From: K. von Heusinger & U. Egli (eds.): *Reference and Anaphoric Relations*.
Dordrecht: Kluwer (Studies in Linguistics and Philosophy), 2000.

JAROSLAV PEREGRIN

REFERENCE AND INFERENCE: THE CASE OF ANAPHORA*

This paper discusses the relationship between the concept of reference and that of inference; the point is to indicate that contrary to the usual view it may be good to see the former as "parasitic" on the latter, not the other way around. The paper is divided into two parts.

In part one, I give an (unsystematic) overview of the development of logical tools which have been employed in the course of the analysis of referring expressions, i.e. definite and (specific) indefinite singular terms, of natural language. I present Russell's celebrated theory of definite descriptions which I see as an attempt to explain definite reference in terms of unique existence (and reference in general in terms of existence *simpliciter*); and I present Hilbert's ε-calculus as an attempt to explain existence in terms of choice. Then I turn to contemporary, dynamic approaches to the analysis of singular terms and point out that only within a dynamic framework can the Russellian and Hilbertian ideas yield a truly satisfactory analysis of singular terms, and consequently of reference and coreference. I call attention to the fact that current results of formal semantics demonstrate the advantages of viewing singular terms as denoting *updates*, i.e. as a means of changing the context (information state), and especially that part of the context which I call the *individuary*.

In part two I turn to the discussion of the nature of such explications; especially to the question whether it forces the acceptance of a representational view of language. I answer the question negatively; I deny that we should see discourse representations, information states, or individuaries, which play the central roles within contemporary semantic theories, as descriptions of a mental reality; I try to show that these entities can, and indeed should, be seen as tools internal to our accounts for singular terms' inferential capacities. Therefore I conclude that we should not take the obscure concept of reference at face value, but rather as parasitic upon the clear concept of inference.

1 THE LOGICAL GRIP ON REFERENCE

1.1 Russell

Some expressions of our language are seen as doing their linguistic jobs by referring to definite things of our world. How do they manage to do this?

The classical analysis of definite descriptions (and of the English definite article, which is their linguistic hallmark), as presented by Russell (1905), consists in explicating definiteness in terms of unique existence. To say that the king of France is bald

* I am grateful to Klaus von Heusinger and Vladimír Svoboda for valuable comments and criticism. This work was supported by the Grant Agency of the Czech Republic, grant No. 401/99/0619.

is to say that there is one and only one entity which is the king of France, and that this entity is bald. Thus, the sentence

(1) The king is bald.

should, according to Russell, be construed as

(1') $\exists x\ (\mathbf{K}(x)\ \&\ \mathbf{B}(x)\ \&\ \forall y\ (\mathbf{K}(y)\ \rightarrow\ (y = x)))$

Thus, Russell's claim is that definite singular terms which are not proper names ('the king') are not in fact referring expressions at all but that they, when properly logically analyzed, give rise to a certain quantificational structure: a sentence consisting of a definite description and a predicate, according to him, says that there is one and only one object satisfying the description and is such that it has the property expressed by the predicate. This claim means reconstructing the definite article as a "syncategorematic term", i.e. as something which is not itself a fully-fledged constituent of the sentence. However, if we employ stronger formal means than those entertained by Russell (which were in fact those of the first-order predicate calculus), we can reconstruct 'the' "categorematically" (in the sense of granting it its own denotation) without violating the spirit of the Russellian analysis.

Thus, helping ourselves to the machinery of lambda-abstraction, we can rewrite (1') as

$$\lambda f\ \exists x\ (\mathbf{K}(x)\ \&\ f(x)\ \&\ \forall y(\mathbf{K}(y) \rightarrow (y = x)))(\mathbf{B})$$

and this further as

$$\lambda g\ \lambda f\ \exists x\ (g(x)\ \&\ f(x)\ \&\ \forall y(g(y) \rightarrow (y = x)))(\mathbf{K})(\mathbf{B})$$

This yields a formula consisting of three parts which may be put into natural correspondence with the three components of the analyzed sentence; the definite article thus gets formalized as $\lambda g \lambda f \exists x(g(x)\ \&\ f(x)\ \&\ \forall y(g(y) \rightarrow (y = x)))$, i.e. as a function which takes sets into sets of sets; or, probably more perspicuously, as a relation between sets. The relation holds between two sets iff the first is a singleton and has a non-empty intersection with the second. This is the analysis which has become standard after Montague (1974); and which has given rise to the so called *theory of generalized quantifiers* (*viz* Barwise & Cooper 1981).

An alternative elaboration of (1') can issue from the following consideration. First, assume that $\exists x(\mathbf{K}(x)\ \&\ \forall y(\mathbf{K}(y) \rightarrow (y = x)))$ is true, i.e. that the extension of **K** is a singleton. Under such an assumption, if we denote the single element of the extension of **K** as **c**, the whole formula becomes equivalent to **B**(**c**). Next, assume that $\exists x(\mathbf{K}(x)\ \&\ \forall y(\mathbf{K}(y) \rightarrow (y = x)))$ is *not* true, i.e. that the extension of **K** is either empty or contains more than one element. Then the whole formula (1') is patently false, i.e. equivalent to $\perp$. This means that if we were able to define a (second-order) function **F** which maps singletons on their single elements, and all other sets on something of which **B** is inevitably false, we could rewrite (1') as **B**(**F**(**K**)); or, writing $\iota x\ \mathbf{K}(x)$ instead of $\mathbf{F}(\mathbf{K})$, as $\mathbf{B}(\iota x\ \mathbf{K}(x))$.

The most straightforward way to devise such a function would be to stipulate an object of which everything is necessarily false and to let this object be the value of the function for all non-singletons; or, which is the same, to let the function be defined only for singletons, and to stipulate that any predicate applied to a term which lacks denotation yields a false sentence. In the latter way, we would reach an analysis which would be clearly effectively equivalent to the previous one, which treated the definite article as a generalized quantifier. However, a modification suggests itself: we can also stipulate that a predicate applied to a denotationless term yields a sentence which is not false, but truthvalueless – in this way we can clearly accommodate the idea of an existential presupposition associated with a definite description (as urged by Strawson 1950).

