently not vague at all, is liable of turning out to be vague *just for the zoologist*. And, as we will see later, if in the former case there are good grounds to consider the scientist the one apt to resolve doubtful cases (and also relevant ones for demandig his assistance), these same grounds should apply here. For the paleontologist the problem is that in the past there has been, or might have been a dubious species. For the philosopher the problem (with the criterion in use) is that *at present* there are remains of which it is doubtful whether they belong to a determinate species.

But I think we could make use again of the forensic clause «innocent, until proven guilty». Thus if a paleontologist finds remains of ?tigers and is unable to determine from these reamains alone whether they belong to the species of tigers, whether they belong to an altogether different species, or whether they belong to a species in between, say, tigers and leopards (and so they belong to a doubtful or vague

species), then the term «tiger» must continue to be counted among the non-vague terms. So, only when the paleontologist finds evidence that ?tigers *reproduced partially* with tigers (for instance, that tigers generally reproduced with ?tigers, and also their immediate descendants, but that in the third generation reproduction was difficult, and the th generation was entirely sterile) is he entitled to take as vague the term «tiger». Since this is usually very difficult to prove for paleontology, I take it that the vast majority of species terms are non-vague. A similar argument runs for artefacts. For example, ten years ago or so some vehicles were made that are in between being a car and a van. The consequence is that the terms «car» and «van» have been vague for a short period: just that period necessary to find a new term for a concept we had in advance (without the concept, without the idea, the product could not have been designed and made), viz. «carvan». Perhaps the next year a vehicle will be released which is in between a carvan and a car. Then –and only then– we will have again problems of vagueness with the term «car», problems which, presumably, will be solved in a similar way.

Now things are radically different when we focus on the remaining predicates. For there are lots of predicate terms that give rise –even if only occasionally– to problems of borderline cases. And others that give rise to this same problem very frequently, in particular certain adjectives (young, tall, short, nice) and adverbs (much, little, enough etc.).

The second part of this article is devoted to critisising the degree-of-truth approach to coping with vagueness in our language[5]. I will put aside many logical difficulties –pretty well brought out by others– that degree of truth theories and fuzzy logic give rise to, and I will concentrate upon philosophical problems, both semantical and ontological.

Many discussions about vague terms take for granted that there is a well established series of standard examples in which the kind of vagueness is more or less equivalent. Among the classical examples are the one of the heap, the one of baldness and the one of the colour patch. All of them have something in common: vagueness is one-dimensional. There is a sort of line, and it is unclear when the term begins to apply, and sometimes is also unclear when the term ceases to apply. In other more complex cases vagueness is multi-dimensional. For instance, «beautiful». For something to be or not to be beautiful many parameters and their relations to each other must be taken into account: shape, size, colour, appearance etc. Some of them may be extrinsic to the object. So, a modern building of steel and glass can be aesthetically of great value, but horrible in the middle of an old town. Words like «nice», «clever», «able» etc. are all multidimensionally vague. So in order to avoid difficulties, let's limit ourselves to the simplest type: one-dimensional vagueness (with the hope that, if there

---

[5] I mainly aim Lorenzo Peña, 1996 (in fact this paper has prompted the present one). This sort of approach has also been defended in K. Machina, 1976. Machina's option is to gradualize the force of the inference in a sorites argument, so that the more numerous the premises, the less the truth transmited from the premises to the conclusion. For that reason, sorites arguments can begin with premises of value 1 and end up with conclusions of value 0.

is any solution to the problem of vagueness in one dimension *whatever*, the same strategy can be used for *every* dimension).

A different case is the one in which we find vague terms, but in which the kind of vagueness is totally context-dependent. Viz. «enough». For a clerk a salary of $ 1,500 per month can be «enough» acceptable; for someone unemployed this will usually be much more than «enough»; for a top football player like Ronaldo it will be totally unacceptable. If we put together terms like «beautiful» and «enough» things become more and more complicated.

If we are interested just in how to cope with semantically vague terms, and if –as Frege and Dummett both think– vagueness is semantically incoherent, a single case of vagueness is enough to create the whole problem. But if we are also interested in how far vagueness is entrenched in our language and our thinking, then both the question of the amount of vague terms and the question of their degree of vagueness become relevant. For that matter, I would like to remark that *much* of apparent vagueness is only contextual dependence. Let's take an example with «far». The University Autonoma of Madrid is 15 km. away from Madrid (from the center at Puerta del Sol). Does the property «being 15 km away from» fall under the predicate «far»? This is *almost* entirely context-dependent. If we are talking about a car race that is running from Paris to Madrid, being 15 km. away from Madrid is, without any doubt, not to be far. If we are talking about a plane that is coming from San Francisco to Madrid, to be 15 km. away from Madrid, is not to be far. If we are talking about a spacecraft coming from Neptune, to be 15 km away from its destination, is not to be far. If, conversely, we are talking about whether it is fair or reasonable for a student living in Puerta del Sol to go and come back by walking every day to the University, surely it is far indeed. If we are comparing the Universidad Autonoma with the Universidad Complutense (which is in the city of Madrid and very close to the center), without any doubt, the Universidad Autonoma, by virtue of its being 15 km away from Madrid, is again far, and so on. Now, imagine we are thinking of running this distance by cycling twice a day; here it seems we have a real problem of vagueness. But, once more, partially context dependent. For a young 20 year old, 30 km by bicycle per day is a perfectly feasible distance; so, for him the University is not far. For the emeritus professor who is over his 70, surely 30 km. by bicycle is too much, so for him to cycle to the University is definitely far. But, what about the Reader who is in his forty? *Cases like this* are the really vague ones (and again, it could be made more precise specifying the contex: for a sportsman, 30 km a day by bicycle is no problem; if he never does sport, the distance will be unsurmountable. Solely if he does sport from time to time, *there* –and only there– we have a genuine doubtful case). In summary, «far» is a vague term because there are some cases of dubious application, but once the context is fixed *entirely*, the vague cases are a minority, fewer than usually thought of. The fact that in ordinary speech acts the context is usually well defined is what makes *serious* troubles with vagueness to be unusual, and that communication normally flows free from obstacles.

Nevertheless, whether minority or majority, cases of vagueness do happen even in a maximally precisified context, and pose a real problem for the philosopher. It is here that some philosophers contend that, to be able to cope with vague terms we must

reject the principle of bivalence and admit that not every statement is simply true or false[6]. There are intrinsically doubtful statements, that is, with a non determinate truth value. Whether we interpret «indeterminate» to be a third truth value or an absence of truth value, there is a general consensus that this move is useless, for now we are faced with finding two clear cuts for the application of a vague term, instead of just one as was the case before. Now there has to be, on the one hand, a sharp cut-off between the cases in which someone, say, is «definitely tall» and the cases in which he or she is «dubiously tall»; and, on the other hand, a sharp cut-off between the cases of being «dubiously tall» and the cases of being «definitely not tall». But, as is well known, higher order vagueness makes equally difficult to draw these two new bounds than the former one, just because it is dubious when someone is «dubiously tall»[7]. True, some philosophers have raised doubts about the existence of higher order vagueness[8]. But I take it to be rather obvious that it does exist, and also take it that its existence has been accepted by most philosophers involved in this topic.

Once is higher order vagueness allowed, we need more truth values in order to be able to come to terms with it. Then there will be cases in which someone is «dubiously definitely tall», «dubiously dubiuosly tall» etc. Now the question is, how many values are necessary for the full range to be covered? It seems uncontroversial than any finite number of truth values is both arbitrary and insufficient. Arbitrary because there is no compelling reason to distribute the degrees of truth or corectness for the predication of terms like «tall», young», «happy» etc. in a certain number rather than in another. Insufficient, because a dubious case can always arise which is intermediate between two sucessive degrees. Futhermore, a finite grading would involve a discontinuous application of a term where the use sets rather a continuous line. It seems, then, we are compelled to admit infinitely many degrees of truth in the application of a vague predicate, where 0 will be the absolute non application of a predicate and 1 its absolute application, leaving in between the infinite interval of real numbers. It will correspond in set theory to a logic that represents concepts as points in logical space, so that to be just on the point is to be a member of the set at degree

---

[6] This move is made by supervaluationism, maybe the approach which has gained maximal recognition to cope with vagueness in spite of all its difficulties. Cf. a recent revision and highly technical statement in McGee & McLaughlin, 1995.

[7] For that reason, to talk of fuzzy cumules with two sharp boundaries, the one of the beginning of fuzziness, and the one of its end (Peña, 1996; p. 134) is problematic. As T. Williamson, among others, has insisted on, higer order vagueness invalidates these kinds of approach (cf. Williamson, 1994, chaps. 4-5 and 7, and his 1999). M. Tye claims: «I do not accept that «is true» is extensionally vague. And the same goes *mutatis mutandis* for «is false» and «is indefinite». Of course, in taking this view I am not commiting myself to the position that these predicates are precise. Indeed, it is crucial to my account that they *not* be classified as precise… they are vaguely vague: there simply is no determinate fact of the matter about whether the properties they express have or could have any borderline instances» (1994, in Kefee, 1996; p. 290). But you could replace just the last sentence by «there is no way to know the determinate fact of the matter…» or even «there is no way to know whether there is a determinate fact of the matter about…».

[8] Among them Crispin Wright in his 1992.

1 (or to possess at degree 1 the property) and as we go away from the point the degree of set membership decreases gradually up to being finally 0. My aim here is to contend that this kind of answer is both technically deficient and philosophically misguided[9].

I begin by classifying one-dimensional and non-context-dependent vague terms in three categories: a) those with two blurred boundaries, b) those with a blurred boundary at one side and a sharp boundary at the other, and c) those with a blurred boundary at one side and no boundary at the other. My claim will be that the gradualist approach works with the first category, but is unnecesary, and that it does not work with the other two. The main problem is: if we are compelled to assign values between 0 and 1, either we are unable to assign coherently the 1, or we are unable to assign coherently the 0, or both.

We start with cases of category (a). Let's take the colour predicate «green». It is clear that grass, pine leaves, peas and spinach are all of them unequivocally green. But bananas and lemons fre**q**uently are rather yellow. Conversely, sea water is normal**l**y blue, although on cloudy days can be rather green. Accordingly, «green» is a predicate that applies to a waveband of the spectrum between blue and yellow. Let's suppose, for the sake of argument, that there is an exact point –even if unknowable– at which blue becomes green, and another exact point at which green ceases to be green and becomes yellow. Let's call these points *a* and *b* respectively. Now, how are we to assign the values 1 and 0? Obviously we cannot decide that $a=1$ and $b=0$, because they are equally little green. Which point owns green at degree 1? Surely the one that is just in the middle between *a* and *b*. In that case you will have two points, *a* and *b*, for zero (0), and one point for one (1), the point just in the middle of both. This seems not to be a big difficulty. You could just assign values differently, and consider for instance 0 to be the perfect possession of a property, and -1 and +1 the perfect non possession of that property. But it is more in accordance with standard usage to assign 1 to the maximal degree of property possession, and 0 to the minimal, even if there are two 0 cases in opposite directions.

The first serious problem is: is this procedure not a bit question begging? A vague concept is one which possesses borderline cases for having blurred boundaries, and we have supposed there to be an exact point at which blue becomes green and another exact point at which green becomes yellow. And this is what is questionable. But the gradualist, infinitist or finitist as it may be, needs to suppose that there is an exact point at which the predicate does apply, and an exact point at which the predicate no longer applies, only the range in between being gradual[10]. Note that this supposition

---

[9] I am not competent to discuss the logical difficulties of this approach. But they do exist (cf. Morgan & Pelletier, 1977).

[10] It is just this supposition which is problematic. As Mark Sainsbury has pointed out (1990, in Kefee 1996; p. 254-6) degree approaches and fuzzy logic seem useless: «one cannot do justice to the phenomena of vagueness, in particular to the phenomena of «higher order vagueness» simply by icreasing the number of sets of individuals associated with a predicate… a predicate which effects such a threefold partition is not vague… This same point is what scuppers the set-theoretic descriptions of vague languages offered by fuzzy logicians and supervaluations theorists… Yet a fuzzy set is a genuine

–also accepted by the epistemic theory– that there is an exact cut-off point (in this case two) where the application of the term begins or ends, to my mind is sound concerning colours.

