Any dictator would admire the uniformity and

obedience of the US media.

–Noam Chomsky

CHOMSKY

AND HIS CRITICS

PHILOSOPHERS AND THEIR CRITICS

General Editor: Ernest Lepore

Philosophy is an interactive enterprise. Much of it is carried out in dialogue as theories and ideas are presented and subsequently refined in the crucible of close scrutiny. The purpose of this series is to reconstruct this vital interplay among thinkers. Each book consists of a temporary assessment of an important living philosopher's work. A collection of essays written by an interdisciplinary group of critics addressing the substantial theses of the philosopher's corpus opens each volume. In the last section, the philosopher responds to his or her critics, clarifies crucial points of the discussion, or updates his or her doctrines.

1 Dretske and His Critics
Edited by Brian McLaughlin

2 John Searle and His Critics
Edited by Ernest Lepore and Robert van Gulick

3 Meaning in Mind: Fodor and His Critics
Edited by Barry Loewer and Georges Rey

4 Dennett and His Critics
Edited by Bo Dahlbom

5 Danto and His Critics
Edited by Mark Rollins

6 Perspectives on Quine
Edited by Robert B. Barrett and Roger F. Gibson

7 The Churchlands and Their Critics
Edited by Robert N. McCauley

8 Singer and His Critics
Edited by Dale Jamieson

9 Rorty and His Critics
Edited by Robert B. Brandom

10 Chomsky and His Critics
Edited by Louise M. Antony and Norbert Hornstein

CHOMSKY

AND HIS CRITICS

Edited by

Louise M. Antony and Norbert Hornstein

Blackwell Publishing

350 Main Street, Malden, MA 02148-5020, USA
108 Cowley Road, Oxford OX4 1JF, UK
550 Swanston Street, Carlton, Victoria 3053, Australia

First published 2003 by Blackwell Publishing Ltd
Reprinted 2003

Library of Congress Cataloging-in-Publication Data

Chomsky and his critics / edited by Louise M. Antony and Norbert Hornstein.
p. cm. – (Philosophers and their critics)
Includes bibliographical references and index.
ISBN 0-631-20020-7 (alk. paper) – ISBN 0-631-20021-5 (pbk. : alk. paper)
1. Chomsky, Noam. 2. Linguistics – Philosophy. I. Antony, Louise M.
II. Hornstein, Norbert. III. Series.
P85.C47 C47 2002
410'.92–dc21

2002074365

A catalogue record for this book is available from the British Library.

Set in 10/12pt Ehrhardt
by SetSystems Ltd, Saffron Walden, Essex
Printed and bound in the United Kingdom
by MPG Books Ltd, Bodmin, Cornwall

For further information on
Blackwell Publishing, visit our website:
http://www.blackwellpublishing.com

Contents

Notes on Contributors

FRANCES EGAN is Associate Professor of Philosophy and a member of the Center for Cognitive Science at Rutgers University. She is the author of numerous articles in the philosophy of mind and the philosophy of psychology, including "Folk Psychology and Cognitive Architecture," and "Computation and Content."

ALISON GOPNIK is Professor of Psychology at the University of California at Berkeley. She is the author of many articles on cognitive development and its relation to philosophical problems, and has written two books, *Words, Thoughts and Theories* with Andrew Meltzoff, and *The Scientist in the Crib* with Andrew Meltzoff and Patricia Kuhl.

PAUL HORWICH is Kornblith Professor of Philosophy at the Graduate Center of the City University of New York. He is the author of *Probability and Evidence: Asymmetries in Time, Truth, Meaning*, as well as numerous essays in the philosophy of language, the philosophy of science, and metaphysics.

PETER LUDLOW is Associate Professor of Philosophy at the State University of New York, Stony Brook. He is author of *Semantics, Tense, and Time: An Essay in the Metaphysics of Natural Language*, as well as a number of articles on topics ranging from semantics and the philosophy of linguistics to cyberspace ethics. He is editor or co-editor of several volumes, including *Readings in the Philosophy of Language*.

WILLIAM G. LYCAN is William Rand Kenan, Jr. Professor of Philosophy at the University of North Carolina. His publications include *Logical Form in Natural Language*, *Knowing Who* (with Steven Boër), *Judgment and Justification*, *Modality and Meaning*, *Consciousness and Experience*, and *Real Conditionals*.

RUTH GARRETT MILLIKAN is a graduate of Oberlin College and Yale University. She is currently Board of Trustees Distinguished Professor and The Alumni Association Distinguished Professor for 2000–2003 at the University of Connecticut. She is author of *Language, Thought and Other Biological Categories*, *White Queen Psychology*, *On Clear and Confused Ideas*, and *Varieties of Meaning: The Jean Nicod Lectures for 2002*, forthcoming. She works in the philosophies of biology, psychology, and language, and on questions in ontology.

PAUL M. PIETROSKI is Associate Professor of Philosophy and Linguistics at the University of Maryland at College Park. He is the author of *Causing Actions*, along with articles in the philosophy of language and semantics.

JEFFREY POLAND is a member of the Department of Psychology and the Center for Curriculum & Instruction at the University of Nebraska, Lincoln. He is the author of *Physicalism: The Philosophical Foundations*, and co-author, with William Spaulding, of two forthcoming books on the analysis and reform of clinical and research practices concerning mental illness.

GEORGES REY is Professor of Philosophy at the University of Maryland at College Park. He is the co-editor, with Barry Loewer, of *Meaning in Mind: Fodor and his Critics*, and the author of *Contemporary Philosopy of Mind: A Contentiously Classical Approach*, as well as of numerous articles in the philosophy of mind and cognitive science.

GALEN STRAWSON is Professor of Philosophy at the University of Reading. He is the author of books on free will, causation, and the philosophy of mind: *Freedom and Belief*, *The Secret Connexion: Causation, Realism, and David Hume*, and *Mental Reality*, and is currently working on a book on the "self."

Acknowledgments

We would like to thank the Anarchy Archives, for compiling our list of works by Noam Chomsky, and the College of Humanities and Department of Philosophy of the Ohio State University for help in deferring the costs of production.

Introduction

NORBERT HORNSTEIN and
LOUISE ANTONY

There is probably no single person within the last century who has had a more dramatic impact on contemporary understandings of language and the mind than Noam Chomsky. Not only did he fundamentally restructure the science of linguistics, but he transformed the science of psychology, rehabilitating the doctrines of mentalism and nativism after their long exile during the reign of behavorism. Philosophers took note: the Chomskian revolution offered both new possibilities for understanding language and mind, and new challenges to entrenched philosophical opinion. This volume brings together ten scholars – philosophers and philosophically minded cognitive scientists – to explore some of the possibilities and to take up some of the challenges.

Any discussion of Chomsky's work must begin with the linguistics. One of the central features of Chomsky's view of language is his emphasis on the biological basis of the human linguistic capacity. Fish swim, birds fly, people talk. According to Chomsky, these various talents are of a piece – all of them rest on specific biological structures whose intricate detail is, to a considerable degree, attributable to the organism's genetic endowment. Human linguistic capacity involves, in particular, a postulated "mental organ," many of whose specific features are innately specified. This posited structure is initially dedicated to the acquisition of linguistic knowledge and is subsequently involved in various aspects of language use, including the production and understanding of utterances. The aim of linguistic theory is to describe the initial state of this faculty and the changes it undergoes with exposure to linguistic data. Chomsky characterizes the initial state of the language faculty as a set of principles and parameters (see Chomsky 1981). Language acquisition then consists in the setting of these open parameter values on the basis of linguistic data available to the child. In this way, Chomsky explains both the commonalities and the differences among humanly natural languages. The commonalities are reflections of the initial state

of the system: it is a universal grammar (UG), a super-recipe for concocting language-specific grammars. The differences are embodied in the various grammars that result when parametric values are fixed.

Linguistic theory, given these views, has a double mission. First, it aims to adequately characterize the particular grammars (and, hence, the mental states) attained by native speakers during the process of language acquisiton. A theory is said to be "descriptively adequate" just in case it achieves this goal. But linguistic theory has a second aim, and that is to explain how grammatical competence is attained – to explain, in other words, how descriptively adequate grammars can arise on the basis of exposure to "primary linguistic data" (PLD) – the data children are exposed to during the period of acquisition. A theory that satisfies this requirement is said to be "explanatorily adequate." Explanatory adequacy will depend on an articulated theory of UG – a detailed theory of the general principles and open parameters that characterize the initial state of the language faculty.

Chomsky's conception of linguistics as a science oriented around these two goals – particularly the latter goal – was one of the revolutionary aspects of his theory of language. In any given domain, there is an intimate relation between how a problem is conceived and the kinds of explanations one should offer. Chomsky proposed that we identify explanation in linguistics with a solution to the problem of how children can attain mastery of their native languages. This is often referred to as "the logical problem of language acquisition." It sets linguistic explanation in the context of the problem of explaining how a child masters a rich and highly structured system on the basis of a rather slender data base. The problem looks like this.

Natural languages pair sounds and meanings over an unbounded domain of structures. Humans typically come to master at least one such mapping in a surprisingly short time, without conscious effort, explicit instruction or apparent difficulty. How is this possible? An adequate answer to this question must take account of the following facts.

First, a human can acquire any language if placed in the appropriate speech community. Get raised in Boston and you'll grow up speaking English the way Bostonians do. Similar remarks hold for Moscow and Russian, Tel-Aviv and Hebrew, and Paris and French. It would seem that this fact could be handily explained by reference to the diversity of the data available to the children raised in these various locales and exposed to these various forms of human language. But here we must confront a second fact, namely, that the PLD, the data actually available to the child, are, taken on their own, vastly insufficient to determine the rich grammars the children inevitably attain. Specifically, there are four kinds of problems with the data that prevent them from determining the outcome:

(a) The set of sentences the child is exposed to is finite, while the knowledge attained extends over an unbounded domain of sentences.

(b) The child is exposed not to sentences *per se* but to *utterances* of sentences.

Utterances are imperfect vehicles for the transmission of sentential information as they can be defective in various ways. Slurred speech, half sentences, slips of the tongue, and mispronunciations are only a few of the ways that utterances can prove unreliable as indicators of sentence structure.

(c) Acquisition takes place without explicit guidance by the members of the speech community. As it happens, children do not make many errors to begin with when one considers the range of logically possible mistakes, and the errors they do commit fall into regular, predictable patterns. Moreover, adults do not engage in systematic corrections of errors that do occur. Finally, even when correction is offered children don't seem to notice or to care. Children seem, in short, surprisingly immune to any form of adult linguistic intrusion (see Lightfoot 1982 for further discussion).

(d) Last, but perhaps most importantly, it is likely that, of all the linguistic evidence theoretically available to the child, only that in simple sentences is absorbed. The gap between in*put* and in*take* is attributable to various cognitive limitations such as short attention span and limited memory.

All of this implies that the acquisition process is primarily guided by the information available in well-formed simple sentences. In contrast to the evidence available to the linguist in theory construction, the information the child uses in building its grammar is severely restricted. The PLD do not contain, for example, either negative data (the information available in unacceptable, ill-formed sentences) or complex data (the information yielded by complex constructions). This suggests that whenever the linguistic properties of complex clauses diverge from simple ones, the acquisition of this knowledge cannot be data-driven. Thus the acquisition of a grammar cannot be a matter of routine induction. The overall argument here has become known as the argument from the poverty of the stimulus.

The general picture that emerges from these considerations is that attaining linguistic competence involves the development of a rule system or grammar and that humans come equipped with a rich innate system that guides the process of grammar construction. This system is supple enough to allow for the acquisition of any humanly natural language grammar, yet rigid enough to yield results even in the face of the deficiencies of the PLD. Linguistic theories must describe both the grammars attained, and the fine structure of the innate capacity that makes possible the attainment of these grammars – they must be, in other words, both descriptively and explanatorily adequate.

Since the very beginnings of generative grammar these issues of descriptive and explanatory adequacy have loomed large. Chomsky's arguments, for example, against Markov models of human linguistic competence were that they were incapable of dealing with long-distance dependencies exemplified by conditional constructions in English and hence could not be descriptively adequate. His argument in favor of a transformational approach to grammar rested on the claim that it allowed for the statement of crucial generalizations evident in the

judgments of native speakers and so advanced the goal of descriptive adequacy (see Chomsky 1957). Similarly, his influential review of Skinner's *Verbal Behavior* (Chomsky 1959) consisted in showing that the learning theory mooted therein was explanatorily inadequate: it either failed, on grounds of vagueness, to provide any clear account of language acquisition, or, if key terms like "stimulus" and "reinforcement" were given even moderately precise meanings, then the theory's predictions were clearly incorrect.

The shift from the early *Syntactic Structures* theory to the one in *Aspects of the Theory of Syntax* (Chomsky 1965) was also motivated by concerns of explanatory adequacy. In the earlier model generalized transformations constitute the recursive engine of the grammar. The successive application of these transformations allows for the generation of more and more complex sentences from the strings produced by the phrase structure component of the grammar. In the *Aspects* theory, recursion is incorporated into the phrase structure component itself and removed from the transformational part of the theory. The impetus for this was the observation that greater explanatory adequacy could be attained by grammars that had a level of "Deep Structure" incorporating a recursive base component. In particular, Fillmore observed that the various optional transformations in a Syntactic Structures theory always applied in a particular order in any given derivation. This order is unexplained in a Syntactic Structures theory. In *Aspects* it is deduced. Central to the deduction is the reorganization of the grammar so that recursion becomes a property of the base component. Thus, the move to an *Aspects* style grammar is motivated on grounds of greater explanatory adequacy: introducing Deep Structure and moving recursion to the base allows for a more restricted theory of universal grammar. Restricting UG is always desirable (all things being equal) as it advances a central goal of grammatical theory: the more restricted the options innately available for grammar construction, the easier it is to explain how language acquisition is possible despite the defects in the PLD.

The same logic motivates various later additions to and shifts in grammatical theory. For example, a major move in the 1970s was to radically simplify transformational operations so as to make their acquisition easier. This involves eliminating any mention of construction-specific properties from transformational rules.

For example, an *Aspects* rule of passive looks like (1):

(1) X - NP1 - V - NP2 - Y → X - NP2 - *be* + *en* V - *by* + NP1 - Y

(The left-hand side gives the structural description (SD) and the right-hand side gives the structural change (SC).)

Observe that the SC involves the constants *be* + *en* and *by*. The SD mentions three general expressions, NP1, V, and NP2 and treats these as part of the context for the application of the rule. In place of this, Chomsky proposed

eliminating the passive rule and replacing it with a more general rule that moves NPs (see, e.g., Chomsky 1977, 1986). The passive rule in (1) involves two applications of the "Move NP" rule; one moving the subject NP1 to the *by* phrase and another moving the object NP2 to the subject position. In effect, all the elements that make the passive rule in (1) specific to transitive constructions are deleted and the rule itself is replaced by the simpler rule "Move NP."

There is a potential empirical cost to simple rules, however. The simpler a transformation, the more it overgenerates unacceptable outputs. Thus a grammar with (1) cannot derive "was jumped by John" from "John jumped" because the requirements of the SD are not met by underlying forms that are intransitive. A grammar eschewing (1) and opting for the simpler "Move NP" rule is not similarly restricted. To prevent overgeneration, therefore, the structure of UG must be enriched with general grammatical conditions that function to rein in the undesired generation (see Chomsky 1973, 1977, 1986). Chomsky has repeatedly emphasized the tension inherent in developing theories with both wide empirical coverage and reasonable levels of explanatory adequacy.

A high point of this research agenda is Chomsky's *Lectures on Government and Binding* (LGB) (Chomsky 1981). Here the transformational component is reduced to the extremely simple rule "Move α," i.e., "Move anything anywhere." To ensure that this transformational liberty does not result in generative chaos, various additions to the grammar are incorporated, many conditions on grammatical operations and outputs are proposed and many earlier proposals (by both Chomsky and others) are refined. Among these are trace theory, the binding theory, bounding theory, case theory, theta theory and the Empty Category Principle. The picture of the grammar that LGB presents is that of a highly modular series of interacting subsystems which in concert restrict the operation of very general and very simple grammatical rules. In contrast to earlier traditional approaches to grammar, LGB witnesses the virtual elimination of grammatical constructions as theoretical constructs. Thus, in Government Binding (GB) style theories, there are no rules of Passive, Raising, Relativization or Question Formation as there were in earlier theories. Within GB, language variation is not a matter of different grammars having different rules. Rather, the phenomena attested in different languages are deduced by variously setting the parameters of universal grammar. Given the interaction of the grammatical modules, a few parametric changes can result in what appear on the surface to be very different linguistic configurations (see Chomsky 1983).

The GB research program has proved to be quite successful both in its descriptive range and explanatory value. Despite this, Chomsky has recently been urging a yet more ambitious avenue of research. He has embarked on the development of a rationalist approach to grammar that goes under the name of "minimalism" (see Chomsky 1995). The term "rationalist" signifies a theory grounded on very simple and perspicuous first principles. A minimalist theory will employ only those notions required by "virtual conceptual necessity" –

concepts that no approach to grammar can conceivably do without while remaining true to the most obvious features of linguistic competence. For example, every theory of grammar treats sentences as pairings of sounds and meanings. Thus, any theory will require that every sentence have both a phonological and an interpretive structure. In GB theories, information of these two kinds is encoded in the Phonetic Form (PF) and Logical Form (LF) phrase markers, respectively.

GB theories recognize two other distinctive grammatical levels: S-structure and D-structure; but a minimal theory, Chomsky argues, should dispense with everything but LF and PF. A minimal theory will, therefore, differ significantly from any current version of GB. It will be based on natural "economy" principles and indispensable primitives. Chomsky has suggested reanalyzing many of the restrictions that GB theories impose in terms of "least effort" notions such as "shortest move" and "last resort movement." For example, the unacceptability of sentences such as (2), he proposes, is ultimately due to the fact that the moved NP *John* need not have moved from the embedded subject position (marked by "t"), as it fulfills no grammatical requirement by so moving.

(2) *John is expected t will win

This work is still in its infancy, but it has already prompted significant revisions of earlier conclusions. For example, with the elimination of D-structure, the recursive engine of the grammar has once again become the province of generalized transformations. Whatever its ultimate success, however, minimalism continues the pursuit of the broad goals of descriptive and explanatory adequacy enunciated in Chomsky's earliest work.

Chomsky's research revitalized two philosophical programs that had fallen into disrepute in the twentieth century: rationalism and mentalism. The debate between rationalists and empiricists is frequently characterized in terms of innate ideas: rationalists believe that innate knowledge is possible, while empiricists believe that all knowledge derives from experience. But this is somewhat misleading. Both camps agree that native mental structure constrains the kinds of knowledge that a subject can acquire; what they disagree about is the richness and task-specificity of that mental structure. Empiricists contended that all human knowledge was derived from primitive sensory experiences, via specified operations of association and abstraction. Rationalists, on the other hand, argued that sense experience alone could not explain crucial features of human knowledge, so that rich innate concepts had to be admitted. Empiricist views of knowledge and the mind dominated Anglo-American philosophy in the early part of the twentieth century, until Chomsky's devastating critique of Skinner's behaviorism (Chomsky 1959) brought a rationalist alternative back into play. Empiricists have not conceded the game, however. Many hold that a new theory of cognitive architecture, connectionism, will reveal that there is considerably

more information available in the sensory stream than was thought, challenging the basis of Chomsky's poverty-of-the-stimulus argument for native linguistic knowledge. Contributor Alison Gopnik argues for a quasi-empiricist approach (chapter 10): while she does not object to the notion of native mental structure, she does take issue with the idea that there is a specialized structure dedicated to the task of acquiring *language*. She argues that a native "theory-building" faculty, needed to explain the development of empirical theorizing in general, particularly scientific theorizing, can also be invoked to account for the learning of one's native language.

The cause of rationalism was not helped by its association with the doctrine of dualism, the view that mind and body are distinct substances. Descartes himself urged the connection, by arguing that the faculty responsible for language must be a faculty of an immaterial substance – matter could not account for the infinite flexibility and creativity manifest in language use (Descartes 1637/1966). Developments in logic over the last century have made vivid to us what Descartes could not conceive: that a finite, material substance can embody an infinite capacity. The idea that mental states could be admitted as scientifically respectable posits, and, correlatively, that the mind itself could be understood as a kind of computational device, owed a great deal to Chomsky's realist construal of computational linguistics. But what about the mind–body problem itself? Did it dissolve once Descartes's *Discourse* argument was answered? Hardly. Many philosophers maintain that, regardless of how useful the computational model was for the understanding of cognitive phenomena (and there is great disagreement about *this*), it still could not provide adequate accounts of two crucial features of mental life: *consciousness* and *intentionality*. A good number of contemporary philosophers are convinced that no adequate materialist account of mind can possibly accommodate these phenomena, and that, as a result, the doctrine of *physicalism* must be abandoned: the mental is an irreducible part of the world.

Probably most parties to these debates expected Chomsky to weigh in on the side of the computationalists, or, at the very least, to oppose the dualists. What surprised philosophers, however, was that Chomsky's response to the new debate about dualism was to declare the entire problem nonexistent. The sentiment that mentalistic theories must be somehow reduced to physical theories in order to be truly legitimate represents, in Chomsky's view, a kind of methodological dualism, and ought to be rejected. First, the notion of a "mind–body" problem presupposes that there is a tenable distinction between the mental and the physical. However, Chomsky argues, since Newton undermined the Cartesian theory of body by showing that "occult" forces were required in an adequate physics, mind–body dualism has lost all grounding. Second, even if reduction were possible, reduction comes in many varieties and there is little reason to believe that the contours of the reducing physical theory would be left unaffected by the process. Since Newton, Chomsky notes, 'physical' is an honorific that

signifies those areas in which we have some non-trivial degree of theoretical understanding. The relevant scientific question is whether some theory or other offers interesting descriptions and explanations. The further insistence that its primitives be couched in physical vocabulary is either vacuous (because 'physical' has no general connotation) or illegitimate (another instance of methodological dualism).

Contributors William G. Lycan, Jeffrey Poland, and Galen Strawson discuss this theme in Chomsky's writings. Lycan argues, against Chomsky, that the mind–body problem can still be formulated, without problematic appeals to outdated conceptions of matter. There is, moreover, hope that the problem can be solved, by the very theory of mind inspired by Chomsky's work – functionalism. Poland (chapter 2) takes seriously Chomsky's charge that philosophical physicalism is vacuous, but argues that a properly constrained version of the doctrine survives Chomsky's challenge. Strawson (chapter 3), too, grants much to Chomsky's critique, but maintains that a contrast can still be drawn between the "Experiential" and the "non-Experiential," in which terms he characterizes his own metaphysical position, "Realistic Monism."

Another aspect of Chomsky's work that has garnered philosophical attention concerns the ontological import of his computational models of language acquisition and use. For a long time, the debate was carried out in terms of the "psychological reality" of the structures and operations adverted to by the theory. Chomsky himself never recognized as legitimate the question whether transformations (for example) were "psychologically real" – the only sensible question, to his mind, was the empirical question whether the theory invoking transformations was *true*. If the theory was accepted, then there could be no further question whether its posits were real. But philosophers have not been so easily satisfied. Even if the computational model is accepted, there is still a question about the nature of the "representations" over which computations are defined. Are elements like "Wh" and "NP" merely formal tokens, or are they names of things like Wh-phrases and noun phrases, as the denomination "representation" suggests? Can the computations subserving the acquisition and deployment of language be characterized completely internalistically, as purely syntactic operations, or should they rather be taken (as Jerry Fodor clearly construed them in *The Language of Thought*) as fully intentional? Chomsky, once again, has thwarted philosophical expectations by insisting that his use of terms like "representation" should not be taken to commit him to representationalism in the philosopher's sense. Contributors Frances Egan and Georges Rey press Chomsky on this question. Egan (chapter 4) argues that Chomsky's theory, like that of David Marr, can be stated in purely syntactic terms, but that to understand the role of the computational mechanism in the human subject's overall behavior, it is necessary to assign interpretations to the elements of the computations. Rey (chapter 5) surveys the many places in Chomsky's work in which he appears to make crucial use of intentional notions. Rey is doubtful that such appeals are

idle, and challenges Chomsky to clarify the alternative, technical sense he claims to attach to the crucial terms.

Another way in which Chomsky's research challenges philosophical tradition concerns the notion of "language" itself. Beginning with Wittgenstein's arguments against the possibility of a "private language," many philosophers have maintained that the focal object of philosophical scrutiny must be public language, a system of shared conventions devised for the purpose of communication. Chomsky's program turns this notion on its head.

