Logic, Language, and Knowledge. Essays on Chateaubriand's ***Logical Forms***
Walter A. Carnielli and Jairo J. da Silva (eds.)

CDD: 149.94

EXPLAINING INTENTIONALITY

PAUL HORWICH

Department of Philosophy
New York University
5 Washington Place
NEW YORK, NY 10003
USA

ph42@nyu.edu

Abstract: The goal here is to demystify the relation of aboutness that associates thoughts and their linguistic expression with particular features of the world. It is argued that the main obstacle to providing a naturalistic account of this relation is a misguided ('inflationary') view of truth. A deflationary perspective, on the other hand, enables us to see how the basic use of a mental or physical term establishes its referent, thereby determining what the sentences containing it are about.

Keywords: Intentionality. Deflationism. Reference. Meaning. Truth. Kripke.

EXPLICANDO A INTENCIONALIDADE

Resumo: O propósito aqui é desmistificar a relação que associa pensamentos e suas expressões lingüísticas com aquelas características do mundo sobre as quais eles são. Argumenta-se que o maior obstáculo à formulação de um tratamento naturalista desta relação é uma visão equivocada ('inflacionária') de verdade. Por outro lado, uma perspectiva deflacionária permite-nos ver como o uso básico de um termo mental ou físico estabelece seu referente, determinando desta maneira sobre o que são as sentenças que o contém.

Palavras chave: Intencionalidade. Deflacionismo. Referência. Significado. Verdade. Kripke.

> Intentionality is a very interesting and fundamental feature of humans and other animals, and attempts to explain it, or define it, either syntactically or behaviorally, seem to me unconvincing. It plays a central role in Kripke's account of reference, which is a fundamental semantic notion, irreducible to syntactic notions. What is so special therefore, about using intentionality as a semantic primitive in developing an account of language and other human cognitive abilities? (Chateaubriand 2005, Chapter 14)

What follows is an attempt to resist the point of view expressed in this passage. I would like to dispel the idea that intentionality, i.e. 'aboutness', is so hard to explain that we must either accept it as an explanatorily fundamental primitive or else question its very existence. Granted, any theory of meaning and reference must deal with a daunting array of puzzles – for example, the relationship between what we mean by a word and how we ought to use it, the way that word-meanings combine to form sentence-meanings, and the nature of understanding, of knowledge of meanings. But these problems are by no means intractable – or so I argue elsewhere. The present essay focuses on what I take to be the most notorious of them:– How can a thought or a linguistic expression be about a specific aspect of reality?

1. THE CENTRAL ISSUE

If we make the naïve assumption that words have distinctive meanings – e.g. that Victor's word, "cão", has the property of meaning DOG – then we might reasonably be concerned with the question of how such phenomena can arise, how the existence of this sort of fact may be explained. Or, more specifically:

- to what, if anything, are meaning-properties, such as 'w means DOG', conceptually (a priori) analysible?

- to what, if anything, are they empirically (a posteriori) reducible?
- which causal processes are responsible for their exemplification?

One may raise such questions about any property – being red, being a dog, etc – and there are established methodologies for answering them. But it is often supposed that peculiar difficulties – special constraints – exist in case of meaning-properties.

The main problem is felt to be that of doing justice to their truth-theoretic (i.e. referential) import. For example, because of what it means, the English word "planet" is true of Mars and is true of Jupiter, but is not true of Aristotle or of the number 3. Thus, in virtue of a word's being given a certain meaning, it 'reaches out' into the world and 'grabs on to' a certain specific collection of things – perhaps infinitely many of them and perhaps things that are inaccessible to us. But how could this so-called 'intentionality' or 'aboutness' be engendered? What sort of mental or behavioral or social activity on our part could result in our 'investing a certain word, w, with a certain meaning', given that this would have to entail a staggering profusion of facts of the forms, 'w is true of x' and/or 'w is not true of x'?