Anyway, we have seen that Russell himself claimed that a sentence consisting of a definite description and a predicate says that there is one and only one object satisfying the description and such that it has the property expressed by the predicate (*viz* (1')). Similarly he claimed that a sentence consisting of an indefinite description ('a king') and a predicate says that there is an object satisfying the description and having the property expressed by the predicate; thus, the adequate analysis of (2) is, according to him, (2').

(2) A king is bald.

(2') $\exists x(\mathbf{K}(x) \ \& \ \mathbf{B}(x))$

In this way, the functioning of a definite singular term gets reduced to the unique existence of the corresponding referent; and that of an indefinite one to existence *simpliciter*. We have seen that by allowing ourselves more powerful logical means than Russell himself, we can extrapolate Russell's analysis to explicate definite singular terms via taking the denotation of the definite article to be a function mapping singletons onto their unique elements (and nonsingletons onto some kind of "nothing").[1] However, no analogous straightforward explication is available for the meaning of the indefinite article, and hence for reference in general.

1.2 Hilbert

Russell's treatment of definite descriptions illustrates the intuitive intimate tie between definiteness and choosability: definiteness, i.e. unique existence, turns out to be a matter of unique choosability. This seems to invite generalization: why not see existence in general as choosability in general? The idea is that something exists if and only if it can be chosen ("picked out"); not necessarily by a particular human subject (whose capacities to actually carry out the choice could be limited), but "in principle", or "by God". That there is an F means that an F can be chosen. If we render the possibility of choosing as the existence of the corresponding choice function, we can say that the existence of an item is tantamount to the existence of the appropriate choice function: to say that there is an F is to say that there is a choice function which chooses an F.

Two kinds of objections can be raised against the identification of the existence of an object with the existence of a choice function choosing the object. First, there is the

"constructivistic" objection claiming that to be choosable is always more than merely to exist – that claims about the former necessarily violate bivalence, while those of the latter do (*ex definitio*) not. Then there are the scruples of set-theoreticians cash out the intuition directly by embracing the axiom of choice. The first objection can be turned into a purely terminological matter: to *say* that something exists is to *acknowledge* that it exists and in this sense to *choose* the thing from among other things. Thus, in this sense, if it makes sense to speak about *existence* of a thing, it makes the same sense to speak about that thing being *chosen* – although we have to keep in mind the broad ("bivalent") sense in which the term *choice* is being used. The second objection invokes the well-known set-theoretical perplexities swarming around the axiom of choice, which we are not going to discuss in this paper.[2]

One of the possible ways of developing this idea is to stipulate the reduction of the axioms of existence to the axiom of choice: this development was carried out by Hilbert. To see the semantic point of the enterprise, let us consider functions mapping nonempty sets onto their elements: a function f is called a *choice function on the set U* iff the domain of f is included in the power set of U and $f(s) \in s$ whenever s is not empty. If M is a model, then the set of all total choice functions on the universe of M will be denoted as $\mathrm{CHF_M}$. It follows that if $c \in \mathrm{CHF_M}$ and if s is a subset of the universe, then c(s) $\in$ s iff s is nonempty, i.e. iff there is an element of s. If s is the denotation of a unary predicate **F** ($[\![\mathbf{F}]\!]^M$ = s), then c(s) $\in$ s iff there is an **F**, i.e. iff $\exists x\, \mathbf{F}x$:

$$[\![\exists x\, \mathbf{F}x]\!] = 1 \text{ iff } c([\![\mathbf{F}]\!]) \in [\![\mathbf{F}]\!] \text{ for an arbitrary}^3\ c \in \mathrm{CHF_M}$$

If we now understand ε as denoting an arbitrary function from $\mathrm{CHF_M}$ in such a way that $\varepsilon x\, \mathbf{F}x$ denotes the value of ε for the set, and if t is an arbitrary term, then the following formulas will clearly be valid:

$$\mathbf{F}t \rightarrow \mathbf{F}(\varepsilon x\, \mathbf{F}x)$$
$$\exists x\, \mathbf{F}x \leftrightarrow \mathbf{F}(\varepsilon x\, \mathbf{F}x)$$
$$\forall x\, \mathbf{F}x \leftrightarrow \mathbf{F}(\varepsilon x\, \neg\mathbf{F}x)$$

Hilbert (1925) showed that if we accept the first of these as an axiom characterizing the ε-operator, we will be able to prove the other two and hence justifiably reduce quantification (and especially existence) to choice.[4]

In this way the ε-operator might seem to be capable of explicating existence *simpliciter*, and consequently the semantics of the indefinite article, in a way parallel to that in which the ι-operator explicates unique existence and hence the definite article. In one sense, this is indeed the case;[5] however, we should notice an essential difference between the two operators: while ι can be seen as a logical constant denoting a definite function (namely the function which maps singletons on their unique elements and is undefined for non-singletons) and thus furnishing a definite model-theoretical explication of the meaning of the definite article, ε is in fact more like an extralogical constant, ranging over the whole set of choice functions and thus *not* furnishing any definite explication of the meaning of the indefinite article. Moreover, one of the unmistakable features of the articles is their interplay: if *the F* follows *an F*, then the two

noun phrases are in the typical case coreferential; and this is something which is not directly reflected by the Russellian and Hilbertian treatments.

Therefore the direct exploitability of the ideas of Russell and Hilbert seems to be limited; and my opinion is that to progress we must "go dynamic" – we cannot have an explication of the meaning of the indefinite article (nor indeed a satisfactory explication of that of the definite one) until we start seeing meanings as "context-change potentials".

1.3 Dynamic semantics

To do this, let us first return to the Russellian analysis: it is clear that as it stands, it is *not* adequate to explicate our everyday use of the definite article. Saying *the F* does not usually involve claiming that there is one and only one F, but rather that there is one and only one "salient" F. With respect to this point two principal candidate ways to amend the analysis seem to emerge. The first possibility is to retain the Russellian analysis as such and to retract the assumption that all evaluations take place with respect to the general, all-embracing universe. This is to assume that the evaluation of a particular sentence may be based on a local, restricted universe, which is the result of the ("pragmatic") circumstances, in particular of the preceding discourse. Thus, we keep assuming that the sentence 'The king is bald' implies that there is one and only one king – not, however, in the general universe, but rather in a local universe determined by the context in which the sentence is being uttered. Hence it implies not that one and only one king exists, but rather that one and only one king is *salient*. The second possibility then is to assume that the Russellian analysis itself must be amended, that the choice function represented by the definite article is defined not only for singletons, but instead that it is capable of using strategies to successfully pick out an element even from some other nonempty sets.