Given the aims of linguistic theory, as Chomsky construes them, the proper objects of study are the "I(nternal/individual)-languages" acquired by native speakers, rather than the public "E(xternal)-languages" used by populations. Indeed, it is only I-languages, according to Chomsky, that are apt objects for scientific study. Because the fundamental achievement for each individual is the attainment of *some* particular I-language, E-languages are, at best, radical idealizations, involving the properties of various I-languages, attained by speakers who manage to understand each other. It is not that speakers communicate because they have an E-language in common; rather, where I-languages overlap sufficiently, communication is possible. Reifying this intersection adds nothing to the observation that speakers often communicate. Contributor Ruth Garrett Millikan challenges both of these claims: first, she argues, E-languages are not arbitrary or interest-relative social constructions, but are rather real kinds of things, apt for scientific investigation; second, communication is not an accidental product of the happy coincidence of I-languages, but is rather the reflection of the natural function – the teleofunction – of the human language faculty.

A large part of contemporary philosophy of language is concerned with questions of meaning and reference. One set of questions concerns intra-linguistic relations: for example, relations of analyticity and compositionality. Another set concerns ostensible relations between words and things in the world: in virtue of what does "dog" denote dogs? What makes a sentence like "snow is white" true, or the sentence "snow is pink" false? Although it is controversial among philosophers whether any systematic theory can be given that will illuminate the words–world relation, Chomsky has consistently sided with the skeptics. Contributors Peter Ludlow, Paul Horwich, and Paul M. Pietroski discuss Chomsky's position on the possibility of referential semantics. Pietroski is the most optimistic, arguing that a semantic theory – a theory of the relation between linguistic form and meaning – is both explanatorily necessary and theoretically tractable, especially within the framework provided by minimalism. Horwich takes the opposite view, arguing that all the legitimate explanatory projects of linguistic theory can be accomplished without a referential semantics in the traditional sense. Ludlow contends that the prospects for a referential theory depend on precisely how one understands the project of explaining reference itself.

There are many, many more ways in which Chomsky's work has stimulated, provoked, and inspired philosophers and other theorists of the mind, more than

any one volume could fairly represent. The papers we've gathered here, we hope, compose a fair sample of the philosophical currents Noam Chomsky has generated.

References

Chomsky, N. 1957: *Syntactic Structures*. The Hague: Mouton.
Chomsky, N. 1959: Review of *Verbal Behavior* by B. F. Skinner. *Language*, 35 (1), 26–58.
Chomsky, N. 1965: *Aspects of the Theory of Syntax*. Cambridge, Mass.: MIT Press.
Chomsky, N. 1973: Conditions on Transformations. In S. R. Anderson and P. Kiparsky (eds.), *A Festschrift for Morris Halle*, New York: Holt, Rinehart and Winston.
Chomsky, N. 1977: *Essays on Form and Interpretation*. Amsterdam: North Holland.
Chomsky, N. 1981: *Lectures on Government and Binding*. Dordrecht: Foris.
Chomsky, N. 1983: Some Conceptual Shifts in the Study of Language. In L. Cauman, I. Levi, C. Parsons and R. Schwartz (eds.), *How Many Questions? Essays in Honor of Sidney Morgenbesser*, Indianapolis: Hackett.
Chomsky, N. 1986: *Knowledge of Language*. New York: Praeger.
Chomsky, N. 1995: *The Minimalist Program*. Cambridge, Mass.: MIT Press.
Descartes, René 1637/1966: *Discourse on Method, Part V*. Tr. E. Anscombe and P. Geach. London: Nelson Publishers.
Fodor, J. A. 1975: *The Language of Thought*. New York: Thomas Crowell Publishers.
Lightfoot, D. W. 1982: *The Language Lottery*. Cambridge, Mass.: MIT Press.

1
Chomsky on the Mind–Body Problem

WILLIAM G. LYCAN

Ontological questions are generally beside the point, hardly more than a form of harassment.

Noam Chomsky[1]

Some people say that the founding document of twentieth-century cognitive science was Chomsky's (1959) review of Skinner's *Verbal Behavior*. (Certainly it converted me.[2]) By any measure, Chomsky was a leading figure in the victory of cognitivism over behaviorism in psychology. In philosophy too, Chomsky led the attack against Quine's behaviorism regarding language and language learning.[3] Moreover, Chomsky's (1957, 1965) expressly computational view of language processing was a major inspiration for Functionalism in the philosophy of mind, as founded by Hilary Putnam and Jerry Fodor.[4]

Thus, when Chomsky turned his attention specifically and explicitly to the mind–body problem, one might naturally have expected him to grant his endorsement to Functionalism, whether or not he were to say anything further.[5] But if one had expected that, one would have been wrong. For in his writings on the mind–body problem, Chomsky has vigorously challenged several of the claims and presuppositions characteristically made by Functionalists. I shall survey some of those challenges and, as a zealous Functionalist myself, try to rebut them *seriatim*.

1 The Metaphysics of Reduction

Functionalists think of themselves as "naturalizing" the mind, as ontologically reducing (though of course not type-reducing) the mental to the physical or material. But Chomsky questions this use of "physical" and "material."

> [T]he notion of "physical world" is open and evolving. No one believes that bodies are Cartesian automata . . . or that physical systems are subject to the constraints of Cartesian mechanism, or that physics has come to an end. It may be that

> contemporary natural science already provides principles adequate for the understanding of mind. Or perhaps principles now unknown enter into the functioning of the human or animal minds, in which case the notion of "physical body" must be extended, as has often happened in the past, to incorporate entities and principles of hitherto unrecognized character. Then much of the so-called "mind–body problem" will be solved in something like the way in which the problem of the motion of the heavenly bodies was solved, by invoking principles that seemed incomprehensible or even abhorrent to the scientific imagination of an earlier generation. (Chomsky 1980: 5–6)[6]

This passage seems to contain two pertinent claims:

(a) that we have no compelling reason to expect that the mental will ever be reduced to the "physical" *as currently conceived*;

and

(b) that even if the mental is reduced only to a hitherto unconceived scientific realm that "incorporate[s] entities and principles of hitherto unrecognized character," the result will still count as a fully adequate naturalization of the mental, a locating of mind entirely within the domain of natural science.

I believe most Functionalists would resist if not balk at (a), and at least wrinkle their noses at (b).

The tension increases when we consider Chomsky's later remarks on Cartesian dualism (unpublished d). He argues that Descartes's own dualism was rejected by the end of its own century, not because it said unacceptable things about mind but because it presumed a "contact mechanics" regarding matter.

> Newton exorcised the machine, not the ghost: surprisingly, the principles of contact mechanics are false, and it is necessary to invoke what Newton called an "occult quality" to account for the simplest phenomena of nature, a fact that he and other scientists found disturbing and paradoxical . . .
>
> These moves also deprive us of any determinate notion of body or matter. The world is what it is, period. The domain of the "physical" is nothing other than what we come more or less to understand, and hope to assimilate to the core natural sciences in some way, perhaps by modifying them radically, as has often been necessary . . .[7]
>
> With the collapse of the traditional theory of "matter" or "body," metaphysical dualism becomes unstateable; similarly, such notions as "physicalism" or "eliminative materialism" lose any clear sense – unless some new notion of "physical" is offered to replace the abandoned Cartesian concept . . . (Chomsky unpublished a: 1–2)

Thus:

(c) Metaphysical dualism is "unstateable" in that it has been given no determinate sense;

and

(d) the labels "physicalism" and "eliminative materialism" are in the same fix.

Chomsky adds,

> (e) "We can continue to distinguish 'physical' or 'material' from 'mental', but recognizing that the usage is only a descriptive convenience, with no metaphysical import" (ibid.: 2).

He accepts the thesis that "human thought and action . . . are properties of organized matter," but the indeterminacy of "matter" keeps that assertion from being a substantive metaphysical claim; I suspect the only "-ism" label of which he would think it worthy is "truism."

Let us consider claims (a)–(e) in order. (a) has been made before, notoriously by Wilfrid Sellars (1962, 1965, 1971, 1981) in mitigation of his outrageous ontology of microphysical sensa. For example:

> The important thing is not to let our reflections on the developing Scientific Image of man-in-the-world be tied too closely to the current institutional and methodological structure of science, or, above all, to its current categorial structure . . . Sensa are not "material" as "matter" is construed in the context of a physics with a particulate paradigm. But, then, as has often been pointed out, the more seriously this paradigm is taken, and the more classically it is construed, the less "matter" there seems to be. (Sellars 1971: 440, 446)

As microphysics continues to get weirder and weirder, it would indeed be idiotic to insist on a nineteenth-, twentieth-, or even twenty-first-century conception of ultimate matter; it is hardly our place to second-guess the physicist. For that reason, (a) as I have stated it is plainly true, since by the time the mental is actually reduced to anything (if ever), physics may well be other than physics as conceived in the 2000s.

But it is a separate question whether the mental *could* be reduced to the physical as currently conceived, holding contemporary physics fixed *arguendo*. Chomsky strongly implies that we have no compelling reason to believe that either. And here I have two qualms. Qualm the first I share with Jack Smart: it is that changes in the physics underlying biology and chemistry should not matter in any way to the mind, however much they matter to matter.

> [I]f it be granted that the brain is essentially a nerve net, then physics enters our understanding of the mind by way of the biochemistry and biophysics of neurons. But neurons are, in Feinberg's sense, "ordinary matter." So whatever revolutionary changes occur in physics, there will be no important lesson for the mind–body problem or for the philosophy of biology generally . . . The situation is not like that in the eighteenth century, when physics was mainly mechanics, and needed to be supplemented by the theory of electricity and magnetism, even for the purpose of understanding the behaviour of ordinary bulk matter. (Smart 1978: 340)

Consider: if we were to take a collection of molecules, assumed to have just the properties they are thought to have at present, we could in principle build a version of a human organism whose behavior, including verbal behavior, would

be like ours under appropriate circumstances. Would such a simulacrum not have a mind? *Maybe* not, but we would have every reason to think it did and no reason I can anticipate for denying that.[8]

The second qualm can be put in the form of a dilemma. If reduction of mind to "matter" requires a reconception and expansion of physics to "incorporate entities and principles of hitherto unrecognized character," then either those entities and principles will be localized where we now take minds to be, viz., in central nervous systems, or like other entities and principles of fundamental physics they will pervade nature. But (1), while perfectly coherent, the former hypothesis is loony; whyever would the new entities occur and principles apply only in regions of spacetime shaped like the heads of sentient creatures, or be specific to neural tissue, which regions and tissue are specified only at a level of organization far higher than that of microphysics? Why would their occurrence depend on their *so much* larger molecular environment? Possible, certainly, but grotesque. (2) The second hypothesis, that the new entities and principles will pervade nature, is far more likely, but it encourages panpsychism. If they are posited out of the need to reduce or explain mental phenomena, and they occur throughout nature, then so, presumably, do the mental phenomena. Again possible, but fanciful.

Of course these arguments do not prove Chomsky wrong. But they lend some support to the idea that reduction of mind to matter, if possible at all, does not in fact wait upon any expansion of physics.[9]

Claim (b), that even if the mental is reduced only to an expanded physics the result will still count as a fully adequate naturalization, confronts the foregoing dilemma also. A naturalization that leaves us either with shy new entities and principles that hide in brains, or with panpsychism, may be fully naturalistic but is not a fully *satisfactory* naturalization. Also, if the new entities and principles are posited out of the need to reduce or explain mental phenomena, it follows that there is something scientifically very *special* about the mind. Perhaps there is – and of course the mind is macroscopically very special – but wholehearted naturalists would have hoped otherwise. As I believe I once put it, there is nothing the mind does that calculators and automobile engines do not do, albeit on a smaller scale.

Claims (c) and (d), of the "unstateability" of mind–body theories on account of the underspecification of "physical," must be taken seriously. As Sellars said, there does not seem to be much "matter" any more, and characterizations of "the physical" should not be lashed to our own century's microphysics. Yet one feels there is still an important distinction to be drawn between the sorts of things Descartes or Hume thought minds were and the sorts of things present-day materialists think they are. The difficulty, of course, is that if we do not draw that distinction in terms of current science, we have to draw it *a priori*, which we cannot very responsibly do either.

Yet here are two ways of putting the materialist or physicalist thesis that are, I

believe, good enough to work with and do distinguish that thesis from the views of Descartes and Hume. The first is comparative: "Creatures with minds are made entirely of the same ultimate components as are ordinary inanimate objects, and their properties are determined by the ways in which those components are arranged and related to external things." Descartes could not accept that claim. (The case of Hume is more problematic, since strictly he did not believe in ordinary inanimate objects at all; but so far as he regarded such "objects" as collections of impressions and ideas, he had to insist that a mind is a very different sort of bundle of impressions and ideas.) For that matter, the materialist claim as I have stated it further rules out the view mentioned above, that the mind involves new entities and principles that are confined within central nervous systems and not found in computers and auto engines; I am happy with that result, for although the parochialist view would still be broadly naturalistic, it is not intuitively materialist, but is a form of what is often called "property dualism."

A second characterization of materialism, perhaps not perfectly coincident with the first, is Sellars's, in terms of spacetime. The physical, we may say for purposes of the mind–body problem, is the spatiotemporal, meaning that to be physical is to be located within the same spacetime as are MIT, North Carolina, the Andromeda galaxy and the Sydney Harbour Bridge. (A qualification may be needed, in case it should turn out under some version of relativity theory or cosmology that there are multiple spacetimes; then we would have to speak of spacetimes similar to ours in such-and-such topological ways.) Descartes and Hume denied that mental states and events occurred in ordinary spacetime, even if there is such a thing as "phenomenal space." (Recall that Princess Elizabeth scored a telling hit against Descartes, when he had attempted to solve the interaction problem by comparing mental causation to gravity: gravitational fields are not physical in the sense that billiard balls are, but unlike Cartesian egos they are thoroughly spatiotemporal).[10] I do not say that the spacetime criterion is philosophically unproblematic, but it saves mind–body dualism, materialism and eliminativism from meaninglessness.[11]

Incidentally, I think we should resist Chomsky's suggestion that "the domain of the 'physical' is nothing other than what we come more or less to understand, and hope to assimilate to the core natural sciences in some way," for that characterization is entirely epistemic, couched in terms of propositional attitudes; and whatever "physical" was supposed to mean exactly, it has always been an ontological term rather than an epistemic one.

Finally, claim (e): that the dichotomy of "mental" vs. "physical" or "material" is of no metaphysical import. There is at least a grain of truth to this, in that naive uses of "mental" to *mean* something nonphysical are misguided. For example, medical science should stop distinguishing between mental illnesses that "have a neurobiological basis" (and so are really physical illnesses potentially treatable with drugs or surgery and carry no social stigma) and mental illnesses

that by implication have no such basis; that distinction really is no more than epistemic, since every mental illness in fact has some neurobiological basis or other, known or unknown. But Chomsky overstates the case considerably in saying that the mind–body distinction has no metaphysical import.

Actually there are two distinctions one might mean in talking of "mental" vs. "physical," and I am unsure which of them Chomsky has in mind here. First, if one is a materialist, one still distinguishes between the physical things that are also mental and those that are *merely* physical. That distinction does not have *a lot* of metaphysical import, but it has some. In particular, physical things that represent the world have whole ranges of interesting properties that merely physical things do not, such as truth-value and entailment relations. Organisms that have (however physical) minds are capable of agency, which affords them yet another wide range of interesting features unavailable to the merely physical, and so on. (Perhaps that is not metaphysical enough, or important enough, for Chomsky.)

The other "mental"/"physical" distinction is deeper, and does not presuppose the truth of materialism in any sense. *Au contraire*, it is the distinction between putative things that are mental in ways that could not be accommodated by any readily imaginable expansion of physics and things that are either not mental at all or mental but also physical in some generous sense. Now, it may be thought that Chomsky would object to, indeed has just been vigorously objecting to, that very distinction; is it not precisely what he has been inveighing against, a second-guessing of physics and such?

No, because it is clear that there are metaphysical conceptions of mind that, even for Chomsky, are too non- or anti-physical on any permissible reading of "physical." He holds, after all, that human thought and action are properties of organized matter, which doctrine taken in any sense of "matter" has some substantive metaphysical implications. Moreover, he even himself quotes some of Descartes's metaphysics with which he certainly disagrees: that "there is within us but one soul, and this soul has not in itself any diversity of parts."[12] That Descartes's thesis is wrong, that whatever soul there is is an *organization* or arrangement of matter and so must have parts, is an ontological thesis and still (I myself regret to say) a somewhat controversial one even in current mainstream American philosophy.[13] Or consider the putative indestructibility of the soul, defended by Socrates. If, as Chomsky believes, soul or mind is really an arrangement of matter in any sense of "matter," it is presumably not indestructible (for the matter could be rearranged in, or disorganized into, a non-mind-realizing state). But if minds or souls are "mental" in Socrates' sense, then they – and we – are immortal and will survive our bodily death, even unto eternity. Is that metaphysical import enough?

2 The Ordo Cognoscendi

Once Chomsky has set aside Functionalists' Procrustean concern with reduction of mind to "matter" while accepting the thesis that human thought and action are properties of matter, he says, "the next question is how organized matter can have these properties"; this is "the new version of the unification problem" (Chomsky unpublished d: 2). And the "mode of connection" between a fairly abstract level of description of a system and other such levels "may involve reduction, expansion, or modification of several levels: the unification problem may take any course" (ibid.: 12). But Chomsky believes philosophical Functionalists have approached this unification problem in an unmotivated, perhaps arbitrary way:

> In the case of language and other cognitive functions, it is common to try to relieve the fear [unpublished c: 14: "the sense"] that something is amiss with such slogans as, "The mental is the neurophysiological at a higher level," taken as a characterization of the mental [unpublished c: 14: "as a kind of definition"] . . . [This is a] strange move . . . From a naturalistic perspective there are just various ways of studying the world, with results of varying firmness and reach, and a long-term goal of unification. We assess the credibility of assumptions at various levels on the basis of explanatory success. In the case of language, the most credible assumptions about what the brain is and what it does are those of the computational theories. We assume, essentially on faith, that there is also an account in terms of atoms and molecules, though without expecting the operative principles and structures to be identifiable at these levels. And with a much larger leap of faith, we also tend to assume that there is an account in neurological rather than vascular terms, though a look at the brain reveals huge amounts of blood [unpublished c: 14: "reveals blood as well as neurons"] . . .
>
> The familiar slogan about the mental and the neurophysiological has the matter backwards: it should not be taken as a characterization of the mental, [unpublished c omits "has . . . it"] but rather as a hypothesis [unpublished c: 14: "tentative hypothesis"] about neurophysiology: perhaps the neurophysiological is the mental at a "lower" level, perhaps not. As of now, we have more reason to feel secure about the mental than about the neurophysiological. (Chomsky unpublished d: 12–13)

So:

(f) we should reject or at least not accept the Functionalist characterization of the mental as the neurophysiological at a higher level of organization or more abstract level of description.

Yes, of course from a naturalistic perspective there are just various ways of studying the world, their respective vocabularies, and the long-term goal of unification. But the rest of what Chomsky argues here has to do with inquiry and with the state of our knowledge, not about the respective natures of mind and brain, so it is not obvious how he means thereby to establish (f). Perhaps the

argument is this: "We know that mental states of kind K are real. Although we know that neurophysiological states are real also and we tend to assume that they realize the mental states, there is no actual evidence indicating any such relation between any neurological state, however abstractly described, and kind K – the assumption is little more than an article of faith. Therefore one cannot fairly build a *basic characterization* or definition of K on its putative type-identity with any functional or other abstract kind."

There is some historical accuracy to Chomsky's suggestion that Functionalism stems from "the fear [unpublished c: 14: 'the sense'] that something is amiss." For we did back into Functionalism; the main reason we believe it is that it is the least bad way of not being a Cartesian dualist. Behaviorism proved inadequate in ways that were happily overcome by the Identity Theory propounded by Place and Smart, but then Putnam noticed the Identity Theory's species chauvinism and suggested characterizing the relevant neurophysiological states abstractly rather than neurophysiologically. We (at least those of us who insist that mental states are real) have been Functionalists ever since.

But if this provenance is slightly pusillanimous, the corresponding argument is not a bad one: for well-known reasons, most notably the interaction problem, Cartesian dualism is repugnant; behaviorism is superior, but inadequate in ways that are uniformly and dramatically improved upon by the Identity Theory; the Identity Theory is excellent but has one troubling flaw that is repaired by Functionalism. That is a fairly strong letter of recommendation for Functionalism.

The argument does ignore at least two alternatives: eliminative materialism, and Davidson's (1970, 1974) "Anomalous Monism," the view that although all mental state-tokens are neurophysiological state-tokens, there is no correct type-identification of the mental with anything else. Each of those competitors emerged after the nervous birth of Functionalism, and each would have to be confronted by a proponent of the foregoing argument. But each has been so confronted,[14] and after due consideration the vast majority of philosophers of mind have chosen to stick by Functionalism.

And there are further, more direct arguments for Functionalism. For one thing, the theory shows how mental types can be natural kinds without being human-biological kinds. For another, the computer analogy is a powerful attraction (but see section 3 below). In addition, Armstrong (1968) and others have argued that mental *concepts* are even functional concepts to begin with.

Finally and despite that last point, we should not be misled by Chomsky's phrase, "taken as a characterization of the mental [unpublished c: 14: 'as a kind of definition']." The Functionalist is far from saying either that mental terms *mean* functional things or that the mental should initially be picked out in functional terms. Epistemically, the Functionalist proposal is like the Identity Theory, in being a metaphysical *speculation* held, for reasons, to be more credible than other such speculations, and it must continue to take its chances against competitors.

3 Computer Models

Unexpectedly, Chomsky distances his own computationalism somewhat from the computation actually performed by computers.

> Another common way to relieve uneasiness about computational theories [in psychology] is to invoke computer models to show that we have robust, hard-headed instances of the kind: psychology studies software problems. That is, again, a strange move, remote from the naturalist perspective. Artifacts raise all kinds of questions that do not arise in the case of natural objects. Whether some object is a key or a computer depends on designer's intent, standard use, mode of interpretation, and so on . . . There is no "natural kind" or "normal case" in the study of keys or computers; virtually anything could be one. The hardware–software distinction is a matter of interpretation, not simply of physical structure . . . Such questions do not arise in the discussion of organic molecules, nematodes, the language faculty, or other aspects of the natural world. (Chomsky unpublished d: 13–14)

(g) Functionalists have erred in appealing to the computer analogy, because computers "compute" only in a different and more suspect sense from that in which biological organisms do.

This is refreshing. I think most people see *teleology* as paradigmatically located in artifacts, and the notion of natural teleology as highly dubious (perhaps as requiring belief in a superhuman artificer). To the contrary, Chomsky seems to see biological teleology as straightforward and that of artifacts as vexed.

I am somewhere in between. I love natural teleology (see Lycan 1987) and agree that it is ontologically prior to artifactual teleology; I also share Chomsky's misgivings about the ascription of functions to artifacts and especially about the "hardware–software" distinction (see again, even more urgently, Lycan 1987: ch. 4). But, sad to say, natural teleology is hardly unproblematic. There is a huge literature devoted to its explication and vindication.[15] Note, vindication is needed, not just explication, because of the understandable suspicion that literal ascriptions of natural function really do presuppose a superhuman designer.

4 Eliminative Materialism (and Connectionism)

Functionalists of course reject eliminative materialism, since in type-identifying the mental with the functional we affirm the reality of each. But we have two things importantly in common with the eliminativists: what Adam Morton (1980) called the "'Theory' theory" – the claim that mental expressions are the theoretical terms of a folk or commonsense theory[16] – and accordingly the belief that even the reality of the mental is an empirical question and needs vindication of some substantive sort. This agreement spills over into a willingness to play

devil's advocate for the eliminativist from time to time, so I shall do a bit of that in this section.

Here is Chomsky on eliminativism:

> The idea that cognitive psychologists should drop the inquiry into computational systems [unpublished c: 14: "rule systems"] in favor of the study of neurophysiology seems about as reasonable [unpublished c: 14: "reasonable today"] as the [unpublished c: 14: "as a"] proposal that embryologists should keep to [unpublished c: 14: "drop their inquiries in favor of"] superstring theory; arguably less so, given the status of the theories. As for eliminative materialism, the very doctrine remains a mystery until some account is given of the bounds of the material, and given that account, some reason why one should take it seriously or care if successful theories lie beyond its bounds. (Chomsky unpublished d: 13)

(h) Eliminative materialism, far from being recommended by serious scientific and scientistic reflection, is methodologically just silly.

The proposal that "cognitive psychologists should drop the inquiry into computational systems [unpublished c: 14: 'rule systems'] in favor of the study of neurophysiology" is more demanding than the official eliminativist thesis, and the latter is much stronger than any claim actually made by self-styled eliminativists in person, but the eliminativists themselves have been deliberately unclear about those distinctions. The official thesis is that no mental ascription has ever been true (or less ambitiously, no propositional-attitude ascription, or no sensation ascription, or whatever, has ever been true – I shall hereafter take such restrictings for granted); nothing has ever been in a mental state. But it is rare for anyone to have asserted that outrageous thesis flatly. The claim is normally just that the official thesis is a strong possibility or at least must be taken seriously. And that claim, Chomsky would be right to insist, does not entail or even suggest that cognitive psychologists should drop their inquiry into computational systems.