The central issue is whether this is a genuine difficulty or not. Does the need to resolve it provide a legitimate and hard-to-satisfy adequacy condition on good answers to our initial questions about how meaning-facts are engendered? Is that adequacy condition impossible to satisfy (as Kripke argues in Wittgenstein on Rules and Private Language)?[1] *And must we conclude, with him, that there cannot be any*

[1] The following account of what we might call "Kripke's paradox" diverges substantially from his own presentation of it. For my aim here is not to give a faithful exposition of his discussion, but to articulate what I take to be the strongest formulation of the problem with which he is

genuine facts of meaning? If so then we have a paradox; for it seems obvious that words do have distinctive meanings.

2. ELABORATION OF THIS APPARENT DIFFICULTY

The explanatory route from our particular meaning-giving activity with, for example, the word "dog" – call it Act_{57}("dog") – to that word's being true of all and only the dogs, would presumably have to take the following form:

1) Act_{57}("dog")

2) Act_{57}("dog") → "dog" bears relation R to every dog and only to dogs

3) ∴ "dog" bears R to every dog and only to dogs

4) Word w is true of x ⇔ w bears R to x

5) ∴ "dog" is true of each dog and only of dogs

Such an explanatory deduction would show how a term comes to be true of dogs and only of dogs, and thereby how it might come to mean what it does. However, it is hard to find any relation, R, able to play the role prescribed. For remember that our basic concern is with what we do (computationally, or neurologically, or behaviorally, or socially, etc.) in order to put words into their referential relations with objects. Thus the sort of verbal 'activity' that we are assuming must entail the instances of R (see line 2) is non-semantic activity – and so R itself would have to be something

---

concerned. In particular, I will not focus, as he does, on issues deriving from the normative import of meaning (– though I will briefly indicate, in footnote 7, how I think they can be dealt with).

that can be articulated in non-semantic terms. But what relation could that be? What non-semantic relation links "dog" to the dogs, "neutrino" to the neutrinos, "plus" to the triples <x, y, z> satisfying x plus y equals z,..., and so on?

A natural candidate is something along the following lines:

We are disposed to apply w to x

However this particular suggestion overlooks the fact that we sometimes make mistakes – sometimes we are disposed to call a given thing "a dog" even when that term is not in fact true of the thing. On a dark night we might wrongly apply it to a large and distant cat.

It is tempting to imagine that this difficulty can be avoided by modifying the original proposal, as follows:

We are disposed in ideal circumstances to apply w to x

But a powerful objection to this new idea about the identity of relation R is that we have absolutely no reason to believe that any such 'ideal circumstances' exist. Why should there be general conditions of inquiry in which, whenever the question arises as to whether a given predicate is true of a given object, we would inevitably reach the correct answer? Certainly, no philosopher has ever come close to specifying what those conditions are.[2]

[2] Even if we were to relax the assumption that a single relation, R, accounts for every instance of 'being true of', no matter which predicate is at issue – that is, even if we were to allow that a variety of relations, R1, R2, ..., Rk, might engender different instances of it (perhaps one relation for color terms, one for substance terms, one for size terms, etc.) – the difficulty of specifying, for each such category, its 'ideal conditions of inquiry', would not be significantly diminished.

This sort of reasoning is the core of Kripke's skeptical argument.[3] For if there is indeed nothing about our relationship to a word that could provide it with its distinctive referential import, and if (as seems obvious) any fact about its reference – hence, its meaning – would have to somehow result from some characteristic neural or computational or behavioral feature, then there can be no such thing as reference or meaning.[4]

---

[3] For something very like the argument just sketched, see pp. 22-32 of Kripke's book. His line of thought is elaborated and extended by Paul Boghossian (1985). It does not, of course, establish that there is no relation R that will do. But, in undermining the initially most attractive candidate, it puts a considerable onus of proof on anyone who continues to maintain that such a relation can nonetheless be identified.