Within the spectrum of wavelengths there is an exact point at which it becomes visible for the human eye (like there is an exact different point at which it becomes visible for dragonflies), and there also is another exact point at which it becomes invisible for the human eye. Along this range we are able to set up divides based upon qualitative differences (qualia) clearly perceptible for all normal human beings. The ordinary use in many languages has assigned seven color terms as basic –the ones of the rainbow– (putting aside black and white). Well, we can then, following this linguistic use, divide the spectrum in seven equally long wavebands and ask the physicist to establish at which wavelength each waveband starts. Let's take the green waveband. Its center point corresponds to maximum greeness or greeness to degree 1, its bounds to greeness to degree 0. In between there are as many degrees as you like. In fact, I have choosen green instead of the usual red, because green is not a primary colour, but the mix of blue and yellow. Then, surely a mix at 50% of perfect blue and perfect yellow is green to degree 1, and possibly pure blue and pure yellow is green to degree 0.

Now, the critic can respond that we have artificially defined colors by using scientific patterns[11], when it is use which bestows meaning upon the terms. Quite naturally, different linguistic communities, with different ways of life, would have different uses and, consequently, different meanings for different terms (that is, the semantic field of terms will not be equivalent). For instance, eskimos usually are able to distinguish up to 10 shades of white, and have names for each one. Let's accept this criticism. But we are talking about what «green» means within our linguistic community –in a different language our «green» can be polisemic, or not exist at all. In our language there is only one term for green things, and all shades are named by qualifying the basic term: «bottle green», «grasp green» etc. Only the bounds of application are dubious. Now: who should establish these bounds? Wright has insisted upon the linguistic community as being entitled to do that[12]. This is a matter of statistics. Chose a group of native English speaking people (better from the same country?), and show them a series of colour patch shades gradually going from green to yellow. When the consensus is broken –that is, as soon as one of the people dissents from the proposition «this patch is green» – then that shade of colour is no longer

---

set, a completely sharp object… the fuzzy logician too will be commited to a threefold partition: the sentences which are true to degree 1, those true to degree 0, and the remainder… But we do not know, cannot know, and do not need to know these supposed boundaries to use language correctly».

11 K. Machina (1972, 27) makes this criticism, claiming that there always are more than one possible –and plausible– translation of observational predicates into scientifical ones, and that any decission in favor of just one precisification would be artificial. I agree with the artificial character of the decission, but you can have good grounds for stipulating one wave lengt instead of another as the corresponding to, for instance, green.

12 Wright, 1976, 232.

definitely green, or clearly green. When they regain consensus by refering all of them for the first time to a patch colour as «yellow», then the shade is definitely yellow[13].

This procedure has a number of well known difficulties. The least serious is that, if this is supposed to be a determinate number of people chosen on a certain occasion, the result will be both random and arbitrary: even the same people are liable to change their minds at a different time. But, if we refer to an «ideal linguistic community», then it is impossible to know the meaning of «green», because we have no way of finding out the verdict of such a community. It is even controversial whether the concept of an «ideal linguistic community» is a coherent one, if it has to include contradictory decisions of the same person at different times.

The more serious difficulty is that in many borderline cases the speaker would not know what to say. It might take some minutes for him to decide; maybe he retracts shortly after having said something concrete; maybe he just whispers his response without saying anything clear aloud etc.[14]. And should the answer of someone who quickly and with resolve says «green», be scored equally with the answer of one who, after thinking carefully about it, says rather shyly and tentatively «green»? The real problem is that the actual actions of fully flesh and blood individuals, including their speech acts, have such a huge number of parameters, many of them relevant to the present case, that no statistic, detailed and complex as it may be, is able to mirror them. In summary, there will be no way of assigning precise values to many boxes in the statistics. The very statistic should have instead of three boxes (green, not green, yellow), either an innumerable ammount of boxes or boxes fuzzy defined («rather a bit more green than nearly yellow» and so on). Either alternative renders the statistic unviable.

Is it not then more sensible let the physicist to be the referee in this kind of situation, and let him decide where lies the exact bound between green and yellow? But in that case, the term «green» is no longer vague, because there no longer are non-decidable borderline cases. Thus, the cases are borderline only concerning everyday use, because of our lacking adequate perceptual discrimination, but with an spectographer we are always able to tell whether some shade is definitely green or yellow, *with so much accuracy as desired*. If this is correct, colour terms are not intrinsically vague, it is only usage that make**s** them vague. Now, let's suppose that meaning supervenes on use. In spite of this, in cases of doubtful use it is still the scientist who is responsible for fixing the correct use. Think for instance of the case of gold. Whether a given pie**c**e of metal is or is not gold is decided not by the linguistic community, but

---

[13] This proposal is made by Crispin Wright in his 1987a (in Kefee, 1996; p. 222). He adds: «We have therefore to acknowledge, surprising as it may seem, that a sorites series of indistinguishable color patches *can* contain a last patch which is definitely green [red]… It may be difficult or impossible to *identify* such a patch in a particular case… it should be noted that there is, a priori, no reason to suppose that «the last definitely red patch» would turn out to have a stable reference; if it did not, that would disclose an element of context-relativity in the concept of green which we normally do not suspect» (loc. cit).

[14] This point is made by M. Tye in his 1994 (in Kefee, 1996; p. 289.)

by the scientific community (the chemists, this time). Why can't this be the same in the case of colours? Normally because it is not very important whether something is definitely green or yellow, while it is extraordinarily important whether a collection of ingots are or are not gold. If in a certain case the verdict concerning a color were of upmost importance (for instance, to decide whether something is a certain gem or just imitation jewelry) we would require the assistance of the scientist (the gemologist, this time).

Now, it might be reasonable to accept that perfect green consists of a perfect mix of blue and yellow. But where along the spectrum does green end? Surely, a uniform path shade of 1% blue and 99% yellow continues to be yellow, and so shall we perceive it. Very likely the same will happen if the mix is made of 2% and 98%, and so on. Does there not continue to be, after all, a blurred boundary for ceasing to be green? If we have decided to divide the spectrum in seven bands, then the band between pure blue and pure yellow should be shared as follows: 25% at the left for blue, 25% at the right for yellow and the remaining 50% in the middle for green (this implies that colors at the end of the spectrum will have a band half wide than the rest, and that their value 1 will coincide with the point at which spectrum becomes visible; so we would have 5 equally wide bands, and two half bands at the spectrum borders).

If this analysis is correct, «green» does not have intrinsically doubtful cases of application, they are only epistemically doubtful for the perceptual abilities of the average people. But –and this is the big question– if it is after all not a vague term, what is the purpose of a gradualist analysis? The gradualist will contend that, even if it is well defined when a shade is green and when it is not, the further questions remain: how green is the green patch? up to what point is it green? And it is *here* where the degree of truth approach has a role to play, establishing degrees of greeness between 1 and 0. But this is, to say the least, a matter for discussion. Look upon a varied garden in winter. You have there lots of plants, with all shades and intensities of green. Does it make sense to ask which of the leaves are the most green? It seems not; the proper answer seems to be something like: «there are many different shades, but all of them fully green, just variously green». If this is right, we have no reason to accept the gradualist approach, because all shades within a certain waveband of the spectrum we have accorded to call «green» are fully green, differing only in shade.

Now it can be objected: could the trouble with vagueness and fuzzy boundaries not arise again at the level of shades? Of course it could, but it does not *necessarily* arise. Suppose we are interested in making a pencil case with pencils of many colours. For green we want, say, 16 pencils. Well, we divide the green band in 16 equally wide stripes, we proceed to match the central shade point of every stripe with the colour shade of one pencil, and then –if you wish– bestow upon it a name (presumably a name related to the green colour shade of some object in nature normally of this shade).

Of course, this example is intentionally simplified. In reality, every determinate colour shade admits also of differences concerning intensity: the same colour shade can be lighter or darker, depending on its mixing with white or black (and also more or less bright or pale etc.). A lemon is between green and yellow, but a lettuce (like most

green vegetables) is between green and white. This shows that if color terms were vague they would in fact be multidimensionally vague: also a blue patch can be not only very close to green or to violet, it can be very close to white (like sky in a sunny summer day) or to black (like dark navy blue). So that in our case, if we wish add to the 16 shades of green for instance three intensities of each, we will need 48 pencils, each one with the name «green», the surname «grass», «pine» etc. and the epithet «light», «normal» or «dark». For if there is a way of fixing the cut-off point between green and blue and between green and yellow, a similar procedure could be used to specify the sharp cut-off between white and green and green and black.

In conclusion, in spite of all contrary appearances, color terms are not intrinsically vague. Or, to put it differently, usage makes some of their ordinary applications troublesome, but there are no objects «vaguely green». That conceded, two options remain open for us. The first is to regard as maximally green (green at degree 1) only color shades being a perfect mixture of pure blue and pure yellow (and perhaps also without any mixture of white or black), and regard all shades that progressively recede from this pattern as gradually decreasing their degree of greeness, until being green at degree 0 (and place this 0 point either at blue at degree 1 and yellow at degree 1 respectively, or rather in some other point, for instance, in a mixture of 75% green (of degree 1) and 25% either blue or yellow (also of degree 1). It seems to me that actual use of ordinary language does not justify this procedure at all. For example, according to the gradualist approach it makes perfect sense to ask: is there something which is perfectly blue? And presumably it will be very difficult to find such a thing (but think of a woman who in coming into a dress shop asked: «do you have a perfectly blue evening dress?»; surely the proper answer will be: «what do you mean by «perfectly blue»?»). But in a case like this, logical grounds justifying the gradualist approach are also lacking. At any rate, it is not *necessary* to tackle vagueness of those predicates, once it is acknowleged that they are not really vague.

Now let's go with cases of the second (b) type: that in which there is a blurred boundary at one side and a sharp one at the other side. Standard examples are the bald man and the heap. With 0 hairs on his head Mark is undoubtably bald, 50,000 unequally distributed hairs could be a dubious case, and to have 200,000 uniformly distributed is not to be bald at all. So, the more hairs on his head and the better they are distributed, the less bald is he. For the degree theorist approach this time we have a clear assignation of the value 1 for baldness: 0 hairs on the head. In such a case Mark is 100% bald. But where begins baldness at degree 0? Here we are supposed to have the fuzzy boundary. Maybe something around 100,000 hairs or so. But this answer is forbidden for the gradualist. He is commited to the claim: the more hairs, the less baldness. And that is the problem, for what is the maximum amount of hairs someone is able to have? We could say, as many as there is room for on the head. But, on what head? Imagine Mark having a head double sized than Jim. Mark may have 400,000 hairs on his head, Jim 100,000, both uniformly distributed. Is not the degree theorist compelled to say that Mark is less bald than Jim? But, quite plainly, neither of them is bald at all. So, maybe, degree of baldness is dependent upon one's head size (maybe also on one's hairs' size?, because the thinner the hairs, the more are able to stand on a head). So we could conclude: the head size puts the limit to baldness 0; when on a

certain head there is no room for just one hair more –independently of the head size– the one who owns this head is bald at degree 0. This analysis has again two counterintuitive results. One is that, since everybody is continuously replacing his or her hairs (at about 50 to 100 hairs a day) no actual head is altogether full of hair, and so *everybody* is at least a bit bald. The other is that, conversely, the one who has no hairs on the top on his head, but has many hairs on the sides, is to be counted as half bald. This is not in accordance with the use of ordinary language nor with ordinary knowledge. According to both, about 10 to 20 % of male adults are bald, another 5 to 10% are becoming bald, and the rest are not bald at all. And among the ones who are definitely bald there are the ones with no hair on the top and thousands of hairs on the back of their heads.

The example of the heap is yet worse. The problem has always been posed as one of identifying the point at which a group of grains put together makes up a heap. No grains are supposed to be a heap at degree 0, and from 1 grain on, the more you add the closer the collection becomes to being a heap. This is controversial. In my view, the right answer to how many grains (are the minimum to) make up a heap is four, because, as a matter of fact, only four piled-up items are liable to form a stable three-dimensional shape[15]. Notice that if this answer seems rather implausible this is because we are thinking of a certain context — a heap of sand grains, in building a house for example – where it seems bizarre to say that four grains make up a heap. But think of a heap of books; here, four can easily form a heap. Moreover, perhaps in this case three is the minimal number (if the criterion for being a heap is to be able to pile up items in a stable manner, then two books could be a heap, but I think our linguistic intuitions tell us that a heap, whatever its components, has to have *more than two* components).

