Alexander Hieke, Hannes Leitgeb

Reduction

Between the Mind and the Brain

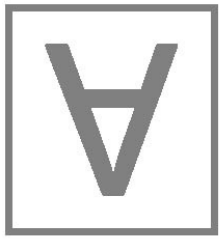
Publications of the
Austrian Ludwig Wittgenstein Society.
New Series

Volume 12

Alexander Hieke, Hannes Leitgeb

Reduction

Between the Mind and the Brain



ontos
verlag

Frankfurt I Paris I Lancaster I New Brunswick

Bibliographic information published by Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliographie; detailed bibliographic data is available in the Internet at http://dnb.ddb.de

Gedruckt mit Förderung des
Bundesministeriums für Wissenschaft und Forschung in Wien
und der Kulturabteilung der NÖ Landesregierung

North and South America by
Transaction Books
Rutgers University
Piscataway, NJ 08854-8042
trans@transactionpub.com

United Kingdom, Ire, Iceland, Turkey, Malta, Portugal by
Gazelle Books Services Limited
White Cross Mills
Hightown
LANCASTER, LA1 4XS
sales@gazellebooks.co.uk

Livraison pour la France et la Belgique:
Librairie Philosophique J.Vrin
6, place de la Sorbonne ; F-75005 PARIS
Tel. +33 (0)1 43 54 03 47 ; Fax +33 (0)1 43 54 48 18
www.vrin.fr

P.O. Box 15 41, D-63133 Heusenstamm
www.ontosverlag.com

ISBN 13: 978-3-86838-046-0

2009

Printed on acid-free paper
ISO-Norm 970-6
FSC-certified (Forest Stewardship Council)
This hardcover binding meets the International Library standard

Printed in Germany
by buch bücher **dd ag**

Introduction

The investigation of the *mind* has been one of the major concerns of our philosophical tradition and it still is a dominant subject in modern analytic philosophy as well as in science. Many philosophers in the scientific tradition want to solve the "puzzles of the mind"; but many philosophers in the very same tradition do regard these puzzles as puzzles of the *brain*, or to put it differently, they suggest we should avoid the reference to the mind altogether. So, whilst the former think of mental entities as something independent and of our speaking about them as irreducible to talk about physical entities, the latter deny that philosophy of mind has to do with anything else but the brain (or related physical entities) and thus rather want reference to mental entities to be eliminated altogether. And then there are those who think that *reduction* is the way to go: even if the language of the mental can not be translated into some language referring to physical entities only, maybe mental entities are still brain-dependent and hence reducible to physical entities in some ontological way.

This volume collects contributions comprising all those points of view. The articles originate from the 31st International Wittgenstein Symposium, August 2008, Kirchberg am Wechsel, which was devoted to *Reduction and Elimination in Philosophy and the Sciences*. We want to thank the authors for their great support, and we are confident that their work will stimulate further philosophical progress in this area of research.

Alexander Hieke & Hannes Leitgeb
Spring 2009

TABLE OF CONTENTS

I. BETWEEN THE MIND…

II. … AND …

III. … THE BRAIN

I.
BETWEEN
THE MIND ...

ENOUGH WITH THE NORMS ALREADY!

JERRY FODOR
Rutgers University

This isn't really a paper. It's more of a temper tantrum. Perhaps it will make up in vehemence what it lacks in arguments.

Everybody goes on about norms. Well, I am fed up with norms. If I never see another norm, that will be soon enough for me. Enough with the norms already. I am not, of course, antithetic to *every and all* norms. To the contrary, there are many norms with which I absolutely concur. For example, Flanders and Swann famously remarked that 'eating people is wrong' (In 'The reluctant cannibal'). Well, eating people *is* wrong (except, perhaps, in the most dire of emergencies.) Speaking for myself, I simply can't imagine eating a person. I honor this norm and I wish it well.