[4] One might hope to escape the paradox by denying that our meaning-giving activity with a word must be specifiable in non-semantic terms – allowing that such activity may contain intentions and other propositional attitudes. It may be thought, for example, that "There's a dog" means what it does in virtue of the speaker's intention to communicate his belief that a dog is present. But, on reflection, this move merely re-locates the problem. Instead of the question of how linguistic expressions acquire their semantic contents, we now face the equally hard, parallel question of how mental/neurological states manage to acquire their particular contents (and thereby come to qualify as the particular intentions, beliefs, etc. that they are). Moreover, the most promising approach to this parallel question is to suppose (with Jerry Fodor) that such states are just 'sentences in the head' – an assumption that would bring us right back to square one!

So one may well be led to the idea, endorsed by Professor Chateaubriand, that semantic facts (concerning both sentences and mental states) are explanatorily fundamental. As I understand it, some such 'primitivist' move is also part of Kripke's own proposed solution to the problem. But we can reasonably object that this idea merely trades one paradox for another. For, on the one hand, the suggestion is that semantics facts are entirely ungrounded in, and uncaused by, physical phenomena. But, on the other hand, such facts are surely capable of influencing physical phenomena – in particular, a person's utterance of a given sentence on some

3. DEFLATIONISM WITH RESPECT TO "TRUE OF"

The thesis of the present paper is that this entire 'problem of aboutness' rests on a misconception about truth. The idea, more specifically, is, first, that no such difficulty can arise from a deflationary perspective; and, second, that we therefore have good reason to adopt that perspective. So let me now summarize the salient features of deflationism, and then proceed to describe their bearing on Kripke's paradox.

According to deflationism, and contrary to traditional thinking, sentential truth is not a deep 'substantive' property – i.e. a property about which one might expect a theory of the form

u is true ≡ u is Q

where "Q" stands for some correspondence property, or some verifiability property, or some pragmatic property, etc. Rather, the fundamental and defining principle is

u means that p → (u is true ↔ p)[5]

---

occasion is surely explained, in part, by the meaning he attaches to that sentence; yet this can be so (assuming 'the causal autonomy of the physical') only if that meaning is engendered by physical phenomena.

Thus, insofar as our aim is to demystify semantic facts, it doesn't help matters to question whether they result from non-semantic activity.

[5] It would perhaps be more accurate to say that the truly fundamental principles – from which this one is immediately derived – are (i) the equivalence schema for propositional truth, '<p> is true ↔ p', and (ii) the obvious definition of sentential truth in terms of propositional truth, 'u means <p> → (u is true ↔ <p> is true)'. But nothing in what follows would be affected by this correction. ("<p>" abbreviates "the proposition that p").

For it is our acceptance of such conditionals – not of any traditional explicit definition – that explains how we deploy the notion of truth. (– Primarily, as a device of generalization).

Similarly, "w is true of x" doesn't stand for a substantive (i.e. potentially-analyzable) relation. Rather, we understand it – relative to a prior understanding of meaning-attributions, such as "w means DOG" – through our acceptance of conditionals such as:

w means DOG → w is true of all and only dogs

w means CAR → w is true of all and only cars

... and so on[6]

4. DEFLATIONISM IMPLIES THAT THE PROBLEM OF EXPLAINING MEANING'S REFERENTIAL IMPORT IS A PSEUDO-PROBLEM

In general, if a predicate "f" (e.g. "ice") is defined in terms of an expression "g" (e.g. "frozen water"), and if something, k-ness, is proposed as the source (or cause, or origin) of g-ness, we do not think that this proposal stands in need of justification by reference to some way of explaining, independently of the definition, why it is

[6] "w means DOG" is an artificial notation introduced to express the distinctive meaning-property possessed by our word "dog" and by synonymous terms such as the Portuguese "cão", the Arabic "caleb", and so on. Thus one might think of it as w's property of meaning what is in fact meant by our word "dog". There are interesting issues as to whether such meaning-properties are really as relational as they seem – each involving the relation, means, together with one or another meaning entity (e.g. DOG, or CAR) – and, if so, as to what those ingredients themselves consist in. But these issues have little bearing on our present concerns.

that if something is k then it is f. Rather, we first justify the proposal on the basis of considerations that involve no use of "f", and we are then entitled to cite it, together with the definition, to explain why that conditional holds.