But what if the official thesis is or were true, as is at least possible? Then it still would not follow that cognitive psychologists should drop their computational inquiries. For even if our ordinary folk mental concepts are ultimately unexemplified and our Functionalist type-identifications are empty, the brain still performs computations and our behavior can be explained in such terms. To motivate dropping computational psychology, the eliminativist would have to eliminate not just commonsense mental items but all computational ones as well, a more taxing project. Chomsky is surely right to find this idea methodologically outlandish. (However, I shall return to this theme shortly.)

I would put in a word in favor of one eliminativist theme: eliminativists are impressed by a disparity in robustness between mental states (especially propositional attitudes) and neurological states. At the current stage of science, there is simply no denying that brains are full of neurons organized into fibers, nets, and larger systems, and that behavior is caused by fantastically complicated electro-

chemical activity in the nervous system. It is less obvious that mental states such as beliefs and desires occupy real physical locations inside our heads and interact causally with each other in order to produce behavior in turn. Forced to bet on one or the other, neural states as inner causes of behavior and mental states as the same, I too would have to choose neurophysiology. (As a Functionalist, of course, I do myself believe that mental states are internal physical states that play characteristic causal roles. But that is a controversial philosophical view, not a plain fact like the fact that coordinated neural firings cause behavior.) Note, eliminativists often motivate their view by hinting or pretending that we do have to choose between commonsense mental states and neurophysiology; that is one place where they go wrong.

Turning to Chomsky's claim that "the very doctrine remains a mystery," I want to register a solid objection, over and above the rebuttals I have made to theses (c) and (d) in section 1. He says that eliminativism remains a mystery "until some account is given of the bounds of the material"; this is evidently a reprise of his complaint that the notions of "matter" and "the physical" are undefined. But here the eliminativist does better than the various psychological realists, for strictly, s/he need not presuppose anything about "matter" or "the physical" but can continue to say merely that no mental ascription has ever been true because nothing has ever been in a mental state. It may be that to *defend* that official thesis, the eliminativist would have to allude to "matter," but that is hardly obvious, and in any case, the thesis itself simply does not presuppose any such notion, so in this case Chomsky's criticism just goes wide.

I agreed that it would be silly to advocate giving up computational ideas and dropping computational inquiry. But there is a weaker suggestion that is not silly and that has been seriously defended by opponents of Chomsky's particular computational approach to natural language. It is the connectionist proposal that we abandon Chomsky's "rules and representations" paradigm and investigate language by means of connectionist modeling, training up networks to perform grammatical feats such as inflection formation, prediction of succeeding words, and detection of well-formed strings.[17] The argument here abandons neither folk psychology nor the idea of computation, but urges a different, more brainlike computational architecture.

Chomsky addresses that proposal as such:

> Perhaps it will turn out that connectionist models are more adequate for a system of 10^{11} neurons than one with 300, but one awaits an argument. The discussion of what it would imply about computational systems [unpublished c: 15: "rule systems"] if such theories were to become available in some distant future [unpublished c: 15 omits "in some distant future"] is as interesting as a debate over what it would mean for embryology if unstructured systems, now entirely unknown [unpublished c: 14: "if it were shown that some unstructured system"], could achieve the explanatory power that biologists seek in terms of their complex notions. (Chomsky unpublished d: 13)

At the very least:

(i) There is (at present) no reason to think that connectionist models of cognitive capacities such as language processing are superior to the classically computational models already being built.

This seems fairly clear. Though connectionist models have done better at certain sorting tasks than has classical AI at those same tasks, e.g. at discriminating sonar echoes produced by undersea mines from those produced by rocks,[18] I know of no evidence that connectionist models of syntax or semantics will ever do better than classical rules-and-representations models.

Also, two caveats: first, even if a network were to start from scratch, with the customary random connection strengths and levels of unit activation, and learn English syntax, in much the way that NETtalk[19] very impressively learns English phonology, it would very likely do this by coming to implement a classical syntax recoverable at a higher level of functional abstraction – in much the way that a mature NETtalk comes to implement all the *phonemes* and phoneme classifications of English phonology.[20] (One must avoid the fallacy of inferring properties of a trained network from properties of its training. That a network is given no classical rules and representations prior to or during training does not entail that it does not grow representations and rules on its own.)

Second, one must be a bit careful about impressionistic claims of connectionist superiority even in domains that are widely thought impervious to classical AI. McLaughlin and Warfield (1994) have argued that contrary to advertising and to widely shared assumptions, connectionist networks have not historically proved to be better than classical programs either at modeling standard pattern-recognition tasks or at learning to perform those tasks. In particular, comparative studies as between connectionist networks and "Top-Down Inductive Decision Tree" algorithms have revealed no such superiority.[21]

5 Mysteries

I close with some remarks on a side issue raised by Chomsky, not because he relates it particularly to the mind–body problem, but because others have done so citing his authority, and because I believe his general idea is philosophically important. Methodologically, Chomsky (1975, 1980, unpublished a, d) distinguishes between mere "problems" and what he calls "mysteries," "the former being questions that we seem to be able to formulate in ways that allow us to proceed with serious inquiry and possibly to attain a degree of understanding, the latter including questions that seem to elude our grasp, perhaps because we are as ill-equipped to deal with them as a rat is with a prime number maze" (Chomsky unpublished a: 41). He suggests, more dramatically, that some of the "mysteries" may be permanently and systematically intractable for us because of innate structure in the human mind.

> The human mind is a biologically given system with certain powers and limits . . . The fact that "admissible hypotheses" are available to this specific biological system accounts for its ability to construct rich and complex explanatory theories. But the same properties of mind that provide admissible hypotheses may well exclude other successful theories as unintelligible to humans. Some theories might simply not be among the admissible hypotheses determined by the specific properties of mind that adapt us "to imagining correct theories of some kinds," though these theories might be accessible to a differently organized intelligence. (Chomsky 1975: 15–56)

> [T]he naturalistic temper . . . takes for granted that humans are part of the natural world, not angels, and will therefore have capacities with specific scope and limits, determined by their special structure. For a rat, some questions are problems that it can solve, others are mysteries that lie beyond its cognitive reach; the same should be true of humans, and to first approximation, that seems a fair conclusion. What we call "natural science" is a kind of chance convergence between aspects of the world and properties of the human mind/brain, which has allowed some rays of light to penetrate the general obscurity, excluding, it seems, central domains of the "mental." (Chomsky unpublished d: 3)

Strong stuff, but plausible. Chomsky makes no definite pronouncement on what "central domains of the 'mental'" he thinks may harbor mysteries. He does speculate that one such domain may be that of "will and choice" (Chomsky 1980: 7): as Descartes said, we human beings are not "compelled" to perform most of the actions we do, as lower animals are compelled, but only "incited and inclined." The human power of choosing, "[t]his essential capacity of the human to act as a 'free agent', able to choose to follow or to disregard 'the rule that is prescribed to it' by nature," is a good candidate for the status of mystery. As another candidate he adds what may be a related point: "Human action is coherent and appropriate, but appropriateness to situations must be sharply distinguished from the causal effect of situations and internal states" (Chomsky unpublished a: 41).

Colin McGinn (1989, 1994) has picked up Chomsky's theme of mysteries, and argued that the mind–body problem is insoluble because, in addition to free will, the qualitative character of experience is a mystery. (To my knowledge, Chomsky nowhere mentions phenomenal experience, though he cites McGinn (Chomsky unpublished d: 3) without evident disapproval.)

This is not the place for me to expound and assess McGinn's arguments.[22] Let me just record my opinion that they are unsound, and add a hearty plug for my own solutions to problems of free will (Lycan 1987: ch. 9) and the qualitative character of experience (ibid.: ch. 8, 1996a). My solutions may be inaccurate, even squarely false.[23] But *if* correct they *would* solve the problems at issue; they are good answers to the corresponding "how-possibly" questions. And I think that suffices to show that the problems are only problems, and not mysteries.[24]

However, I offer two further candidate groups: (1) central questions in the philosophy of time, and (2) the puzzles of predication. It seems likely that

because of the complex ways in which we are ourselves located in time, we will never gain an objective philosophical and conceptual understanding of time (even though some of us can understand the mathematical representation of time employed by contemporary physics). It seems even likelier that subject-predicate structure is somehow fundamental to our thought, and we will never be able to look under or into it and figure out its ontological workings. I stopped thinking about time in around 1975, and I have tried fairly hard never to think about predication at all. If those topics are indeed mysteries in Chomsky's sense, then he has given me not only an excuse but a good reason for my omissions, and I am very grateful.

Chomsky's reply: pp. 255–63.

Notes

1 Chomsky (unpublished c: 14), slightly out of context.

2 However, see also Craik 1943; Piaget 1954; Bruner, Goodnow and Austin 1956; Miller 1956; Broadbent 1958; Newell, Shaw and Simon 1958; and Miller, Galanter and Pribram 1960.

This is a good opportunity for me to thank Professor Chomsky for his kindness thirty-plus years ago, when I was an Amherst undergraduate writing an honors thesis on the then new discipline of mathematical linguistics. (My advisor, Robert Tredwell, a Frederic Fitch Ph.D., had pressed a copy of *Syntactic Structures* upon me, and two days later I had abandoned mathematics for the philosophy of language.) I contacted Chomsky and asked for further materials; he sent an entire microfilm copy of his then unpublished 1955 monsterpiece, *The Logical Structure of Linguistic Theory*.

3 But for similar criticisms, see also Rosenberg 1967 and Landesman 1970. Quine (1968) replies to Chomsky, and Quine (1970, 1972) turns to address syntax; Chomsky (1975, 1980, unpublished b, and elsewhere) rejoins.

4 Putnam 1960; Fodor 1968. I am using the label "Functionalism" with its original meaning (and with a proud capital "F"), as naming the *a posteriori* scientistic speculation that mental states and events are functional states and events in either a computational or a teleological systems-theoretic sense of "functional." This is the doctrine called "Psychofunctionalism" by Block (1978), as opposed to the *a priori* commonsense Causal Theory rooted in Sellars and developed by David Armstrong and David Lewis, which Block viciously neologized as "Functionalism" and which has since come to be called, only slightly more appropriately, "Analytical Functionalism." (It is analytical – and thereby incurs some convincing objections – but there is nothing functionalist about it in either the computational or the teleological sense.)

5 There are encouraging passages, too, e.g.:

> When I use such terms as "mind," "mental representation," "mental computation," and the like, I am keeping to the level of abstract characterization of the properties of certain physical mechanisms, as yet almost entirely unknown. There is no further

> ontological import to such references to mind or mental representations and acts. (Chomsky 1980: 5)
>
> [W]e may think of the study of mental faculties as actually being a study of the body – specifically the brain – conducted at a certain level of abstraction. (ibid.: 31)

6 There will be many such long quotations in the expository sections of this paper, as Chomsky speaks so well for himself.

7 A stronger claim is made in Chomsky (unpublished a: 38–9):

> [The terms] "body" and "the physical world" refer to whatever there is, all of which we try to understand as best we can and to integrate into a coherent theoretical system that we call the natural sciences . . . If it were shown that the properties of the world fall into two disconnected domains, then we would, I suppose, say that that is the nature of the physical world, nothing more, just as if the world of matter and anti-matter were to prove unrelated.

(Yet on the same page (39) of the same article, Chomsky repudiates Thomas Nagel's (1979) ascription to him of "the prediction that mental phenomena will eventually come to be counted as physical, once we understand them systematically – even if they are not reduced to terms already admitted as physical." Chomsky's objection is that he makes no such *prediction*, but holds only that *if* mental phenomena should come to be understood systematically, they would or should then be counted as physical. The "if" is a big one; see section 5 below.)

8 I made a similar argument against Sellars in Lycan 1987: 103, in terms of supervenience. Roughly: since molecules are made of atoms, molecular facts supervene on already familiar microphysical facts; biological facts supervene on molecular facts plus ordinary macroscopic surroundings; psychological facts supervene on biological facts plus ordinary macroscopic surroundings; so, given transitivity of supervenience, psychological facts supervene on microphysical facts of the sort that are already fairly well known.

9 Actually there is a third qualm, but I am not equipped to enforce it by means of argument: I have heard at least one respected physicist aver that "physics is finished," meaning that even microphysics is already empirically adequate and its *physical* ontology, its ontology of substances, is reasonably well understood; the remaining projects of microphysics – positing superstrings, constructing a unified field theory and the like – are only matters of interpreting and mathematizing the physical ontology. If that is so, then there is no reason to think that physics will expand its ontology in so fundamental a way as to afford a reduction of the mental that was not already available. But I am unqualified to judge whether it is so.

10 See her letter from The Hague, June 10–20, 1643.

11 Anent eliminativism, actually, I believe claim (d) is just mistaken; see section 4 below.

12 Chomsky 1980: 30; the quotations are from Article XLVII of *The Passions of the Soul* and *Meditation* VI.

13 See, e.g., Chisholm 1976.

14 For a review of the most compelling arguments against eliminative materialism, see Lycan 1996b; for a sustained defense of Functionalism as superior to Anomalous Monism, see Lycan 1981.

15 To hit only a few high spots: Wimsatt 1972; Wright 1973; Cummins 1975; Millikan 1984; Bigelow and Pargetter 1987; Neander 1991; Godfrey-Smith 1994.
16 On which at length, see Lycan 1996b.
17 See, e.g., Rumelhart and McClelland 1986; McClelland and Kawamoto 1986; Elman 1991, 1992. There are hybrid and compromise proposals, such as Horgan and Tienson's (1989, 1996) format of "representations without rules."
18 Gorman and Sejnowski 1988.
19 Sejnowski and Rosenberg 1986; Rosenberg and Sejnowski 1987.
20 This is defended in Lycan 1991.
21 See, particularly, Shavlik, Mooney and Towell 1991.
22 Regarding the one directed at phenomenal experience, however, see Flanagan 1992.
23 In fact, they are no such thing.
24 But that is a bit hasty, because in Chomsky's vocabulary a "problem" is a question which has been well enough formulated that we can "proceed with serious inquiry." If by that phrase he means, proceed with serious *empirical* enquiry, I am not ready to claim such for my solutions to the mind–body problem or for almost anyone's philosophical theory on any subject. Perhaps there is a middle category, of, say, philosophical problems as opposed to scientific problems.

References

Armstrong, D. M. 1968: *A Materialist Theory of the Mind*. London: Routledge and Kegan Paul.

Bigelow, J. and Pargetter, R. 1987: Functions. *Journal of Philosophy*, 84, 181–97.

Block, N. J. 1978: Troubles with Functionalism. In W. Savage (ed.), *Perception and Cognition*, Minnesota Studies in the Philosophy of Science, Vol. IX, Minneapolis: University of Minnesota Press.

Broadbent, D. E. 1958: *Perception and Communication*. New York: Pergamon Press.

Bruner, J. S., Goodnow, J., and Austin, G. A., 1956: *A Study of Thinking*. New York: Wiley.

Chisholm, R. M. 1976: *Person and Object*. London: George Allen and Unwin.

Chomsky, N. 1957: *Syntactic Structures*. The Hague: Mouton.

Chomsky, N. 1959: Review of *Verbal Behavior* by B. F. Skinner. *Language*, 35, 26–58.

Chomsky, N. 1965: *Aspects of the Theory of Syntax*. Cambridge, Mass.: MIT Press.

Chomsky, N. 1968: Quine's Empirical Assumptions. *Synthèse*, 19, 53–68.

Chomsky, N. 1975: *Reflections on Language*. New York: Pantheon Books.

Chomsky, N. 1980: *Rules and Representations*. New York: Columbia University Press.

Chomsky, N. unpublished a: Linguistics and Cognitive Science: Problems and Mysteries. (An edited version of this paper was published in A. Kasher (ed.), *The Chomskian Turn*. Oxford: Blackwell, 1991.)

Chomsky, N. unpublished b: Language and Interpretation: Philosophical Reflections and Empirical Inquiry. (An edited version of this paper was published in N. Chomsky. *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press, 2000.)

Chomsky, N. unpublished c: Language from an Internalist Perspective. (An edited version of this paper was published in N. Chomsky. *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press, 2000.)

Chomsky, N. unpublished d: Naturalism and Dualism in the Study of Language and Mind. (An edited version of this paper was published in N. Chomsky. *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press, 2000.)

Craik, K. J. W. 1943: *The Nature of Explanation*. Cambridge: Cambridge University Press.

Cummins, R. 1975: Functional Analysis. *Journal of Philosophy*, 72, 741–65.

Davidson, D. 1970: Mental Events. In L. Foster and J. W. Swanson (eds.), *Experience and Theory*, Amherst, Mass.: University of Massachusetts Press.

Davidson, D. 1974: Psychology as Philosophy. In S. C. Brown (ed.), *Philosophy of Psychology*, London: Macmillan.

Elman, J. 1991: Distributed Representations, Simple Recurrent Networks, and Grammatical Structure. *Machine Learning*, 7, 195–225.

Elman, J. 1992: Grammatical Structure and Distributed Representations. In S. Davis (ed.), *Connectionism: Theory and Practice*. Oxford: Oxford University Press.

Flanagan, O. 1992: *Consciousness Reconsidered*. Cambridge, Mass.: Bradford Books/MIT Press.

Fodor, J. A. 1968: *Psychological Explanation*. New York: Random House.

Godfrey-Smith, P. 1994: A Modern History Theory of Functions. *Noûs*, 28, 344–62.

Gorman, R. P. and Sejnowski, T. J. 1988: Learned Classification of Sonar Targets Using a Massively-Parallel Network. *IEEE Transactions: Acoustics, Speech, and Signal Processing*, 36, 1,135–40.

Horgan, T. and Tienson, J. 1989: Representations without Rules. *Philosophical Topics*, 17, 147–74.

Horgan, T. and Tienson, J. 1996: *Connectionism and the Philosophy of Psychology*. Cambridge, Mass.: Bradford Books/MIT Press.

Landesman, C. 1970: Skepticism about Meaning. *Australasian Journal of Philosophy*, 48, 320–37.

Lycan, W. G. 1981: Psychological Laws. *Philosophical Topics*, 12, 9–38.

Lycan, W. G. 1987: *Consciousness*. Cambridge, Mass.: Bradford Books/MIT Press.

Lycan, W. G. 1991: Homuncular Functionalism Meets PDP. In W. Ramsey, S. P. Stich, and D. Rumelhart (eds.), *Philosophy and Connectionist Theory*, Hillsdale, NJ: Lawrence Erlbaum.

Lycan, W. G. 1996a: *Consciousness and Experience*. Cambridge, Mass.: Bradford Books/MIT Press.

Lycan, W. G. 1996b: Folk Psychology and Its Liabilities. In M. Carrier and P. K. Machamer (eds.), *Philosophy and the Sciences of the Mind*, Pittsburgh: University of Pittsburgh Press.

McClelland, J. L. and Kawamoto, A. H. 1986: Mechanisms of Sentence Processing: Assigning Roles to Constituents of Sentences. In McClelland and Rumelhart 1986.

McClelland, J. L. and Rumelhart, D. (eds.) 1986: *Parallel Distributed Processing 2*. Cambridge, Mass.: MIT Press.

McGinn, C. 1989: Can We Solve the Mind–Body Problem? *Mind*, 98, 349–66.

McGinn, C. 1994: The Problem of Philosophy. *Philosophical Studies*, 76, 133–56.

McLaughlin, B. and Warfield, T. A. 1994: The Allure of Connectionism Reexamined. *Synthèse*, 101, 365–400.

Miller, G. A. 1956: The Magical Number Seven, Plus or Minus Two. *Psychological Review*, 63, 81–97.

Miller, G. A., Galanter, E., and Pribram, K. 1960: *Plans and the Structure of Behavior.* New York: Holt, Rinehart, and Winston.

Millikan, R. G. 1984: *Language, Thought, and Other Biological Categories.* Cambridge, Mass.: Bradford Books/MIT Press.

Morton, A. 1980: *Frames of Mind.* Oxford: Oxford University Press.

Nagel, T. 1979: Subjective and Objective. In *Mortal Questions*, Cambridge: Cambridge University Press.

Neander, K. 1991: The Teleological Notion of "Function." *Australasian Journal of Philosophy*, 69, 454–68.

Newell, A., Shaw, J. C. and Simon, H. A. 1958: Elements of a Theory of Human Problem-Solving. *Psychological Review*, 65, 151–66.

Piaget, J. 1954: *The Construction of Reality in the Child*. New York: Basic Books.

Putnam, H. 1960: Minds and Machines. In S. Hook (ed.), *Dimensions of Mind*, New York: NYU Press.

Quine, W. V. 1968: Reply to Chomsky. *Synthèse*, 19, 274–83.

Quine, W. V. 1970: Philosophical Progress in Language Theory. *Metaphilosophy*, 1, 2–19.

Quine, W. V. 1972: Methodological Reflections on Current Linguistic Theory. In D. Davidson and G. Harman (eds.), *Semantics of Natural Language*, Dordrecht: D. Reidel.

Rosenberg, C. R. and Sejnowski, T. J. 1987: Parallel Networks that Learn to Pronounce English Text. *Complex Systems*, 1, 145–68.

Rosenberg, J. 1967: Synonymy and the Epistemology of Linguistics. *Inquiry*, 10, 405–20.

Rumelhart, D. and McClelland, J. L. 1986: On Learning the Past Tenses of English Verbs. In McClelland and Rumelhart 1986.

Sejnowski, T. J. and Rosenberg, C. R. 1986: NETtalk: A Parallel Network that Learns to Read Aloud. Baltimore: Johns Hopkins University EE and CS Technical Report JHU/EECS–86/01.

Sellars, W. 1962: Philosophy and the Scientific Image of Man. In R. Colodny (ed.), *Frontiers of Science and Philosophy*, Pittsburgh: University of Pittsburgh Press.

Sellars, W. 1965: The Identity Approach to the Mind–Body Problem. *Review of Metaphysics*, 18, 430–51.

Sellars, W. 1971: Seeing, Sense Impressions, and Sensa: A Reply to Cornman. *Review of Metaphysics*, 24, 391–447.

Sellars, W. 1981: Foundations for a Metaphysics of Pure Process. *The Monist*, 64, 3–90.

Shavlik, J., Mooney, R., and Towell, G. 1991: Symbolic and Neural Learning Algorithms: An Experimental Comparison. *Machine Learning*, 6, 111–43.

Smart, J. J. C. 1978: The Content of Physicalism. *Philosophical Quarterly*, 28, 339–41.

Wimsatt, W. C. 1972: Teleology and the Logical Structure of Function Statements. *Studies in the History and Philosophy of Science*, 3, 1–80.

Wright, L. 1973: Functions. *Philosophical Review*, 82, 139–68.

2
Chomsky's Challenge to Physicalism

JEFFREY POLAND

Introduction

Physicalism[1] is usually conceived of as a doctrine which expresses a certain view of the world, roughly the view that everything bears one or another important ontological or explanatory relation to physical entities of one sort or another. For example, typical physicalist theses are:[2]

1 Every object is a physical object or is constituted by physical objects.
2 Every property is a physical property or is realized by physical properties.
3 Every truth (fact) is supervenient on physical truths (facts).
4 Every event has a physical explanation.

According to what I shall refer to as the "standard metatheory of physicalism," such claims as 1–4 are also understood in the following ways:

5 Physicalist theses have a truth value and are accepted as true.
6 Physicalist theses are empirical hypotheses.
7 Physicalist theses play a methodological role in science and philosophy.
8 Physicalist theses have human significance.

Physicalists tend to believe that, in the sciences, physicalism is a component of many scientists' philosophy of science, expressing implicitly or explicitly held beliefs which guide thought and other research activities. In philosophy, physicalism has been debated in metaphysics, ethics, aesthetics, philosophy of mind, philosophy of science, and philosophy of language. Among philosophers at least, physicalism, in one form or another, enjoyed considerable significance during the twentieth century. Beyond any implications physicalism has for understanding

modern science, it represents a continuation of the intellectual tradition which aims at the development of a unified "world view," according to which everything is grounded in a core set of fundamental constituents. Such an aim has value insofar as it provides a form of knowledge which deepens our understanding of ourselves and our place in nature or insofar as it has some other practical, ethical, or religious significance.