By contrast, the norms with which I am fed up are the kind that are alleged to block certain philosophical projects of naturalization to which I am professionally committed. In particular, I'd very much like there to be a naturalized account of the kinds of concepts (or terms, or constructs, or properties, or predicates, or whatever) that figure centrally in semantic and in intensional psychological explanations. Among the former, I tentatively Include 'is true', 'is false', 'is necessary', 'entails', 'means', 'refers to' and so forth. Among the latter, I tentatively include 'believes', 'intends', 'desires', 'acts', 'thinks', 'sees as' and so forth. In neither case do I know exactly what belongs on the lists; suffice it that the notions I would like to see naturalized include all the ones that occur ineliminably in such psychological explanations as we take to be true (or will take to be true.when/if we finally arrive at reflective equilibrium).

In particular, I'm interested in the prospects for constructing a naturalistic propositional attitude psychology. The idea is that believing, intending, and other states that figure in typical explanations of cognitive phenomena, informal or in the laboratory, would be treated as relations to a certain class of mental particulars ('mental representations'); and mental processes would be defined over these. This approach has been around, in one form or other, at least since Hume, according to whom mental repre-

sentations are ‘Ideas’ (something like mental images), and mental processes are causal (associative) relations among Ideas.

But a lot of philosophers think, on the one hand, that all the concepts, properties,…, whatever that such an account of cognition would require (for economy’s sake, I will henceforth call them all ‘whatevers’ because that avoids having to decide about just what ontology a naturalistic theory of cognition might be supposed to postulate), are, as it were, quasi- or crypto-normative. And these philosophers also think that practically as a point of definition, that what is normative can’t be naturalized. So, because of the normative character of whatevers, the naturalization project can’t but fail in precisely the areas where I most want it to succeed. In a nutshell, the norms I’m fed up with are the ones that are supposed to be incompatible with a naturalistic account of the psychology of cognition.

I strongly suspect that all the issues about the naturalization of whatevers have to do, in one way or other, with questions about symbols. I shall therefore assume that any explanation that is remotely plausible in the psychology of cognition will have to endorse (not just some notion of mental representation but also) some notion of ‘mental representations’. In the core cases, mental representations are discursive (non-iconic) symbols. That's to say that their tokens must be susceptible of semantic evaluation and they must have causal powers. Except for the ‘discursive’ part, this would come as no surprise to Hume, for whom Ideas are typically *of* something (in effect, they have referents) and association is a process of (mental) causation. Unlike Hume, however, I doubt that mental representations are anything like images. The reasons for denying that they are are familiar in both the philosophical and the psychological literatures, so I won’t rehearse them here.

In short, I want there to be a language of thought. I may, of course, be ill advised to want this; but, if so, I want to be ill advised on empirical grounds; I do not want to be ill advised a priori. So, assuming that what's normative can't be naturalized, I don’t want accounts of mental symbols or mental processes to be ipso facto normative.

Where did all this normative stuff come from? Some of it must surely be blamed on Hume, who claimed (or, at least, is claimed to have claimed) that you can't derive ‘ought’ from ‘is’. But I imagine that its modern incarnations started with, on the one hand, G.E. Moore’s formulation of the ‘open question’ argument and, on the other hand, Wittgenstein's suggestion that the use of symbols is a kind of rule-governed behavior. The open question argument went something like this: Whatever naturalistic

account you propose for goodness (or for 'good'), the question will remain intuitively open whether something that is good according to that account is, in actual fact, good. I think this is plausible because I suppose (and so, I take it, did Moore) that any account that doesn't leave the open question open would have to be not just *necessarily* true but a priori; and that got such an account to be a priori, it would have to be true by definition (hence analytic). Unlike Moore, however, I very much doubt that there are *any* definitions or analyticities, normative or otherwise. And even where there is definitional equivalence between a naturalistic expression and a normative expression, the two might nonetheless differ in their pragmatics, which may or may not count as a parameter of their 'use'. (As far as I know, no one has any serious proposal on offer as to what aspects of the use of an expression constitutes its 'use' in the technical sense where, we're told, use either is meaning or is what meaning supervenes on. That being so, issues like whether perlocutionary force is a parameter of use are moot as things stand.)