In both cases the analysis rests on one or another formalization of the notion of context. In a context, some objects of the universe are salient, while others are not. The first approximation of the formalization of the concept of context could be thought of as simply a set of objects, a subset of the universe. This kind of context is utilizable by both of the above mentioned strategies: in the first case, we apply the Russellian analysis not to the universe, but rather to the context-delimited subset; in the second case we let the Russellian operator choose not *the only element* of the extension in question, but *the only element which is in the context-delimited subset*.

Taking the concept of context seriously entails subscribing to some version of the dynamic view of semantics, as proposed by a number of semanticists.[6] It leads to reconstructing meaning as resting on (or at least involving) a kind of "context-change potential",[7] of mapping of contexts on contexts. Contexts may be, and indeed are, captured in a variety of ways – for the problems discussed here it is nevertheless vital that a context somehow "contains" a class of (salient) individuals. In this way we can reconstruct definiteness not as presupposing unique existence, but rather as presupposing unique "referential availability".

To sensitize the denotation of a sentence to context, the denotation must become a "context-consumer", it must become functionally dependent on the context (formal-

ized in one or another way). For the semantics to become really dynamic, we must turn denotations not only into "context-consumers", but also into "context-producers" – so that an utterance might consume a context produced by a preceding one. From the point of view of anaphora it is again vital that the context somehow contains individuals – so that one utterance can introduce (raise to salience) an individual, to which other utterances may then refer back. Our current interests lie only in that part of the context which functions as such an individuals-container (which other parts a context can or should have is not a theme for us now) – to have a name for it, let us introduce the term *individuary* (devised in analogy to the well- established term *bestiary*).

In the simplest case, an utterance introduces a single object by using an indefinite description, and a subsequent utterance uses a definite description, or a pronoun, to pick it up. This is the case of

(3) A man walks. He (the man) whistles.

For this example, it suffices to construe the individuary as a single slot which can be occupied by an object, or be empty. The phrase 'a man' fills the slot with a ("fixed but arbitrary") individual, and the phrase 'he' is then interpreted as referring to this very individual. (There are essentially two ways of formalizing this "fixed but arbitrary": we can either represent the assignment of such an "arbitrary" object by means of a whole class of assignments of "real" objects, or we can introduce some kind of genuine "arbitrary objects", pegs, which are capable of being identified with real objects. The former is the way of Groenendijk & Stokhof 1991, the latter is that of Kamp 1981, and also of Groenendijk & Stokhof & Veltman 1996).

The situation is more complicated once there is more than one object which is passed from one utterance to another. Take

(4) A man meets a woman. He greets her.

In this case, when it comes to the anaphoric reference, it is necessary to choose the right referent from among more salient objects. This is not to say that there is *always* the right choice, but in many cases, such as in this one, there clearly is. This possibility of the right choice implies that the items in the individuary, the salient individuals, have to be in some way characterized, for it is only on the basis of some characteristic specification that we can distinguish between them and carry out the choice.

1.4 The structure of the individuary

Usually, approaches to dynamic semantics take the resolution of anaphora to be a matter of "coindexing"; they in fact assume that the "real" semantic analysis begins only *after* coreference has been settled. Thus, they assume that getting the right semantic analysis of (4) is a matter of being given (4a), and not, say, (4b), as the input of the semantic analysis.

(4) a. A man$_1$ meets a woman$_2$. He$_1$ greets her$_2$.
b. A man$_1$ meets a woman$_2$. He$_2$ greets her$_1$.

Under such an approach, the salient item stored into the individuary is characterized in such a way that it is associated with an index, or with some other formal item working to the same effect (Groenendijk & Stokhof's *discourse marker*) – the individuary can thus be constructed as a unique assignment of objects to indices.

However, in this way semanticians get rid of a part of the burden which is evidently *their own*. It seems obvious that the semantic analysis resulting from (4a) is the *right* one, while that resulting from (4b) is a *wrong* one – and this comes from understanding (4) – hence, it is the matter of the *semantics* of (4) and it has to be brought out by the semantic analysis of (4). Thus, the semanticians should not wait for someone to give them a coindexing, they should aim at yielding the right analysis directly. And this requires a more *substantial* characterization of the items in the individuary.

The idea that comes to mind is to store the individual with the "attribute" which is employed to introduce it: to store the item that is raised to salience by means of the phrase 'a man' with the attribute *man*, and, more generally, that raised to salience by means of the phrase $a(n)N$ with the attribute N. This would turn the individuary into an assembly of the $\langle$individual, attribute$\rangle$ pairs – a phrase *the N* would then look for the individual which is paired with the attribute N. Thus, the first sentence of (4) would fill the individuary with the pairs $\langle \mathrm{I}_1, \textit{man}\rangle$ and $\langle \mathrm{I}_2, \textit{woman}\rangle$; and for the resolution to succeed it would be enough to secure that the pronoun 'he' seeks an individual with the attribute *man* (i.e. it is in effect equivalent to 'the man') and the pronoun 'she' seeks one with the attribute *woman* (it is equivalent to 'the woman').[8]

The question now is how to construe the word "attribute". We may take attributes simply as *words* – and thus form the individuary as a collection of individual-term pairs. However, this would block the resolution in intuitively clear cases, where one uses a slightly different word, like

A man walks. The guy whistles.