It is of some importance, then, that although physicalism apparently has profound significance and has been in the spotlight to a considerable extent, proponents of physicalism have not adequately responded to a frequently voiced objection, viz., that the theses of physicalism are vacuous because the critical concept of *the physical* has no well-defined content. Although stated in different ways by different critics (cf. Feigl 1969: 21, Chomsky 1972: 98, Hempel 1980: 194–5, Crane and Mellor 1990: 186–7), the core objection is a deep and rather complicated one with far-ranging implications for physicalism. If the theses of physicalism are vacuous, then they cannot properly be conceived of as empirical hypotheses which have a truth value and which play a role in science and philosophy. And if the theses are vacuous, it is unclear how physicalism can have any human significance.

Physicalists have, of course, attempted to clarify what is meant by "physical" in their various theses. There are two strategies that have been employed, an *a priori* strategy and an *a posteriori* strategy, and each has taken several forms. An *a priori* strategy is one which attempts to identify the essence of the physical via an analysis of the meaning of the term in either commonsense or philosophical usage (e.g., a "physical" object is one which is essentially located in space). An *a posteriori* strategy is one which does not necessarily assume that there is such a thing as the essence of the physical but rather attempts to identify physical entities in some way other than by linguistic or conceptual analysis (e.g., by appeal to physical theory or to exemplars of the physical). It is a contention, explicit or implicit, of the critics of physicalism that neither of these strategies can be successfully pursued.

Chomsky's criticisms of physicalism along these lines are among the most important, both because of their penetrating and persuasive character and because they contain within them the seeds of a positive defense of physicalism. In this paper I shall examine a number of claims and arguments which Chomsky has advanced over the past thirty years, not always to a fully comprehending physicalist audience. My objectives are twofold: first, to clarify the depth of Chomsky's objection, and second, to suggest a way physicalists might respond to it. In section 1 I shall attempt to reconstruct Chomsky's case for the claim that physicalism is vacuous because the concept of *the physical* lacks content. Building on this reconstruction, I shall identify some critical constraints to which physicalists must be responsive if they are to meet the objection effectively. In section 2, drawing on some of Chomsky's ideas about unification problems in science, I shall present a version of physicalism ("Methodological Physicalism") and show how it responds to the objection.

1 Chomsky's Challenge

What I shall refer to as "Chomsky's challenge" to physicalism is expressed in such passages as the following: "Returning to the critique of materialism . . . it seems to face several problems. The supposed concepts 'physical' or 'material' have no clear sense . . . There seems to be no coherent doctrine of materialism and metaphysical naturalism, no issue of eliminativism, no mind–body problem" (Chomsky 1994: 195–6). As an initial formulation, Chomsky's challenge to physicalism is the contention that physicalist theses (e.g. 1–4) are meaningless because the term "physical" has no definite content. And, of course, if physicalist theses are meaningless then they lack truth value, they are not empirical hypotheses, they can play no serious role in inquiry, and they have no human significance. What are the grounds for this challenge? As I understand it, Chomsky's argument involves two critical premisses. The first is that there is no *a priori* conception of the physical grounded in natural language, folk science, or metaphysics which provides the required content. The second is that, given the character of scientific inquiry, and in particular, inquiry in physics, physicalists cannot hope to identify a definite meaning for the term "physical" from a consideration of physical theory. If these two premisses are combined, then, unless there is some further way of identifying the content of either "physical" or physicalist theses, Chomsky's challenge to physicalism would appear to be insurmountable.

In support of the first premiss of the argument (viz., that there is no *a priori* conception of the physical which can provide the required content), Chomsky's primary consideration is the idea that commonsense understanding does not impose constraints upon science:

> whatever may be learned about folk science will have no relevance to the pursuit of naturalistic inquiry into the topics that folk science addresses in its own way, a conclusion taken to be a truism in the study of what is called "the physical world" but considered controversial or false (on dubious grounds, I think) in the study of the mental aspects of the world. (Chomsky 1995: 14)

Here Chomsky takes as obvious that in the study of "the physical world" naturalistic (i.e., scientific) inquiry proceeds unencumbered by commonsense concepts which might be expressed in natural language or used in folk theories. Although we might begin scientific investigation of nature with antecedent commonsense or folk understandings, these have no special status. For present purposes, this means that even if there is a commonsense or folk concept of *the physical*, it is not anything which constrains developments in science. This is bad news for physicalists who are hoping for an *a priori* solution to the problem of providing content to the term "physical" *and* who conceive of physicalism as a

doctrine which plays a role in science or is an expression of a scientifically based "world view," conceptions with which most physicalists would agree. The upshot is that attempts to analyze the term "physical" in order to identify its ordinary-language or folk-scientific meaning are strategically misguided attempts to rescue physicalism from Chomsky's challenge unless one is prepared to abandon physicalist commitments to science.

And what applies to common sense applies equally to *a priori* metaphysics. Chomsky simply closes the door on the idea that philosophy has any special contribution to make to scientific inquiry into nature. Hence, no *a priori* analysis of the physical, no *a priori* identification of the essence of the physical, deriving from philosophical thought, has any privileged status in scientific inquiry. Science is a bootstrapping operation which is not substantively constrained by *a prioristic* forms of speculative or analytic philosophy and which can in principle help itself to anything that can meet the test of scientific rationality. Thus, again, insofar as physicalism is a doctrine which plays a role or is grounded in science, *a priori* analysis of the physical cannot be a part of it.

What are Chomsky's reasons for the second premiss in his argument? Assuming the rejection of an *a priori* conception of the physical, an alternative strategy is to turn to scientific practice for some sort of *a posteriori* conception. Although there are different forms such a strategy can take, all depend essentially upon an appeal to theories within physics as the source of the required conceptual content. And it is this aspect of the strategy on which Chomsky has frequently focused in his critiques of physicalism.

One of Chomsky's important contributions to studies of methodology in linguistics and psychology is his insistence that we reflect carefully on the history of science, specifically the history of physics, so that we might discern how best to understand and pursue emerging and relatively immature scientific studies. For present purposes, reflection on the history of science yields an observation which Chomsky wants to parlay into a devastating objection to the physicalist program. To see this, consider the following passage:

> the Cartesian concept of *body* was refuted by seventeenth-century physics, particularly in the work of Isaac Newton, which laid the foundations for modern science. Newton demonstrated that the motions of the heavenly bodies could not be explained by the principles of Descartes's contact mechanics, so that the Cartesian concept of body must be abandoned. (Chomsky 1988: 143)

Thus, it is an historical fact that physics underwent a transition from the Cartesian to the Newtonian framework as a consequence of an explanatory shortcoming of the former. The significance of this episode according to Chomsky is at least threefold. First, the episode demonstrates that physical theory undergoes change: it is *evolving*. Second, the resistance to the change generated by commitments to the Cartesian theory was not sufficient to prevent

its rejection in the face of the demands of scientific inquiry. Theories in physics evolve *openly*,[3] subject only to the demands of good scientific practice and the phenomena they are concerned with. This point is an extension of the idea discussed above that science is not substantively constrained by *a priori* commitments: neither common sense nor metaphysics *nor prior scientific commitments* are strong enough to override the demands of scientific rationality. Finally, the rejected physical theory, the so-called "mechanical philosophy," had provided the basis for a well-defined conception of the physical world (the world of "body"), relative to which significant metaphysical theses could be formulated. Its rejection created a gap, and a new conception of the physical world was needed for the purpose of framing metaphysical theses.

Given these points, Chomsky draws the following conclusion:

> What is the concept of body that finally emerged? The answer is that there is no clear and definite concept of body. If the best theory of the material world that we can construct includes a variety of forces, particles that have no mass, and other entities that would have been offensive to the "scientific common sense" of the Cartesians, then so be it: We conclude that these are properties of the physical world, the world of body. The conclusions are tentative, as befits empirical hypotheses, but are not subject to criticism because they transcend some *a priori* conception of body. There is no longer any definite conception of body. Rather, the material world is whatever we discover it to be, with whatever properties it must be assumed to have for the purposes of explanatory theory. (Chomsky 1988: 144)

Although Chomsky is here concerned with the issue of whether there is a definite conception of body which can underwrite the mind–body problem, the points he makes are directly relevant to the physicalist's more general interests in identifying a definite conception of the physical. And his conclusion is clear: the gap created by the rejection of the mechanical philosophy cannot be filled, given the open and evolving character of scientific theory and the lack of any *a priori* conception of the physical to constrain developments in science. Conceptions of the physical are, at best, contingently tied to tentative theories in physics. Since such theories are open and evolving, the concept of the physical is unstable and, hence, not sufficiently well defined for the purpose of framing empirical or metaphysical theses. There simply is no *definite*[4] *a posteriori* concept of the physical available for use by the physicalist. The significance of this conclusion for physicalism is also clear: if our conception of the physical is tied to open and evolving theories in physics and there is, therefore, no well-defined *a posteriori* conception of the physical, it follows that it is pointless to inquire about the content of the theses of physicalism since they too have no well-defined content (cf. Chomsky 1988: 178).

Now, before we agree to such a dire (for physicalists) conclusion, we should ask, granting that there is no relevant *a priori* conception of the physical and granting that physical theory is open and evolving, why is it that an *a posteriori*

conception of the physical cannot be developed which can underwrite the formulation of significant physicalist theses? Why, for example, can't the physical be identified in terms of current physics or some future or ideal physics? The answer is that the argument, as stated so far, does not logically preclude these options. Clearly, the step from the claim that *a posteriori* conceptions of the physical are unstable because they are contingently tied to tentative, open, and evolving theories in physics to the claim that the concept of the physical is not sufficiently well defined for the purpose of framing physicalist theses requires further defense.[5] As I understand him, Chomsky has not given (nor does he think he has given) a logical demonstration of the conclusion he suggests. Rather, he is laying down a challenge to physicalists, a challenge to which physicalists have not yet adequately responded. Further, the challenge does not simply rest on the two claims discussed so far (viz., that there is no relevant *a priori* conception of the physical and that given the open and evolving character of scientific theory there is no definite *a posteriori* conception). The difficulties that physicalists have had in developing a viable response[6] suggest that more is involved.

In addition to buttressing the idea that there is no *a posteriori* conception of the physical, the open and evolving character of physical theory indicates two further problems for physicalists to cope with. First, it provides grounds for thinking that physicalism is a trivial doctrine. A common reading of certain passages in Chomsky's writings (e.g., Chomsky 1972: 98; 1980: 5–6) is that, given the open and evolving character of physical theory and hence of the notion of "the physical world," phenomena recalcitrant to integration into a physicalist system can and will inevitably be brought into the fold via an appropriate revision of our conception of the physical world. And given the general availability of this option, physicalist principles are guaranteed true.[7] As we shall see below, this is *not* the best way to understand Chomsky, but physicalists interested in meeting Chomsky's challenge need to have a response to such a line of objection.

Second, the open and evolving character of physical theory suggests what seems to be an even more serious implication for physicalists: viz., that a revision of physics might involve an incorporation into the physical realm of precisely those recalcitrant phenomena about which physicalists have typically been most concerned (i.e., mental entities). The open and evolving character of physical theory means that over the course of history physicists have postulated a wide variety of different sorts of entities that fail to exhibit any shared feature which makes them all nonetheless "physical." Since the only constraints on what gets postulated in physics are considerations of explanatory adequacy, empirical adequacy, and any other demands on the rational conduct of scientific inquiry, the class of physical entities viewed historically (i.e., those entities that have been, are, or will be postulated by physicists) exhibits no integrity beyond being the class of entities postulated by physicists in the course of doing their work.

This feature of physical inquiry is what opens up the possibility of downward incorporation of the mental into physics.

Chomsky, in suggesting this possibility, has identified and applied pressure to a sensitive nerve in the physicalist position, and many friends and foes of physicalism agree that if physicalism is to be a substantive doctrine retaining significant content, it must not allow for the possibility of downward incorporation of the very phenomena about which physicalists have been historically so exercised. For many, incorporation of the mental into physics would be the demise of physicalism. Chomsky, of course, is not troubled by accepting such revisions as part of the normal course of science[8] and giving up on the significance of physicalism, if that is a consequence. But, physicalists cannot be so sanguine; and they need to have a cogent framework for dealing with the complex set of issues underlying this line of objection.

A further problem which confronts physicalists interested in pursuing an *a posteriori* strategy for defining the physical is the following: insofar as the approach depends upon an appeal to theories in *physics* for the purpose of identifying physical entities, "physics" must pick out a well-defined research program in science. Typically, physicalists have in mind a specific branch of physics (viz., "fundamental physics") rather than all of what is called "physics," but either way the problem of individuation is critical. For if physics (or the relevant branch of physics) is not itself well defined and well individuated from other research programs in science, the strategy of providing a definite conception of the physical in terms of theories within physics is undermined from the start.[9]

So what, then, is the full thrust of Chomsky's challenge to physicalism? Consider what has emerged in the discussion so far. First, the challenge is centered on the following argument:

(P1) There is no relevant *a priori* conception of the physical.
(P2) There is no definite *a posteriori* conception of the physical.
(C1) Thus, neither an *a priori* nor an *a posteriori* approach to assigning definite content to "physical" is viable.
(P3) There are no other approaches for assigning definite content to "physical."
(C2) Thus, there is no definite content assignable to "physical."
(P4) If C2 then C3.
(C3) Thus, there is no definite content assignable to physicalist theses.
(P5) If C3 then C4.
(C4) Thus, no aspect of the standard metatheory of physicalism is tenable (i.e., physicalist theses are not true, are not empirical hypotheses, play no useful methodological role, and have no human significance).

Second, discussion of the first two premisses of this argument revealed a number of points which constrain physicalist replies:

There are no *a priori* substantive constraints on scientific theorizing.
Revision of physical theory is subject only to the demands of scientific rationality.
Physical theory is open and evolving.
Physicalist theses should not be trivially true.
Downward incorporation of the mental appears to be a possibility in science.
Fundamental physics must be well individuated if it is to play a role in defining the physical.

The full thrust of Chomsky's challenge, then, is that an effective reply to the argument must be responsive to these various points as well.

I suggest that one reason why physicalists have not yet successfully met Chomsky's challenge is that they have not understood well enough the problem they have to solve. It is important to see that negotiating Chomsky's challenge involves coming to terms with a fundamental tension in the standard physicalist framework which Chomsky has effectively flushed out: viz., the tension between (a) commitment to empirically based scientific rationality, (b) commitment to unification of a certain sort, and (c) commitment to understanding physicalist theses in certain ways (i.e., the standard metatheory and the idea that mind is not fundamental). The first commitment implies that there is no *a priorism* in science, that scientific theory is open and evolving guided only by principles of scientific rationality, and that downward incorporation is possible. The second commitment is to a certain sort of structure: i.e., ontological and explanatory grounding of all entities and principles in some basic set, where this basic set is derived from theories developed in physics. What Chomsky's challenge brings out is that the combination of these first two commitments appears to breed problems for the third, i.e., the commitment to understanding the theses as contingently true empirical hypotheses which play a role in inquiry and which presuppose that mind is not included in fundamental physics. For, if the content of physicalist theses is not tied down by *a priori* conceptual analysis of *the physical*, then it is tied to open and evolving physical theory, in which case the theses have no definite content, they appear to be trivial and of no use, and it is possible for "mental" phenomena to be downwardly incorporated into physics. At issue for physicalists is whether they can resolve this tension and negotiate the demands of the challenge while retaining both a strong commitment to science and the standard metatheory of physicalism.

2 Methodological Physicalism

Given Chomsky's challenge, how might physicalists respond? There are, at least, three strategies to consider: identifying and defending the viability of an *a priori* analysis of the physical; identifying and defending an *a posteriori* analysis of the

physical; and *reconstructing* physicalism by altering some aspect of the standard metatheory, thereby giving Chomsky's challenge its due, but retaining the most important features of physicalism in the process. The first two approaches, respectively, attempt to take on P1 and P2 directly, while retaining the standard metatheory. I think the first approach is hopeless because it does not take seriously physicalist commitments to science. The second approach, although more consilient with physicalist commitments to science, is not very promising either, in light of the full complement of problems and issues to which it must be responsive. Thus, in the remainder of this chapter, I shall explore the prospects of a reconstructive approach.

It will be instructive to begin by considering two other recent attempts at such a strategy, one by Hellman and the other by van Fraassen. In responding to a version of "Hempel's Dilemma," Hellman (1985: 609–10) proposes that physicalists accept the fact that their doctrine is probably false. His idea is that, assuming that the only way to provide content to the term "physical" is via reference to (mild extensions of) current physical theory, and since (a mild extension of) current physical theory is very likely to be false, "there is every reason to think" (ibid.: 609) physicalist theses are false, although they have some redeeming qualities (e.g., usefulness as a criterion of completeness in science, usefulness in guiding research, and near or exact truth in specific subdomains). This is an instance of a reconstructive strategy, since, although Hellman allows that physicalist principles have content and are empirical theses which play a role in science, he is evidently willing to abandon the idea that they are true, something which is central to the standard metatheory of physicalism. What is significant about Hellman's approach here is that, although he seems to want to retain most components of the standard metatheory, he is emphasizing the importance of the role of physicalist theses in science over their truth value.[10] Physicalist theses express a certain sort of structure which scientific theories and their associated ontologies can exhibit, and the theses play a role in guiding science toward the discovery of that structure, at least in some domains.

More recently, van Fraassen (1996), in a discussion of the relations between metaphysics and science, offers a critique of physicalism focusing on the open and evolving character of physical theory and the problem of identifying any enduring content for such terms as "material" and "physical." In the process, he distinguishes between the specific content of particular formulations of physicalist theses and the "spirit of physicalism," and he proposes the following: physicalism should be identified, not with a theory about what there is, but rather with such *attitudes* as deference to science in matters of opinion about what there is and inclination to accept science as (approximately) complete at any given time. Thus, to think that science requires a presumptive physicalism "which constrains scientific theories to consistency with certain determinate factual theses" (ibid.: 170) is a matter of false consciousness based on a confusion of the attitudes

expressed (the real content) and the theses held (transitory forms of expression of the attitudes tied to current science). Van Fraassen is offering, though not endorsing, another reconstruction of physicalism: to be a physicalist is really to have certain *attitudes*. Such an approach to the content of physicalist theses means that their real content is non-cognitive and hence they are neither truth-valuable nor empirical. This also means, according to van Fraassen, that physicalist theses can play no role in science: since they lack factual content, they can impose no factual constraints on theories.

Van Fraassen's idea that the content of physicalism involves certain attitudes is promising, and he has more or less correctly identified one such attitude. Insofar as physicalists view physicalism as a program grounded in science, they do have an attitude of deference to science in matters of what there is, but it is neither scientism (the view that science is the only legitimate source of knowledge) nor a simple appeal to authority: science involves a sort of rationality worthy of deference in certain matters concerning the study of nature. However, the other component of the "spirit of physicalism" which van Fraassen identifies is more problematic. Talk about the (approximate) completeness of science is, I think, both vague and misleading. In the absence of clarification of the relevant completeness claims, it is difficult to assess whether physicalists are inclined to accept them. So, although van Fraassen has a promising idea regarding the content of physicalism, he has not adequately specified that content. With respect to the conclusion that physicalism can play no role in science, this follows only if the imposition of factual constraints is the only possible role for physicalist principles to play. As we shall see below, there are other ways for physicalism to influence scientific decision making.

A third reconstructive strategy of reply to Chomsky's challenge is what I shall call "methodological physicalism."[11] On this view, physicalism is understood as embodying a certain "spirit" which involves the deference to science already identified and a commitment to a certain sort of program of *unification*,[12] and it is these commitments which are expressed by physicalist principles and which remain constant as physical theory and our conception of the physical evolves. According to methodological physicalism, physicalist principles delineate a set of ontological and explanatory relations which structure a certain sort of idealized, hierarchically organized system. These relations constitute a certain sort of unity which, according to the physicalist, is of value because it contributes to serving human interests (e.g., understanding). But methodological physicalism departs from the standard metatheory of physicalism with respect to how physicalist principles are to be construed. The principles are not best viewed as having a truth value or, hence, as being empirical hypotheses. Rather, they function as *regulative ideals*[13] which both call for, and aid in the development of, a system of theories and associated entities which exhibits the specified structure. As such, they play a methodological role in inquiry in at least two ways: they help frame certain questions (e.g., questions concerning the physical basis

of some phenomenon) and they play a role in structuring certain sorts of decisions (e.g., decisions about how to resolve crises when unification is difficult to achieve).

It is at this point that Chomsky's critical discussion of physicalism provides valuable clues for how to further understand this reconstructive strategy in defense of physicalism. There are three themes in Chomsky's writings that are relevant here. The first concerns "the unification problem" in science and will help clarify the physicalist's commitment to unification. The second concerns the ways in which specific unification problems can be resolved and will assist in clarifying the methodological role physicalist theses play. The third concerns how to approach unification problems when they arise and will give further content both to physicalist commitments to science and to the role of the theses in scientific inquiry.

With respect to the physicalist commitment to unification, Chomsky emphasizes the importance of unification in science, by which he means a certain sort of integration of "ways of looking at the world" (Chomsky 1993: 80–1). Given that the study of nature proceeds at multiple levels of description and explanation, unification involves showing how theories (and associated entities) at relatively abstract levels are related to theories (and associated entities) at more fundamental levels, and ultimately to theories and entities at the level of basic physics. The type of unification with which Chomsky is concerned involves showing "how . . . abstract structures and processes are realized or accounted for in some concrete terms" (Chomsky 1972: 14),[14] or how "organized matter can have these [abstract] properties" (Chomsky 1992: 2). Thus, solving a "unification problem" requires relating entities at higher levels of organization and abstraction to entities at lower levels in explanatory ways that clarify both which lower-level entities (objects, properties, processes, mechanisms) are involved in the realization of higher-level entities and how the realization is effected. Similarly, solution of a unification problem provides explanations of how higher-level explanatory principles are "grounded" in lower-level "mechanisms" (Chomsky 1991: 5).[15]

This picture is in substantial agreement with the thinking of many physicalists, and provides an initial characterization of the sort of unification to which physicalists are committed and which physicalist theses should express as an ideal to be sought. One point of clarification is to distinguish between what I shall call "the general unification problem" (GUP) and "a local unification problem" (LUP) as follows. The GUP is the problem of constructing a system of theories and entities which exhibits the idealized structure characterized by physicalist theses. LUPs, on the other hand, concern specific ontological or explanatory relations between more or less adjacent theories and entities within the larger system. Solution of the GUP requires solution of a multitude of LUPs. Chomsky's focus has typically and appropriately been on the articulation and solution of LUPs relevant to contemporary linguistics and psychology. This is

quite compatible with the physicalist's interest in articulating more general principles concerning the larger system of which linguistics and psychology are a part.

With respect to clarifying the methodological role of physicalist principles, Chomsky's discussion of how unification can be achieved in particular cases provides important insights. In his writings (cf. Chomsky 1992: 12; 1993: 44, 80–1; 1995: 1–2) he has frequently emphasized the idea that there is no one way that unification can be achieved, if it can be achieved at all. In fact, he has identified at least four main options relevant to the solution of a local unification problem: "reduction," "expansion," "modification of all accounts," and no successful unification. With respect to talk of *reduction*, Chomsky is not explicit regarding how narrowly or broadly to construe it. For present purposes, I shall take him broadly as meaning by reduction a successful ontological and explanatory integration of higher-level and lower-level theories *in their current forms*, where the nature of the integration is abstractly specified by some appropriate set of physicalist principles. By *expansion* Chomsky has in mind some theoretical modification. On the one hand, lower-level theories can be modified, possibly by introducing "outlandish" ontological posits or novel explanatory principles, in such a way as to allow for successful integration of a target higher-level theory with the modified lower-level theory. This option could include a downward incorporation as an extreme case. On the other hand, expansion might involve a modification of the higher-level theory. And although Chomsky does not usually feature *elimination* of the higher-level theory as a serious option (he is often at pains to point out how inappropriate elimination is in many cases), it too falls within the space of possibilities to consider. *Modification of all accounts*, of course, is the possibility of making modifications at multiple levels of theory. Finally, Chomsky is quite adamant about taking seriously the possibility that humans might *fail to achieve integration* in some cases. There are three scenarios to consider here: that we have a difficult unification problem, solvable by humans in some way or another, but not yet solved for contingent reasons; that we have a "mystery," a situation in which a solution to the unification problem, although possible, is not cognitively accessible to humans; and finally, that we are faced with a situation in which a solution is simply not possible (i.e., the call for unification *cannot* be answered in any of the ways just identified).

In light of these distinctions, the following reconstruction of Chomsky's conception of a local unification problem will help clarify the role of physicalist theses in inquiry: given some higher-level theory T and some body of lower-level theory P, a local unification problem is the problem of integrating T with P in accordance with physicalist principles (i.e., by establishing certain ontological and explanatory relations between elements of P and elements of T). There are, in this picture, at least four components which need to be kept in mind: T, P, the physicalist principles which express the general character of the desired unifying relations, and the specific explanatory and/or ontological relations involved in

particular cases. When we speak of how a unification problem can be resolved, we are concerned with which component of the picture requires attention. Thus, the options are as follows:

1 Integrate T with P by identifying an appropriate set of specific explanatory or ontological relations.
2 Revise P in a way which permits integration of T with the revised theory, P′.
3 Modify T in a way which permits integration, or eliminate T.
4 Combine modifications of T and P in a way which permits integration.
5 Interpret a failure to integrate T with P as signaling either a problem, a mystery, or an impossibility.