Anyhow, the open question argument is one plausible source of the notion that theories of whatevers are ineliminably normative. Another is Wittgenstein's suggestion that using symbols (mental or otherwise) is a species of rule-governed behavior. If that is granted, then the question arises whether, on a certain occasion, a symbol is used correctly or incorrectly. The idea is roughly that the semantics of symbols emerges from conventions for using them properly. And it seems plausible that notions like *proper* and *improper* are normative and ipso facto incapable of naturalization.

It's easy enough to tell a story that makes this seem true for natural languages (English, as it might be). Natural languages are vehicles of communication, and there is no communication without synchronization. If, for example, you and I are to communicate in English, you must mean by 'giraffe' and 'blue' what I mean by 'giraffe' and 'blue' and vice versa; otherwise we won't understand one another when either says that giraffes are blue. It's natural enough to gloss this as 'we must both be following the same rules for using 'giraffe' and 'blue'', where the normative force of the 'must' is instrumental; it means something like 'on pain of failing to communicate.'

But then, it would seem that the sort of story that seems plausible enough for English breaks down if you try to apply it to mental representations (unless mental representations are expressions in natural languages; a question that I also wish not to be settled a priori). For one thing, we don't

use the language of thought (hereinafter 'LOT' or 'Mentalese'). Not, at least, in the sense in which our using it would involve our having intentions with respect to how we use it. We mean to refer to giraffes when we say 'giraffe'. But (supposing that 'swiggle' is the word that refers to giraffes in Mentalese), we don't use 'swiggle' to refer to giraffes with the intention of so doing. In fact, we have no intentions at all in respect to 'swiggle'. Tokenings of expressions in Mentalese don't count as *actions*; they're things that just happen; presumably as a causal consequence of prior thoughts or of perceptual promptings. Least of all do we use 'swiggle' with the intention of acting in accord with norms for its use.

So, to gather all this together, English and LOT may well be different in that here are norms in accordance with which we use English, but (so far at least) none in accordance with which we use Mentalese. And maybe (*maybe*) a normless language is a contradiction in terms.

Notice, however, that here the normativity comes from the (instrumental) demands that *communication* makes, not from any demands (instrumental or otherwise) that *reference* makes. And there is, nothing so far, that shows that an expression's being used for communication is, as it were, *constitutive of* its being a referring expression. So, suppose that there aren't any *rules* for using a Language of Thought; psychology might still be naturalizable even though it claims that 'squiggle' refers to squiggles.

So now the question whether it is possible to naturalize Mentalese comes down to the question whether Mentalese is used as a vehicle of communication. Which, of course it isn't. Nobody ever used 'squiggle' to communicate anything to anyone; not even to themselves. Rather, the assumption is that Mentalese is used as the vehicle of *calculation*, (which, according to the kind of psychological theories I have in mind, is what many mental processes consist of). So, finally, the question about naturalization comes down to whether or not there can be a language that is used to calculate but not to communicate; a *de facto* private language. Or at least I shall assume that it does in the rest of this discussion.

I suppose that the burden of proof is on anybody who argues: 'no language without norms; no norms without communication; hence no de facto private languages', hence no Mentalese. I assume that, in any such argument, questions about normativity and questions about de facto publicity are inextricably tangled together: either normativity is inferred from publicity or that publicity is inferred from normativity, and both are taken to be essential properties of languages. It would be nice to have a reason to believe that this sort of argument is sound, but it's remarkably hard to find

one in the literature. What follows is a sketch of some of the candidates that have occurred to me. I don't claim that these exhaust the options; but I do claim that all of them are, in all likelihood, fallacious.