This may lead us to abandon taking attributes as terms in favor of taking them more as *meanings* of terms – where meanings can be, in turn, taken to be extensions, intensions, or something else. (Then, however, we may have to face the opposite problem: the problem of "overgeneration" of anaphora resolutions. In general, the optimality of the account for anaphora is a matter of fine-tuning the fine-grainedness of the attributes – and due to the heterogeneity of language there is little hope that we can find one universally optimal solution.[9])

Anyway, within the dynamic approach, both the meaning of an indefinite noun phrase and that of a definite one get explicated as *updates*, as means of innovating the current context and especially its individuary. An indefinite noun phrase changes the individuary by introducing a new inhabitant characterized in a certain way; a definite noun phrase does not change the individuary (in this sense it is a *trivial* update, a "test"), but searches it for the existence of an individual with a certain specification, thereby triggering a presupposition that such an individual is indeed present there, i.e. that it is "referentially available".[10] This opens the possibility of adequately explicating the meanings of the indefinite and definite articles: both get explicated as functions which map properties onto corresponding updates. Thus, the denotation of

'a' is the function which maps the denotation $[\![N]\!]$ of the common noun N onto the update which stores $\langle \mathrm{I}, [\![N]\!]\rangle$ into the current individuary and makes the whole singular term refer to the I thus introduced; and the denotation of 'the' is the function which maps $[\![N]\!]$ on the update which searches the current individuary for an $\langle \mathrm{I}, [\![N]\!]\rangle$ and, if successful, makes the whole singular term refer to the I thus found.

One possibility is to take the attributes as sets of potential referents (this is straightforward if we stay on the level of extensions; but attributes can be constructed as sets of individuals even when we embrace intensions – in this case they are sets of not only actual, but rather also possible, individuals[11]). In that case, individuary is explicated as a choice function taking sets of individuals into their members – this is the framework introduced by Peregrin & von Heusinger (1995). In this setting, *choosing the right referent* is reconstructed as bringing the choice function to bear on the relevant subpart of the class of potential referents.

2 REFERENCE AS RESTING ON INFERENCE

2.1 The nature of the individuary

In this second part of the essay we now descend to a more foundational level and turn to a different question concerning the dynamic semantic framework: what are the individuaries, and more generally the "discourse representation structures" of all sorts, supposed to be? Few people practicing dynamic semantics seem bothered by this problem: they apparently assume it to be straightforwardly answerable in terms of "mental representations" or "cognitive states". I do not think they are right: I am convinced that to explain the linguistic by means of the mental is to explain the clearer by the more obscure; and, moreover, it is to block the requisite possibility of going the other way around, namely to use the linguistic to account for the mental. Therefore, I want to propose an alternative answer: the answer that the individuary, and indeed the relation of reference connecting words with the inhabitants of the individuary, is our (i.e. of us, theoreticians) way of accounting for the inferential properties of anaphoric expressions.

This question is also essentially relevant for the proper understanding of the analysis of definiteness in terms of choice outlined above: if we see the apparatus of choice functions as descriptive of mind or cognition, we will be likely to see choice functions as reports of actual "mental actions" carried out by speakers and hearers and we will be inclined to pose questions such as how they carry out the relevant choices or why they choose as they do and not otherwise; whereas if we see it as our way of accounting for certain valid inferences, such questions do not really make sense. In the former case, ultimate answers will appear to be buried within people's heads and we will have to set upon the slippery path of introspection; whereas in the latter case we shall be able to rest on the relatively solid notion of inference as based in the norms of verbal behavior.

To illustrate what I mean, let us compare the "semantics of anaphora" with a more ordinary semantics of logical connectives, such as **&** (the regimentation of natural language 'and'). It is well known that **&** can be characterized either by saying that

it denotes a certain binary truth function (namely the familiar function assigning T to two T's and F to any other pair of truth values), or by stating that it is governed by certain inferential rules, namely that $X\mathbf{\&}Y$ can be inferred from X and Y, and that X, as well as Y can be inferred from $X\mathbf{\&}Y$. The inferential characterization amounts to the following "introduction" and "elimination" rules:[12]

$$(\&\mathrm{I})\quad \frac{X \quad Y}{X\&Y}$$

$$(\&\mathrm{E}_1)\quad \frac{X\&Y}{X}$$

$$(\&\mathrm{E}_2)\quad \frac{X\&Y}{Y}$$

It is clear that the denotational and the inferential ways are equivalent from the formal point of view; but it should be also clear that from the point of view of analyzing natural language the second is superior. The point is that if somebody asks why we should regiment the natural language 'and' as **&**, two answers, corresponding to the two characterizations of **&**, are possible: it is possible to say either that 'and' indeed denotes the truth function stood for by **&**, or that it is indeed governed by the inferential rules governing **&**. However, the first answer can hardly be made sense of otherwise than as resting upon the second: if someone goes to ask *how do we know* that 'and' denotes such function, we can hardly do anything else than refer him to the inferences which sentences with 'and' license. (We can hardly *show* him - inside the speakers' heads or wherever – 'and' to be associated with the function.) On the other hand, the second answer does make a direct sense: we indeed can find out what are the inferential patterns governing the proper use of 'and' – it is enough to study the (publicly accessible) ways English speakers use their language.

Thus, from the point of view of the analysis of natural language, propositional logic is best seen as the theory of the inferential behavior of our basic "logical" vocabulary: of those particles and connectives which get regimented as $\neg$, **&**, $\vee$, $\rightarrow$ etc. (Further logical calculi then can be seen as theories of the inferential behavior of more advanced "logical" words and aspects of our language; thus, e.g. modal logic as the theory of the inferential behavior of adverbs like 'necessarily' and 'possibly', yielding the operators $\Box$ and $\Diamond$). And what I want to claim here is that dynamic logic, and consequently dynamic semantics of the kind envisaged in the first part of the paper, should be seen as the theory of the inferential behavior of further items of our "logical" vocabulary, namely of pronouns and articles – rather than a theory of how certain words refer to things via our mental repositories. (In the case of pronouns this is about the first systematic attempt at such an account; in the case of articles, especially the definite one, it is an improvement on previous attempts, like the Russellian one.)