Of course, in any given situation, it may be quite unclear which of these options is most appropriate to pursue. This list simply identifies the range of possible ways of resolving a particular unification problem and clarifies how physicalism helps structure the decision to be made. Specifically, physicalist principles delineate the ontological and explanatory relations which constitute one *ideal* for integration of T with P.

The third theme in Chomsky's discussion of these matters concerns how unification problems should be approached, and provides a way of understanding the physicalist's commitment to scientific rationality. Succinctly put, Chomsky rejects any sort of dogmatism with respect to approaching issues of unification in science (cf. Chomsky 1988: 190; 1993: 86–7; 1995: 11 for representative expressions of this point). The rejection of dogmatism in this context means at least three things. The first is that there is no one way in which unification problems can be resolved; the history of science has seen resolutions in many, if not all, of the possible ways. The second point is that determination of an appropriate solution is a matter of empirical inquiry and scientific problem solving; *a prioristic* speculation and argumentation are of no value with respect to this sort of problem. And, finally, both reduction of T to P and elimination of T are rare and neither should be viewed as the presumptive solution.

This last point is based on the deeper claim that there should be no presumption to solve local unification problems in a way which is biased in favor of the lower-level theory P. This is so because the *ontological primacy* of P (i.e., the fact that it is concerned with more fundamental entities and principles) does not guarantee either that P is *epistemically superior* to T (i.e., that it is better tested and more highly confirmed) or that P is *sufficiently mature* to provide the basis of a solution (i.e., that it is sufficiently developed with respect to its target domain). For example, in the context of discussing unification problems with respect to linguistics and psychology, Chomsky (1995) has argued, in effect, that if P (e.g., a theory in neuroscience) is not epistemically superior to T (e.g., a theory in linguistics/psychology) then certainly elimination of T is not a serious option and, assuming reduction cannot be achieved, P could well be a more likely

target of revision than T. With respect to maturity considerations, if P is not sufficiently developed with respect to its target domain, then, even if it is well developed and well confirmed with respect to some phenomena, it may have nothing to say about phenomena in its domain which are relevant to solving a given unification problem. It simply might be the case that the P-level resources required for successful integration are not on hand and that the proper course of action for seeking a solution is further development, or revision, of P. Indeed, with respect to integration of theories in contemporary linguistics/psychology with lower-level theories, Chomsky suggests that there is no good reason to suppose that contemporary neuroscience is sufficiently mature to underwrite a solution to the unification problem.

The main points bearing on how to proceed in approaching unification problems are as follows. Whether and how to proceed depends essentially upon the epistemic and developmental status of the relevant theories. It is pointless to consider issues of unification when T is relatively immature or untested. Only if T is sufficiently mature and well tested can serious work begin, because only under such conditions will the constraints upon solution of the unification problem be known. Further, it is only under such conditions that researchers know what sorts of things they are looking for at lower levels and, hence, have a basis for assessing whether P is sufficiently mature (cf. Chomsky 1988: 186). Finally, when revisions are required in either T or P, their epistemic and developmental characteristics play a central role in deciding which is a more plausible target. Thus, at any given point in time, the epistemic and developmental status of T and P determine whether the unification problem should be seriously studied at all and, if so, what directions are most plausible to pursue. Because of its central commitment to science, methodological physicalism sees the role of physicalist theses in inquiry as a modest one (i.e., to frame and call for a certain kind of structural unity) subject to exactly the sorts of considerations just outlined.

For present purposes, I want to suggest that it is commitment to views such as those articulated by Chomsky on unification which constitutes the spirit of physicalism. Physicalists of this sort are committed to scientific rationality as characterized above (i.e., rejection of dogmatism, rejection of *a priori* substantive constraints, rational consideration of epistemic and developmental characteristics of theories, etc.) and to a unification program which aims at developing a certain ontological and explanatory structure ultimately grounded in a set of entities and principles which are the focus of inquiry in fundamental physics. Methodological physicalism is one way of embodying this spirit which is distinct from standard versions of physicalism and from the non-standard versions discussed by Hellman and van Fraassen.

Its first distinctive feature is that it rejects two key components of the standard metatheory (viz., 5 and 6). Physicalist principles, according to the methodological physicalist, are not declarative sentences which express certain general facts about

the world. Because of the irrelevance of *a priori* conceptions of the physical and the open and evolving character of physical theory, the term "physical" should be viewed as not having a definite content (i.e., one which is associated with a particular, well-defined class of objects, attributes, and principles), just as Chomsky contends. Thus, contra both Hellman and the standard metatheory, physicalist theses do not have a truth value and are not empirical hypotheses.

However, in granting this, the methodological physicalist is not committed to claiming, with van Fraassen and Chomsky, that "physical" and physicalist theses have no content at all. The second distinctive feature of the view, then, is that it provides an alternative account of the content of physicalist theses. With respect to the term "physical," it is to be understood as standing for whatever class of entities and principles fundamental physics identifies; without an identification of a particular physical theory, this is an indefinite specification, but it is not entirely vacuous. Rather, the term "physical" refers indefinitely to entities and theories which are postulated within a certain research program. That is, it functions as an indexical: when employed in a particular context (e.g., a particular moment in the history of science) it picks out a specific class of entities, although it picks out nothing apart from a particular context of application. The content of the term is what allows it to function in this way.

With respect to physicalist theses, since "physical" lacks a definite content, they also lack a definite content; and hence they do not make factual claims about the structure of reality which are either true or false. But they too are not vacuous. Rather, they delineate an ideal abstract structure which a system of theories and their associated ontologies can exhibit: viz., ontological and explanatory relations between theories (entities) at different levels within a hierarchically organized system which is grounded in the theories (entities) of fundamental physics. Further, such principles, rather than being factual assertions, have a prescriptive force which is an expression of the physicalist's commitment to building a system which exhibits the ideal abstract structure expressed by the theses. In effect, a principle such as 2 above means something like: Strive to do science in such a way that every property is a physical property or is realized by physical properties. The specific meaning and potential utility of such a prescription exist only when the principle is applied in a particular context and the "physical" properties are identified in terms of some specified physical theory. Apart from such particular contexts of application, physicalist principles are general prescriptions for developing a system of theories in a particular way. As such, they should be construed as "regulative ideals" (i.e., principles which identify some feature and call for its production) rather than empirical hypotheses.[16]

The third distinctive feature of methodological physicalism is its construal of the role of physicalist principles in inquiry. As regulative ideals such principles call for certain kinds of unifying ontological and explanatory relations. As a consequence, in particular research contexts, they lead to the formulation of research questions such as "What is the physical basis of human language

capacity?", where "physical" takes on a specific meaning based upon the evolving theoretical understanding characteristic of the context. And, within such contexts, physicalist principles provide some structure to the decision problems which arise when unification of the sort called for by the theses is not readily forthcoming. Both pursuit of the questions and attempts to resolve the decision problems guide research in the direction of greater unification, subject to the demands of scientific rationality as outlined above. This role for physicalist principles is an appropriately limited one which is *not* (contra van Fraassen) based on the idea that they impose *general factual constraints* on theory construction. Rather, in their concrete application, they introduce structural constraints within a particular context of problem solving, and their relative strength depends upon various contingent features of the context as partially described above. This limited role of physicalist principles should be understood within a more general framework of rational inquiry, according to the methodological physicalist.

So, finally, how does methodological physicalism meet Chomsky's challenge? Recall that meeting the challenge requires developing a framework which is responsive both to Chomsky's central argument and to the six additional points which emerged from discussion of premisses P1 and P2. As part of its reconstructive strategy, methodological physicalism assigns content to physicalist principles without taking on either *a priori* or *a posteriori* commitments regarding the character of the physical. In this way, physicalist principles are understood as lacking definite content but not being vacuous. And, given the content they have, physicalist principles, although lacking truth value and empirical status, play a limited but nonetheless real role in inquiry. Thus, methodological physicalism responds to the argument by denying P5 (viz., if there is no definite content assignable to physicalist theses, then no aspect of the standard methatheory of physicalism is tenable). This denial effectively blocks the slide to the rejection of *all* aspects of the standard metatheory while granting Chomsky the main thrust of his argument (viz., physicalist theses lack definite content, have no truth value, and are not empirical hypotheses). What remains are the "spirit of physicalism" (which Chomsky admirably exemplifies), a modest role in inquiry for physicalist principles, and at least the hope of their having human significance (e.g., by contributing to the construction of a system which exhibits physicalist unity).

With respect to the six additional points, methodological physicalism as articulated here is directly responsive to the first five, whereas the sixth is an acknowledged burden.[17] The commitment to science at the heart of methodological physicalism means that there are no substantive *a priori* constraints on scientific research and that the evolution of scientific theories is based only on the demands of scientific rationality. Thus, methodological physicalism is quite compatible with both the open and evolving character of scientific theory and the possibility of downward incorporation of the mental into physics. Physicalists who maintain strong commitments to science must allow that *physics* is unconstrained in the resources to which it can appeal for explanatory purposes.[18] Such

considerations override any contingent concerns about the mind with which physicalists have been historically preoccupied.

Regarding the alleged triviality of physicalism in light of the open and evolving character of physical theory, there is nothing within the framework of methodological physicalism which permits *ad hoc* modification of theory designed to save physicalism: as Chomsky rightly emphasizes, the possibility that unification might not be attainable needs to be countenanced, and any theoretical modifications are always subject to the demands of scientific rationality. This is why it is a mistake to interpret Chomsky as suggesting that physicalism is trivial because it is always possible to revise physics and our conception of the physical; such an interpretation ignores his fundamental scientific commitments. For methodological physicalism, this denial of triviality means that although there is a *call* for a system exhibiting physicalist unification, there is no guarantee that such a system will be successfully constructed (i.e., that the GUP will be solved).

Whether methodological physicalism can be developed further and defended effectively remains to be seen. What I hope is clear is that Chomsky's challenge to physicalism is quite serious and that it has targeted a weak point in the standard physicalist position (viz., a tension among fundamental commitments). One reason why physicalists have failed to adequately respond to Chomsky's challenge is that they have, for the most part, not been sensitive to this tension, and have attempted to preserve the metatheory of physicalism intact while tinkering with how to define "the physical" along either *a priori* or *a posteriori* lines. The main moral of this paper, and of Chomsky's criticism of physicalism, is that something has to give: in particular, our conception of the content and the empirical status of physicalist principles must be closely re-examined and in all likelihood modified.[19]

Chomsky's reply: pp. 263–5.

Notes

1 In this paper I shall be focusing on physicalism conceived along the lines indicated in the text. The relations between "physicalism" and "materialism" will be ignored for present purposes. The issues to be addressed below are largely unaffected by distinctions between these terms, and in discussing these issues I shall sometimes use the term "physicalism" in contexts where authors I am commenting on use "materialism."

2 I intend 1–4 to be examples only; in this paper I shall not be concerned with issues concerning how best to formulate physicalist principles.

3 In the text, I will use Chomsky's form of words (viz., "open and evolving") when discussing this feature of scientific theories.

4 By a "definite" conception of the physical is meant one which has a well-defined extension.

5 It is here that discussion of what has been called "Hempel's Dilemma" fits into an attack on physicalism. The dilemma concerns which version of physical theory is presupposed for an understanding of the meaning of "physical": current physics or future/ideal physics. Either way there are supposed to be problems since appeals to current physics lead to the conclusion that physicalist principles are "obviously false" whereas appeals to future/ideal physics lead to the conclusion that physicalist principles are vacuous (or trivial). See Hempel 1980, Crane and Mellor 1990, and Crane 1991 for versions of this dilemma. The arguments for the various prongs of the dilemma are controversial and have been disputed in the literature (cf. Smart 1978; Post 1987; Papineau 1990 and 1991; Pettit 1994; Poland 1994; Melnyk 1997; Ravenscroft 1997; and Jackson 1998).

6 See Daly 1998 for a critical review of some extant attempts to pin down the notion of a physical property.

7 Daly (1998: 199) cites both passages mentioned in the text in support of this interpretation of Chomsky.

8 I believe that Chomsky would not be troubled about this only if such revision was part of the "normal" course of science; *ad hoc* modifications of physics to save physicalism are not acceptable.

9 In Poland 1994 I identify a number of other dimensions of the problem of developing an *a posteriori* conception of the physical: e.g., anti-realism about physics, underdetermination of theory in physics, issues of interpreting the language of physics, and issues of justifying the privilege accorded to physics. See also Post 1987 for valuable discussion of the problem.

10 It is, perhaps, arguable that Hellman is not pursuing a reconstructive strategy at all since he does seem to want to view the theses of physicalism as empirical hypotheses which express a factual content and play a role in inquiry. However, it is his apparent tolerance of the actual falsehood of the theses which makes the approach a reconstructive one. Melnyk's (1997) idea that physicalist theses are not likely to be true, but that it is not unusual in science to hold theses which one views in this way, is clearly not part of a reconstructive strategy. This is because his view is only that, although physicalist theses are more probable than the relevant alternatives, they are improbable relative to the available evidence; he does not accept the idea that the theses are actually false.

11 In several of his recent discussions, Chomsky (1992, 1993, 1994, 1995) has presented a view he calls "methodological naturalism." Methodological physicalism is so named partly to suggest its affinity with methodological naturalism.

12 Melnyk (1997: 633 fn 21), while not specifically developing the theme of unification, suggests that "the spirit of physicalism" might involve commitment to the idea that there is a basic science which occupies a certain place in relation to all other sciences, a relation defined by physicalist theses. Although this idea is compatible with the approach taken here, his view of physicalism and his clever treatment of the problem of the physical are quite different.

13 I first endorsed the idea that physicalist principles are regulative ideals in Poland 1983, although I subsequently abandoned it in Poland 1994, when I believed that an *a posteriori* response to the problem of the physical was viable.

14 See also Chomsky 1988: 145, 185–7; 1993: 52.

15 See Poland 1994: ch. 4 for discussion of ontological and explanatory grounding within a physicalist system.

16 Such "regulative ideals" can generate specific empirical hypotheses in particular contexts along the lines suggested in Giere 1988: 78–81 (e.g., the hypothesis that a certain set of theories in fact possesses the unified structure identified by the regulative ideal). Thanks to Barbara Von Eckardt for pointing this out.

17 In Poland 1994: ch. 3, I attempted to provide a characterization of "physics" which would carry the burden. But see Daly 1998 for misgivings about that characterization. The issue cannot be pursued in this paper and must await another occasion.

18 David Lewis once wrote: "Materialist metaphysicians want to side with physics, but not to take sides within physics" (Lewis 1983: 364). I do not think that Lewis, at the time he wrote this passage, appreciated its full significance. His own position seems not to adequately deal with the tension between its commitments to certain forms of physical theory and physicalist commitments to science. See Poland 1994: 160 for discussion.

19 Thanks to Karl Hostetler, David Pitt, and especially Barbara Von Eckardt for comments and helpful discussions concerning earlier drafts.

References

Chomsky, N. 1972: *Language and Mind*. New York: Harcourt, Brace, Jovanovich.

Chomsky, N. 1980: *Rules and Representations.* New York: Columbia University Press.

Chomsky, N. 1988: *Language and Problems of Knowledge.* Cambridge, Mass.: MIT Press.

Chomsky, N. 1991: Linguistics and Adjacent Fields: A Personal View. In A. Kasher (ed.), *The Chomskyan Turn*, Oxford and Cambridge, Mass.: Blackwell.

Chomsky, N. 1992: Evans Lecture, MIT.

Chomsky, N. 1993: *Language and Thought*. Wakefield: Moyer Bell.

Chomsky, N. 1994: Naturalism and Dualism in the Study of Language and Mind. *International Journal of Philosophical Studies*, 2 (2), 181–209.

Chomsky, N. 1995: Language and Nature. *Mind*, 104 (413), 1–61.

Crane, T. 1991: Why Indeed? Papineau on Supervenience. *Analysis*, 51, 32–7.

Crane, T. and Mellor, H. 1990: There Is No Question of Physicalism. *Mind*, 90, 185–206.

Daly, C. 1998: What Are Physical Properties? *Pacific Philosophical Quarterly*, 79, 196–217.

Feigl, H. 1969: The Origin and Spirit of Logical Positivism. In P. Achinstein and S. Barker (eds.), *The Legacy of Logical Positivism*, Baltimore: Johns Hopkins University Press.

Giere, R. 1988: *Explaining Science: A Cognitive Approach*. Chicago: University of Chicago Press.

Hellman, G. 1985: Determination and Logical Truth. *Journal of Philosophy*, 82 (11), 607–16.

Hempel, C. 1980: Comments on Goodman's *Ways of Worldmaking. Synthèse*, 45, 193–9.

Jackson, F. 1998: *From Metaphysics to Ethics*. Oxford: Clarendon Press.

Lewis, D. 1983: New Work for a Theory of Universals. *Australasian Journal of Philosophy*, 61, 343–77.

Melnyk, A. 1997: How to Keep the "Physical" in Physicalism. *Journal of Philosophy*, 94 (12), 622–37.

Papineau, D. 1990: Why Supervenience? *Analysis*, 50, 66–71.

Papineau, D. 1991: The Reason Why: Response to Crane. *Analysis*, 51, 37–40.

Pettit, P. 1994: Microphysicalism without Contingent Micro-macro Laws. *Analysis*, 54, 253–7.

Poland, J. 1983: Problems and Prospects for the Physicalist Program in Science. Doctoral dissertation, MIT.

Poland, J. 1994: *Physicalism: The Philosophical Foundations.* Oxford: Clarendon Press.

Post, J. 1987: *The Faces of Existence.* Ithaca: Cornell University Press.

Ravenscroft, I. 1997: Physical Properties. *The Southern Journal of Philosophy*, 35, 1–13.

Smart, J. 1978: The Content of Physicalism. *Philosophical Quarterly*, 28, 339–41.

Van Fraassen, B. 1996: Science, Materialism, and False Consciousness. In J. Kvanvig (ed.), *Warrant in Contemporary Epistemology: Essays in Honor of Alvin Plantinga's Theory of Knowledge*, Lantham, Md.: Rowman and Littlefield.

3
Real Materialism[1]

GALEN STRAWSON

Trinculo might have been referring to modern physics in the words, "This is the tune of our catch, played by the picture of Nobody."

Eddington[2]

Love like Matter is/Much odder than we thought.

Auden[3]

1 Introduction

Materialism is the view that every real, concrete[4] phenomenon[5] in the universe is physical. It is a view about the actual universe, and for the purposes of this paper I am going to assume that it is true.

It has been characterized in other ways. David Lewis once defined it as "metaphysics built to endorse the truth and descriptive completeness of physics more or less as we know it,"[6] and this cannot be faulted as a terminological decision. But it seems unwise to burden materialism – the view that every real concrete phenomenon in the universe is *physical* – with a commitment to the descriptive completeness of *physics* more or less as we know it. There may be physical phenomena which physics (and any non-revolutionary extension of it) cannot describe, and of which it has no inkling, either descriptive or referential.[7] Physics is one thing, the physical is another. "Physical" is a natural-kind term – it is the ultimate natural-kind term[8] – and no sensible person thinks that physics has nailed all the essential properties of the physical. Current physics is profoundly beautiful and useful, but it is in a state of chronic internal tension.[9] It may be added, with Russell and others, that although physics appears to tell us a great deal about certain of the general structural or mathematical characteristics of the physical, it fails to give us any further insight into the nature of whatever it is that has these structural or mathematical characteristics – apart from making it plain that it is utterly bizarre relative to our ordinary conception of it.

It is unclear exactly what this last remark amounts to (is it being suggested that physics is failing to do something it could do?), but it already amounts to

something very important when it comes to what is known as the "mind–body problem." Many take this to be the problem of how mental phenomena can be physical phenomena *given what we already know about the nature of the physical*. But those who think this are already lost. For the fact is that we have *no* good reason to think that we know anything about the physical that gives us any reason to find any problem in the idea that mental phenomena are physical phenomena. If we consider the nature of our knowledge of the physical, we realize that "no problem of irreconcilability arises."[10] Joseph Priestley saw this very clearly over 200 years ago, and he was not the first. Noam Chomsky reached essentially the same conclusion over thirty years ago, and he was not the last.[11] Most present-day philosophers take no notice of it and waste a lot of time as a result: much of the present debate about the "mind–body" problem is beside the point.

2 Terminology

I am going to use the plural-accepting, count-noun form of the word "experience" for talking of experience*s* as things (events) that may (and presumably do) have non-experiential being as well as experiential being. And I am going to reserve the adjective "experiential" and the plural-lacking form of the noun "experience" for talking about the qualitative character that experience*s* have for those who have them as they have them, where this qualitative character is considered wholly independently of everything else. The phenomenon of experiential[12] qualitative character is part of what exists – it is part of reality, whatever its ontological category – and it is essential to have some unequivocal way of referring to it and only to it.

One could express this terminological proposal by saying that "experiential phenomena" and "experience" (plural-lacking form) refer in a general way to that part of reality which one is left with when, continuing to live and think and feel as one does, one engages in an old skeptical thought experiment and imagines that the "external world," including one's own body, does not exist. They refer to the part or aspect of reality one has to do with when one considers experience*s* specifically and solely in respect of the experiential qualitative character they have for those who have them as they have them, and puts aside the fact that they may also be correctly describable in such non-experiential terms as "a 70-20-30 Hertz coding triplet across the neurons of area V4."[13]

It is easy to forget the force of this ruling, and I will mark it by giving "experiential" and "experience" capital letters.

3 Realistic Materialism

Realistic materialists – realistic anybodies – must grant that Experiential phenomena are real, concrete phenomena, for nothing in this life is more certain.[14] They must therefore hold that they are physical phenomena. It may sound odd to use the word "concrete" to characterize the qualitative character of experiences of color, gusts of depression, thoughts about diophantine equations, and so on, but it isn't, because "concrete" simply means "not abstract."[15] For most purposes, one may take "concrete" to be coextensive with "possessed of spatiotemporal existence," although this will be directly question-begging in some contexts.[16]

It may also sound odd to use "physical" to characterize mental phenomena like Experiential phenomena: many materialists talk about the mental and the physical as if they were opposed categories. But this, *on their own view*, is like talking about cows and animals as if they were opposed categories. For every concrete phenomenon in the universe is physical, according to materialists. So all mental phenomena, including Experiential phenomena, are physical phenomena, according to materialists; just as all cows are animals.

So what are materialists doing when they talk, as they so often do, as if the mental and the physical were entirely different? What they may mean to do is to distinguish, within the realm of the physical, which is the only realm there is, according to them, between the mental and the non-mental, or between the Experiential and the non-Experiential; to distinguish, that is, between mental (or Experiential) features of the physical, and non-mental (or non-Experiential) features of the physical.[17]

It is this difference that is in question when it comes to the "mind–body" problem; materialists who persist in talking in terms of the difference between the mental and the physical perpetuate the terms of the dualism they reject in a way that is inconsistent with their own view. I use the words "mental" and "non-mental" where many use the mords "mental" and "physical" simply because I assume, as a (wholly conventional) materialist, that every real concrete phenomenon is physical, and find myself obliged to put things in this way.[18]

There is tremendous resistance to abandoning the old mental/physical terminology in favor of the mental/non-mental, Experiential/non-Experiential terminology, although the latter seems to be exactly what is required. Many think the old terminology is harmless, and a few are not misled by it: they consistently use "physical" to mean "non-mental physical." But it sets up the wrong frame of thought from the start, and I suspect that those who are never misled by it are members of a small minority.

When I say that the mental, and in particular the Experiential, is physical, and endorse the view that "experience is really just neurons firing," I mean something completely different from what some materialists have apparently meant by saying such things. I don't mean that all aspects of what is going on, in the case

of conscious experience, can be described by current physics, or some non-revolutionary extension of it. Such a view amounts to radical "eliminativism" with respect to consciousness,[19] and is mad. My claim is different. It is that the Experiential (considered just as such)[20] – the feature of reality we have to do with when we consider experiences specifically and solely in respect of the Experiential character they have for those who have them as they have them – that "just is" physical. No one who disagrees with this is a remotely realistic materialist.