But the truly paradoxical character of the heap case lies at the other side. 100,000 grains of sand are a heap, but 10^{14} are not a heap, they are a hill or maybe a mountain –it depends also on its compactness. The gradualist is able to assign with sense the property of being a heap at degree 0 to zero grains, but there is no way to assign the degree 1[16]. Of course, he cannot admit four grains to be a heap at degree 1, because in that case 2 grains would be half a heap. The perfect heap should presumably lie somewhere between 10,000 and a million grains, perhaps. But here the gradualist is as desperate as any. So, to have a definite proper answer to offer, he is obliged to say that the greater the number of grains, the more something is a heap, so that to be a heap at degree 1 the heap has to contain an infinite number of grains. But this is absurd. And if, instead of infinite we choose a very high but physically feasible (on the Earth) number of grains, we get Everest, which cannot properly be called a heap.

---

[15] This has been argued by W. Hart in his 1991, and to my mind it has been surprisingly neglected.

[16] Being sensitive to this kind of problems, D. Eddington (1996) has recently and interestingly proposed to take the gradualist approach as merely instrumental, and to develop it's logic in terms of probability calculus, rather than in terms of some or other deviant logic. But she is pretty aware that you cannot replace vagueness with perfect precission.

Maybe it can be argued that cases like this are better served by a finitist gradualist approach. After all, the gradualist approach can boast of being able to solve *the* sorites paradox concerning heaps. And so it does when, as our first premise, we assume that a certain determinate number of grains *does* in fact form a heap:

1) 10,000 grains (piled up together) make up a heap

2) If n grains make up a heap, n-1 grains also make up a heap

…….

Then, zero grains make up a heap.

Applying here a finitist gradualist approach, we can say with sense that «17 grains make up a heap» has a degree of truth of 17/10,000, or –if instead of gradualizing truth we rather gradualize the vagueness of a concept– we can say that 9,983 grains make up a heap at 9,983/10,000 (that is: 0.9983). But the root of the trouble lies in that the first premise is totally arbitrary, just as arbitrary as the number 10,000 as paradigm of what it is to be a heap: we are entitled to regard this premise neither as the exact point at which a number of grains is *already* definitely a heap, nor as the exact point at which a certain entity is just in the middle of being between zero grains and a hill. The problem for any gradualist, finitist or infinitist, is that he continues to need an exact number of grains that turn a lot of grains into being a heap at degree 1 (the finitist may choose 10,000 and the infinitist the infinite number: both are troublesome, although for different reasons). So we are not a single step ahead.

Now we are going to take an example of the third (c) case (recall: a fuzzy boundary at one side, no limit at the other). Let's our example to be «tall». In this case, as in many others, much of its vagueness is context-dependent. Viz. among Pygmies, a man 1.50 m. tall, is indeed tall, but among the Masai he is not tall at all. If we are talking about basketball, a player of 2.04 m. is normal, but among the population of his city surely he is tall. As accepted, not every case of vagueness is context-dependent: when talking about basketball it is unclear whether a player of 2.08 m. is tall or not: possibly he can play as 4, as 5 or as 3, according to circumstances.

Now, suppose the context is maximally wide: mundial population *in toto*. We are talking about people in general to determine when they are tall. Well, the gradualist infinitist analysis seems to commit you to saying that «Peter is short» is perfectly true only if Peter measures 0 cm. And conversely, «John is tall» is absolutely true only if John measures infinite meters. Both cases alike are impossible, so to predicate truth is impossible: that is, no predication applied to finite beings is absolutely true.

Nevertheless, if we take a predicate like «tall», even if we are unable to determine exactly when a certain person is tall, to say «John is tall» when John is 2 meters tall, seems *absolutely* true, and to say «Peter is tall» when Peter is 1.50 meters tall seems absolutely false. That is, from the case that we *are* unable to precisify the application of a predicate to a subject in doubtful cases (cases of fuzzy or blurred boundaries), even allowing the fact (if it is a fact) that *every* concept or predicate *would have* some cases of dubious applicability, it does not follow that there do not exist any

precise cases, in which the result of applying that same predicate to a subject results in a statement uncontroversially true or uncontroversially false *simpliciter*.

The infinitist approach is more attractive only when we are concerned with predications on indeterminate subjects. So, to say of *something* that it is «big» seems to be «more true» the bigger the thing (a molecule more than an atom, a cluster of galaxies more than a single galaxy etc.), where big «in all truth» will be solely that which is absolutely infinite in size, and big at degree zero a (geometrical) point in space. The same goes for predicates like «heavy», «large» etc. But when predication is about a *particular* subject, like for instance a person, there is no infinite margin of application, in spite of it always being possible for someone to be a bit taller than the one at present who is most tall. How to fix, then, the point at which tallness of a human person is 1? Let's survey quickly three possibilities.

A) **that «tall» at degree 1 means «what in fact measures the tallest human being now alive»**

Let's suppose it is Robert Robertson, and that he is 2.70 m tall. This move has very inplausible consequences.

1) That people 2 m. tall are not very tall, because they are far away from Robert Robertson tallness. Naturally, how tall they are depends on where we fix tallness at degree 0.

1.1.) If we establish tallness at degree 0 in people being 0 cm. tall, then, people being 2 m. tall are rather tall indeed, because they are closer to Robert's 2.70 m than to 0 cm. But this has the dramatically bizarre consequence that people 1.30 m tall have tallness at degree 0.5 (they are half tall) just for being in the middle point in between Robert's height and 0 cm. And this is not to to mention the yet weirder consequence that we should postulate people of 15 or 20 cm to be qualified as extraordinarily little tall (perhaps: to be describable as tall, but only a very tiny bit tall).

1.2.) But if we wish to fix tallness at degree 0 in any other point, then we get two problems instead one: a) how to fix that point in a non-arbitrary way (this has no solution); b) suppose the point is fixed at 1.90 m. Then people who are 2m tall are not very tall. Even Kareem Abdul Yabbar is not perfectly or clearly tall, because he is *only* 2.17 m tall, and, as a result, he is closer to 1.90 (the point where anyone begins to be tall) than to 2.70, Robert Robertson's tall –where lies tallness at degree 1.

2) This option has a even more counterintuitive consequence. Let's suppose that the second tallest human in the world is Thomas Thomason, who is 2.30 m tall. That is, just in the middle between 1.90 and 2.70. And suppose also that Robert Robertson hurts his head in getting out of a lift, with fatal consequences. But then Thomas, without any effort and without be aware of anything has passed from being half tall to be perfectly tall; that is, the predicate «tall» no longer applies at degree 0.5 to him, but at degree 1. But all this seems ridiculous. Under any standard, Thomas is a person determinately, definitely and clearly tall, quite independently of the continuous fluctuations in the rest of the world population and their tallness, basketball players included.

B) **That tall at degree 1 means «the tallest human being of all times»** (up to now).

The first result of this move is merely epistemical: we don't know who, if any, of the present day humans is tall at degree 1. There is an exact point for tallness of human beings, but it is unknowable to us. This point is liable to vary, and probably will vary, in time. Now, almost certainly, this point must be placed at some point between 2.50 m and 3 m, and very likely closer to the first than to the second. Let us suppose that *as a matter of fact* the human being tallest in the history has measured 2.70 m. Then, except for the last of the former problems (the one posed by the passing away of giant Robert Robertson) all the other problems of the former analysis remain here in the same way.

C) **That tall at degree 1 means «the tallest possible human being»**.

This is the assignment that fits best with the infinitist approach, and for which a degree theorist must have been waiting for some time. But then the assignment of tallness at degree 1 is left totally undetermined, because we don't know how tall a human being could possibly be, either physically or conceptually. Should a «human» 500 m tall be counted as human? For presumably such huge changes in size will have an impact on his longevity, strength, intelligence (just imagine his brain size) and so on. Once more, the only coherent way out is to assign value 1 to someone measuring infinite meters. But, quite obviously, this is as absurd as regarding only people of 0 cm. tall as tall at degree 0.

The mistake seems to lie in treating adjectival terms like «tall» («young», «thin» etc.) which in ordinary language either does apply to an object or does not apply (or we are *in doubt* whether it does or does not) as if they were logically comparative terms: «taller», «younger», «thinner» etc. So the error lies in thinking that if «*x* is more *F* than *y*» is true, then «*x* is *F*» cannot be totally true. But if the notion of degree of truth is spelled out in comparative terms, then the occurrence of degrees of truth between perfect truth and perfect falsity in no way implies the occurrence of vagueness. Something can be clearly true or clearly false without being so in the infinitist sense[17].

Curiously enough, perhaps the infinitist gradualist analysis may work more properly with some sortal predicates. Take the example of «table»[18]. Let's suppose that table means: «*perfectly* horizontal plank supported by legs». Now pose the problem of the horizontalness of the plank (if we posed the problem of maximal and minimal tallness of its legs, the infinitist would again be helpless). Surely no actual table is *perfectly* horizontal, although most of them are very close to being. Well, concerning the horizontalness of the (plank of the) table, this time we have at least an intelligible and coherent way of appliying the values 0 and 1. 1 When the table is perfectly horizontal, 0 when it is perfectly vertical (it is 90° over the horizontal). In this case we have values fixed in a manner exact, clear and non-arbitrary, and also in agreement with our intuitions. Or so it seems. But this is only an appearance, for according to this, four legs supporting a plank with an inclination of 45° is a table at degree 0.5. But such

---

17 This has been remarked by Williamson (1994; p. 127).

18 This line of argument stems from a lecture of Lorenzo Peña slighltly prior to his 1996.

an object is not «somewhat a table» or a «half table»: is not a table at all, by the simple reason of being unable to fulfill the function tables usually have: be usable to eat, to study, to write or to play a chess game.

When, then, does a table cease to be a table because of the excessive inclination of its plank? My answer is, it depends upon the kind of table (once more we have context-dependence). The normal table, for instance a table suited *to eat*, ceases to be a table so suited when its inclination results, for instance, in the spillage of soup from a standard soup dish placed on the table (let's take the «standard dish soup» to be the average dish presently on sale; and the standard level for a soup dish to be «full» of soup to be the maximal amount of soup for the plate to be moved by ordinary people without spilling its liquid– maybe 85% of its volume, or so). Now if we are talking about a table for study (a desk), very likely the inclination allowed is higher (in fact some desks *are* inclined). Up to what point is it permissible? Just up to the point at which a sheet of paper or a book slides down without being touched (let's suppose 20°). The key is that it is the object's function that puts a limit to the margins of variability of its properties.

At this point the gradualist can take advantage of my manouevre and reply: O.K. if for any given putative vague term we are able in the end to set a point where the term no longer applies, why not consider this point the degree 0 of the application of the term? For instance, concerning the desk, why not say that with an inclination of 21° a plank with legs is a desk at degree 0 and with an inclination of zero degrees is a desk at degree 1? There are three reasons to the contrary: a) because once more we get counterintuitive linguistic and semantical implications (viz. that a plank with an inclination of 10° is a desk at degree 0.5; but many desks are made up with just that inclination!). b) Because degree theorists always tend to think that their procedure is able to solve once and for all *every* kind of case susceptible to sorites. c) Because many-valued finitist logics, infinitist logics, and fuzzy logics all of them are faced with insurmontable logical difficulties which are lacking in approaches attached to classical logic (like the epistemic theory, to name one).