In particular, if deflationism (in the above sense) is correct – i.e. if "true of" is defined, as suggested, in terms of "means" – then the explanatory route leading from a word's meaning-engendering property (e.g. that "dog" has non-semantic property $Act_{57}(w)$) to its extension (e.g. that "dog" is true of exactly the dogs) must proceed via the intermediate fact that the word means what it does (e.g. that "dog" means DOG).

Therefore, unless deflationism has somehow been excluded, one has no right to insist that the phenomena responsible for a word's meaning DOG must provide a direct explanation of why it is true of the dogs (i.e. an explanation that does not presuppose the pertinent meaning-to-truth conditional). In other words, one has no right to require (as was done in Section 2) that the non-semantic source of 'w means DOG' entail something of the form '$(x)(wRx \leftrightarrow$ x is a dog)' – where R-ness either constitutes 'being true of', or is, in some other way, necessary and sufficient for that semantic relation to hold. Therefore, absent some refutation of deflationism, the above-argued non-existence of any such relation, R, coheres perfectly well with the reality of meanings, and with their being the product of our mental and/or behavioral activity.

Indeed, this coherence, together with the apparently insuperable difficulties that plague the initially-presupposed (inflationary) approach, provide strong evidence in favor of our deflationary alternative!

To repeat: the deflationist position is not to deny that a word's meaning-giving property fixes its extension, but to recognize that it does so only because it first fixes the word's meaning. For we can then invoke the definition of "true of" in terms of "meaning" – in particular

w means DOG → (x)(w is true of x ↔ x is a dog)

to infer (by transitivity) that the word is true of exactly the dogs.[7]

5. THIS PERSPECTIVE PRESUPPOSES THAT THERE IS SOME WAY OF DISCOVERING – INDEPENDENTLY OF TRUTH-THEORETIC CONSIDERATIONS – WHICH FACTS UNDERLIE THE MEANINGS OF WORDS. BUT NO CLUE HAS BEEN GIVEN AS TO HOW THAT MIGHT BE DONE

I would suggest that we address this issue by reference to the normal methodology, familiar from outside semantics, for settling matters of empirical reduction. In general, the question of how a property (e.g. being made of water) is constituted is approached by

---

[7] Kripke's own skeptical argument trades, not merely on the referential import of meaning, but also on its normative import. For example, it is presumably in virtue of what "dog" means that one ought to apply it only to dogs. So one might expect to be able to test any candidate meaning-constituting property by whether it would have that normative consequence. But – so the argument goes – we can't easily find anything that would pass such a test.

However, it seems to me that the solution to this problem is somewhat analogous to what has just been said about the truth-theoretic properties of terms. We can and should explain why it is that

If Act_{57}(w), then w should be applied only to dogs

by reference to the following pair of explanatorily more basic facts:

(i) If Act_{57}(w), then w means DOG
(ii) If w means DOG, then w should be applied only to dogs

Facts of type (i) are discovered via the methodology sketched immediately below. And facts of type (ii) are instances of the 'value of truth'. For discussion of whether and how the latter might be explained, see my "The Value of Truth". *Nous*, 40(2): pp. 347-360, 2006.

looking for an underlying property (e.g. being made of H_2O molecules) that explains the characteristic symptoms (e.g. boiling at 100 degrees Centigrade) of the superficial property. Now – turning to the question of how the constitutors of meaning-properties are to be identified – the main symptoms of a word's meaning are its various uses. So, we should be looking for whatever underlying property of the word will play a core role in explaining its overall deployment.