When aspiring materialists consider the living brain, in discussion of the "mind–body problem," they often slide into supposing that the word "brain" somehow refers only to the brain-as-revealed-by-current-physics. But this is a mistake, for it refers, just as it says, to the living brain, i.e., the living brain as a whole, the brain in its total physical existence and activity. Realistic – real – materialists must agree that the total physical existence and activity of the brain of an ordinary, living person, considered over time, is *constituted* by Experiential phenomena (if only in part) in every sense in which it is constituted (in part) by non-Experiential phenomena characterizable by physics. A real (realistic) materialist cannot think that there is something still left to say about Experiential phenomena, once everything that there is to say about the physical brain has been said.

4 Materialism Further Defined

Materialism, then, is the view that every real concrete phenomenon is physical in every respect, but a little more needs to be said, for Experiential phenomena – together with the subject of experience, assuming that that is something extra – are the only real, concrete phenomena that we can know with certainty to exist,[21] and as it stands this definition of materialism doesn't even rule out idealism – the view that mental phenomena are the only real phenomena and have no non-mental being – from qualifying as a form of materialism! Now there is a sense in which this consequence of the definition is salutary (see, e.g., sections 14–15 below), but it would none the less be silly to call an idealist view "materialism." Russell is right to say that "the truth about physical objects *must* be strange,"[22] but it is reasonable to take materialism to be committed to the existence of non-Experiential being in the universe, in addition to Experiential being, and I shall do so in what follows.

It is also reasonable to take materialism to involve the claim that *every* existing concrete phenomenon has non-mental, non-Experiential being, whether or not it also has mental or Experiential being. Applied to mental phenomena, then, materialism claims that each particular mental phenomenon essentially has non-mental being, in addition to mental being. This is, I think, the standard view.[23]

I will assume, then, that all realistic materialists take it that there is both

mental and Experiential being and non-mental, non-Experiential being. Must all realistic *monists* also take it that there is non-mental, non-Experiential being? Many would say Yes, on the grounds that it is not remotely realistic to suppose either that there is, or might be, no non-mental or non-Experiential being at all. But the question of what it is to be (metaphysically) realistic is far harder here than it is when it is merely the existence of Experience that is in question. For the purposes of this paper I will *assume* that any realistic position does take it that there is non-mental or non-Experiential being in addition to mental and Experiential being, for this assumption accords with ordinary conceptions, and my main argument does not require me to challenge it. But it is at best an assumption. Idealists, of course, reject the assumption that realistic monism requires acknowledgment of non-mental, non-Experiential phenomena, and I will enter a number of reservations along the way.[24]

It is clumsy to oscillate between "mental" and "Experiential," or constantly double them up, and in the next few sections I will run the discussion in terms of the mental/non-mental distinction (such as it is). This said, all my *examples* of mental phenomena will be Experiential phenomena, for they suffice to make the relevant point and are, in the present context, what matter most.

It may be added that the reference of the term "Experiential" is much clearer than that of the essentially contestable term "mental," and that the latter may in the end deserve the treatment proposed for the term "physical" in section 15 below. Nevertheless it seems best to begin in this way.[25]

I will quote Russell – post-1926 Russell – frequently when discussing materialism, for my views converge with his in certain respects, and he has been wrongly ignored in recent discussion.[26] He was still inclined to call himself a "neutral monist" at that time, but he is equally well read as a thoroughgoing materialist.[27] He rejects materialism in name, pointing out that "matter has become as ghostly as anything in a spiritualist séance" – it has, he says, disappeared "as a 'thing'" and has been "replaced by emanations from a locality"[28] – but he grants that "those who would formerly have been materialists can still adopt a philosophy which comes to much the same thing. They can say that the type of causation dealt with in physics is fundamental, and that all events are subject to physical laws."[29] And this, in effect, is what he does himself.[30]

5 "Mental" and "Non-mental"

It may seem odd to take "mental" as the basic positive term when characterizing materialism. But one is not a thoroughgoing materialist if one finds it so. For all materialists hold that every concrete phenomenon in the universe is physical, and they are neither sensible nor realistic if they have any inclination to deny the concrete reality of mental phenomena like Experiential phenomena.[31] It follows

that they have, so far, no reason to find it odd or biased to take "mental" rather than "non-mental" as the basic term.

Surely it would be better, even so, to start with some positive term "T" for the non-mental physical, and then define a negative term, "non-T," to cover the mental physical; or use a pair of independently positive terms?

There are two good reasons for taking "mental" as the basic positive term, one terminological, the other philosophical. The terminological reason is simply that we do not have a convenient positive term for the non-mental (obviously we can't use "physical," and there is no other natural candidate). The philosophical reason is very old: it is that we have direct acquaintance with – know – fundamental features of the mental nature of (physical) reality just in having experience in the way we do, in a way that has no parallel in the case of any non-mental features of (physical)[32] reality. We do not have to stand back from experiences and take them as objects of knowledge by means of some further mental operation, in order for there to be acquaintance and knowing of this sort: the having is the knowing.[33]

This point has often been questioned, but it remains immovable. Russell may exaggerate when he says that "we know *nothing* about the intrinsic quality of physical events except when these are mental events that we directly experience,"[34] or that "as regards the world in general, both physical and mental, *everything* that we know of its intrinsic character is derived from the mental side,"[35] for it is arguable that the spacetime character of the world is part of its intrinsic character, and, further, that we may have some knowledge of this spacetime character. I don't think he exaggerates much, however. He is on to something important, and the epistemological asymmetry between claims to knowledge of Experiential being and claims to knowledge of non-Experiential being is undeniable, however unfashionable.

The asymmetry claim that concerns me is not the claim that all epistemic contact with concrete reality involves experience, and that we are inevitably a further step away from the thing with which we are in contact when it is a non-Experiential phenomenon. It is, rather, the claim that we are acquainted with reality *as it is in itself*, in certain respects, in having Experience as we do. This second claim revolts against the tendency of much current epistemology and philosophy of mind, but there is no reason why it should trouble thoughtful materialists, and I will offer a brief defense of it in section 13. Here it is worth noting that it is fully compatible with the view that there may also be fundamental things we don't know about matter considered in its Experiential being.[36]

6 Aside: "As It is in Itself"

Does one need to defend the phrase "as it is in itself," when one uses it in philosophy? I fear one does, for some think (incoherently) that it is somehow incoherent. Still, it is easy to defend. The supposition that reality is in fact a certain way, whatever we can manage to know or say about it, is obviously true. To be is to be somehow or other. Nothing can exist or be real without being a certain way at any given time.[37] And the way something is just is how it is in itself. This point is not threatened by the suggestion that our best models of the behavior of things like photons credit them with properties that seem incompatible to us – e.g., wave-like properties and particle-like properties. What we learn from this is just that this is how photons affect us, given their intrinsic nature – given how they are in themselves, and how we are in ourselves. We acquire no reason to think (incoherently) that photons do not have some intrinsic nature at any given time. Whatever claim anyone makes about the nature of reality – including the claim that it has apparently incompatible properties – just is a claim about the way it is. This applies as much to the Everett "many-worlds" theory of reality as to any other.

Some think that what we learn from quantum theory is that there is, objectively, no particular way that an electron or a photon is, at a given time. They confuse an epistemological point about undecidability with a metaphysical claim about the nature of things. The problem is not just that such a claim is unverifiable. The problem is that it is incoherent. For whatever the electron's or photon's weirdness (its weirdness-to-us: nothing is intrinsically weird), its being thus weird just is the way it is.

So we may talk without reservation of reality as it is in itself. Such talk involves no (allegedly dubious) metaphysics of the Kantian kind. Its propriety derives entirely and sufficiently from the thought that if a thing exists, it is a certain way. For the way it is just is how it is in itself.

7 Structure and Structured

So much, for the moment, for our theoretical conception of the mental: it has some securely anchored, positive descriptive content, and we can know that this is so; for whatever the best general account of the mental, it includes Experiential phenomena in its scope; and Experiential phenomena are not only indubitably real; they are also phenomena part of whose intrinsic nature just is their Experiential character; and their Experiential character is something with which we are directly acquainted, however hard we may find the task of describing it in words. This is so even if we can make mistakes about the nature of our experiences, and even if we can do so even when we consider them merely in

respect of their (Experiential) qualitative character.[38] It is so even if we differ dramatically among ourselves in the qualitative character of our experiences, in ways we cannot know about.

Our theoretical conception of the mental, then, has clear and secure descriptive content. (Don't ask for it to be put further into words; the anchoring is sufficiently described in the last paragraph.) Our theoretical conception of the non-mental, by contrast, remains, so far, a wholly negative concept. It has, as yet, no positive descriptive content.

Is Russell right? Something needs to be said about his use of the word "intrinsic." It is potentially misleading, and it helps to consider other ways in which he puts the point. Thus he talks regularly of the "abstractness" of physics. The knowledge it gives is, he says, "purely formal." It reveals the abstract "structure" of physical phenomena while saying nothing about their "quality."[39]

Can anything be done about this? On one reading, Russell thinks not: the science of physics is our fundamental way of attempting to investigate the non-mental being of physical reality, and it cannot help us. "Physics is mathematical," he says, "not because we know so much about the physical world, but because we know so little: it is only its mathematical properties that we can discover. For the rest, our knowledge is negative." "We know nothing about the intrinsic quality of physical events except when these are mental events that we directly experience." On this view, neither physics nor ordinary experience of physical objects gives us any sort of knowledge of the intrinsic nature of non-mental reality.[40]

I am not sure that the distinction between structure and quality is clear, or fundamental in such a way that it holds "all the way down,"[41] but (putting that doubt aside) it seems that the fundamental distinction that Russell has in mind can be expressed by saying that it is a distinction between *how X is structurally disposed* and *what X is apart from (over and above) its structural disposition*.[42] Physics gives the structure, but not the structure-transcendent nature, of the thing that has the structure. If we say that truths about how X is structurally disposed have purely *structure-specifying* content, while truths about what X is over and above its structural disposition also have *structure-transcendent* content, or, more simply, *non-structural* content, then we may say that "non-structural" covers everything that Russell has in mind when he talks of the "intrinsic" nature of things.[43]

One might dramatize Russell's idea by saying that physics can be thought of as a formal system which remains, in a peculiar sense, an *uninterpreted* formal system, even though we know that it *applies* to something = x – reality, the universe – and even though it is elaborated specifically in causal response to x. On this "Ramseyfied" view, we may suppose that the universe has features that are *structurally isomorphic* to the structures delineated in the equations of physics, but we have no account of the non-structural nature of the thing that has the structure(s) in question.[44]

So we are (to pursue the metaphor) in the peculiar position of having a known,

concrete *application* (and so, in one sense, an *interpretation*) for a formal system, without that application constituting a *model* (in the sense of model-theoretic semantics) that can confer positive descriptive meaning on its terms. In being the subject matter of physics, the universe provides it with a merely referential model or object, of which it gives a merely structure-specifying description. Physics is *about* the physical, and may give a correct abstract representation of its structural disposition as far as it goes; but it does not and cannot tell us anything about what the physical actually is, over and above the fact that it exemplifies a certain formal structure.[45]

8 The Non-mental – Space

Back now to the question whether physics endow our general theoretical conception of the non-mental with any positive descriptive (not merely referential) content. Russell in 1927 thinks not. I disagree because correct structural description of a thing is already description of a feature of its intrinsic nature. But this disagreement is merely terminological, and the real question is this: can one go any further than structure-specifying content, when attempting to give a satisfactory theoretical characterization of the non-mental? Again, Russell in 1927 thinks not. It seems to me, however, that we may be able to go a little further. For I think that our ordinary conception of space may get something fundamental right about the nature of reality as it is in itself, and hence about the intrinsic nature of reality – something that survives even after the finite-but-unbounded curved gravity-constituting spacetime of relativity theory (or the ten- or eleven- or twelve-dimensional spacetime of one of the leading versions of string theory) has been granted to be closer to the truth.

I am tempted to hold up my hands, like G. E. Moore, and to consider, not my hands, but the space – by which I mean only the spatial extension – between them, and to say: "This is space (spatial extension), and it is real, and I know its nature, in some very fundamental respect, whatever else I do not know about it or anything else (e.g., the fact that it is an aspect of spacetime)." On this view the ordinary concept of space, or indeed the concept of spacetime, in which (I claim) a fundamental feature of our ordinary conception of space survives, has correct non-structural descriptive content. It does not relate only to "what we may call the causal skeleton of the world,"[46] if to say this is to say that it does not capture any aspect of the non-structural nature of the world. It has non-structural content, and can transmit this content to our more general conception of the non-mental.[47]

Russellians may object as follows. This line of thought is profoundly natural, but it depends on a fundamentally false imagining. It involves the conflation of "objective" space, space (spatial extension) "as it is in itself" (where this is taken as a merely referential, structural-equivalence-class specifying term with no

pretension to non-structural content) with the phenomenological space (or spaces) associated with perception. It involves an almost irresistible but entirely fatal failure to "realize what an abstract affair form really is."[48] All those, like yourself, who think that it is viable are "guilty, unconsciously and in spite of explicit disavowals, of a *confusion in their imaginative picture of matter*."[49]

I think that some of those who take this line are suffering from excessive empiricism. They take it that the notion of spatial extension – or indeed shape – that we possess is essentially informed by the character of our sensory experiences, and in this I think they are mistaken. It may well be true that sensory experiences of specific kinds are necessary for the acquisition of concepts like SHAPE or SPACE, in the case of beings like ourselves.[50] Such concepts can nevertheless float free of the different possible sensory bases of their acquisition and subsequent deployment, without *ipso facto* becoming "merely" formal or structure-specifying in character. It is easy to see that grasp of the content of SHAPE (say) does not require essential reference to any specific sensory experience. It suffices to point out that exactly the same concept of shape – i.e., *the* concept of shape, for there is only one – can plausibly be supposed to be fully masterable by two different creatures A and B on the basis of sensory experiences in entirely different sensory modalities familiar to us – sight and touch.[51] One has to endorse a rather crude form of meaning-empiricism or concept-empiricism to suppose that A and B do not – cannot – have the same concept, as they do geometry together. A concept is not a faint copy or transform of a sensory experience. It is, precisely, a concept.

That's one point. Another, crucial in this context, is that the concept of shape or space that A and B have in common is not an entirely abstract or purely formal concept, as the supporters of Russell seem to suggest. There is more to A and B sharing the specific concept SHAPE or SPACE than there is to their sharing mastery of the principles of an uninterpreted formal system that is in fact suitable for the expression of shape configurations or spatial relations although they know it only as an uninterpreted formal system. It is precisely because pure form is such a *very* "abstract affair," as Russell says, that the concept of shape or space that A and B can have in common in spite of their different sensory experiences cannot be supposed to be a matter of pure form. To think that it is a matter of pure form is to miss out precisely their grasp of the *spatiality* of space – of that which makes their grasp of the concept of space more than grasp of (say) an abstract metric. The concept has non-structural content.

It is true that this content is abstract in one sense: it is abstract relative to all the particularities of sensation, in a way that is sufficiently indicated by reference to the fact that different creatures can acquire it (the very same concept) on the basis of experience in entirely different sensory modalities. It is indeed, and essentially, a *non-sensory* concept.[52] But it is not purely abstract in Russell's sense, because (to repeat) it involves grasp of the spatiality – rather than what one might call the mere abstract dimensionality – of space.[53] Spatiality is not abstract

dimensionality. The nature of abstract dimensionality can be fully captured by a purely mathematical representation. The nature of spatiality cannot.

Obviously questions arise about the precise nature of the non-structural content of concepts like SHAPE and SPACE, about what it is, exactly, to grasp the spatiality of space, given that SHAPE and SPACE may be fully shared by A, B, superbats, and others. But in the present context I am inclined just to hold up my hands again.[54]

Russellians may be unimpressed. Michael Lockwood, in particular, is sympathetic to the idea that knowledge of spacetime structure is not knowledge of any feature of the "intrinsic" or non-structural nature of reality. In doing physics, Lockwood says, we may grasp the abstract structure exemplified by space while having "no conception of its content: i.e. what it is, concretely, that fleshes out this structure. (For all we know, on this view, Henry More and Newton may be right in equating space with God's sensorium!)"[55]

But I am prepared to grant this. I am prepared to grant that we cannot rule out the possibility that space is God's sensorium,[56] or something even more unknown, and that there is therefore a sense in which we may have no idea of what it is that "fleshes out" the abstract structure exemplified by space. For it may still be true that one grasps something fundamental about the non-structural nature of space in thinking of it as having, precisely, spatiality, rather than mere abstract dimensionality. If space is God's sensorium, so be it: God's sensorium may really have the property of spatiality. Between a fat-free, purely mathematical and thus wholly abstract representation of the structure of space and a partly structure-transcending conception of space as God's sensorium (or some such) lies a third option: an ostensibly less rich but still structure-transcending conception of space as specifically spatial in its dimensionality. Some may think this a fine point, but it is (I take it) a huge step away from Russell's claim that we know *nothing* about the intrinsic quality of non-mental events.[57]

I am not claiming that we do know something about the non-structural nature of space, only that we may (I hold up my hands, I move them apart – but my sense of the vulnerability of this claim has increased since I wrote this paper in 1997). This claim allows, as it should, that there may well be more to space than we can know. SPACE, like PHYSICAL, is a natural-kind concept, and there are some atrociously good reasons for thinking that there is more to space than we know or can understand. Putting aside the (already weighty) points that physical space is non-Euclidean, and is itself something that is literally expanding,[58] and the non-locality results,[59] and questions about the nature of the vacuum, and widespread agreement that "there is no good a priori reason why space should be a continuum,"[60] I for one can't fully understand how space and time can be interdependent in the way that they demonstrably are. We are also told on very good authority that gravity is really just a matter of the "curvature" of space, and that string theory is an immensely promising theory of matter (especially after the "second superstring revolution" and the growth of M-theory, and especially

when it comes to understanding gravity) that entails that there are at least ten spatial dimensions . . .

These points reopen the connection to the mind–body problem. For as they pile up, one can't reputably hold on to the old, powerful-seeming Cartesian intuition that there is a "deep repugnance" or incompatibility between the nature of conscious experience and the nature of spatial extension – the intuition that "the mental and the spatial are mutually exclusive categories."[61] We have direct acquaintance with fundamental features of conscious experience – Experiential features – just in having it; but we really have no good reason to think that we know enough about the nature of space – or rather, about the nature of matter-in-space-considered-insofar-as-it-has-non-mental-being – to be able to assert that there is any repugnance.[62] And if conscious experience is in time, as almost everyone agrees, then it is in spacetime, given the way in which space and time are demonstrably interdependent – in which case it is in space in every sense in which it is in time.

Note that it follows that even if our notion of space can confer some non-structural content on our best theoretical conception of the non-mental, it cannot confer any content that is guaranteed to distinguish it from any fully articulated theoretical conception of the mental, although we still intuitively feel it to fit with the former conception in a way in which we don't feel it to fit with the latter.[63]

9 The Non-mental – Spin, Mass, and Charge

I have proposed that our theoretical conception of the non-mental may be able to acquire some non-structural content from its first lieutenant, the concept of space. Can it acquire any more? Well, I think that our more particular spatial concepts of shape, size, position, distance, and local motion (I raise my hands and bring them together) *may* also get something right about reality as it is in itself, and so contribute to the non-structural content of our general theoretical conception of the non-mental; I think Locke may be essentially right in his view that some of our ideas of primary qualities correctly represent how things are in themselves, although his account needs recasting.[64] It may also be that our ordinary conception of time gets something right about the nature of reality (both Experiential and non-Experiential) – even if we need to conceive time as part of spacetime in order to think about it properly. I just don't know.[65]

Going on from space, time, extension, shape, position, distance, and motion, in the attempt to give a positive characterization of the non-mental, one may want to mention properties like spin, mass, charge, gravitational attraction, "color" and "flavor" (in the quantum-theoretic sense). But one will have to bear in mind that our grasp of these things – any grasp of them over and above that which is conveyed by their intimate relation to spatiotemporal concepts – is

expressed merely in equations,[66] and the truth in Russell's remark that physics is mathematical not because we know so much about the physical world, but because we know so little. So although I like to think that spatiotemporal concepts carry non-structural content, I do not think this can be true of any of these other concepts considered independently of their relations to spatiotemporal concepts. Here Russell is right: we know nothing of the non-mental non-structural nature of – for example – electrical phenomena apart from their spacetime structure; all we have are equations.[67]

But even if knowledge of spacetime structure is all we have, in the way of non-structural knowledge of the nature of the non-mental, it makes a huge difference to the case. Consider the difference between a characterization of the forces of electrical attraction and repulsion in which their spatial character (the way they decrease with increasing distance) is given a purely abstract-dimensional interpretation, and one in which it is given a genuinely spatial interpretation. Consider any account of anything in which time relations have a merely abstract representation, and one in which the temporality of time is genuinely represented.

10 Hens' Eggs

I want now to give a further characterization of what it is to be a genuine materialist. But I must first answer one more objection that occurs to many.

It seems to follow, from your claim that we have no knowledge of the non-structural, intrinsic or as we may say N-intrinsic *nature of things, that we cannot know that there are tables and chairs and hens and hens' eggs and "that hens' eggs are generally laid by hens."*[68] *But this is a chair I'm sitting in, and it's made of wood, and this is a hen, and this is a hen's egg, and this hen laid it. These are all facts I know, and they are N-intrinsic facts – ultimate, absolute truths – about the nature of reality. They must be included in any true and full account of the history of the universe.*

My reply to this objection is similar to Moore's a hundred years ago. I agree that we know many such truths, but I take it, as a materialist, that hens are wholly made of the fundamental constituents of matter that physics discusses, and that when we consider our knowledge of these fundamental constituents we encounter the crucial and entirely general sense in which we know nothing about the fundamental N-intrinsic nature of matter. As far as I can see, this ignorance is entirely compatible with the sense in which we do have knowledge of the N-intrinsic nature of reality in knowing that there are hens, and what hens are, and what wood is, and so on. And this compatibility is no more surprising than the fact that I can know that this is a statue without knowing what it is made of.

But we know what hens are made of – carbon, hydrogen, and oxygen, mostly – and we know what carbon, hydrogen, and oxygen are made of – electrons and quarks with various characteristics. Physics gives us knowledge of the properties of these things. If you think that it fails to give us any knowledge of their ultimate, N-intrinsic nature that's because you think that a thing is more than its properties. But that's just bad old metaphysics. A thing is not in any sense more than its properties.

I agree that there is an inescapable sense in which it is true to say that a thing is not more than its properties – I agree that "in their relation to the object, the properties are not in fact subordinated to it, but are the way of existing of the object itself"[69] – but the present claim is not that a concrete phenomenon must be more than its properties, but that it must be more than its purely formal or structural properties. If you say that this is more bad metaphysics, a yearning for lumpen stuff, our disagreement will be plain. My reply will be that you have forgotten "what an abstract affair form really is." A concrete phenomenon must be more than its purely formal or structural properties, because these, considered just as such, have a purely abstract mathematical representation, and are, concretely, nothing – nothing at all. It is true that we get out of the realm of the purely abstract when we add in spatiotemporal properties, on my account, but a thing's non-structural properties can't consist only in its spatiotemporal properties – at least so long as spacetime is conceived, as it generally is, merely as a dimensional manifold with no physical or substantial nature.[70]

Here, then, we return to the point that – the sense in which – we have no knowledge of the N-intrinsic nature of things in spite of the sense in which it is true to say that we know what hens and hens' eggs are.

11 True Materialism

I have suggested that our general theoretical conception of the mental has substantial non-structural descriptive content, because we have acquaintance with fundamental features of the mental nature of reality just in having experience in the way we do. Our general theoretical conception of the non-mental has substantial structure-specifying content, and I have suggested, with some hesitation, that it may also have crucial and correct *non*-structural content deriving from spatiotemporal concepts. Apart from this, though, it is arguable (subject to note 47) that we know nothing about the intrinsic or non-structural nature of non-mental reality.

With this in place, we may ask what it is to be a genuine materialist. The first thing to do is to intone once more that realistic or real materialism entails full acknowledgment of the reality of Experiential phenomena: they are as real as rocks, hence wholly physical, strictly on a par with anything that is correctly

characterized by physics.[71] They are part of fundamental reality, whatever is or is not the case.

It follows that current physics, considered as a general account of the general nature of the physical, is like *Othello* without Desdemona: it contains only predicates for non-Experiential being, so it cannot characterize Experiential being at all (recall the definition in section 2). It cannot characterize a fundamental feature of reality at all.

No one who doubts this is a true materialist. Partly for this reason, I think that genuine, reflective endorsement of materialism is a considerable achievement for anyone who has had a standard modern Western education. Materialism must at first provoke a feeling of deep bewilderment in anyone contemplating the question "What is the nature of the physical?" The occurrence of such a feeling is diagnostic of real engagement with the materialist hypothesis, and hence with the thought that Experiential phenomena are physical phenomena, just like extension phenomena and electrical phenomena insofar as they are correctly characterized by physics (or indeed common sense). I think Russell is profoundly right when he says that most are "guilty, unconsciously and in spite of explicit disavowals, of a confusion in their imaginative picture of matter." I suspect that some will be unable to shake off the confusion, although Locke made the crucial move long ago. Some may say that modern science has changed the situation radically since Locke's time. It has – but only insofar as it has massively reinforced Locke's point.