> *First Try:* Nothing is a symbol unless there is a difference between using it correctly and using it incorrectly. But if *E* is an expression in a de facto private language, then there is no difference between using it correctly and using it incorrectly. So the expressions of Mentalese aren't symbols. So Mentalese isn't a language.
> *Reply:* Who says there's no difference between using its expressions correctly and not using them correctly?
>
> *Second try:* Excuse me, I misspoke. What I meant was that, if a language is *de facto* private, then there must be a *verifiable* difference between using its expressions correctly and not using them correctly; there must be 'criteria' for their use.
> *Reply:* I'm not a verificationist. Assuming that 'squiggle' is an expression that refers to giraffes, it is wrong (mistaken, incorrect) to use it to refer to trees. But nothing *epistemological* follows as far as I know.
>
> *Third try:* The normativity of the rules of English is instrumental. It derives from the use of Engllish as a vehicle of communication. Since Mentalese is de facto *not a* vehicle of communication, its putative rules have no normative force. So there's an essential difference between Mentalese and English.
> *Reply:* Strictly speaking, this begs the question whether normativity is an *essential* property of rules of language. But put that aside. The story is that the normative force of the rules of English derives from the use of English to communicate, which is a project in which we have an interest. But it seems perfectly possible that there is some *other* project in which we are likewise interested, and that the normativity of the rules of Mentalese derives from it. For example, we're interested in having true beliefs, so the instrumental value of Mentalese may derive from its *de facto* necessity for our doing so. It's one thing to say that there must be norms. It's quite another that to say that the norms must derive from the exegencies of communication.

Fourth try: Our acquiescence in the rules of Mentalese is merely tacit.
Reply: So too is our acquiescence in the rules in English.

Fifth try: Equivalence is defined for the rules of Mentalese only up to extensional equivalence. Whereas there can be a matter of fact about which of two extensionally equivalent rules a speaker of a natural language is following.
Reply: Maybe there is no choosing between extensionally equivalent rules for using Mentalese even if the extensions of counterfactual tokenings are included (which they should be). But I can't think of any reason for believing that, and some such reason is owing.

Actually I *can* think of a reason; but it presupposes that rules of natural languages are, as it were, 'written down' in the heads of its speakers. If they are, then the choice between extensionally equivalent rules might be a choice between intensionally distinct ways of mentally representing them. On pain of a Lewis Carroll regress, however, nothing like this could be true of the rules of Mentalese, so this line of argument is unavailable to anyone who denies that Mentalese expressions are symbols (and hence have no referents).

Sixth try: Expressions in Mentalese have no perlocutionary force. For a symbol to have perlocutionary force is for its tokens to be intended to have a certain effect on their hearers. Expressions in a private language (even expressions in a merely de facto private language) don't have hearers and its users don't have intentions with respect to them.
Reply: But this begs the question whether its being used with perlocutionary intent is an *essential* property of something a symbol. I deny that it is; or, anyhow, that it has been shown to be. (See above)

Seventh try: Languages have to be learned. Learning requires instruction by someone. So, de facto, Mentalese can't be learned; so it's not a language.
Reply: It's true that English has to be learned (or, anyhow that it has to be 'picked up'); and that Mentalese can't be (again on pain of a Lewis Carroll regress.) This shows that anyone who posits a language of mental representation will have to be, to some very appreciable extent, a nativist. What's wrong with that?

Eighth try: Languages have to be translatable from the epistemic position of a radical translator. Since Mentalese is de facto private, it is not so translatable. So it isn't a language.
Reply: I know of no reason to suppose that a language must be translatable from the epistemic position of a radical translators. Come to think of it, I know of no reason to suppose that English is.

Ninth try: Tokenings of a language have to have (or any how, to be capable of having) interpreters. Since Mentalese is de facto private, there is no interpreter of its tokenings. So Mentalese can't be a language.
Reply: The premise needs arguing for (on some grounds other than the assumption of verificationism. See the reply to *second try*). But suppose we pass that. Still, the question remains why a *counterfactual* interpreter ('If there *had been* an interpreter, he would have taken the token to mean such and such') wouldn't meet the specifications. Is there some reason why there shouldn't be counterfactual interpreters of Mentalese? (In particular, counterfactual interpreters who know about the causes and effects of its tokenings.) See Fodor 2008 for a sketch of how 'triangulation' (which, at least according to Donald Davidson, is par excellence what interpreters do for a living) might be worked out in terms of such counterfactuals.