Nevertheless, there is indeed a sphere, the one of adjectives involving perfections, at which the infinitist approach seems to work at its best. Let's take «wise». Here it seems sensible to say that one who knows absolutely nothing, who does not know the truth value of any proposition (or statement, or assertive sentence) at all, is «wise» at degree 0. And someone who knows everything, who knows the truth value of every proposition, is «wise» at degree 1 (if there exists non-propositional knowledge, a «knowing how» different in nature from propositional knowledge, then one will be «wise» at degree 1 who also possesses this practical knowledge at degree 1). The same runs for «good», «powerful», intelligent» etc. In short, in all kind of terms called by the tradition «pure perfections» is where the gradualist approach scores at its best. Plainly, these predicates do apply at degree 1 only to God. That is not a problem (at worst there will be no instances of such predicates at degree 1). The real problem is that following this conception of meaning, Aristotle and Einstein had an intelligence near to zero, for compared with a being capable of posing and solving any decidable question, their minds were certainly poor. But if, as accepted above, meaning supervenes on use, the former approach clearly is not in consonance with actual use of

these terms in English. Ordinary language applies «intelligent» or «wise» to humans and Aristotle or Einstein are paradigms of intelligence, like Socrates or Buddha are paradigms of wisdom, clearly closer to degree 1 than to 0. The trouble, again, is the same as before: we may assign the 0 (to the baby just born who knows nothing), but, how to assign 1 to something less than infinite? The troubles we found with «tall» recur here with «wise», «good» etc.

I see the gradualist approach as a sort of ontological argument multiplied and generalized for various categories. For objects *simpliciter*, there is a perfect table (cat, pine, car…), and all the rest are mere approximations; for qualitative properties, there is perfect wisdom (goodness, intelligence…), and all the rest are mere approximations; for quantitative properties: there is a perfect/maximal tallness (wideness, size, wieght…), and all the rest are mere approximations. But nothing of this is either useful for or relevant to accounting for vague terms in ordinary language.

Now, let's imagine the gradualist replying like this: you are compelling me to accept that 1 must always take an infinite value, and so, as a result, never applies to actual cases. But what is infinite is only the interval of real numbers between any two points on the straight line. Just call «0» the point at which the predicate no longer applies and «1» the point at which it absolutely applies. Then, taking again the example of the heap (possibly the most typical one), zero grains make up a heap at degree 0; and there is a finite number of grains k, at which at last we have a heap at degree 1. And that's it!

But this gives rise to a dilemma; k will be either a determinate number or an indeterminate number. If k is intrinsically indeterminate, there is no way to assign degree 1 of heapness; and this amounts to saying: concerning heaps we don't know and never will be able to find out whether there is just a single true heap at all (i.e. at degree 1). But it is this that is highly implausible: we *do know* not only that there are some heaps that are real, genuine and true heaps (of sand, of salt, of books) but that there are thousands of them!

And if, on the other hand, k is a determinate number, say 10,000 (for a heap of sand, for instance), then a heap of 15,000 grains is no more a heap than the one of 10,000, because it is not possible to exceed the value of 1 (and, incontestably, 15,000 grains do not make a hill). But this latter point is in total consonance with the epistemic theory: there is an exact point (to my mind contextually-dependent, and that for sand, sugar, boulders etc. is four, and in the case of bricks, books etc. is three) at which a number of items piled up makes a heap, and it is as much of a heap as one of 10,000 items or one of 12,000 or of 15,000. The difference with the gradualist is that for him, when we focus on the upper side he has to draw different consequences: «5,000 grains make up a heap» has a truth value of 0.5, «2 grains make up a heap» has a truth value of 2/10,000. In such a case 1) the gradualist applies a different criterion to the cases placed on the upper side of degree 1 (10,000 grains), than to the cases placed on the lower side. 2) It is as implausible to say that a heap of 15,000 grains is more of a heap than one of only 10,000, as it is to say that a heap of one grain is less of a heap that one of two grains: the former two are equally a heap; the later two are equally a non-heap.

I wish to conclude by regarding the same matter from the viewpoint of set theory. The spatial representation of a set corresponding to a vague predicate is that of a circle with fuzzy boundaries. The more vague the concept, the more fuzzy or blurred the boundaries. Having a fuzzy boundary, *there is no way of assigning the center of the circle to a determinate point* (for the same reason as it is senseless to try to assign the value 1 to a single point). But, in having boundaries, blurred as they may be, there are cases unequivocally clear of set membership and (more importantly) of non set membership. For as vague a concept can be, it cannot occupy all logical space. If a concept extended to all logical space it would be infinitely vague (and so useless). But just this is the representation that should follow from fuzzy logic.

Mark Sainsbury has considered regarding vague concepts not as circles with fuzzy boundaries, but as poles of attraction[19]. This conception is tempting. In that case there is a single assignation of 1: just the pole of attraction. And as we go away from it attraction decreases and *tends to zero*. But, and this is essential, it *never becomes zero*. For logical attraction of concepts within logical space is like the universal law of gravitation of masses in physical space: it spreads indefinitely towards the borders of the universe. The consequence of this analysis is that, for any vague term *x*, *absolutely everything* is *x* to some –even if minute– extent, because the pole *x* exerts its attraction all over the logical space. But it seems clearly false that everything (my umbrella, England, the least prime number and my uncle Sarah) be at a higher or lower degree tall, big, happy, beautiful, red etc. And if, as it is the case for many infinitist degree theorists, we think that all terms are in the end vague, so much the worse, because then everything is to some extent anything.

In summary, if by «object» we mean what ordinary language takes it to be (and not what the metaphysician determines as the fundamental components of reality or as genuine individuals) then there are not so many vague objects as it is usual these days to claim (hard luck for the librarian if for most books it were doubtful, not just under what category they should be classified, but whether they are books or not!). Concerning words, certainly there are many vague words, although maybe not that many. And third, and fundamentally, semantical problems raised by such words are not to be solved with gradualist approaches, neither finitist nor infinitist, nor in terms of fuzzy logic. On the contrary, these approaches produce really weird results. I think problems of vagueness should be tackled case by case, first, restricting maximally the context; second, artificially stipulating a sharp boundary when we are dealing with artifacts, for having created the object we are entitled to create the concept with so much sharpness as desired (the maximal, for logical ends); third, in natural cases nature itself has established in great measure its own sharp boundaries (think of atoms, molecules or minerals); fourth, where troubles with vagueness yet remain, this may be

---

[19] Sainsbury, (1990, in Kefee, 196; p. 259). He has claimed –rightly, to my mind– that the notion of «fuzzy boundary» is incoherent, because any limit or boundary splits logical space into two. According to that, «fuzzy boundary» actually means «no boundary at all». And that is the point: a concept with no boundaries pervades *all logical space*, and this seems to be not so much incoherent, as absurd. Maybe you could say of all *physical objects* that are more or less green, even if very very slightly green; but how could you sensibly say that also triangles, feelings, God and justice have some of greeness?

due to ignorance of our own (perhaps avoidable, perhaps not), maybe to a lack of conceptual precision (normally due not to intellectual lazyness, but to a non-necessity of precision, or even to a necessity of looseness). I guess that in this, as in many other philosophical questions (like «do theological statements have sense»?) there is no single answer valid for all cases alike, but we instead should proceed step by step.

References

Dummett, M. (1975): «Wang's Paradox». *Synthese*, 30; 301-24 [in Kefee, 1996; 99-118].

Edgington, D. (1996): «Vagueness by Degrees». In Kefee, 1996; 294-316.

Evans, G. & McDowell, J. (eds) (1977): *Truth and Meaning: Essays in Semantics*. Oxford, Clarendon.

Fine, K. (1975) «Vagueness, Truth and Logic». *Synthese*, 30; 265-300 [in Kefee, 1996; 119-50].

Hart, W.D. (1991-2): «Hat-Tricks and Heaps». *Philosophical Studies* (Dublin), 33; 1-24.

Horgan, T. (ed) (1995): *Vagueness*. Spindel Conference, 1994. *The Southern Journal of Philosophy*, Vol. XXXIII, Supplement.

Kefee, R. & Smith, P. (e ds.) (1996): *Vagueness: a Reader*. Cambridge, Mass. The MIT press.

Machina, K. (1972): «Vague Predicates». *American Philosophical Quarterly*, 9; 225-33.

Machina, K. (1976): «Truth, Belief and Vagueness». *Journal of Philosophical Logic*, 5; 47-78 [in Kefee, 1996; 174-203].

McGee, V. & McLaughlin, B. (1995): «Distinctions Without a Difference?». In Horgan, 1995; 203-52.

Morgan, C.G. & Pelletier, F.J. (1977): «Some Notes concerning Fuzzy Logics». *Linguistics and Philosophy*; 79-97.

Peña, L. (1996): «Grados, franjas y líneas de demarcación». *Revista de filosofía*, IX; 121-50.

Sainsbury, M. (1990): «Concepts without Boundaries». In Kefee, 1996; 251-64.

Sainsbury, M. (1995): *Paradoxes*. Cambridge, Cambridge Uni. Press.

Sainsbury, M. (1995b): «Why the World Cannot Be Vague». In Horgan, 1995; 63-82.

Tomberlin, J.E. (Ed.) (1994): *Philosophical Perspectives* 8: *Logic and Language*. Atascadero, Ca, Ridgeview.

Tye, M. (1994): «Sorites Paradoxes and the Semantics of Vagueness». InTomberlin, 1994; 189-201 (in Kefee, 1994; 281-93).

Williamson, T. (1990): *Identity and Discrimination*. Oxford, Blackwell.

Williamson, T. (1994): *Vagueness*. London, Routledge.

Williamson, T. (1999): «On the Structure of Higher Order Vagueness». *Mind*, 127-44.

Wright, C. (1976): «Language-Mastery and the Sorites Paradox», in Evans & McDowell, 1977.

Wright, C. (1987a): «Further Reflections on the Sorites Paradox». *Philosophical Topics*, 15; 227-90 [in Kefee, 1996; 204-50].

Wright, C. (1992): «Is Higher Order Vagueness Coherent?». *Analysis*, 52; 129-39.

Enrique Romerales

Universidad Autónoma de Madrid

<Enrique.Romerales@uam.es>

SORITES (ΣΩΡΙΤΗΣ), ISSN 1135-1349

Issue #11. December 1999. Pp. 66-81.

Are There Mental Entities? Some Lessons from Hans Reichenbach

ARE THERE MENTAL ENTITIES? SOME LESSONS FROM HANS REICHENBACH

by Jeanne Peijnenburg

0. Introduction

'The mental and the physical are not made for each other', wrote Davidson, echoing Brentano's famous thesis that the intentional idiom is irreducible. But if mental terms cannot be translated into physical terms, how can they be translated? What is the meaning of mental terms such as 'belief', 'desire', 'intention'? Or to put it ontologically: what sort of entities are beliefs, desires, intentions?

These questions are core issues in the contemporary philosophy of mind, and the answers are many. Beliefs and desires have been related to actions, to brain processes and to computer programs. The relations in question have been pictured as reductions, as superveniences, as emergence relations, as type- and as token-identities. Nowadays there exists a whole gamut of different positions on the mental: there is physicalism, eliminativism, behaviourism, functionalism, parallelism, epiphenomenalism, interactionism, anomalous monism, and, last but not least, transcendentalism.

In this paper I propose to discuss Hans Reichenbach's views on the mental, especially his views on *abstracta* and *illata*, and compare them with some ideas of Carnap. It was Daniel Dennett who, while explaining his own views on the nature of mental entities, drew attention to Reichenbach's abstracta and illata in Dennett 1987 (cf. Dennett 1991a and Dennett 1991b). However, Dennett is not a historian of philosophy. He never aimed to present Reichenbach's theory in full detail, but only cited it in passing. As a result, the great potential of Reichenbach's ideas concerning mental entities still remains largely unnoticed.

Reichenbach is often considered to be an adamant logical positivist, propagating ideas on the mental that are far too 'physicalistic'. Like Carnap, he is taken for an unrefined behaviorist who perceives the psychological realm as consisting only of gross stimuli and raw responses. In fact, however, his ideas on mental entities are quite sophisticated and by no means the crude positivistic products that some make of them. For instance, as we will see, Reichenbach makes ample room for private experiences and for the first person view, matters that after all are *Fremdkörper* in behavioristic theories of the more simple sort.

Hence I am going to undertake a journey that is rare in analytic philosophy: I propose to go back in time and examine ideas that are more than fifty years old. Analytic philosophers are in general not historically oriented. They tend to forget that an occasional excursion into the past may be worthwhile, especially when the jaunt includes a visit to one's very own roots, as is the case here. Carnap and Reichenbach are early representatives of analytic philosophy, and they commented thoroughly on the nature of abstract entities and the meaning of abstract terms. As we will see, a comparison of their views yields lessons about the mental that might still be of worth today.