If this view is to be tenable, then we have to be able to specify the basic inferential patterns governing pronouns and articles. I propose that we think about those like the following ones (all of them are to be seen as bidirectional inferences, i.e. so that not only the consequent is inferable from the antecedent, but also *vice versa*; M stands for masculine terms, F for feminine ones, and P, Q, R, S for predicates):

$$\frac{\textit{M Ps. He Qs}\ \text{(John walks. He whistles)}}{\textit{M Ps and Qs}\ \text{(John walks and whistles)}}$$

$$\frac{\textit{Somebody Ps. He Qs}\ \text{(Somebody walks. He whistles)}}{\textit{Somebody Ps and Qs}\ \text{(Somebody walks and whistles)}}$$

$$\frac{\dfrac{\textit{M Ps and F Qs. He Rs and she Ts}}{\textit{M Ps and Rs. F Qs and Ts}}\ \text{(John walks and Mary sits. He whistles and she sings)}}{\text{(John walks and whistles. Mary sits and sings)}}$$

$$\frac{\textit{An R Ps. The R Qs}\ \text{(A man walks. The man whistles)}}{\textit{An R Ps and Qs}\ \text{(A man walks and whistles)}}$$

2.2 The emergence of a context-change potential

Let us now show how contemplating inferential patterns like the above may lead us directly to thinking about terms as denoting context-change potentials and thus to positing an individuary on which the potentials could rest. Let us restrict ourselves, for simplicity's sake, to the first of the above patterns (ignoring, moreover, the gender-dependence of the pronoun) and let us see how we could accommodate it within a logical calculus.[13] To do so, we need a constant corresponding to 'he'; let us use the sign $\leftarrow$. To say that the constant should correspond to 'he' as governed by the first inferential pattern above (stripped of the gender-dependence) is to say that $P(A)\ \&\ Q(\leftarrow)$ should be equivalent to $P(A)\ \&\ Q(A)$ for every singular term A and all predicates P and Q, i.e. to characterize $\leftarrow$by the following natural deduction rules:[14]

$$(\leftarrow\text{I})\quad \frac{P(A)\ \&\ Q(A)}{P(A)\ \&\ Q(\leftarrow)}$$

$$(\leftarrow\text{E})\quad \frac{P(A)\ \&\ Q(\leftarrow)}{P(A)\ \&\ Q(A)}$$

Now let us think about a semantics (or about a model theory) for the resulting calculus; for the sake of simplicity let us assume that the calculus contains, besides $\leftarrow$, nothing more than what is needed to spell out ($\leftarrow$I) and ($\leftarrow$E), i.e. individual constants, unary predicate constants and conjunction. We soon discover that we *cannot* make do with the simple semantics of elementary logic, namely with treating terms as names of objects of the universe. (It is not difficult to see that this would, given ($\leftarrow$I) and ($\leftarrow$E), imply that all terms denote a single object.) Similarly we could eliminate other candidate semantics; and after some trial-and-error searching we are likely to realize that since the inferential role of $\leftarrow$ is such that it "behaves like *a* if it follows a clause containing *a*, it behaves like *b* if it follows a clause containing *b* etc.", we could make do with letting $\leftarrow$ denote something like the identity function from individuals to individuals.

This idea can be put to work by making the semantic values of statements into functions from individuals to individuals, and letting **&** work as a concatenation in the following way.[15] First, we change the semantics of our fragment of the predicate calculus by letting statements denote mappings of a class C onto itself: we let true statements denote some function defined for every element of C, and false statements

denote the function which is defined for no element of C. (This is clearly an entirely trivial move: it is clear that we can have any two distinct objects playing the role of the two truth values. However, note that in this setting, **&** denotes *functional composition*.) Then, we let different true statements denote *different* functions defined everywhere on C, and treat *any* such function as the truth value *the truth*. We can, for example, identify C with the universe and define $[\![P(a)]\!]$ to be such function that for every $x \in$ C, if $[\![a]\!] \in [\![P]\!]$, then $[\![P(a)]\!](x) = [\![a]\!]$, and if $[\![a]\!] \notin [\![P]\!]$, then $[\![P(a)]\!](x)$ is undefined. $[\![S\&S']\!](x)$ can then be defined as $[\![S']\!]([\![S]\!](x))$ (where this is meant to be undefined where $[\![S]\!](x)$ is undefined). All of this is still trivial in the important sense of not tampering with the logical properties of the calculus. Now, however, we can easily provide an adequate semantics for statements containing $\leftarrow$: we can define $[\![P(\leftarrow)]\!]$ to be such function that $[\![P(\leftarrow)]\!](x) = x$ if $x \in [\![P]\!]$, and is undefined otherwise. And if we do this, the inferences ($\leftarrow$I) and ($\leftarrow$E) hold.

The final step is then to project the new semantics onto terms. We can define the semantics e.g. in the following way (we can take S to be true if $[\![S]\!](x)$ is defined for every element x of the universe; or else if it is defined for at least one element x of the universe – the difference affects only formulas with a "free occurrence" of $\leftarrow$, i.e. formulas in which $\leftarrow$does not follow an individual constant):

$[\![a]\!]$ is a constant function defined everywhere on U (where a is an individual constant)

$[\![\leftarrow]\!]$ is the identity function defined everywhere on U

$[\![P(t)]\!]$ is a function such that $[\![P(t)]\!](x) = [\![t]\!](x)$ if $[\![t]\!](x) \in [\![P]\!]$, and is undefined if $[\![t]\!](x) \notin [\![P]\!]$ (where t is an individual constant or $\leftarrow$)

$[\![S\&S']\!]$ is the composition of $[\![S']\!]$ and $[\![S]\!]$, i.e. $[\![S\&S']\!](x) = [\![S']\!]([\![S]\!](x))$

In this way we have turned denotations of sentences and terms of our fragment into *context-change potentials*; and we have created an *individuary* (a particularly tiny one, which can contain at most one individual). What is important is that this creation has resulted from our attempt to account semantically for certain inferences – not to depict some mental or real machinery. In other words, the semantics based on this individuary has been employed as a *tool* of our account; not as a *picture*.[16]

The introduction of the whole machinery of definite and indefinite singular terms, which leads to individuaries of more complex kinds, is now only a more complicated version of the same process. We have more inferences to account for: inferences like that of *an R Ps and Qs* ('A man walks and whistles') from *an R Ps and the R Qs* ('A man walks and the man whistles'); hence we need more "slots" to store individuals, and we need labels to tell different slots apart. However, the individuary is again no more than a creature of our theory of drawing inferences.