I'm out of candidates. I have no proof that the only possible normativity/privacy arguments against the naturalization of Mentalese are of the kind that I've been surveying. But, I can't think of any others, and these do all seem to be to be distinctly dubious. I intend, therefore, to proceed on my way, assuming that there aren't any good arguments against a de facto Mentalese that turn on issues of privacy/normativity. If you think of one, do please let me know right away. I have an email account at which I can almost always be reached.

REFERENCES

Fodor, J. *LOT2: The Language of Thought Revisited.* Oxford University Press, Oxford, 2008.

INTENTIONALITY, INFORMATION, AND EXPERIENCE

JOHANNES L. BRANDL
University of Salzburg

INTRODUCTION

Descartes claimed that the essence of the mind is thinking (*cogitare*), to which Brentano added that the essence of thinking is to be mentally directed at objects of some kind.[1] These are controversial assumptions about the nature of the mind. A more modest starting point would be to say that intentionality is an important feature of a large class, but not necesarrily of all mental phenomena. Though this claim too has been challenged in the behaviourist tradition, it is now widely accepted that having a mental life involves mental states with a mental content. Cognitive scientists call it the representational power of the mind. Like Descartes and Brentano, we therefore face the task of explaining this important feature of the mind. Where does its representational power come from?

It is also widely agreed today that the project of reducing intentionality to language has failed. The representational power of the mind does not derive from our capacity to speak a language. It is rather the other way round: linguistic expressions derive their meaning and their referential power from the mental states of speakers and hearers which guide their linguistic behaviour. Mental representation is the foundation of linguistic representation. The foundations on which the power of mental representation rests must lie elsewhere. Can we dig here any deeper? Sceptical philosophers like Quine have resisted that demand. That, however, is not a comfortable position, if one has agreed that intentionality is a real feature of thought. How could there be no further explanation of how the mind acquires its representational power?

---

[1] Brentano gives credit here to the scholastic doctrine of "intentional (or mental) inexistence", i.e. the existence of objects "in" psychic phenomena (see Brentano 1874/1973, p.88). Though the term "intentionality" derives from Brentano and therefore still echoes these historical roots, the term "object" no longer denotes what exists "in" mental acts but has been replaced by the term "mental content".

Fortunately, there is a better option available. In fact there are *two* broadly conceived programs for explaining intentionality that compete with each other. One of them is *informational semantics*.[2] It takes the category of information to be fundamental and tries to explain how intentional mental states arise in cognitive systems from tracking the information that is available in their environment. Theories that propose to naturalize intentionality typically follow this program. The other project might be called *phenomenological semantics*.[3] It rests on the claim that intentionality is founded in conscious experience. Accordingly it tries to explain how conscious experience generates a phenomenal content from which the conceptual content of thoughts can be derived. This too might be seen as providing intentionality with a natural foundation, but it is clearly a very different form of "naturalization". Thus we face a difficult choice: where should we put our money?

The goal of this paper is modest in several ways. First, I will not argue for the claim that informational semantics and phenomenological semantics are the only two games in town. I do think, however, that on a broad understanding of these terms most theories of intentionality may be regarded as belonging to one of these frameworks.

Secondly, I cannot rule out the possibility that there is a deeper level of explanation at which informational semantics and phenomenological semantics might be reconciled with each other. There are theories, like Castañedas guise theory, that might be interpreted along these lines.[4] However, I think that such an integrative theory, if it is successful, would be more like a third approach that shares some features with the others but also gives up some of their basic assumptions.

---

[2] I use the label 'informational semantics' here in a broad sense to include any theory that assigns a fundamental role to natural meaning. This holds for informational theories that emphasize causal relations, like the semantics 'Wisconsin style' of D. Stampe, B. Enç and F. Dretske, as well as for other naturalistic accounts of mental representation by A. Denkel, R. Millikan, D. Papineau, K. Sterelny, and others (see Macdonald & Papineau 2006).