And, quite plausibly, there is such a property. We feel, as just noted, that our verbal output is, in part, the result of what we mean by our words. And, if that is right, then there should be, at the non-semantic level, some property of each word that accounts for this causal capacity.

What sort of property might that be? Arguably, it's a basic propensity of use, a law-like tendency to accept certain specified sentences containing the word in certain specified conditions. For instance, "dog"'s being governed by some such 'law' may (in virtue of the ability of that fact to explain the word's overall deployment), constitute its meaning what it does -– i.e. its meaning DOG.[8]

---

[8] For clarification of this proposal, for arguments in favor of it, and for responses to objections, see my *Meaning* (Oxford University Press, 1998, chapter 3); and my *Reflections on Meaning* (Oxford University Press, 2005, chapter 2).

I can see no non-trivial conceptual analysis of 'w means DOG'. One might try "w means what my word "dog" actually means", or "w has the same meaning as my word "dog"". But these are trivial (given the convention that allows us to name the meanings of our expressions by writing them in capital letters).

A non-trivial proposal would be, "w has the same basic use property as my word "dog"". But although this is, quite plausibly, a priori equivalent to "w means DOG (to me)", there is good reason to doubt that it provides an

6. TELLING ANALOGY

Suppose we define the relational term, "schmoo", by the stipulation that if something is made of plastic then it is schmoo of the dogs. Thus a credit card qualifies as schmoo of Pooch, but a nickel does not.

Would it then be reasonable to complain to a chemist who offers a reductive theory of plastic – e.g. 'plastic = XYZ' – that his theory can be accepted only if he provides a direct explanation (i.e. one that does not simply combine the theory with the stipulation) of why things made of XYZ are schmoo of precisely the dogs – and not (say) all the dogs except for Fido who lives on Alpha Centauri?

Of course not! No such direct explanation is conceivable. Given the definition of "schmoo", the only possible route from "XYZ" to "schmoo of the dogs" goes via "plastic". So the legitimacy of the first of these two steps cannot rest on anything to do with "schmoo". The chemist is perfectly entitled to accept his theory on other grounds, and then to explain the relationship between 'XYZ' and 'being schmoo of dogs' by deriving it from that theory in combination with the definitions of "schmoo".

7. THREE OBJECTIONS

Objection A

> There is a striking disanalogy between our hypothetical "schmoo" example and the case of "true of". For "schmoo" is introduced by means of a stipulation about how it is to be used in relation to the words "plastic" and "dog", which are already understood. However, it almost certainly was not the case that, only after we began to deploy ordinary terms, such

---

analysis of it, rather then an a priori specification of how to go about identifying the correct a posteriori reduction.

> as "dog", and predicates of meaning-attribution, such as "w means DOG", did we introduce "true of" via conditionals such as, "w means DOG → (x)(w is true of x ↔ x is a dog)".

Granted. But this difference, although real enough, is irrelevant. To see this, remember, that we are quite happy to say that "bachelor" abbreviates "unmarried man" – despite the historical absence of any explicit stipulation to that effect – because of how we use these expressions: more specifically, because we can see that the best way to explain our use of "bachelor" is in terms of our treatment of it as intersubstitutable with "unmarried man". Still, it's clear that, if someone did happen incorporate the word into his vocabulary via an explicit stipulation, it would acquire precisely the basic use that it actually has, and hence the same meaning. Similarly, in order for the meaning of "true of" to depend – in the way that deflationists claim it does – on the meanings of such terms as "dog", "w means DOG", "car", "w means CAR", etc., it suffices that the best explanation of our overall use of "true of" be that anyone who understands a predicate, "f", and its corresponding meaning-attribution, "w means F", accepts the conditional, "w means F → (x)(w is true of x ↔ x is an f)". That this explanatory hypothesis is correct would perhaps be more obvious if "true of" were introduced, in a "schmoo"-like way, via the stipulation that these conditionals hold. But it can be correct – and plausibly <u>is</u> correct – in the absence of any such mode of introduction.