1. Reichenbach: reduction and projection

In *Experience and Prediction* Hans Reichenbach distinguished between *direct* and *indirect* propositions. At first sight, the distinction is an unalloyed neo-positivistic product. Direct propositions are the familiar observation sentences capable of direct verification; indirect propositions are indirectly verified, which means that they are reducible to other propositions capable of direct verification.

The interesting question, of course, is what exactly 'reduction' means here. What does it mean for example to say that a proposition about an event horizon, i.e. the border of a black hole where the escape velocity equals the speed of light, 'is reducible to' a class of observation sentences? Cosmologists in the entourage of Stephen Hawking detect an event horizon by measuring electromagnetic radiation emitted from a shrinking star, and by comparing the measured signals with the predictions of quantum field theory and general relativity. The cosmologist's claim that in an event horizon the photons 'hover', i.e. neither escape from the hole nor fall back into it, is based on various sentences concerning outcomes of measurements made with miscellaneous instruments. Each of those instruments, we assume, is placed on our planet, thousands of millions of miles removed from what they are observing: the happenings in an event horizon. What is the relation between the (indirect) statement that the photons in an event horizon hover, and the (direct) statements about results of measurements?

Reichenbach's answer to this kind of question is often put on a par with that of the early positivists. However, the differences are considerable. The early positivists regard any relation between direct and indirect statements as an equivalence: an indirect statement (*IS*) is true if and only if the set *SD* of direct statements is true, where *SD* can contain conjunctions, disjunctions, negations etc. Reichenbach, on the other hand, finds this view too simple. He points out that often *IS* has a surplus meaning compared to the meaning of the propositional function of the statements in *SD*. In those cases *IS* can be true while one or more statements in *SD* may be false, and vice versa. Hence the relation might be not an equivalence but a probability connection: *IS probably* implies *SD* and vice versa.

Reichenbach calls the probability connection a *projection*, and the equivalence relation a *reduction*. An example of a reduction is the relation between (1) «The species of wallabies has its home in Australia» and (2) «All wallabies descend from ancestors that lived in Australia» (the example is a modified version of Reichenbach's example). (1) is an indirect statement, for it contains indirectly verifiable terms: 'the species' does

not denote a concretum, and neither does ‘home’. It is however completely reducible to (2), which contains, besides logical terms such as ‘all’, only terms that refer to what Reichenbach calls concreta, i.e. physical objects or processes supposedly accessible to direct observation. In Reichenbach’s words, (1) denotes a non-concretum that is a *reductive complex* and the expressions in (2) refer to the *internal elements* of this complex (Reichenbach 1938, 110). Another example of a reduction is the relation between a wall and the bricks of which it is built. Every statement about the wall (the reductive complex) can be translated into a statement about the bricks (the internal elements). Of course, the bricks can only form a wall if they are arranged in a particular way: the wall is not dependent upon just bricks, but upon a certain configuration of the bricks. Thus Reichenbach says that the reductive complex is equivalent to the internal elements together with a «constitutive relation».

On the other hand, if an indirect statement is connected to direct statements through a projection rather than through a reduction, then the indirect statement denotes a *projective complex* and the direct statements refer to *external elements*. Reichenbach gives the following example of a projection:

> We imagine a number of birds flying within a certain domain of space. The sun rays falling down from above project a shadow-figure of every bird on the soil, which characterizes the horizontal position of the bird. To characterize the vertical position also, let us imagine a second system of light rays running horizontally and projecting the birds on a vertical plane which may be represented by a screen of the kind employed in the cinemas. We have, then, a pair of shadows corresponding to every bird … every proposition concerning the movement of the birds is co-ordinated with a proposition about the changes of the pairs of shadows. (Reichenbach 1938, 108).

In this example, every single bird is represented by a unique system of marks, in the sense that each movement of the bird corresponds to a movement of the shadows. The birds are however not identical to the shadow pairs, no matter how the latter are arranged with respect to each other. Instead, the birds are only *projected* on to the screen and the soil: they constitute projective complexes of which the shadows are the external elements. This means that no proposition about a bird is completely reducible to a proposition about a shadow pair, and hence that between propositions about the birds and propositions about the shadows only *probability connections* exist:

> if we see the marks only, we may infer with a certain probability that they are produced by birds, and if we see the birds only, we may infer with a certain probability that they will produce the marks. … there is no strict relation between the truth values of the co-ordinated propositions. The proposition about the birds may be true, and that about the marks may be false; conversely, the proposition about the birds may be false, and that about the marks may be true. (Reichenbach 1938, 109).

Projective complexes such as the birds are called *illata*, i.e. ‘inferred things’ (Reichenbach 1938, 212) — other examples of illata are radio waves, atoms, and all sorts of invisible gases. In general, illata exist not only in time but also in space.

Reductive complexes, on the other hand, are *abstracta* (Reichenbach 1938, 93; Reichenbach 1951, 263). Abstracta mostly have no spatial qualities at all, although one could say that they have an existence in time. Thus the species of wallabies and a family's furniture are abstracta, as are the political state, the Bodleian Library, the spirit of the nation, and the financial crisis.

At this juncture, an important observation must be made. It concerns the so-called *internal projection*, a notion that will prove to be significant in Section 5, where we deal with beliefs and desires. Since illata are projective, whereas abstracta are reductive complexes, the elements of illata are of course external while the elements that constitute abstracta are internal. However, Reichenbach stresses that one and the same entity may function as an element or as a complex, depending upon the viewpoint. Thus atoms may be *internal elements* out of which concreta are *built up*, or they may be *projective complexes* that are *inferred from* concreta. In the first case the concreta actually are abstracta (they are complexes that can be completely reduced to atoms), in the second case the concreta are the (external) elements from which the atoms, as projective complexes, are probabilistically inferred. Since in the latter case the projection has a somewhat peculiar character («it leads to things which are the internal elements of the things from which the inference started»), Reichenbach calls it an *internal projection* (Reichenbach 1938, 216). In Section 5 we will see that Reichenbach, in the end, envisages beliefs and desires as internal projections.

2. Carnap: pure dispositions and theoretical primitives

It is interesting to see that Reichenbach's distinction between abstracta and illata has a striking parallel in Carnap's distinction between *pure dispositions* and *theoretical constructs* (Carnap 1956). As is well known, the latter distinction concerns two kinds of scientific concepts; basically it relies on the distinction between an observation language, L_O, and a theoretical language, L_T. Theoretical terms cannot be explicitly defined in L_O and are thus introduced in L_T by means of postulates. Pure disposition terms, on the other hand, occupy an intermediate position between observation terms and theoretical terms. They belong neither to L_O nor to L_T, but are part of a language in between the two: Carnap's extended observation language L'_O.

As do the terms that denote abstracta and illata, disposition terms and theoretical terms signify non-observable or non-concrete complexes. Moreover, the criterion for distinguishing theoretical and disposition terms is the same as that by which illata are separated from abstracta. In Carnap's view, a disposition *D* ascribed to an object *X* by an investigator *Y* is a *pure disposition* if and only if there exist an *S* and an *R* such that:

(i) *S* is a process that affects *X* and is observable by *Y*,

(ii) *R* is a reaction of *X* and likewise observable by *Y*,

(iii) *D* is identical to (a certain combination of) *S* and *R*.

On the other hand, *D* is a theoretical primitive or, as I shall call it, a *theoretical disposition* if *D* is manifested by *S* and *R*, but does not coincide with *S* and *R*. Thus *D* is a theoretical disposition if (i) and (ii) are true whereas (iii) is false. Consequently, theoretical dispositions are only probabilistically connected to concreta. It is precisely

the existence of probabilistic relations that constitutes the difference between pure and theoretical dispositions:

> The decisive difference is this: on the basis of the theoretical interpretation, the result of this or of any other test or, generally, of any observations, …is not regarded as absolutely conclusive evidence for the state in question; *it is accepted only as probabilistic evidence, hence at best as a reliable indicator, i.e., one yielding a high probability for the state*. (Carnap 1956, 71; my emphasis).

Thus Reichenbach's distinction between abstracta versus illata and Carnap's distinction between pure versus theoretical dispositions are based on the same criterion. Both are grounded in *the absence or the presence of probability relations*. Carnap's theoretical constructs as well as Reichenbach's illata are probabilistically connected to concreta. Similarly, Carnap's pure dispositions as well as Reichenbach's abstracta coincide with sets of observable things or events. Of course, the *terms* signifying Reichenbach's illata and Carnap's theoretical entities are also alike: both have a surplus meaning over terms that refer to observable things or events. Thus both may be applied even if the corresponding sentences about concreta are false, or not applied even if those sentences are true. On the other hand, terms denoting abstracta or pure disposition lack this surplus, and are completely translatable into the vocabulary of observables.

Apart from the absence or presence of probability relations, there is another important similarity between the two approaches. Both the Carnapian and the Reichenbachian distinction are *time- and theory-dependent*. In either case, the nature of a particular non-concrete complex is not clear *a priori*: it has to be identified on the basis of empirical findings which in turn are based on our theory. As a result, previous decisions may be reconsidered in view of new evidence, so that, in Reichenbach's case, an illatum can become an abstractum and vice versa. The same goes for the Carnapian distinction. Time and again Carnap stressed that scientists have a certain liberty in regarding non-concrete terms as being either purely dispositional or genuinely theoretical terms; in the end, their decision is guided by considerations of empirical usefulness and efficiency, in combination with a theory at hand.

3. The question of existence

Until now we only talked about sentences and terms. We explained that the relations between sentences are either probabilistic or not, and that meanings of terms are either surplus meanings or not (all dependent upon empirical findings as well as on a theory). However, we have been reticent about the actual things in the world. To what exactly do the terms we have spoken of refer? What, if any, is the pukkah *existence* of non-concrete or non-observable complexes? In the present section we will address that question with respect to Reichenbach's abstracta and illata, but what we say will also apply to Carnap's pure and theoretical dispositions. As far as existence is concerned, Carnap and Reichenbach roughly held the same opinions. Both underscored that 'existence' should be read as 'existence-according-to-a-theory'. And both maintained that the choice of a theory has a conventional element to it (without, of course, being totally a matter of convention).

Consider again the term 'species of wallabies'. According to Reichenbach this term denotes an abstractum, but does this abstractum really exist? Reichenbach's answer here is a yes-and-no. On the one hand we may say that the species of wallabies exists, meaning that many wallabies exist and that they have certain biological qualities in common which distinguish them from other animals. On the other hand, we may also say that it does not exist, meaning that many wallabies exist and that any proposition containing the term 'the species of wallabies' can be translated into propositions concerning those wallabies (Reichenbach 1938, 96). For Reichenbach the question whether or not abstracta exist is settled by a decision rather than being a matter of truth-character. The decision may be an affirmation, or a denial, or neither of them. For instance, of a family's furniture we probably will say that it exists, of the height of a mountain that it does not, and in the case of human society the decision will be somewhat indeterminate. But whatever its outcome, it remains a decision and thus a matter of convention; on no account may the abstract term be taken to have a surplus meaning. The question of whether or not an abstractum exists therefore is a practical affair; regarding the matter as a theoretical topic is to raise a pseudo-problem.[1]

Illata, on the other hand, form a different kettle of fish. Illata do have an existence of their own, and terms denoting them have a surplus meaning which goes beyond the meaning of the terms for the (external) elements. As Reichenbach phrases it: «The relation of the illata to the concreta is a projection … The illata have, therefore, an existence of their own…» (Reichenbach 1938, 212).

The question can be illustrated on the basis of the term 'atom' (cf. Reichenbach 1951). Propositions about atoms can be connected to propositions about macroscopic bodies, albeit only probabilistically: the propositions about atoms may be true whereas those concerning macroscopic bodies may be false, and vice versa. For this reason, most people will deny that the term 'atom' is just an abbreviation for certain relations between macroscopic bodies. Instead, they will maintain that it refers to some thing from which those relations can be explained.