[3] I use the label 'phenomenological semantics' here for any theory that assigns a fundamental role to experience in the constitution of mental content. This includes Husserl's theory of meaning constituting acts, but also a conceptual role semantics that starts with phenomenal intentionality, and other recent contributions to the 'phenomenal intentionality research project' by T. Horgan, U. Kriegel, B. Loar, G. Strawson, and others (see Bayne & Montague, forthcoming).

[4] More recently, Edward Zalta has suggested that his theory of abstract objects might allow for a similar reconciliation; see Zalta 2000.

Thirdly, I am not going to suggest a method or criterion of how one might rationally choose between informational semantics and phenomenological semantics. Clearly, both programs have their advantages and disadvantages, and it would take considerably more space than is available here to evaluate them and weigh their respective pros and cons.

The modest task I set myself here is to show how the conflict between informational semantics and phenomenological semantics *cannot* be resolved. It cannot be resolved, so I shall argue, by demonstrating that one of these programs is "deeply flawed" or perhaps even inconsistent. Arguments that try to find such a flaw in informational semantics have been proposed by Jonathan Lowe (1995/97) and Uriah Kriegel (2007). These arguments, so I shall argue, are ultimately question-begging.

In sections 1 and 2 I do the stage-setting by introducing the main ideas and the attractive features that one finds in informational semantics and phenomenological semantics respectively. In section 3 I present and criticize Lowe's argument against informational semantics. A brief summary of the core of Kriegel's argument is given in section 4, followed by my criticism of this argument in section 5. In the final section I indicate how these results may be helpful – even if not decisive – in finding the right method for explaining intentionality.

1 From information to intentionality

Information is a commodity of our daily life. We continuously receive, transmit and store information of all kinds. Although this information processing is very familiar to us, it is hard to say exactly what this comes to. What is this curious thing called 'information' that exists in our brains, in books and on TV, in our computers, and in many other places?

An answer to this question has been offered by Fred Dretske in his book *Knowledge and the Flow of Information* (1981). This seminal work launched the project of informational semantics in explaining the foundations of cognition. Dretske's conception of information takes its lead from Claude Shannon's probabilistic notion of 'amount of information' that a signal can carry in a communication process. It diverges from Shannon, however, by also appealing to the semantic notions of 'meaning' and 'content'. Thus Dretske arrives at a definition of *informational content*. It defines the content that a signal *s* carries, relative to the knowledge *k* of an agent who receives that piece of information. The definition says: "A sig-

nal *r* carries the information that *s* is *F* $=_{df}$ The conditional probability of *s*'s being *F*, given *r* (and *k*), is 1 (but, given *k* alone, less than 1.)" (Dretske 1981, p.65).

This technical notion seems at first sight to be far removed from our commonsensical understanding of information. We take it that a piece of information can be more or less accurate and that a person can have misleading or even completely wrong information. This is not compatible with Dretske's definition. It is therefore a bit surprising when he claims that his notion of informational content "corresponds strikingly well with our ordinary, intuitive understanding of information: […] it enables us to understand the source […] of the semantic character of information; and it reveals the extent to which, and the reason why, the information one receives is a function of what one already knows." (Dretske 1981, pp.81f.) Where do we find the correspondence that Dretske is speaking of here?

The intuitions that fit Dretske's proposal can be found in the veridical usage of terms like 'perception', 'remember' and 'know'. This usage is constrained by the condition that a subject *S* can perceive an object *O*, or perceive that *O* has a certain property *F*, only if *O* exists and actually has that property. Equally, one can remember that something happened only if it really happened, and one can know something only if it is true. In these contexts we use the notion of informational content (or simply 'information') in the way in which Dretske defines it. We say, for instance, that in perceiving something we pick up information, that in remembering something, we retrieve information from our memory, and that a person who is thus informed about a subject matter knows something about it.

The point here is that perception and memory bring us in contact with reality. This contact becomes more elusive when we consider mental operations in general, not just cognitive operations like veridical perceptions and successful cases of memory retrieval. The faculties of perception and memory can also deceive us. They may deceive us about being in contact with reality. In this case we still have thoughts and these thoughts are about reality, but in a different sense: they merely *purport* to bring us in cognitive relation with real objects and states of affairs. They create an appearance of contact that is not actually there.