<u>Objection B</u>

> In order to justify the hypothesis that 'being water' is constituted by 'being made of H_2O molecules', it was vital to show that certain known implications of something's having the superficial property (e.g. the implication that it boils at 100 degrees) would be explained by its having the proposed underlying property. So, why should we not think, similarly,

> that in order to justify the hypothesis that a word's meaning DOG is constituted by its having 'Act$_{57}$(w)', it would be vital to show that a certain prominent implication of a word's meaning DOG – namely, its being true of the dogs – would be explained by its having 'Act$_{57}$(w)'?

The answer lies in a crucial difference between the two cases. On the one hand, "boils at 100 degrees" is not defined in terms of "water". Rather, we have an independent understanding of it. And that is why we can make it an adequacy condition of the theory, 'water = H_2O', that there be a direct account of how being made of H_2O gives rise to that particular boiling point. But, on the other hand – assuming deflationism is correct – "true of" is defined in terms of "means". So it cannot be supposed that a theory of how 'w means DOG' is constituted will be credible only relative to a prior account of how the alleged constituting property gives rise to w's extension.

Objection C

> Someone can perfectly well accept that "w means DOG → w is true of the dogs" is both true by definition and explanatorily fundamental, and yet suppose – in stark opposition to deflationism – that this conditional helps to define "w means DOG" in terms of a prior notion of "w is true of the dogs" (rather than the other way around). And from this point of view, we can reasonably impose as a constraint on any analysis of 'w means DOG' that it square with a plausible analysis of 'w is true of the dogs'.

No doubt this approach deserves the serious consideration that (in effect) Kripke gave it. But, as he showed – and as indicated in Section 2 – it just doesn't pan out. We are not able to come up with a decent

direct explanation of how our activity with a word could result in its being true of the dogs.[9]

There remains, however, a theoretical option that is far more plausible than either meaning-skepticism or meaning-'primitivism' (a la footnote 4). For the deflationist order of definition can be vindicated. We are able to see how our acceptance of instances of "w means F → w is true of the fs" would, relative to an understanding of "f" and 'w means F", explain our overall use of "true of". And we are able see how a word's non-semantic meaning-giving activity (– the reductive ground of w's meaning F –) might be identified independently of any truth-theoretic considerations. So we can adequately defend the deflationist idea that the explanatory route from the meaning-giving activity with a word to its extension goes via its meaning-property. Indeed it would seem that deflationism about truth, when combined with a use-theory of meaning, provides the only viable perspective on these phenomena.[10]

REFERENCES

CHATEAUBRIAND, O. *Logical Forms. Part II: Logic, Language, and Knowledge*. Campinas: Unicamp, Centro de Lógica, Epistemologia e História da Ciência, 2005. (Coleção CLE, v. 42)

[9] Nor can we say what is the further characteristic of a word that, when added to its being true of the dogs, suffices for it to mean DOG.

[10] I would like to thank Oswaldo Chateaubriand for the stimulus provided by his opposing point of view. I must acknowledge that his book contains anti-reductionist considerations that I have not even mentioned here, let alone addressed or refuted. My hope though is to have achieved at least some degree of demystification of intentionality and to have removed one of the obstacles to achieving a fully naturalistic account of it.

BOGHOSSIAN, P. "The Rule Following Considerations'. *Mind*, 93, pp. 507-549, 1989.

HORWICH, P. "The Value of Truth". *Nous*, 40(2), pp. 347-360, 2006.

———. *Meaning*. Oxford University Press, 1998.

———. *Reflections on Meaning*. Oxford University Press, 2005.

KRIPKE, S. Wittgenstein on Rules and Private Language. Oxford: Basil Blackwell, 1982.