Perhaps I am generalizing illegitimately from my own experience, revealing my own inadequacy rather than the inadequacy of recent discussion of the "mind–body" problem. I don't think so. Materialism requires concerted meditative effort. Russell recommends "long reflection."[72] If one hasn't felt a kind of vertigo of astonishment, when facing the thought, obligatory for all materialists, that consciousness is a wholly physical phenomenon in every respect, including every Experiential respect – a sense of having been precipitated into a completely new confrontation with the utter strangeness of the physical (the real) relative to all existing commonsense and scientific conceptions of it – then one hasn't begun to be a thoughtful materialist. One hasn't got to the starting line.[73]

Some may find that this feeling recurs each time they concentrate on the mind–body problem. Others may increasingly think themselves – quietistically, apophatically, pragmatically, intuitively – into the unknownness of the (non-mental) physical in such a way that they no longer experience the fact that mental and non-mental phenomena are equally physical as involving any clash. At this point "methodological naturalism" – the methodological attitude to scientific inquiry into the phenomena of mind recommended by Chomsky – will become truly natural for them, as well as correct.[74] I think it is creeping over me. But recidivism is to be expected: the powerfully open state of mind required by true materialism is hard to achieve as a natural attitude to the world. It involves a profound reseating of one's intuitive theoretical understanding of things.[75]

I say "intuitive theoretical understanding," but it isn't as if there is any other kind, when the stress falls on the word "understanding." For (briefly) what we think of as real *understanding* of a natural phenomenon is at bottom a *feeling*, and it is *always* and *necessarily* relative to other things one just takes for granted, finds intuitive, feels comfortable with. This is true in science as it is in common life. I feel I fully understand why this tower casts this shadow in this sunlight, given what I take for granted about the world (I simply do not ask why light should do *that*, of all things, when it hits stone). I may also feel I understand – see – why this billiard ball does *this* when struck in this way by that billiard ball. But in this case there is already a more accessible sense in which I don't really *understand* what is going on, and it is an old point that if I were to ask for and receive an explanation, in terms of impact and energy transfer, this would inevitably invite further questions about the nature of impact and energy transfer, starting a series of questions and answers that would have to end with a reply that was not an explanation but rather had the form "Well, that's just the way things are."[76]

The true materialist outlook may become natural for some, then, but many will find they can maintain it only for relatively short periods of time. It is not a small thing. To achieve it is to have evacuated one's natural and gripping commonsense ± science-based conception of the nature of the physical of every element that makes it seem puzzling that Experiential phenomena are physical. I think it is to be at ease with the idea that consciousness is a form of matter.[77]

It can help to perform special acts of concentration – focusing one's thought on one's brain and trying to hold fully in mind the idea that one's experience as one does so is part of the physical being of the brain (part of the physical being of the brain that one may be said to be acquainted with as it is in itself, at least in part, because its being as it is for one as one has it just is what it is in itself, at least in part). It is worth trying to sustain this – it is part of doing philosophy –, forcing one's thought back to the confrontation when it slips. At first one may simply encounter the curious phenomenological character of the act of concentration, but it is useful to go on – to engage, for example, in silent, understanding-engaging subvocalizations of such thoughts as "I am now thinking about my brain, and am thinking that this experience I am now having of this very thinking – and this subvocalization – is part of the physical activity and being of my brain." It is also useful to look at others, including young children, as they experience the world, and to think of the common-or-garden matter that is in their heads (hydrogen, oxygen, carbon, iron, potassium, sodium, and so on). It is useful to listen to music, and focus on the thought that one's auditory experience is a form of matter.[78]

12 Knowledge of Ignorance

Finding it deeply puzzling how something could be physical is not the same as finding something that one takes to be physical deeply puzzling. It is often said that quantum theory is deeply counterintuitive – e.g., in its description of the wave-like and particle-like behavior of fundamental particles, but no one seems to find it puzzling to suppose that it deals wholly with physical phenomena.[79]

The main reason for this seems to be as follows: WAVE and PARTICLE engage smoothly with standard physics concepts of shape, size, position, motion, and so on. There is, so far, a clear sense in which the two concepts are *theoretically homogeneous*, or at least non-heterogeneous; they operate on the same, single conceptual playing field of physics.[80] But when we try to integrate conscious-experience terms with the terms of physics (and commonsense physics), we find that they entirely lack any such felt theoretical homogeneity, or non-heterogeneity. To this extent, they force constantly renewed bewilderment – in a way quite different from the way in which quantum-mechanical phenomena do – on materialists who like to think they have *some* sort of coherent, theoretically unified understanding of the overall nature of the physical, however general that understanding may be, and however incomplete in its details.

But this is the central mistake: to think that one has some sort of theoretically unified understanding of the overall nature of the physical. Once one realizes that this cannot be true, if materialism is true, things change.[81] It begins to look as if there is actually *less* difficulty in the suggestion that physical phenomena have both Experiential and non-Experiential being than in the suggestion that photons (e.g.) behave like both particles and waves. For in the case of Experiential terms and non-Experiential terms there is no direct clash of concepts of the sort that occurs in the case of the wave–particle duality. Being a wave is incompatible with being a particle, but there is nothing in the possession of non-Experiential being that we know to be intrinsically inimical to the possession of Experiential being: we simply do not know enough about the nature of non-Experiential being to have any good reason to suppose that this might be so. Thus the Experiential terms and the non-Experiential terms do not in fact *actively clash*, as the wave and particle terms do. Rather, they fail to connect or engage. One is making progress as a materialist when one has lost all sense of an active clash. It has no scientific or philosophical justification. As Russell says, "the physical [sc. non-mental] world is only known as regards certain abstract features of its spacetime structure – features which, because of their abstractness, do not suffice to show whether the physical world is, or is not, different in intrinsic character from the world of mind."[82]

Arnauld made the essential point in 1641, in his comments on Descartes's *Meditations*, and he was not the first.[83] Locke in 1690 "did not apprehend that there was any real inconsistency between the known properties of body, and

those that have generally been referred to mind."[84] Algarotti observes in 1737 that "we are as yet but Children in this vast Universe, and are very far from having a compleat Idea of Matter; we are utterly unable to pronounce what Properties are agreeable to it, and what are not,"[85] and Hume in 1739 shows a very clear understanding of the point.[86] Priestley in 1777 argues, with unanswerable force, and by appeal to a scientific conception of the physical that (in essence) still holds good today, "that we have *no reason* to suppose that there are in man two substances so distinct from each other as have been represented."[87] Kant concurs in 1781, although his special terms of debate preclude him from agreeing directly with Priestley's further materialist claim that "mind . . . is not a substance distinct from the body, but the result of corporeal organization"; that "in man [thought] is a property of the nervous system, or rather of the brain"; that "sensation and thought do necessarily result from the organization of the brain."[88] The quality of the mind–body debate is in many ways lower today than at any other time in the last 300 years.

Substance dualism may have looked like a plausible response to the mind–body problem in Descartes's time, for classical *mechanistic* materialism, according to which the physical world consists entirely of small, solid, intrinsically inert particles in motion, was then the dominant view, and Leibniz's famous image of the mill seemed hard to counter.[89] But the strict mechanist understanding of the physical world was fatally undermined by 1687, when Newton published his *Principia*.[90] Since then we have had no good scientific reason to think that mind is not physical. And even before Newton, in the high days of "contact mechanics," there were no *philosophically* respectable grounds for claiming that mind is not physical. The mechanists or "Cartesians," as Hume calls them, made a wholly unjustifiable move: they "established it as a principle that we are *perfectly acquainted* with the essence of matter."[91] That is, they not only assumed that their fundamental theory of matter was sound as far as it went; they also assumed that it went all the way – that it was complete. It is the second of these two false assumptions that causes most trouble, for even if the Cartesians had been right that all physical change is a matter of the motion, contact, and impact of solid particles, they still would not have been justified in claiming that this fact was definitely – knowably – incompatible with some of it also being a matter of conscious goings-on. Many today make exactly the same sort of mistake.

13 The Reality of Appearance

I have claimed that thoughtful materialism requires draining one's conception of the non-Experiential physical of any element that, in a puzzling world, makes it seem especially puzzling that the Experiential is physical. Many philosophers – all those legions who tried, for most of the twentieth century, to reduce the mental to the non-mental in some way – think this is the wrong way round.

They think we have to drain our conception of the Experiential of any element that produces special puzzlement, leaving our existing conception of the non-Experiential physical in place. But no substantial draining can be done on the Experiential side. In having Experience in the way we do, we are directly acquainted with certain features of the ultimate nature of reality, as Russell and others have remarked – whether or not we can put what we know into words in any theoretically tractable way. And this is so whatever it is best to say about any non-Experiential (e.g., dispositional) aspects of the mental that there may be. We may certainly hope to *develop* our understanding of the nature of the Experiential; but we can do this only by adding to what we already know of it by direct acquaintance.

But in having experience we only have access to an appearance of how things are and are not cognizant, in the mere having of the experience, of how anything is in itself.[92]

The reply is immediate. Here, how things appear or seem is how they really are: the reality that is at present in question just is the appearing or seeming. In the case of any Experiential episode E there may be something X of which it is true to say that in undergoing E we only have access to an appearance of X, and not to how X is in itself. But serious materialists must hold that E itself, the event of being-appeared-to, with all the qualitative character that it has, is itself part of physical reality. They cannot say that it too is just an appearance, and not part of how things are, on pain of infinite regress. They must grant that it is itself a reality, and a reality with which we must, in plausibility, be allowed to have some sort of direct acquaintance. As Russell says, we must "treat 'seeming' with respect."[93]

At this point some may try to adapt Ryle-type arguments for the "systematic elusiveness of the 'I'" to the present case.[94] They may argue that anything that can count as *knowledge* of experience involves an operation of taking experience as an object that necessarily precludes apprehending it in such a way that one can be said to have access to how it is in itself, rather than merely to an appearance of it. Now I suspect that this form of argument is invalid even in its original application, where it is used to argue that the putative mental subject of experience can never directly apprehend itself.[95] But even if this is not so it has no valid application to the present case – to things like pain and color-experience. The way a color-experience is Experientially, for the subject of experience that has it, is part of its essential nature – its ultimate reality – as a physical phenomenon. When we claim (with Russell) that to have an experience is *eo ipso* to be acquainted with certain of the intrinsic features of reality, we do not have to suppose that this acquaintance involves standing back from the experience reflectively and examining it by means of a further, distinct experience. It doesn't. This picture is too cognitivist (or perhaps too German-Idealist). The having is the knowing.

14 The Radiance of Reality

I have argued that the first thing that one needs to do, when it comes to the mind–body problem, is to reflect on one's ignorance: one's ignorance of the non-Experiential. One's intuitive theoretical attitude to the nature of the non-Experiential needs to evolve until any sense that there is an active clash between Experiential terms and non-Experiential terms has disappeared, leaving only the awareness that they fail to connect in a way that brings a sense of intuitive understanding. This awareness ought not to be merely a matter of book learning.

At this point at least two paths open up for materialists. The first goes deeper into reflection on the nature of understanding in physics. Proceeding down this path, one encounters one's sense that at least some of the terms of physics (both commonsense and scientific) connect up with one another in a way that justifies a feeling of intuitive understanding of at least some of what goes on in the world. One is then asked to examine (possibly at length) the question of what exactly one supposes this to amount to. Does it really amount to anything very solid? Is it more than a certain kind of feeling one is disposed to get (either innately or as a result of training) when considering some but not other co-occurrences of features in the world? What exactly is its significance?[96]

Well, one probably has to go down this path, as a materialist, returning to the questions raised on p. 63. But I will choose another, which has a sunnier aspect. Here one confronts the deep puzzlement one still feels when one considers Experiential properties and non-Experiential properties and fails to see how they coexist, and, also, one's persisting feeling that this puzzlement has, in a puzzling world, a very special if not unique status.

The question is whether one can do anything about this. I think the answer is Yes. I think physics can help us – it has already helped us a great deal – by diluting or undermining features of our natural conception of the physical that make non-Experiential phenomena appear *toto coelo* different from Experiential phenomena.

The basic point is simple, and can be elaborated as follows. At first, perhaps, one takes it that matter is simply solid stuff, uniform, non-particulate: Scandinavian cheese. Then, perhaps, one learns that it is composed of distinct atoms – particles that cohere more or less closely together to make up objects, but that have empty space (to put it simplistically but intelligibly) between them. Then, perhaps, one learns that these atoms are themselves made up of tiny, separate particles, and full of empty space themselves.[97] One learns that a physical object like the earth or a person is almost all empty space. One learns that matter is not at all what one thought.

Now one may accept this while holding on to the idea that matter is at root solid and dense. For this picture retains the idea that there are particles of

matter: minuscule grainy bits of ultimate stuff that are in themselves perfectly solid (in Locke's phrase), "continuum-dense." And one may say that only these, strictly speaking, are matter: matter as such. But it is more than 200 years since Priestley (citing Boscovich) observed that there is no positive observational or theoretical reason to suppose that the fundamental constituents of matter have any perfectly solid central part.[98]

In spite of this, a fairly robust conception of truly solid particles survived all the way into pre-1925 quantum mechanics. It suffered its most dramatic blow only in modern (1925 on) quantum mechanics, in which neither the nucleus nor the electrons of an atom are straight-up solid objects, and are much more naturally thought of as fields. It may be said that the basic idea of the grainy particle survives even here, at least in as much as the nucleus and its components are still fairly well localized within a small central region inside the atom (albeit with small "tails" that go out to infinity), and in as much as the probability of finding one of the (far less localized) electrons is significant only within a volume that is normally considered to be the dimensions of the atom. But this commitment to the localization of particles does not in itself amount to any sort of commitment to continuum-dense solidity, but only to fields and repulsive forces that grow stronger without any clear limit when one travels in certain directions (i.e., towards the center of the field associated with a particle). And whatever is left of the picture of ultimate grainy bits is further etiolated in quantum-field theory, in which the notion of the field more fully overrides the picture of grainy particles.[99] In this theory it becomes very hard to treat "bound" systems like atoms at all. As for what I've been calling "empty space" – the supposed vacuum – it is understood to be simply the lowest energy state of fields like the electron, proton, and photon fields. It turns out to be something which "has structure and can get squeezed, and can do work."[100]

It may be said that quantum-field theory is complicated and ill understood, but there is a clear sense in which grainy, inert bits of matter, naively conceived, are already lost to us independently of quantum-field theory, given only the fact that matter is a form of energy, and interconvertible with it. This fact of interconvertibility is widely known, however little it is understood, and it seems to me that it further, and utterly, confounds any understanding of matter that takes it to be in any obvious way incompatible with consciousness. To put it dramatically: physics thinks of matter considered in its non-Experiential being as a thing of spacetime-located forces, energy, fields, and it can also seem rather natural to conceive of consciousness (i.e., matter apprehended in its Experiential being) as a spacetime-located form or manifestation of energy, as a kind of force, and even, perhaps, as a kind of field.[101] We may still think the two things are deeply heterogeneous, but we have no good reason to believe this.[102] We just don't know enough about the nature of matter considered in its non-Experiential being; and doubtless there are things we don't know about matter considered in its Experiential being. Those who think speculations like this are enjoyable but

not really serious haven't really begun on the task of being a materialist; they haven't understood the strangeness of the physical and the extent of our ignorance. It is a long time since Russell argued that "from the standpoint of philosophy the distinction between physical and mental is superficial and unreal," and it seems that physics can back philosophy on this question.[103] In fact – and it had to come back to this – we really don't know enough to say that there is any non-mental being. All the appearances of a non-mental world may just be the way that physical phenomena – in themselves entirely mental phenomena – appear to us; the appearance being another mental phenomenon.[104]

Whatever you think of this last proposal, lumpish, inert matter, dense or corpuscled, stuff that seems essentially alien to the phenomenon of consciousness, has given way to fields of energy, essentially active diaphanous process-stuff that – intuitively – seems far less unlike the process of consciousness. When Nagel speaks of the "squishy brain," when McGinn speaks of "brain 'gook'" and asks how "technicolour phenomenology . . . can . . . arise from soggy grey matter," when the neurophysiologist Susan Greenfield describes the brain as a "sludgy mass," they vividly and usefully express part of the "imaginative . . . confusion" in the ordinary idea of matter.[105] But physics comes to our aid: there is a clear sense in which the best description of the nature of the non-Experiential *in non-technical, commonsense terms* comes from physics. For what, expressed in commonsense terms, does physics find in the volume of spacetime occupied by a brain? Not a sludgy mass, but an astonishingly (to us) insubstantial-seeming play of energy, an ethereally radiant vibrancy.

It finds, in other words, a physical object; which, thus far examined, is like any other. Examined further, this particular physical object turns out to have a vast further set of remarkable properties: all the sweeping sheets and scudding clouds and trains of intraneuronal and interneuronal electrochemical activity which physics (in conjunction with neurophysiology) apprehends as a further level of extraordinarily complex intensities of movement and (non-Experiential) organization.

All this being so, do we have any good reason to think that we know anything about the non-mental physical (assuming it exists) that licenses surprise – even the very mildest surprise – at the thought that the Experiential is physical? I do not think so. Brains are special, but they are not strange. The ghost in the machine is special, but it is certainly in the machine, and the machine, like the rest of the physical world, is already a bit of a ghost – as ghostly, in Russell's view, "as anything in a spiritualist séance."[106]

So when David Lewis says that "the most formidable opposition to any form of mind–body identity comes from the friends of qualia"[107] there is no reason to agree. The main opposition to (realistic) mind–body identity comes, paradoxically, from its most passionate proponents, who are so strongly inclined to think they know more about the nature of the non-mental physical than they do. Lewis exemplifies the great mistake in his well-known summary account of his position

in the philosophy of mind: "Remember," he says, "that the physical nature of ordinary matter under mild conditions is very well understood."[108] There is no reason to believe this, and every reason to disbelieve it. "What knowledge have we of the nature of atoms that renders it at all incongruous that they should constitute a thinking object?" asks Eddington, who took the existence of Experiential phenomena – qualia – for granted, and meant an Experiencing object: "science has nothing to say as to the intrinsic nature of the atom." The atom, so far as we know anything about it,

> is, like everything else in physics, a schedule of pointer readings [on instrument dials]. The schedule is, we agree, attached to some unknown background. Why not then attach it to something of spiritual nature of which a prominent characteristic is *thought*. It seems rather *silly* to prefer to attach it to something of a so-called "concrete" nature inconsistent with thought, and then to wonder where the thought comes from. We have dismissed all preconception as to the background of our pointer readings, and for the most part can discover nothing as to its nature. But in one case – namely, for the pointer readings of my own brain – I have an insight which is not limited to the evidence of the pointer readings. That insight shows that they are attached to a background of consciousness.[109]

The point is still negative. It may destroy one common source of puzzlement, but it doesn't offer any sort of positive account of the relation between the play of energy non-Experientially conceived and the play of energy Experientially apprehended, and some may find it no help at all. Others may say that it is a positive mistake to think that it is especially helpful, on the grounds that there is in the end no more difficulty in the thought that the existence of matter naively and grossly conceived involves the existence of consciousness than there is in the thought that matter quantum-mechanically conceived does so.[110]

We can grant them their objection for their own consumption (they are likely to be fairly sophisticated philosophers). Many others – not excluding philosophers – are likely to find the negative point rather useful, and I will conclude this section by relating it to three currently popular issues.

(1) Eliminativism. Consider any philosopher who has ever been tempted, even momentarily, by the 'eliminativist' suggestion that one has to question the reality of the Experiential in some way in order to be a thoroughgoing materialist. It is an extraordinary suggestion,[111] and what is most striking about it in the present context is that it constitutes the most perfect demonstration in the history of philosophy of the grip of the very thing that it seeks to reject: dualist thinking. The eliminativists make the same initial mistake as Descartes – the mistake of assuming that they understand more about the nature of the physical than they do – but their subjugation to dualist thinking is far deeper than Descartes's.[112] They are so certain that the physical excludes the Experiential that they are prepared to deny the reality of the Experiential in some (admittedly unclear) way – i.e., to make the most ridiculous claim ever made in philosophy – in order to

retain the physical. (The mistake of thinking one has grasped the essential nature of the physical is perhaps forgivable in the early seventeenth century, but not now.)

(2) The hard part of the mind–body problem. It can be seriously misleading to talk of the "hard part of the mind–body problem,"[113] or "the hard problem,"[114] for this suggests that the problem is clearly posed. It is not, as Chomsky has observed. One might say that it is not sufficiently well defined for us to be able to say that it is hard; for although we have a clear and substantial positive fix on the non-structural nature of Experiential reality, we have no substantial positive fix on the non-structural nature of non-Experiential reality, apart, perhaps, from its spatiotemporal characteristics. To this extent we have no good reason to think that the mind–body problem is a harder problem than the problem posed for our understanding by the peculiarities of quantum physics, or indeed – as Chomsky might say – by the phenomenon of motion. The problem is the nature of the physical, and in particular, perhaps, of the non-mental physical.

(3) Zombies. It is, finally, a mistake to think that we can know that "zombies" could exist – where zombies are understood to be creatures that have no Experiential properties although they are *perfect physical duplicates* (PPDs) of Experiencing human beings.[115] The argument that PPD-zombies could exist proceeds from two premisses: (i) it is conceivable that PPD-zombies exist, (ii) if something is conceivable, then it is possible. The argument is plainly valid, and (unlike many) I have no great problem with (ii). The problem is that we can't know (i) to be true, and have *no* reason to think it is. To be a genuine materialist is precisely to hold that (i) is false, and while materialism cannot be known to be true, it cannot be refuted *a priori*, as it could be if (i) could be known to be true. PHYSICAL is a natural-kind concept, and since we know that there is much that we do not know about the nature of the physical, we cannot claim to know that an Experienceless PPD – a perfect physical duplicate, no less – of a currently Experiencing human being is even conceivable, and could possibly exist. One needs to be very careful how one embeds natural-kind terms in "it is conceivable that" contexts.[116]

It is worth adding that anyone who holds that it is as a matter of *physical* fact impossible for a PPD of an actual, living normally, Experiencing human being to have no Experience must hold that PPD-zombies are *metaphysically* (if not *logically*) impossible. Physical impossibility entails metaphysical impossibility in this case, because the question is precisely what is possible given the actual nature of the physical.

15 Realistic Monism

In section 1 I pointed out that the word "physical," as used by genuine materialists, entails "real and concrete": given that one is restricting one's

attention to concrete phenomena, as we are doing here, to say something is a physical phenomenon is simply to say that it is a real phenomenon. But then why bother to use "physical"?[117] It has become an entirely empty or vacuous term, in so far as it is supposed to mean anything more than "real." So why not simply use "real"? And why bother with "real," given that we are talking about whatever (concretely) exists, whatever it is? It is redundant. All one strictly needs, to mark the distinctions centrally at issue in the unfortunately named "mind–body problem," are "mental" and "non-mental," "Experiential" and "non-Experiential."[118] One can simply declare oneself to be an *Experiential-and-non-Experiential* monist: one who registers the indubitable reality of Experiential phenomena and takes it that there are also non-Experiential phenomena.

I nominated this position for the title "realistic monism," having explicitly assumed that any realistic position must take it that there is non-Experiential being. Now this assumption can be backed by an argument that seems quite strong – (1) Experience (Experiential content) certainly exists; (2) Experience (Experiential content) is impossible without a subject of experience; (3) a subject of experience cannot itself be an entirely Experiential (Experiential-content) phenomenon; so (4) the existence of Experience (Experiential content) entails the existence of non-Experiential phenomena.[119] But one can have no deep confidence in the correctness of the assumption if one accepts the general principles of ignorance defended in this paper, for it invites the reply that even if a subject of experience must have non-Experiential being relative to its own Experience, its non-Experiential being may be the Experience of some other, lower-order, subject or group of subjects, and so on down. I am not sure this reply is cogent, in fact, or that premiss (3) is solid,[120] but I propose to leave the assumption as an assumption: one cannot really know what is "realistic," at this point.

You say we can do without the word "physical." But if one can do without "physical," then "materialist," used so diligently in this paper, is just as superfluous – vacuous. You have already stated (note 14) that you make no distinction between materialism and physicalism, and the word "materialist" is deeply compromised by its history.