The crux of this kind of semantic treatment is that the link between a singular noun phrase and a following pronoun (or between an indefinite noun phrase and a following definite noun phrase) is established by linking both phrases to the same inhabitant of

the individuary. Now insofar as reference is taken as the link between a word and an inhabitant of an individuary, and coreference consequently as the property of being linked to the same item, saying that two noun phrases are *coreferential* thus becomes a short way of stating that their inferential roles are in a certain way interconnected (to say that within 'A man walks and he whistles', 'a man' is coreferential with 'he' is to say that the sentence entails, and is entailed by, 'A man walks and whistles'), and talking about reference in turn becomes only a particularly illustrative way of rendering coreference. If we recognize individuaries and their inhabitants as mere tools to account for inferences, then the talk about reference becomes essentially parasitic upon the talk about inference – a *referent* is nothing more than an illustrious clamp holding certain inferentially related expressions together.

2.3 Capturing inference as reference

We have given an example, admittedly oversimplified, of how talk about an individuary, and about expressions' referring to elements of the individuary, can be rendered as talk about inferential patterns. Now if we want to make a more general claim concerning the reducibility of "referential talk" to "inferential talk", such an example is surely not enough. We have to indicate that rendering talk about reference as talk about inferential patterns is possible in general, for all "referring expressions", and besides this, we have to indicate that we do not need the concept of reference to underpin language in the first place. The former task will be the topic for the present section, the latter will be left for the next one.

Let us first introduce some terminology in order to be able to talk about the inferential roles of expressions. In general, we can see two expressions as inferentially connected if, informally stated, the inferences licensed by one of them are licensed also by the other. There are two levels of inferential relationships between expressions, the first level concerning *(material) implication* (licensing inferences "here and now"), and the second concerning *entailment* (licensing inferences "everywhere and always"). We say that a statement S *implies* a statement S' if S' is not false unless S is, i.e. if $S \rightarrow S'$; and we say that S *entails* S' if S' *cannot* be false if S is not; i.e. if $\models S \rightarrow S'$. Let us then call an expression e *weakly (inferentially) subordinated* to an expression e' iff for every atomic statement S, S implies $S[e'/e]$ (where $S[e'/e]$ is the statement which arises from S by replacing e by e'); and let us call e *strongly (inferentially) subordinated* to e' iff for every atomic statement S, S entails $S[e'/e]$. Finally, let us call e and e' *weakly (inferentially) equivalent* iff e is weakly subordinated to e' and e' is at the same time weakly subordinated to e (i.e. iff $S \leftrightarrow S[e'/e]$ for every statement S); and analogously for *strong (inferential) equivalence*.

With the help of this terminology, we can characterize the inferential behavior of expressions which are traditionally considered as "referring", i.e. of names (in a broad sense). First, names are characterized by the fact that no name is subordinated to another name without being equivalent to it. This means that the inferential structure of the domain of names is more or less trivial (in contrast to that of predicates which constitutes a Boolean algebra).[17] As the consequence, we inferentially characterize

a name simply by saying with which other names it is weakly equivalent. If we use the term *coreferential* as a synonym of *weakly equivalent*, then we can say that to inferentially characterize a name is to state with which other names it is coreferential; and as the relation of coreferentiality is clearly an equivalence relation which thereby decomposes the class of names into the corresponding equivalence classes, this is to specify the *coreferentiality class* to which the name belongs. And if we further use the term *what it refers to* as a shortcut for *which coreferentiality class it belongs to*, we can say that by inferentially characterizing a name we pinpoint what it refers to – thus explicating "referent" as "that which is shared by all coreferential expressions".

We can also make the "inferential sense" of distinctions which are usually drawn from the referential perspective: such distinctions as those between proper names ("rigid designators"), descriptions ("contingent designators") and pronouns ("context-dependent designators"). First, we can single out the class of *proper names* as the maximal subclass of the class of names which has the property that any two of its members are weakly equivalent (coreferential) if and only if they are strongly equivalent. Coreferentiality is thus a *standing* property for proper names, and the inferential behavior of a proper name is thus *exhaustively* characterized by specifying its (standing) coreferentiality class, i.e. by its (standing) referent.

Then we can single out the subspecies of *descriptions* analogously by stating that two descriptions can be coreferential without being strongly equivalent, and no description can ever be strongly equivalent to a proper name (although it can be coreferential with it). The relation of coreferentiality among descriptions is a fluctuating, contingent matter. Thus, the exhaustive inferential characterization of a description cannot consist simply in pointing out its (momentary) coreferentiality class; we must somehow say to which coreferentiality class it belongs when. This could be done by specifying its coreferentiality class relative to the truth-valuation of sentences, i.e. to a possible world. Thus, a description can be inferentially characterized by being assigned a function from possible worlds to referents; as it is the case within Montagovian frameworks.

Finally we can characterize the third basic kind of names, *pronouns*, by stating that no pronoun is coreferential with a proper name nor with a description (which may suggest that pronouns are "in fact" not names and have no referents), that, however, pronouns are what can be called "strongly locally" equivalent with names – sentences containing pronouns are often strongly equivalent with sentences containing proper names in their place (*viz* ($\leftarrow$I) and ($\leftarrow$E)). One way to accommodate this is to make coreferentiality, and hence reference, somehow relative to the context – to associate pronouns with functions from contexts to referents. It then becomes vital to articulate the notion of context appropriately (as discussed in section (1.4) and to embody this into a tractable compositional semantics.

2.4 The essentiality of inference

Thus, we can, at least in principle, render the concept of reference, and referential links between expressions and occupants of "slots" of individuaries, as a means of

characterizing inferential behavior. However, it may be objected that reference is primarily a relation between expressions and real-world things which is only *mediated* by inhabitants of mental individuaries; and that as such it is something which underpins language in the first place. Is thus not explaining reference in terms of inference putting the cart before the horse? Is reference not the thing via which language "hooks on the world" and without which no language could exist; and is thus reference not the key to everything else in language, including inference?