Informational semantics requires acknowledging the fact that mental states can deceive us, while also respecting the veridical usage of epistemic terms. In this way the Dretskian notion of information can be sustained. The insight here is that carrying information (in the Dretskian sense) is not essential for a mental state, just as it is not essential for it to bring us in

contact with reality. This insight gives rise to an important question: Why is it that cognitive systems are necessarily fallible? Why is it that any system that has the power to receive and store information by perceiving and remembering things, also has the power of forming false beliefs and false memories about reality?

This question takes us beyond the notion of information, as Dretske defines it. The central term now becomes 'representation'. With this notion we enter the familiar territory of a theory of intentionality. The fallibility of mental operations is their central feature from an intentional point of view. There are three aspects to it that may be distinguished:

(1) Thoughts can represent existing objects as well as objects that do not exist.
(2) Thoughts can represent some object *O* that has property *F* without representing the fact that *O* has *F*.
(3) Thoughts can represent some object *O* although the subject of this thought does not believe to have thoughts about *O*.

How does informational semantics explain these features? That question is too complex to be answered succinctly. There are several attractive ideas that one may pursue here. I can only mention some of these ideas without offering any details.

First, there is the idea of *teleological function* and *cognitive fitness* championed in the work of Ruth Millikan (see Millikan 1989/93). According to this idea, thoughts have the power to represent because in doing so they enhance the cognitive fitness of the system in which they occur. For instance, it is of great importance for an organism to know when an enemy is near. It is therefore part of the proper function of its cognitive apparatus to make the system aware of the presence of enemies. Yet objects may appear to be dangerous for the system without actually being so. When the system is thus deceived, its cognitive system still performs its proper function of indicating the presence of a dangerous object. This is, in a nutshell, the teleological explanation how our thoughts can come to represent non-existing objects.

A second idea is "nomological control". Jerry Fodor has proposed this idea for explaining which properties enter the content of a concept and thus become part of the thoughts we have about objects having those properties (see Fodor 1990). For instance, why do we think of an animal with a tail when we think of a dog? This cannot be explained by saying that all

dogs that we have seen actually had tails since we may have seen dogs without tails. Yet we might try the following explanation: a concept *C* refers to objects with a certain feature *F* iff *C* is under the nomological control of *F*, i.e. if there is a counterfactual-supporting causal relation between tokens of *C* and instances of *F*. This might explain why "having a tail" is part of our concept of dog and why other features, like their color, are not. This is, very briefly stated, the nomological explanation of how we can represent objects without also representing all the properties they have.

A third idea that plays an important role in informational semantics is *etiology*. This idea may be used to explain why our thoughts can represent objects different from those that we believe them to represent. Such cases have become prominent with the Twin-Earth thought-experiments. My twin on Twin-Earth may believe that he is drinking water and that he has thoughts about water, while in fact he is drinking and thinking about a quite different substance. Causal theories of reference explain this possibility along the following lines: A concept *C* refers to objects of kind *K* iff an object of kind *K* has been the incipient cause for forming the concept *C*. That explanation is useful also in cases of real life. We all have many false beliefs about the environment we live in. This need not prevent us from having thoughts about the real substances we interact with. Only etiology, so it seems, can explain this peculiar aspect of intentionality.

These ideas have often been taken to be in competition with, or even opposed to informational semantics (see Fodor 1990; Millikan 1990/93). If one thinks of it as a project that addresses a number of different problems, however, these ideas may also be fruitfully combined with each other. Explanatory power can thus be gained by combining a teleological background theory with the facts of nomological co-variation and incipient causation. This certainly looks like a highly promising research program for a reductive explanation of intentionality.[5]

But doubts remain. These doubts do not just concern the details of the program. There are reasons to think that informational semantics might be completely on the wrong track. Before I consider some of these objections, however, I want to introduce the main alternative approach to explaining intentionality, namely phenomenological semantics.

[5] For a recent example of such a combined approach see Jesse Prinz 2002.