The atom example also illustrates another point that we have made above, namely that the distinction between abstracta and illata is time-dependent. The theory of the atom emerged as a pure speculation from the philosophy of Democritus in the fourth centry B.C., after which it took another twenty-two centuries before it was subjected to an empirical test. About 1800 it was found that compounds (such as for instance sucrose) consist of chemical elements (carbon, hydrogen and oxygen), of which the weights make up a fixed proportion that can be expressed in whole numbers. The English chemist Dalton realised that these fixed and quantitative relations require

---

[1] An example of such a pseudo-problem is provided by the traditional controversy between nominalists and realists. According to Reichenbach, nominalists and realists disagree with respect to the existence of abstracta: the former deny, and the latter assert that abstracta exist (Reichenbach 1938, 93-98). The debate in question was also seen by Carnap as a pseudo-problem, basically for the same reasons. In Carnap's view, both factions battle about a so-called 'external existential statement', this is a statement in which it is claimed that certain entities exist *as such* rather than exist 'internally' according to the rules of a certain linguistic framework. An external existential statement is a pseudo-statement; it embodies a practical decision rather than a theoretical claim (Carnap 1950, 1963).

an explanation at the microscopic level. It turned out that all macroscopic bodies are made of microscopic particles, viz. atoms (in the case of sucrose, twelve atoms of carbon combine with twenty-two atoms of hydrogen plus twelve atoms of oxygen). By the end of the nineteenth century most philosophers and physicists agreed that atoms have an existence of their own, although there still were dissidents such as Ernst Mach, who kept believing that the word 'atom' is just an umbrella term for a reducible complex.

To summarize: when taken theoretically, the question whether or not abstracta or pure dispositions exist is a standard pseudo-question. It can be answered by 'yes' and 'no' alike, depending on where you wish to lay the stress. If you wish to stress that sentences about abstracta resp. pure dispositions can be completely reduced to sentences about observables, then you are likely to come up with 'no'. But if you wish to say that the observables in question exist as a group, then your reaction will be that abstracta do exist. All this is quite different from the illata case. There we encounter entities that do have an existence of their own, a fact that is revealed by the probability relations between sentences about illata and sentences about concreta.[2]

4. Are there mental entities?

Beliefs, desires, and the like are non-concrete complexes of a mental kind. As such, they can be either abstracta or illata, either pure dispositions or theoretical primitives. What does Reichenbach say about them?

Reichenbach's view on beliefs and desires is quite sophisticated and certainly not the naive neo-positivistic thing that many see in it. Even scholars who sympathise with it seem to overlook how subtle and ingenious his view actually is. We could think here of Daniel Dennett, who first referred to Reichenbach's abstracta and illata in Dennett 1987. In order to get rid of the «mixed bag» of folk psychological notions, Dennett proposes «a divorce» between abstracta and illata (ibid., 57). This separation should enable us to create two tidy new theories on the mental:

> one strictly abstract, idealizing, holistic, instrumentalistic — pure intentional theory — and the other a concrete, microtheoretical science of the actual realization of those intentional systems — what I will call sub-personal cognitive psychology (*ibid.*).

By suggesting a split and directing abstracta to the one theory and illata to the other, Dennett ignores the essence of and the interesting thing about beliefs and desires, namely that they have a mixed nature. Beliefs and desires are neither plain flesh nor pure fowl. They stand «somewhere midway between abstracta and illata», and are being «pulled in two directions» (ibid., 55, 57).[3] Reichenbach, for his part, was fully aware

---

[2] As was pointed out to me by Huw Price in private communication, the distinction between abstracta and illata resembles a distinction that is common in functionalist literature, viz. that between role states and realiser states. It seems to me that role/realiser states are not entirely the same as abstracta/illata, although there are indeed similarities. The matter needs to be examined more closely, and I will not elaborate on it here.

[3] On the dual nature of beliefs and desires, see also: Peijnenburg and Hünneman, to appear.

of this dual nature of beliefs and desires. Rather than eliminating it by heading for a split, he tries to incorporate it by showing how abstracta and illata merge together in the mental dispositions that make up our «higher psychical life» (Reichenbach 1938, 239). Let us now see how he tries to accomplish this task.

Reichenbach's starting point is the common opinion that psychology is the science about our inner world. Next to 'the higher psychical life' of beliefs and desires (with which we will deal in Section 5), our inner world entails 'lower' psychic experiences such as impressions or sensations. The latter Reichenbach describes as «phenomena occurring within my mind but produced by physical things outside my mind» (Reichenbach 1938, 89-90). Examples of such phenomena are: seeing a bird flying by, feeling a man touching your elbow, hearing Donald Davidson chuckle.

Apart from the words 'impressions' or 'sensations', Reichenbach also uses the term 'presentations' to denote the phenomena in question. What is presented by presentations are 'immediate things' rather than 'objective things'. Objective things are the observable things around us: tables, trees, tunes, trains. Immediate things, on the other hand, are the things experienced by the senses. The two are by no means the same. Taken as an objective thing, a pair of train rails is parallel, but taken as an immediate thing the two rails converge. Similarly, a flying bird as an objective thing differs from a *seen* flying bird that is an immediate thing. Objective things somehow remain the same at all times and in all spaces, whereas immediate things change with persons and perspectives: a flying bird looks smaller or larger depending on the distance, it can be taken for an airplane, a piece of paper, a drifting balloon, an UFO et cetera. Immediate things can correspond to objective things, but they can also be dreams or hallucinations. What we have called concreta are objective things; they are the physical objects or processes accessible to direct observation that form the basis of all the sciences. Immediate things, on the other hand, are the objects of psychology; they are the sort of things psychologists try to describe by referring to the basis of concreta.

As Reichenbach sees it, both the outer and the inner world can be reconstructed on the basis of concreta. This means that not only the outer world of the physical scientists, but our own inner world too can be erected on the basis of observable objects and processes. Phrased in this way, the idea is not a particularly novel one: it is shared by empiricist philosophers of all times and of all places. However, in the hands of Reichenbach this familiar thought gets an original twist. For according to Reichenbach, psychology «is a science which infers illata from concreta» (ibid., 247). To see what this means, let us take a closer look at both the concreta and the illata in question.

The concreta in psychology are observable objects or processes that can be either *outside* or *inside* your body. The outer concreta can function in two different ways, as stimuli or as responses. Typically they are *stimuli* whenever we are working within a first person perspective, whereas they will be *responses* when the third person view prevails. Thus if a car riding up causes you to believe that a car is riding up, then, from your first person perspective, the approaching car functions as a stimulus for your belief and for your subsequent action of jumping aside. The driver, on the other hand, who from his third person's perspective sees you jumping aside, may conclude that you do

not want to get hit and are believing that by jumping aside you will not get hit. He describes your psychic life in terms of your reactions while you yourself are inclined to report it in terms that are taken from the stimulus sphere. Of course, the roles can be reversed: one can look upon oneself as an outsider: from the fact that you are jumping aside you yourself may safely infer that, apparently and perhaps surprisingly, you are not yet tired of life. These reversions are however exceptions. Normally one describes one's inner life by referring to stimuli, whereas the psychic life of others is mainly described by citing their observable reactions. (Mainly but not exclusively: the driver probably observed that I saw a car approaching, i.e. he noticed that a stimulus was acting upon me.)

Concreta need not be outer processes; they can also occur inside your body. In fact, Reichenbach describes two classes of inner concreta. The first is the more interesting one, since it reveals a difference between psychology and physics. It is the class of inner concreta that can only be felt by the person in whose body they occur. A physicist would certainly banish such processes as being unscientific, but in psychology they function as stimuli that are perceived only by the person who has them. An example of such an inner stimulus is the pulsation of the heart, but also a bodily awareness such as the feeling of hunger. According to Reichenbach, hunger is an inner process that is accessible to «the inner tactile sense» (Reichenbach 1938, 238). It is a concretum that is «directly observed in the same sense that we observe, say, a movement of our legs with the tactile sense» (ibid., 236). As the mentioning of leg-movement already indicates, the distinction between processes that are observed by the 'inner tactile sense' and outer reactions is often not clear. Some processes, such as blushing, may be described in the reaction language as well as in the language of inner self-observation.

Concreta that function as inner stimuli differ from objects and processes that are observed by a physiologist; the latter we call inner concreta of the second class. Pictures on the retina, changes in the optic nerve, transformations of the brain, convulsions in the stomach, secretions of the salivary gland: we ourselves do not observe any of those processes. Yet they all take place in our own body, and they all can be directly observed. However, they are only observed by outsiders. Rather than being described in the language of reactions or of (inner or outer) stimuli, they are reported in the language of the physiologist who can observe the interior of bodies directly. Again, one and the same process may be described as an inner concretum of either the first or the second class, e.g. a certain process might be described as hunger or as convulsions of the stomach.

Inner concreta, whether of the first or the second class, should never be confused with illata. The confusion is easily made, since in psychology illata too are inner processes. Yet the two processes differ greatly: inner concreta can, whereas illata cannot be directly observed. Rather than being observed, illata are inferred from (inner or outer) concreta. This inference takes place along the lines of classical probability theory, since, as we have seen, between sentences about illata and sentences about concreta only probability relations exist.

What sorts of things are the illata in psychology? We have already indicated Reichenbach's surprising answer: sensations. Against the received opinion, Reichenbach argues that an optical or an acoustical sensation is not observed, but inferred. A man is exposed to an objective thing in front of him; as a consequence, he *sees* an immediate thing before him and *has* a sensation. He does not observe this sensation as he observes the thing before him or as he observes the pulsation of his heart. He must infer it, since he «does not know anything about its qualities, except that it has a certain correspondence to the immediate thing he observes. It is an unknown, *X*, determined as a function of the immediate thing observed» (ibid., 237).[4]

As our explanations have hopefully shown, Reichenbach's theory of the mental cuts across the standpoints of behaviourists and non-behaviourists alike. In conformity with the habit of holding him for a logical positivist of the simple minded sort, Reichenbach is often pictured as a rigid behaviourist. It should be clear by now that this is a grave mistake. Rigid behaviourists describe people's minds in terms of their reactions to certain stimuli. They only have an eye for the outer concreta, and, since they prefer the third person view, especially focus on those outer concreta that function as reactions. Hence behaviourists have little or no interest in the essentials of Reichenbach's theory: the stimuli (especially the inner stimuli), the immediate things, the allowance of the first person view.

Needless to say, Reichenbach's theory also differs from that of the non-behaviourists, which in his case are mainly traditional psychologists fond of introspection. The proponents of introspection make the mistake mentioned above: they fail to distinguish between inner concreta and illata. If introspection means that stimuli can be inside your body (as is the case with some inner concreta), or that some things are best described by the person who has them (as is the case with immediate things), then Reichenbach has nothing against it. If, however, by introspection is meant that you can observe psychical phenomena, then you are on the wrong track. For psychical phenomena are not directly accessible by an inner sense. They are illata, that can only be indirectly inferred and never be directly observed:

> The mischief of psychology does not arise from [the method of self-observation], but from a false interpretation that have been given to it. It is the concept of introspection which marks this misinterpretation, as it is meant to indicate a direct view of psychical phenomena. The interpretation developed by us, in the sense of a stimulus language, is free from such misconception … The method of self-observation, if it is conceived as the method of stimulus

---

[4] Reichenbach's words here might remind us of what God said to the Mortal in a paper by Raymond Smullyan that became justly renowned:

> You can no more see me than you can see your own thoughts. You can see an apple, but the event of your seeing an apple is itself not seeable. And I am far more like the seeing of an apple than the apple itself (Smullyan 1977, 330).

The idea sounds worth exploring: God as an immediate thing, triggered by objective things around us such as apples. About His qualities and existence, however, we do not know anything. Those we have to infer, using the laws of probability and perhaps taking our inspiration from some mediaeval philosophers who attempted to prove His existence.

> language, is not less objective than reaction language. However, it opens up possibilities for observation which do not exist for the reaction method. (ibid., 243-244).
>
> The idea of introspection is an illusion if we understand by introspection an observation of 'psychical' phenomena; what we observe are physical phenomena, and the inner processes corresponding to them are only inferred. They are *illata*; and the basis from which we infer them is the totality of concrete objects of the physical world. (ibid., 227).

As we have seen, these 'concrete objects of the physical world' exist either outside or inside our body. In the first case they are stimuli or responses, in the second case they are either stimuli or processes accessible to a physiologist. In neither case are they things that one discovers by introspection.