Without being able to go into detail here, let me point out that the picture of language centered around the concept of reference (justified by the claim that reference is what gives language its hook on the world) can be counterposed to a picture centered around the concept of inference; justified by the fact that what distinguishes a language is its capability of serving as the medium of the human "game of giving and asking for reasons" (Brandom 1994). The latter perspective denies that language would be, as the referential perspective seems to have it, only a rich and complexly interrelated system of names; it sees the referential view as a misguided "museum myth" (Quine 1969). According to this inferential picture, drawn most vividly by Wilfrid Sellars and Robert Brandom (but in fact going back to Kant), the distinctive feature of language is that it is capable of conveying *propositional content* – and hence that the constitutive characteristics of words is not that they name (refer to, represent, stand for) things, but rather that they can add up to propositionally contentful utterances. If we subscribe to this story, then we have to conclude that it is the relation of *inference*, not that of *reference*, which is the backbone of language on which everything else should be seen as supervening – for it is only *inferential articulation* which can confer propositional contents on sentences. Even the relation of reference is then seen as parasitic on the relation of inference – which inevitably leads to a *deflationary* view of reference (in the sense of Horwich 1990, Sec. 39).[18]

The obvious objection is that this leads to an absurdly idealistic picture of language, to a picture in which language is completely cut loose from the world. But this is a misunderstanding. Of course if we want to see language generally as a matter of inferences, then we have to construe the term *inference* broadly enough to comprise what Sellars (1974) calls *language entry transitions* (roughly, inferences from situations to statements) and *language exit transitions* (inferences from statements to actions). (However, equally of course, only inferences in the narrow sense, inferences from statements to statements are capable of being the subject of semantic *theory*. The other ones do not yield a nontrivial *theory* over and above trivialities like *We (correctly) say 'A rabbit runs' iff a rabbit runs.*) So according to this view, language, of course, *is* connected to the world, but not via denotative word-thing links, but rather via normative statement-occasion links. According to this view, the unequivocal aim of semantics is to account for the relation of inferability among statements – and any semantic value which semantic theory associates with an expression should be seen not as a depiction of a real thing on which the expression is claimed to (causally or otherwise) hook, but rather as a kind of hypostasis of the way the expression functions within inferences.[19] And it is important to see that urging this view is not merely an

exercise in speculative philosophy, that it has most important consequences for the *practice* of semantic analysis:

1. It implies that the way of semantics is essentially parallel to the way of logic – for logic is precisely the account for inferability. The main difference is that semantics endeavors to see the inferential behavior of an expression materialized in its *meaning*, i.e. in an object associated with the expression. (In contrast to this, the view which makes reference basic is bound to see semantics allied to cognitive psychology investigating the ways in which we attach names to things.)
2. It suggests that semantic formalisms should not be seen as *describing* platonistic entities or mental representations "behind" expressions, but rather as *explicitly articulating* expressions' inferential properties. That is to say, a formula or a schema associated with a sentence should be seen as explicating the "inferential potential" of the sentence, i.e. as encapsulating what is implied by the sentence and what implies it; and the formula associated with a subsentential expression as that which adds up to the inferential potentials of the sentences in which it occurs.
3. It entails that the criteria of adequacy of a formal semantic theory are a matter of success in capturing inferences. Such success then can be checked by finding out about the publicly accessible rules of correct usage of the language in question. (In contrast, the mentalistic approach is bound to see such criteria as a matter of faithfully depicting the hopelessly private mind or cognition).

What, then, about the common temptation to see semantic representations, individuaries and the like as depicting some structures of the minds or brains of the speakers who draw the inferences? In a certain weak sense, this need not be incompatible with the vantage point advocated here: if we accept the inseparability of language and thought, we have to see *any* account of one's usage of language as *eo ipso* an account of her thinking. However, what I cannot find any substantiation of is seeing the referential apparatus as a *depiction* of structures and processes going on within speakers' minds/brains (and thus holding that to assess semantic theory we should observe what is going on within our heads).

Chomsky (1986, 45) writes: "One can speak of 'reference' and 'coreference' with some intelligibility if one postulates a domain of mental objects associated with formal entities of language by the relation with many of the properties of language, but all of this is internal to the theory of mental representations; it is a form of syntax." If we use the term *syntax* in such a way that we see inference as a syntactic matter (which is usual, although perhaps misguiding), then the standpoint advocated here can be seen as consisting precisely in taking the theory of reference as "a form of syntax". However, Chomsky's pronouncement is problematic, it seems to me, because he speaks about "*mental* objects" and "*mental* representations" – which I find simply unwarrantable. I suggest that we replace the term "mental" in claims of this kind simply with the neutral term "semantic": perhaps some of the people speaking about *the mental* in this context really do *not* mean anything over and above *the semantic*.

NOTES

[1] Russell himself introduced the ι-operator as a mere notational short-cut governed by a contextual definition, not as a fully-fledged term (see Russell & Whitehead 1913).

[2] However, notice that, as Lavine (1994, 104) points out, the axiom of choice is problematic only under a certain specific notion of set, namely under the "logical" notion, claiming that items can be collected into a set only if they can be delimited by a criterion.

[3] The definition of CHF_M clearly guaranties that this holds for at least one $c \in CHF_M$ if and only if it holds for every $c \in CHF_M$.

[4] Hilbert's treatment of the ε-operator was purely axiomatic; he did not consider any kind of model theory. Nevertheless, choice functions clearly represent the straightforward way to put his ideas into the semantic cash. (Notice, however, that Hilbert's axiom allows for "intensional" choice – the value of $\varepsilon x\, \mathbf{F}x$ may differ from that of $\varepsilon x\, \mathbf{F}'x$ even if **F** and **F**' are coextensional. See, e.g. Meyer Viol 1995).

[5] In fact, it seems to be this idea which lays the foundations of Egli's and von Heusinger's exploitation of the Hilbertian ideas for semantic analysis (see their contributions in Egli & von Heusinger 1995).

[6] The pioneers of the kind of dynamic semantics which is relevant here are especially Kamp (1981) and Heim (1982); but other kinds of dynamic semantic theories were proposed earlier (e.g. by Hintikka 1973).

[7] As far as I know, this term is due to Irene Heim.

[8] The most straightforward way of exploiting this idea is perhaps Heim's (1982) *File Change Semantics*.

[9] This is analogous to the case of objects of propositional attitudes: they also seem to be sometimes like propositions, whereas sometimes rather like sentences.

[10] An individual can, of course, enter the individuary also in a "non-linguistic" way, e.g. via ostension.

[11] It is not without interest to note that this is in fact the way in which the concept of intension was approached by Rudolf Carnap (see esp. Carnap 1955), who is mostly responsible for its current dissemination.

[12] For the general notion of introduction and elimination rules see, e.g. Prawitz (1965).

[13] For a more detailed elaboration see Peregrin (1998).

[14] From the purely logical viewpoint, these rules, and consequently ⇐, are clearly not of great interest, for what they introduce is in fact nothing but certain notational variants of certain conjunctive statements.

[15] From the formal point of view, the enterprise is, of course, only a minor variation on the basic theme of dynamization of semantics as presented by Groenendijk & Stokhof (1991).

[16] For the discussion of this dichotomy see Peregrin (1999).

[17] Making a similar point, Brandom (1994, 372) says: "Singular terms are grouped into equivalence classes by the good substitution inferences in which they are materially involved, while predicates are grouped into reflexive, transitive, *a*symmetric structures or families. That is to say that some predicates are simply inferentially weaker than others, in the sense that everything that follows from the applicability of the weaker one follows also from the applicability of the

stronger one, but not vice versa. (...) Singular terms, by contrast, are not materially involved in substitution inferences whose conclusions are inferentially weaker than their premises."

[18] In some of my recent writings (see esp. Peregrin 1995) I have urged a distinction between two views of language, which I have called the *nomenclatural* and the *structural* view, respectively (Brandom 1994 speaks about the *representational* and the *inferential* view to the same effect). The former view is based on a view of language as a nomenclature of some kind of things; the latter view sees language rather as a kind of toolbox. What I tried to indicate, and what I am trying to indicate also here, is that it is the latter which can help us gain real insight into the workings of language.

[19] As Brandom (1994, 84) puts it, "an association [of abstract objects with strings] amounts to specifically *semantic* interpretation just insofar as it serves to determine how those strings are correctly used". This also implies that within logic, proof theory is in an important sense primary to model theory – see Peregrin (1998).

REFERENCES

Barwise, J. & Cooper, R. 1981. Generalized Quantifiers and Natural Language. *Linguistics and Philosophy 4*, 159-219.

Brandom, R. 1994. *Making It Explicit.* Cambridge/MA: Harvard University Press.

Carnap, R. 1955. Meaning and Synonymy in Natural Languages. *Philosophical Studies 7*, 33-47.

Chomsky, N. 1986. *Knowledge of Language*. Westport: Praeger.

Egli, U. & von Heusinger, K. 1995. The Epsilon Operator and E-Type Pronouns. In: U. Egli et al. (eds.). *Lexical Knowledge in the Organization of Language*. Amsterdam: Benjamins, 121-141.

Groenendijk, J. & Stokhof, M. 1991. Dynamic Predicate Logic. *Linguistics and Philosophy 14*, 39-100.

Groenendijk, J. & Stokhof, M. & Veltman, F. 1996. Coreference and Modality. In: S. Lappin (ed.). *The Handbook of Contemporary Semantic Theory*. Oxford: Blackwell, 179-213.

Heim, I. 1982. *The Semantics of Definite and Indefinite Noun Phrases*. Ph.D. dissertation. University of Massachusetts, Amherst. Distributed by Ann Arbor: University Microfilms.

Hilbert, D. 1925. Über das Unendliche. *Mathematische Annalen 95*, 161-190. [Translated as "On the Infinite" in: J. van Heijenoort (ed.) 1971. *From Frege to Gödel: A Source Book from Mathematical Logic*. Cambridge/MA: Harvard University Press, 367-392.]

Hintikka, J. 1973. *Logic, Language-Games and Information.* Oxford: Clarendon Press.

Horwich, P. 1990. *Truth.* Oxford: Blackwell.

Kamp, H. 1981. A Theory of Truth and Semantic Representation. In: J. Groenendijk & T. Janssen & M. Stokhof (eds.). *Formal Methods in the Study of Language*. Amsterdam: Mathematical Centre, 277-322. [Reprinted in: J. Groenendijk & T. Janssen & M. Stokhof (eds.) 1984. *Truth, Interpretation and Information*. Dordrecht: Foris, 1-41.]

Lavine, S. 1994. *Understanding the Infinite*. Cambridge/MA: Harvard University Press.

Meyer Viol, W. 1995. *Instantial Logic*. Ph.D. dissertation. University of Amsterdam. Distributed by Institute for Logic, Language and Computation (ILLC), University of Amsterdam.

Montague, R. 1974. *Formal Philosophy. Selected Papers of Richard Montague*. Ed. and with an introd. by R. Thomason. New Haven: Yale University Press.

Peregrin, J. 1995. *Doing Worlds with Words*. Dordrecht: Kluwer.

Peregrin, J. 1998. *The Logic of Anaphora*. Ms. Prague University.

Peregrin, J. 1999. Linguistics and Philosophy. *Theoretical Linguistics 24*.

Peregrin, J. & von Heusinger, K. 1995. Dynamic Semantics with Choice Functions. In: U. Egli & K. von Heusinger (eds.). *Choice Functions in Natural Language Semantics*. Arbeitspapier 71. Fachgruppe Sprachwissenschaft, Universität Konstanz. [Also in: H. Kamp & B. Partee (eds.) 1997. *Proceedings of the Workshop "Context Dependence in the Analysis of Linguistic Meaning"*. Vol. I. Institut für maschinelle Sprachverarbeitung, Universität Stuttgart, 329-353.]

Prawitz, D. 1965. *Natural Deduction*. Stockholm: Almqvist & Wiksell.

Quine, W. 1969. *Ontological Relativity and Other Essays*. New York: Columbia University Press.

Russell, B. 1905. On Denoting. *Mind 14*, 479-493.

Russell, B. & Whitehead, A. 1913. *Principia Mathematica*. Vol. I. 2nd ed. Cambridge: Cambridge University Press.

Sellars, W. 1974. Meaning as Functional Classification. *Synthese 27*, 417-437.

Strawson, P. 1950. On Referring. *Mind 59*, 320-344.