5. Beliefs and desires

We have seen that for Reichenbach psychology is the inference of illata from concreta, and we have explained what illata and concreta are. Our main question, however, is still unanswered. What is the nature of those entities that make up «the higher psychic life»? What are beliefs and desires? Are they the same things as sensations, viz. illata, or should we take a different view? It is time to deal with these issues, and to explain why Reichenbach's view is so interesting for us today.

Like Carnap, Reichenbach regards beliefs and desires as dispositions. And as we have explained, both Reichenbach and Carnap hold that there are two sorts of dispositions: Carnap distinguished between pure and theoretical dispositions, while Reichenbach argued that dispositions can be either abstracta or illata. However, Reichenbach claimed that *psychological* dispositions such as beliefs and desires are always abstracta. What is more, they are abstracta of *a special kind*. For the internal elements that compose beliefs and desires are not only concreta (as in ordinary abstracta) but also illata:

> Psychology is a science which infers illata from concrete objects. The inferred objects are projective complexes of these concrete objects. Since some of the objects of psychology such as bodily feelings are accessible to the inner tactile sense, the inferred illata in such cases are internal elements of the observed concrete objects; it is therefore the process of internal projection which plays a role here. The 'higher' psychological objects, and just those most frequently occurring in practical psychology, i.e., psychology as needed for daily life, are abstracta, built up of concreta and illata. (ibid., 1938, 247).

Thus the situation seems to be as follows. In psychology we aim at knowledge of people's minds, including our own mind. Since we are unable to read minds directly, we must start by looking at people's bodies, including our own body. What we then see is a number of objective things: a motorcar approaching rapidly, a man next to us jumping aside, a sharp pain in our left arm, a bone sticking out of the man's leg. The example is not very pleasant, and perhaps I should apologise for that, but it illustrates clearly the four objective things that we have distinguished: outer stimuli, outer responses, inner stimuli, and the objects of physiological observation.

Of course, I am the one who senses these objective things, and therefore a number of immediate things is present too. There is a motorcar and a seen motorcar, a jumping man and a seen jumping man, a bone and a seen bone. This does not mean, however, that objective and immediate things coincide. The psychologist who afterwards is going to treat me for the trauma caused by witnessing this accident is not interested in the objective events. She primarily wants to know what went on in my mind. That is, she wants to know what my impressions are, and those are characterised by the immediate rather than by the objective things. The objective events are only important in so far as they teach us something about my impressions. At the same time, however, objective things are the only things my psychologist can rely on. Nothing else than physical concreta can guide her — or me — in the search for what exactly went on in my mind. But since physical concreta do not fully determine my mental impressions, we need probability relations to infer the immediate from the objective things.

And now we are able to draw an interesting conclusion. In Reichenbach's view, the sensations, feelings, impressions or 'presentations' that make up the 'lower' psychic life are all illata: they are immediate things that are inferred from concreta. Together with (other) concreta, these illata constitute our 'higher' psychic life, notably our beliefs and desires. From this it follows that beliefs and desires are abstracta. However, they are abstracta of a special kind. For they are composed not just of concreta, but of concreta and illata. Hence the illata in question, that were originally inferred from concreta, now function as internal elements of abstracta, c.q. of beliefs and desires (cf. Section 1). This means that beliefs and desires have a mixed nature. They are neither pure abstracta nor pure illata, neither plain reductions nor pure projections. Rather they are examples of what in Section 1 were called *internal projections*.

Those are the outlines of Reichenbach's theory of the mental. What lessons can be drawn from them?

The first lesson concerns the assumption, also entertained by Carnap, that the mental and the physical are related in a probabilistic way. Reichenbach developed this assumption in his notion of probability meaning, whereas Carnap made it the basis for his theories of inductive logic. Neither of the two projects proved to be entirely successful, but that does not mean that research into probability connections is doomed to disappointment. After all, the idea that probabilistic features play a role in the relation between mental and physical features is by no means unrealistic. It is therefore somewhat surprising that the notion of probability seems to be entirely forgotten whenever one talks about relations of emergence or supervenience. Elsewhere we have written about probability (Atkinson & Peijnenburg 1999), and we will not dwell upon the subject here.

The second lesson pertains to Reichenbach's explanation of the first person view. Self-observation has always been a problem for empiricists, naturalists, physicalists and, in general, all philosophers who are scientifically oriented. On the one hand they cherish the idea that outer, verifiable events make up the basis for science, and, in fact, for any meaningful statement. On the other hand, they deplore an all too rigid approach, in which any special access of a person to (part of) his mental life is bluntly

denied. Reichenbach offers a way out. He runs with the hare and hunts with the hounds by making a distinction between inner concreta (that can be self-observed) and illata (that are only inferred). In this manner introspection becomes a fact, but it is no more extraordinary than the fact that I can see your back whereas you cannot. Of course, Reichenbach is not the only one who offered an empiricist solution for the problem of first person authority. It cannot be denied, however, that his approach is rather original and worthy of further exploration. In any case it is a welcome antidote for the increasing number of approaches that have abandoned empiricism altogether (cf. the transcendentalism of White in White 1991).

The third lesson is the most important one. Simply put, it boils down to the advice that we should be tolerant. By this I do not mean Carnap's famous adage about freedom in the choice of language systems («Let us be cautious in making assertions and critical in examining them, but tolerant in permitting linguistic forms»), although Reichenbach no doubt would have heartily endorsed this maxim. What I mean is that we should not try to fit beliefs and desires into the straitjackets of being *either* an ordinary abstractum consisting of concreta *or* an illatum inferred from concrete events. Beliefs and desires have a dual character, which Reichenbach tries to acknowledge by saying that they consist of concreta and illata alike. We have seen that Dennett too notices this dual character (Dennett 1987); however, rather than broad-mindedly accepting this fact, he tries to get rid of it by suggesting a divorce, culminating in two totally different theories, one about abstracta and one about illata.

Once the dual nature of beliefs and desires is taken seriously, some strong hunches are easily accounted for. For example, it is highly unlikely that all beliefs and desires are either abstracta or illata. It is much more plausible that they, like all dispositions that make up the higher psychic life, exhibit gradual differences. Thus some will be very close to pure abstracta, while others depend for the greater part on theoretical entities or illata. It seems only natural to regard for instance politeness and prosperity as abstracta or pure dispositions: it is unlikely that they will ever be more than abbreviations for a cluster of responses which appear under certain circumstances. Aggressivity and claustrophobia, on the other hand, presumably are illata. It is quite possible that future research will find that frequent aggressive behaviour corresponds to sensations caused by a chemical substance or a physical entity (the pugnacity lobule? the truculence particle?). By taking a tolerant stance and accepting that the nature of beliefs and desires is mixed, we can make these intuitions plausible. Hence we can avoid an all too monolithic approach to the higher psychic life, and learn to see reliefs in the mental map.

The old empiricists divided the mental into impressions and ideas. Modern empiricists have adopted this division by distinguishing between feelings or sensations on the one hand and pro-attitudes on the other. The criteria for the division correspond to two major themes in the contemporary philosophy of mind, viz. consciousness and content. According to almost everybody in this field, impressions or sensations are things of which we are conscious or aware; they are characterised by *qualia*. Ideas and pro-attitudes, on the other hand, are said to have content; they are characterised by intentionality or 'aboutness'. In general, consciousness is seen as the fundamental phenomenon, upon which intentionality ultimately depends. There are however dissident

philosophers, such as Dennett, who think that the order should be reversed. Be that as it may, the two great problems in the philosophy of mind are exactly about these two features: how to give an account of qualia and what exactly is intentionality? Reichenbach's position, old though it may be, might shed new light upon both of them. For as we have seen, Reichenbach regards sensations, impressions and thus qualia, too, as things that are probabilistically *inferred* rather than directly felt. Moreover, he conceives pro-attitudes as being composed of concreta and illata, thus making it more easy to understand that beliefs and desires have a mixed nature, and that some are 'more real' than others.[5]

Bibliography

Atkinson, D. and J. Peijnenburg (1999), 'Probability as a Theory Dependent Concept', *Synthese*, 118, 207-228.

Carnap, Rudolf, (1936-1937), 'Testability and Meaning', *Philosophy of Science 3(4)*, 1936, 420-471, and *4(1)*, 1937, 1-40.

——— (1950), 'Empiricism, Semantics, and Ontology', *Revue Internationale de Philosophie 4*, 20-40. Reprinted in: Carnap 1947, 205-221.

——— (1947), *Meaning and Necessity. A Study in Semantics and Modal Logic*. Chicago: University of Chicago Press, 1947. Enlarged edition 1956, second impression 1958.

——— (1956), 'The Methodological Character of Theoretical Concepts', in: H. Feigl and M. Scriven (eds), *The Foundations of Science and the Concepts of Psychology and Psychoanalysis*. Minneapolis: University of Minnesota Press, 38-76.

——— (1963), 'Intellectual Autobiography', in: Schilpp, P.A. (ed.), *The Philosophy of Rudolf Carnap*. La Salle, Ill.: Open Court; London/Cambridge: Cambridge University Press, 1963, 3-87.

——— (1966), *Scheinprobleme in der Philosophie: das Fremdpsychische und der Realismusstreit*. Frankfurt a. M.: Suhrkamp.

Dennett, Daniel (1987), *The Intentional Stance*, Cambridge (Mass.), MIT Press.

——— (1991a), *Consciousness Explained*, Boston: Little Brown. Published in 1993 by Penguin Books, Harmondsworth, Middlesex, England.

—— (1991b), 'Real Patterns', *Journal of Philosophy 89*, 27-51.

---

[5] Recently, Dennett has called himself a «mild realist» with respect to beliefs and desires (Dennett 1991b). He places mild realism somewhere between the «industrial-strength Realism» (with a capital 'R') of Fodor and the «milder-than-mild irrealism» of Rorty (Dennett 1991b, 30). Without going into the question of what mild realism actually is (it has to do with the reality of patterns), I only wish to stress here that Dennett applies it to all beliefs and desires alike. By contrast, I think it is more fruitful to claim that some beliefs or desires are real in the sense of Fodor's industrial-strength Realism whereas others are closer to Rorty's irrealism.

Peijnenburg, Jeanne, and Ronald Hünneman, 'Translations and Theories', to appear in *Ratio*, March 2001.

Reichenbach, Hans (1935), *Wahrscheinlichkeitslehre: eine Untersuchung über die logischen und mathematischen Grundlagen der Wahrscheinlichkeitsrechnung.* Leiden: Sijthoff.

––– (1938), *Experience and Prediction. An Analysis of the Foundations and the Structure of Knowledge*. Chicago: The University of Chicago Press, seventh impression, 1970.

––– (1951), 'Are There Atoms?', in: *The Rise of Scientic Philosophy*. Berkeley: University of California Press, Chapter 11.

Smullyan, Raymond (1977), 'Is God a Taoist?', in: *The Tao is Silent*, Harper & Row. Reprinted in: Douglas R. Hofstadter and Daniel C. Dennett (composers and arrangers), *The Mind's I*, New York: Basic Books, 1981, 321-341.

White, S. (1991), *The Unity of the Self.* Cambridge, MA: MIT Press.

Jeanne Peijnenburg

University of Groningen

A-weg 30

9718 CW Groningen. The Netherlands

<jeanne@philos.rug.nl>

SORITES (ΣΩΡΙΤΗΣ), ISSN 1135-1349
Issue #11. December 1999. Pp. 82-85.
Benardete's Paradox

BENARDETE'S PARADOX

by Michael B. Burke

We are indebted to Graham Priest[1] for focusing attention on an intriguing but neglected paradox posed by José Benardete in 1964.[2] Benardete, who evidently was the first to notice this Zenoesque paradox, presented it as a threat to the intelligibility of the spatial and temporal continua. Priest views it, perhaps less plausibly, as a paradox of motion.[3]

Benardete gave, rather informally, several versions of the paradox. Priest has selected one of those versions and, with five postulates, formalized it. Although Priest has succeeded nicely in sharpening the paradox, the version he chose to formalize has distracting and potentially problematic features that are absent from some of Benardete's other versions. In particular, the selected version involves an infinitude of gods, intentions, and distinct acts performed in a finite time.[4] Suspicion is sure to fall on Priest's fifth postulate, which is the one needed to accommodate those complicating but dispensable features.

I propose to offer a Priestly formalization of a *simpler* version of the paradox, the one that presents most plainly Benardete's challenge to the spatial continuum. Proposed resolutions of Benardete's paradox should address this version of the paradox as well as the one formalized by Priest.

---

1 «On a version of one of Zeno's paradoxes», *Analysis* 59 (1999), pp. 1-2.

2 *Infinity: An Essay in Metaphysics* (Oxford: Clarendon Press, 1964), pp. 236-39, 254-61, 271-79.

3 To generate the paradox we need to assume both the reality of motion and the continuity of space. If to solve the paradox we need to surrender one or the other of those assumptions, presumably we will prefer to surrender the latter.

4 Each of the «acts» consists in (a) determining whether a certain man does or doesn't reach a certain point in space and (b) depending upon that determination, either erecting or refraining from erecting a barrier at a certain other point in space. Priest suggests that the acts might all be performed by a single god rather than the infinitude of gods supposed by Benardete. But there are strong reasons for doubting the logical possibility of an agent's performing an infinitude of distinct acts in a finite time. (See B. Burke, 'The Impossibility of Superfeats', *The Southern Journal of Philosophy*, forthcoming'.) It is undesirable, and unnecessary, to make the new paradox dependent upon that possibility.

The version to be formalized[5] may be stated informally as follows: Point -1 is one meter west of point 0, which is one meter west of point 1. The ground between -1 and 1 is smooth and level. A ball at -1 is rolling eastward with sufficient momentum to reach 1 and beyond, if nothing (other than friction) impedes it. But rising from the ground between 0 and 1 (as they have from all eternity) are infinitely many barriers. Specifically, there are barriers at points ½, ¼, ⅛, and so on. (The barriers are equal in height and width, but they differ in thickness. The barrier at ½ is one centimeter thick. Each of the other barriers is half as thick as the first barrier to its east.) Each barrier is strong enough to stop the rolling ball. (This does not seem problematic logically. But if it were, we could replace the ball with a massless particle, such as a photon.) Now here is the problem: It seems obvious that the ball cannot progress beyond point 0, since to do so it would have to get past an infinitude of barriers, none of which it is able to get past. But since there is no *first* barrier, the ball does not *reach* any barrier (since it can't get past the preceding barriers) and thus is not *stopped* by any barrier. But there's nothing to stop the ball *other* than a barrier. And it may be assumed, in accordance with Newton's first law, that the ball will not stop unless something stops it. Thus we arrive at a contradiction — and a paradox.

In formalizing this version of the paradox, I will use as many of Priest's symbols and postulates as possible (so as to facilitate comparison of the two versions).

First, the symbolization key: *x* and *y* range over the set of spatial points belonging to the line segment containing -1 as its westernmost point and 1 as its easternmost point; *Bx* = there is (the western surface of) a barrier at *x*; *Rx* = the (foremost point of the) ball reaches *x*; Sx = the ball is stopped by the barrier at *x* (from ever going further than that barrier); $x<y$ = *x* is west of *y*.

Four postulates are needed, none of which is an analogue of Priest's fifth postulate. The second and fourth are the same as two of Priest's (except that for Priest, *Bx* = a barrier is created at *x* while the moving object is west of *x)*. Like Priest, I have suppressed universal quantifiers.

(1) $Bx \leftrightarrow x \in (\ldots ⅛, ¼, ½)$ (There are barriers at all and only these points: … ⅛, ¼, ½.)

(2) $(Rx \ \& \ y<x) \rightarrow Ry$ (The ball reaches a point only if it reaches every point to its west.)

(3) $Sx \leftrightarrow (Bx \ \& \ Rx)$ (The ball is stopped by a barrier iff the ball reaches the barrier.)

(4) $\neg\exists x(x<y \ \& \ Sx) \rightarrow Ry$ (The ball reaches a point unless stopped by a barrier to its west.)

Let *p* be any point east of 0. Given 1, it follows that there is a barrier west of *p*. But then, given 2, the ball will reach *p* only if it reaches that barrier. Since, given 3, the ball will be stopped by that barrier if it *does* reach it, the ball will not reach *p*. But now consider any barrier *b* west of *p*. Given 1, it follows that there is a barrier *b*»

---

[5] Benardete, op. cit., pp. 237-38.

west of *b*. But then, given 2, the ball will reach *b* only if it reaches *b*». Since, given 3, the ball will be stopped by *b*» if it *does* reach it, the ball will not reach *b*. Therefore, given 3, the ball is not *stopped* by *b*. It follows, by universal generalization on *b*, that the ball is not stopped by *any* barrier west of *p*. So, given 4, the ball *does* reach *p*. And we have reached a contradiction.

Which postulate might we reject? Neither (2) nor (3) seems a promising target. In the unlikely event that we should feel driven to deny the possibility of motion (as per Priest's suggestion), we would reject (4). (But neither (2) nor (3). If motion were impossible, the left side of (2), and both sides of (3), would be necessarily false [on every valuation of «*x*» and «*y*»]. In standard logic, that would assure the necessary *truth* of (2) and (3).) Benardete[6] suggested an *alternative* basis for denying (4): that the ball might be stopped, not by any one barrier, but by the infinite *sequence* of barriers. The ball stops at point 0, despite having encountered no barriers, because it would otherwise have to overcome an infinitude of barriers, none of which it is able to overcome. But as Benardete soon acknowledged,[7] his suggestion doesn't suffice to resolve the paradox. It merely reiterates the *proof* that the ball will stop; it does not provide a dynamical *explanation* of its stopping. At least until further ideas are forthcoming, suspicion will fall on (1). And Benardete's paradox will stand as a substantial challenge to a *presupposition* of (1): the continuity of the spatial continuum.[8]

Michael B. Burke

Indiana University School of Liberal Arts. Department of Philosophy

425 University Boulevard

Indianapolis. IN 46202. USA

<mburke@iupui.edu>

---

6 Ibid., p. 258.

7 Ibid., p. 261.

8 There is no apparent basis for objecting to (1) solely because of the infinity of *barriers* it postulates. If there is an objection to «actual infinities», the objection would apply not just to the barriers but to the actual infinity of points and line segments contained within any continuous spatial interval. In general, it's hard to imagine why the infinity of barriers might be logically objectionable if the requisite infinity of spaces is available to accommodate them. As Benardete notes on p. 255, the barriers might have sprung into being spontaneously and simultaneously, or might have been created one per year over the course of an infinite past, or might simply have existed from all eternity.

SORITES

An Electronic Quarterly of Analytical Philosophy

ISSN 1135-1349

COPYRIGHT NOTICE AND LEGAL DISCLAIMER

Please, read!

---

[1] The reader may find an excellent discussion of copyright-related issues in a FAQ paper (available for anonymous FTP from rtfm.mit.edu [18.70.0.209] /pub/usenet/news.answers/law/Copyright-FAQ). The paper is entitled «Frequently Asked Questions about Copyright (V. 1.1.3)», 1994, by Terry Carroll. We have borrowed a number of considerations from that helpful document.

(7) This issue of ***SORITES*** may be sold for profit or incorporated into any commercial material only with the previous explicit consent granted by the SORITES team. Otherwise, no fee may be charged for its circulation. An exception is granted to non-profit organizations, which are hereby authorized to charge a small fee for materials, handling, postage, and general overhead.

(8) Private copying of single papers by any lawful means is allowed only when done in good faith and for a fair use, namely for purposes of teaching, study, criticism or review; but no part of this issue of ***SORITES*** may be conveyed to another individual or to a gathering — whether in writing or through oral teaching or by any other means — unless the source is clearly and explicitly acknowledged.

(9) In particular, no part of this issue of ***SORITES*** or of any paper therein included may be conveyed to others by means of reproduction, quotation, copy or paraphrase, without a clear and explicit acknowledgement of the issue of ***SORITES*** and its date, the author's name and the paper's full title. Whenever the quotation occurs within a publication, it is also mandatory to mention the official pages (as shown within the Copyright box on top of the paper), the ISSN (1135-1349) and the official home site of electronic display, namely **http://www.ifs.csic.es/sorites/**.

(10) Any perpetration of, or complicity with, unfair use of copies or partial copies of this issue of ***SORITES***, or of papers therein included, especially forgery or plagiarism — being, as it is, an infringement of the authors' and the electronic publisher's rights — is in any case a civil tort, but may even be a crime under current legislation.

(11) This issue of ***SORITES*** is provided «as is», without any guarantee of any kind. The electronic publisher, the **SORITES** team, disclaims all warranties, whether expressed or implied, including, without limitation, the implied warranties of fitness for any particular purpose with respect to the papers included in this issue. By furnishing this document, the **SORITES** team does not grant any license or endorses any commitment except in so much as explicitly set forth in the present Copyright Notice.

(12) The electronic publisher, the **SORITES** team, does not necessarily agree with the authors' views or arguments. The electronic publisher cannot certify the accuracy of any quotations or references contained in the papers.

(13) Each author vouches alone for the originality of the papers they submit to ***SORITES*** and for their compliance with established Copyright laws. Acceptance of a manuscript is done in good faith under the assumption the originality claim is truthful. The electronic publisher — i.e. the **SORITES** team — does not pledge itself for the accuracy of such declarations.

(14) The **SORITES** team cannot be responsible for any real or imaginary damages suffered as a result of downloading, reading, using, quoting or circulating any materials included in this issue of ***SORITES***. The user assumes, at their own risk, full responsibility for the proper use of this issue of ***SORITES***.

(15) Downloading, reading or in any other way using this issue of ***SORITES*** or any part thereof entails full acceptance of the hereinabove stated terms and conditions. If, after downloading a file containing this issue of ***SORITES*** or a part thereof, a user fails to agree to the conditions and terms contained in this Notice, they must discontinue using the material and irrecoverably erase or destroy the downloaded file, so as not to occasion any third-part's unfair use thereof.

(16) Although, thanks to a permission kindly granted by the system's administrators, this electronic journal is displayed at the internet host <www.ifs.csic.es/sorites/> (hostname whose current official IP is 161.111.210.19), which belongs to the Spanish institution

CSIC, the journal is not published or sponsored or endorsed by the CSIC, the only owner and publisher being the **SORITES** team.

(17) A specific licence is hereby granted for this issue of ***SORITES*** — and all future issues of the journal as well — to be freely displayed by any BBS and any Internet node or site, provided all conditions stated above are fully aheared to. No previous consent of the **SORITES** team is required for such a display, which may be in the form of FTP, Gopher, http-WWW or any other electronic means.

Madrid. April 10, 1995

(Updated: December 31, 1996; April 27 1998)

The **SORITES** Team

RELEASE NOTICE

This issue of ***SORITES*** is made available in several formats, but its only official version is that released with filename:

sorite11.wp

which is the only file within the archives: **sori11wp.zip**, **sori11wp.arj**, **sori11wp.gz**, etc. A print-file (**sorite11.ps**), also released, and generated from the file **sorite11.wp** can be found in the archive **sori11ps.zip** and printed with a PostScript printer.

Two whole «doc» versions of this issue of ***SORITES*** are provided, but they cannot truly or faithfully reflect the official WordPerfect 5.1 version, departing as they do from the authorized WP 5.1 document — in a more or less severe way, depending on the particular case. One of them, **sorite11.wp.html** (or its equivalent, **sorite11.htm**), is an hypertext HTML version chiefly destined to be desplayed at the InterNet Web and viewed with http browsers. The other, **sorite11.txt**, is an empoverished version, with only ASCII symbols <Alt-32> through <Alt-126> being used, and a CR at the end of each line. Those two versions are archived, respectively, as **sor11htm.zip** and **sor11txt.zip**.

Several of those files are made available in MIME (Base64), UUenCODed (.UUE) and BinHexed (.HQX) translations, in order for them to be attached to e-mail messages.

Although each version, whether official or not — as initially released today (27 December 1999) by the **SORITES** team — is an entire seamless file, it may be splitted down into chunks in order to facilitate downloading, browsing, transferring or e-mailing. In such cases, the unity of this issue of **SORITES** as a whole must be preserved by keeping the ensuing collection intact.