History is two-faced, and I think that "materialist" – an adjective formed from the natural-kind term "matter" – can be harmlessly and even illuminatingly retained.[121] What, after all, is matter? As a materialist, I take it that it is whatever we are actually talking about when we talk about concrete reality. I fix the reference of the term "matter" in this way – giving a chair a kick, perhaps – independently of any reference to theories. I can be certain that there is such a thing as matter, as a realistic materialist monist – one who takes it that Experience is wholly material in nature – because I can be certain that there is such a thing as concrete reality: Experience, at the very least. What a materialist may still wish to add to this is the insistence that nothing can count as matter unless it has some sort of non-Experiential being (section 4); together with the working

presumption that current physics is genuinely reality-representing in certain ways, even if any correctness of representation is only a matter of the holding of certain relations of structural correspondence between the nature of matter and the equations of physics.

In so far as I am a realistic materialist monist, then, I presume that physics' best account of the structure of reality is genuinely reality-representing in substantive ways, and that the term "materialist" is in good order. I sail close to the wind – by which I mean the charge of vacuousness, and the charge that it may be hard to distinguish my position from idealism – in my use of the word "matter" because that is exactly what one has to do at this point. Kicking another chair, I grant that the term "materialist" has traveled far from some of its past uses, but there is no good reason to think that its meaning is especially tied to its past uses rather than to the current understanding of matter.[122] And there is a sense in which its past use makes it particularly well worth retaining: it makes the claim that the present position is materialist vivid by prompting resistance that turns out to be groundless when the position is properly understood.[123]

That is all I have to say about the word "materialist," and some will probably think that I would do better to call myself a "neutral monist," or just a "monist." But what about "monist"? There is serious unclarity in this notion. Monists hold that there is, in spite of all the variety in the world, a fundamental sense in which there is only one basic kind of stuff or being. But questions about how many kinds of stuff or being there are are answerable only relative to a particular point of view or interest; and what point of view is so privileged that it allows one to say that it is an absolute metaphysical fact that there is only one kind of stuff or being in reality? Materialists call themselves monists because they think that all things are of one kind – the physical kind. But many of them also hold that there is more than one kind of fundamental particle, and this claim, taken literally, entails that there isn't after all any one basic kind of being out of which everything is constituted. For it is the claim that these particles are themselves, in their diversity, the ultimate constituents of reality; in which case there is kind-plurality or stuff-plurality right at the bottom of things.

But these particles are nevertheless all physical, *and in that sense of one kind.*

To say that they can be classed together as single-substanced in this way is question-begging until it is backed by a positive theoretical account of why it is correct to say that they are all ultimately (constituted) of one kind (of substance). To claim that their causal interaction sufficiently proves their same-substance-hood is to beg the question in another way, on the terms of the classical debate, for classical substance-dualists simply deny that causal interaction entails same-substancehood. The claim that they are all spatiotemporally located also begs the question. For how does this prove same-substancehood?

It may be replied that all the particles are just different forms of the same stuff

– energy. And it may be added that the so-called fundamental particles – quarks and leptons – are not strictly speaking fundamental, and are in fact all constituted of just one kind of thing: superstrings. And these monist approaches deserve investigation – to be conducted with an appropriately respectful attitude to panpsychism.[124] But one can overleap them by simply rejecting the terms of the classical debate: one can take causal interaction to be a sufficient condition of same-substancehood.

I think that this is the right dialectical move in the present context, if one wants to retain any version of the terminology of substance. Dualists who postulate two distinct substances while holding that they interact causally not only face the old problem of how to give an honest account of this interaction. They also face the (far more difficult) problem of justifying the claim that there are two substances. As far as I can see, *the only justification that has ever been attempted* has consisted in an appeal to the *intuition* that the mental or the Experiential is utterly different in nature from matter. But this intuition lacks any remotely respectable theoretical support, if the argument of this paper is even roughly right. The truth is that dualism has nothing in its favor – to think that it has is simply to reveal that one thinks one knows more about the nature of things than one does – and it has Occam's razor (that blunt sharp instrument) against it. This is not to rule out the theoretical possibility that substance dualism – or pluralism – is in fact the best view to take about our universe for reasons of which we know nothing.[125] The fact remains that the objection to dualism just given remains decisive when dualism is considered specifically as a theoretical response to the "mind–body problem."

But why persist with "monist"? You might as well call yourself a "neutral pluralist," for all the difference it makes, and "monist" carries bad baggage. Why not simply call yourself a "non-committal naturalist," or, with Chomsky, a "methodological naturalist"? Or a "?-ist"?[126]

This section stirs up large questions, but I'm not too troubled. In some moods I am prepared to call myself an Experiential-and-non-Experiential ?-ist and think no more about the word "monist"; there is no decidable issue here, as the old decriers of metaphysics (e.g., Locke, Hume, Kant) knew. At the moment, though, the physics idea (the ancient idea) that everything is made of the same ultimate stuff – that the deep diversity of the universe is a matter of different states or arrangements of the same fundamental *ens* or *entia* – that "in the whole universe there is only one substance differently modified"[127] – seems to me as compelling as it is remarkable, and I choose to register my attraction to it with the word "monism." If realistic monism is Chomskian methodological naturalism ontologized, good.

Eddington summarizes his position as follows: "To put the conclusion crudely – the stuff of the world is mind-stuff." "As is often the way with crude

statements," he continues, "I shall have to explain that by 'mind' I do not here exactly mean mind and by 'stuff' I do not at all mean stuff"; "the mind-stuff of the world is, of course, something more general than our individual conscious minds." Nevertheless, he believes, "we may think of its nature as not altogether foreign to the feelings in our consciousness."[128] This seems to me to be the most parsimonious, plausible and indeed "hard-nosed" position that anyone who is remotely realistic about the nature of reality can take up in the present state of our knowledge.[129]

Chomsky's reply: pp. 266–8.

Notes

1 This paper is an attempt to elaborate on "Agnostic Materialism" (Strawson 1994: 43–105); trailers appeared in Strawson 1998 and 1999a. Since writing it I have come across several expressions of similar views and have added a considerable number of quotations.

2 Eddington 1928: 292, quoting Shakespeare, *The Tempest*, Act III, scene ii. When I cite a work I give the original publication date, while the page reference is to the newer edition listed in the bibliography.

3 Auden 1940.

4 By "concrete" I simply mean "not abstract." It is natural to think that any really existing thing is *ipso facto* concrete, non-abstract, in which case "concrete" is redundant. But some philosophers like to say that numbers (for example) are real things – objects that really exist, but are abstract.

5 I use "phenomenon" as a completely general word for any sort of existent that carries no implication as to ontological category (the trouble with the perfectly general word "entity" is that it is now standardly understood to refer specifically to things or substance); and suppress its meaning of *appearance*.

Note that someone who agrees that physical phenomena are all there are but finds no logical incoherence in the idea that physical things could be put together in such a way as to give rise to non-physical things can define materialism as the view that every real, concrete phenomenon that there is *or could be* in the universe is physical.

6 Lewis 1986: x.

7 Physics is trivially referentially complete, according to materialism, insofar as its object of study is the universe, i.e., the whole of concrete reality. There may nevertheless be specific, smaller-scale phenomena of which physicists have no descriptive or referential inkling.

8 Failure to recognize this simple point, long after the existence of natural-kind terms has been generally acknowledged, is one of the more disastrous legacies of positivism. (Compare the survival of the "regularity theory of causation" after the abandonment of phenomenalism.)

9 I have in mind the old quarrel between general relativity theory and quantum mechanics, but there is also turmoil in cosmology.
10 Eddington 1928: 260.
11 Chomsky 1968: 6–8, 98; 1988: 142–7; 1994: passim; 1995: 1–10; 1996: 38–45; 1998: 437–41; compare Crane and Mellor 1990.
12 "Qualitative" has to be qualified by "experiential" because experiences also have non-experiential qualitative character, according to materialists (every non-relational property of a thing contributes to its qualitative character). Having made the point, I will either bracket "experiential" or follow common practice and omit it.
13 Churchland 1995: 202. Obviously "correctly describable" does not entail "fully describable." Note that one also puts aside the fact that they can be correctly described in such non-experiential terms as "perception of the Eiffel Tower."
14 I make no distinction between "materialism" and "physicalism."
15 If "immaterial souls" existed, they would of course be concrete phenomena.
16 Experiential phenomena would be concrete phenomena even if space and time were not really real – were somehow mere forms of experience.
17 One needs to distinguish between mental and Experiential phenomena because, although all Experiential phenomena are mental, not all mental phenomena are Experiential, on the ordinary view of things: certain *dispositional* states – beliefs, preferences, and so on – are mental phenomena although they have no Experiential character. There are also powerful reasons for saying that there are *occurrent* mental phenomena that are non-Experiential.
18 See Chomsky 1968: 98.
19 Some readers doubt this, but it follows from the fact that current physics contains no predicates for Experiential phenomena at all, and that no non-revolutionary extension of it could do so.
20 The parenthesis is redundant given the definition of "Experiential" in section 2.
21 Unless the existence of Experiential phenomena of kinds that we know to exist entails the existence of non-Experiential phenomena. See note 24 below.
22 Russell 1912: 19.
23 In the case of experiences, it amounts to saying that they are not just Experiential phenomena, although Experiential phenomena are of course part of what constitutes their existence. Note that to distinguish between mental being and non-mental being is not to claim to know how to draw a sharp line between them. The starting situation is simply this: we know there is mental being, and we assume, as materialists, that this is not all there is.
24 Elsewhere (Strawson 1994: 134–44) I argue that there could not be Experiential or Experiential content phenomena of the sort with which we are familiar unless there were also non-Experiential phenomena; and if it is true (1) that a subject of experience cannot itself be a wholly Experiential phenomenon (ibid.: 144), and (2) that "experience is impossible without an experiencer" (Frege 1918: 27), then the conclusion that the existence of Experience entails the existence of non-Experiential phenomena is guaranteed. The argument stalls, however, if one substitutes "mental" for "Experiential," if only because of the vagueness of the term "mental" (Strawson 1994: 140–2 and ch. 6).
25 I discuss the difference between "Experiential" and "mental," and the vagueness of

"mental," in Strawson 1994 (see, e.g., pp. 136–44 and ch. 6). Here I am trying to avoid the issue as far as possible.

26 Largely, perhaps, because of the looseness of his use of the word "see," and the reactive excesses (which led to exegetical insensitivity) of the first wave of twentieth-century "direct realists." See, however, Lockwood 1981.

27 See, e.g., Russell 1927b: 110, 119, 123, 126, 170. I do not understand everything Russell says and may misrepresent him. I aim to take what I think is right from his views without attempting exegesis, and I will sometimes detour from the main argument in Russellian directions.

28 Russell 1927b: 78, 84. N. R. Hanson spoke similarly of the "dematerialization" of matter, and Priestley (1777) made essentially the same point. See also Lange (1865).

29 Russell 1927b: 126–7.

30 In his introduction to Lange's *History of Materialism*, Russell notes that "physics is not materialistic in the old sense, since it no longer assumes matter as permanent substance" (Russell 1925: xix), and he may have the following passage from Lange in mind: "We have in our own days so accustomed ourselves to the abstract notion of forces, or rather to a notion hovering in a mystic obscurity between abstraction and concrete comprehension, that we no longer find any difficulty in making one particle of matter act upon another without immediate contact. We may, indeed, imagine that in the proposition, 'No force without matter', we have uttered something very Materialistic, while all the time we calmly allow particles of matter to act upon each other through void space without any material link. From such ideas the great mathematicians and physicists of the seventeenth century were far removed. They were all in so far still genuine Materialists in the sense of ancient Materialism, that they made immediate contact a condition of influence. The collision of atoms or the attraction by hook shaped particles, a mere modification of collision, were the type of all Mechanism and the whole movement of science tended towards Mechanism" (Lange 1865: 1.308, quoted in Chomsky 1996: 44).

31 This is so even if "eliminativism" about other candidate mental phenomena – dispositional phenomena like preferences, beliefs, and so on – is worth serious discussion.

32 The word "physical" is bracketed because it is redundant, here as elsewhere. See section 14.

33 Compare Shoemaker's idea (rather differently applied) that many mental states and goings-on are "constitutively self-intimating" (Shoemaker 1990). See also Maxwell 1978: 392, 396.

34 Russell 1956: 153; my emphasis.

35 Russell 1927a: 402; my emphasis. See Lockwood 1989: 159: "Consciousness . . . provides us with a kind of 'window' on to our brains, making possible a transparent grasp of a tiny corner of material reality that is in general opaque to us . . . The qualities of which we are immediately aware, in consciousness, precisely *are* some at least of the intrinsic qualities that the states and processes that go to make up the material world – more specifically, states and processes within our own brains. This was Russell's suggestion."

36 Not only facts about Experience in sense modalities we lack, or (e.g.) about the brightness-saturation-hue complexity of seemingly simple color-Experience, but

also, perhaps, murkier facts about its composition, and also, perhaps, about the "hidden nature of consciousness" postulated by McGinn (1991: chs. 3 and 4).

37 If you are worried about the concept – or reality – of time, drop the last four words.

38 See Dennett 1991, e.g., ch. 11.

39 Russell 1927a: 392, 382, 388.

40 Russell 1927b: 125; 1956: 153. Lockwood 1989 (ch. 10) contains some illuminating pages on Russell and a useful historical note on versions of the idea that precede Russell's. See also Maxwell 1978, whose Russellian approach is treated sympathetically in Chalmers 1996: 153–4 (and see index), and Chalmers 1997: 405–6. Jeremy Butterfield and Bas van Fraassen have pointed out to me the link here to John Worrall's "structural realism"; see, e.g., Worrall 1989 and Ladyman 1998.

41 Structure is a matter of quality because a thing's qualitative character, exhaustively considered, is a matter of *all* aspects of how it is, and its structural character is an aspect of how it is. The converse claim – that quality (in spacetime) is in some sense a matter of structure – sounds a bit mystical, but it can on further reflection begin to seem hard to rebut, even when one maintains, as one must, a sharp distinction between epistemology and metaphysics. (The distinction between form and content may seem more robust, but may also succumb.)

42 It seems (subject to the doubt expressed in the previous note) that this distinction must be a real one – that if there is structure, there must be something structured. Only extreme positivistic irresponsibility, or failure to "realize what an abstract affair form [or structure] really is" (Russell 1927a: 392), can make this seem questionable.

43 At one point Russell also takes it that position in spacetime is an intrinsic property of things. Considering the relation between a perception and the object it is a perception of, he remarks that "we cannot say whether or not it resembles the object in any intrinsic respect, except that both it and the object are brief events in spacetime" (Russell 1927b: 118).

44 When thinking of structural isomorphism, it is helpful to consider a version of an old example: the structural isomorphism between (1) sound waves produced by an orchestra playing Sibelius's "Valse Triste" that are registered as (2) vibrations of a condenser plate in a microphone and sent as (3) electrical signals to a recording device that stores them as (4) pits on the surface of a compact disk that is then read as (5) digital information by a machine that transmits this information in the form of (6) radio waves to (7) a receiver that puts it through (8) an amplifier to (9) speakers that give rise to (10) sound waves that give rise to (11) electrical impulses in the auditory nerve that give rise to (12) neural occurrences in the auditory cortex and elsewhere that are conscious auditory experiences. There is a structural description that captures the respect in which all these phenomena are the same (assuming no significant loss of information even at the stage of conscious hearing). The abstract character of this description is revealed precisely by the fact that this is what it does: capture the respect in which all these substantially different phenomena are – structurally – the same. Compare Wittgenstein 1922: 4.0141.

45 In 1928, a year after the publication of *The Analysis of Matter*, Max Newman published a conclusive objection to the pure form of this view, as Russell immediately acknowledged (Russell 1967–9: 413–14). See Demopoulos and Friedman 1985.

46 Russell 1927a: 391.

47 Cp. Hirsch 1986: 251–4. I will not here consider the "direct realist" view that we

may have some real insight into the non-mental nature of force, say, or causation, as a result of experiencing pushes and pulls and so on in the way we do.

48 Russell 1927a: 392. One could say that it is this point that Newman turns back against Russell (see note 45).

49 Russell 1927a: 382; my emphasis.

50 I will use small capitals for names of concepts.

51 One may contrast the case of a congenitally blind person with the hypothetical case of a fully sighted person congenitally paralyzed and devoid of tactile sensation – before thinking of superintelligent echolocating bats and aliens with other sensory modalities.

52 See Evans 1980: 269–71; McGinn 1983: 126.

53 Even if no finite sensory-intellectual being can possess the concept of shape or space without having, or without at least having grasp of the nature of, some form of sensory experience, it does not follow that specification of the content of the concept it possesses necessarily involves reference to any features of sensory experience.

54 If empiricists press me further I will offer (a) the suggestion that sensory modalities that differ qualitatively at first order (i.e., in the way that sight and touch do) may be said to be crucially similar at second order in as much as they are "intrinsically spatial" in character, (b) the speculation that this similarity can itself be understood as a kind of similarity of (Experiential) *qualitative* character, (c) the acknowledgment that it may be that one must be capable of experience in some "intrinsically spatial" sensory modality or other (even if only in imagination) in order to possess SHAPE or SPACE, (d) the reservation that even if a *non-conceptual* experiential modality must be in play, it is not obvious that this must be a *sensory* modality. This, however, is too simple (I discuss the question further in "Knowledge of the World").

55 Personal communication. Eddington agrees: "We know nothing about the intrinsic nature of space" (Eddington 1928: 51–2).

56 After setting aside the problem of evil.

57 "We know nothing about the intrinsic quality of physical events except when these are mental events that we directly experience" (Russell 1956: 153). Perhaps Russell takes this distancing step himself in his 1928 reply to Newman (see note 45): "It was quite clear to me, as I read your article, that I had not really intended to say what in fact I did say, that *nothing* is known about the physical world [the non-mental world as opposed to the mental world, in my terminology] except its structure. I had always assumed that there might be co-punctuality between percepts and non-percepts, and even that one could pass by a finite number of steps from one event to another compresent with it, from one end of the universe to the other . . . spacio-temporal continuity of percepts and non-percepts was so axiomatic in my thought that I failed to notice that my statements appeared to deny it" (Russell 1967–9: 413).

58 In such a way that the correct answer to the question "Where was the Big Bang taking place at the first moment in which it made sense to say that it was taking place anywhere?" is "Right here," wherever you are.

59 Bell 1964; for an informal illustration see Lockwood 1996: 163–4.

60 Isham and Butterfield 2000: 1712.

61 McGinn 1995: 221.

62 Foster (1982: ch. 5) and McGinn (1995) give forceful presentations of the repug-

nance intuition. At one point McGinn makes the funky suggestion that consciousness might be a manifestation of the non-spatial nature of pre-Big Bang reality (223–4). I think he moves in a better direction when he shifts to the very different claim that "consciousness tests the adequacy of our spatial understanding. It marks the place of deep lack of knowledge about space" (230).

63 I am grateful to Mark Sainsbury for encouraging me to make this point more explicit.

64 I argue for this in "Triangle and TRIANGLE." Locke's talk of "resemblance" between primary qualities and ideas of primary qualities is unfortunate in as much as it suggests a (mere) picturing relation, and Russell (1927a: 385) holds that Locke is definitely wrong.

65 Perhaps it gets something right in an Augustinian sense, according to which we can be said to know what time is even though we find we don't know what to say when someone asks us what it is.

66 Unless some "direct realist" account of our understanding of force is defensible. See note 47. (Note that no sensible direct realist view can suppose that we derive understanding of the nature of force directly from the sensory character of experiences of pushes, pulls, and so on; that would be like thinking that we can get some real insight into the nature of electricity from the qualitative character that experiences of electric shocks have for us (compare Evans 1980: 270). Somehow, the sensory experiences have to be the basis of an abstract, essentially cognitive, general, non-sensory concept of force.)

67 The word "non-mental" is not redundant in the last sentence, for it seems very plausible to suppose that consciousness is an electrical phenomenon, whatever else it is; in which case it may be said that we do have some knowledge of the non-structural nature of electrical phenomena just in having conscious experience.

68 Moore 1905–6: 64.

69 Kant 1781: A414/B441. I have substituted "object" and "property" for "substance" and "accident" respectively.

70 For the importance of this qualification, see note 116 below.

71 "As characterized by physics" is a necessary qualification; see the remarks about "brain" above.

72 Russell 1927b: 112.

73 The only alternative, I think, is that one has a very rare and beautiful intellect.

74 See Chomsky 1994; 1995: 1–10. Chomsky is a clear example of someone who is, methodologically, a true materialist in my sense. I am not sure that he would accept the title, however; he avoids the term "materialist" because of the point made by Lange in note 28 above, which I try to counter on page 74.

75 In fact one doesn't have to be a materialist to hold that no defensible conception of the physical contains any element that gives one positive reason to doubt that Experiential phenomena are physical. One can hold this even if committed to dualism.

76 See, e.g., van Fraassen 1980: ch. 5; Strawson 1994: 84–93.

77 I think that it requires realization that this claim is inadequately expressed by saying "consciousness is a property of matter," or even "consciousness is a physical property of matter," given the almost irresistible incentives to metaphysical misunderstanding that are – I argue elsewhere – already built into the word "property."

78 Perhaps intuitive materialism is not always an achievement, and comes easily, and without positive error, in certain Eastern schools of thought. The requirement that there be no positive error of conception is, however, important.

79 Some may object that there is a compelling description of quantum-mechanical phenomena that completely eliminates the air of mystery attaching to wave–particle duality (see, e.g., Deutsch 1997: ch. 2); but it does so at the cost of another large strangeness, because it requires one to accept Everett's many-worlds hypothesis; and although it may be that this is what one should do, I will continue to use the case of wave–particle duality as an example for the purposes of discussion. (I will also put aside the view that the real intuitive difficulty resides in the phenomenon of superposition rather than in the wave–particle duality.)

80 I try to give more content to the idea of theoretical homogeneity in Strawson 1994: 88–93. Note that one can have a sense that a group of terms is theoretically homogeneous, or at least not problematically heterogeneous, without feeling that one *understands* the phenomena these terms are used to describe.

81 Although there are plenty of deep puzzles in physics even when mind is put to one side, and it is an old point that we do not really even understand why one billiard ball has the effect it has on another (see, e.g., Locke 1690: II. xxiii. 28; Hume 1739: 246–8).

82 Russell 1948: 240; see also 247.

83 In Descartes 1641: 141. Lange (1865) discusses many precursors.

84 Priestley 1777/1965: 115. Locke doesn't fully carry through his point that our ignorance of the nature of the physical means that we lack any good reason to doubt that consciousness is wholly physical; for at one point he says that "matter . . . is evidently in its own nature void of sense and thought" (Locke 1690: IV. iii. 6; see also IV. x. 5, 10). But he also says that we "possibly shall never be able to know whether any material being thinks, or no," and holds that an omnipotent being could give "some systems of matter, fitly disposed, a power to perceive and think" (IV. iii. 6). The force of the second quotation is less than one might suppose: it does not conclusively establish that Locke thought that God could make matter have such a power *intrinsically* or in and of itself – i.e., without any special wizardry. But Locke's correspondence with Stillingfleet strongly suggests that his considered view is that our ignorance of the nature of matter is in the end too great for us to have any good reason to claim that matter could not have the power of thought in and of itself (Locke 1696–9: 459–62).

85 Algarotti 1737: 2.194.

86 Hume 1739: 246–8.

87 Priestley 1777 (1881: 219). Priestley observes, correctly, that there is no evidence for absolute solidity: "I . . . define . . . matter . . . to be a substance possessed of the property of *extension*, and of *powers of attraction and repulsion*. And since it has never yet been asserted, that the powers of *sensation* and *thought* are incompatible with these (solidity, or impenetrability only, having been thought to be repugnant to them), I therefore maintain that we have no reason to suppose that there are in man two substances so distinct from each other as have been represented."

88 Kant 1781: A358–60; Priestley 1777/1818: 220, 244, 303. John Toland in 1704 "obviously regards thought as a phenomenon which is an inherent accompaniment of the material movements in the nervous system" (Lange 1865: 1.329).

89 "We must admit that perception, and whatever depends on it, cannot be explained on mechanical principles, i.e. by shapes and movements. If we pretend that there is a machine [e.g. a brain] whose structure makes it think, sense and have perception, then we can conceive it enlarged . . . so that we can go inside it as into a mill. Suppose that we do: then if we inspect the interior we shall find there nothing but parts which push one another, and never anything which could explain a perception. Thus, perception must be sought in simple substances, not in what is composite or in machines" (Leibniz 1720: 150 [*Monadology* section 17]).

90 Locke saw this pretty clearly after reading Newton (see, e.g., Locke 1690: 309, 559–60; 1696–9: 467–8). Chomsky (1995: 4) quotes tellingly from Leibniz and Huygens, who condemned Newton for abandoning sound "mechanical principles" and reverting to mystical "sympathies and antipathies" and "inexplicable qualities." And yet Leibniz's considered view is arguably in harmony with the present view.

91 Hume 1739: 159; my emphasis.

92 See Dennett 1991: 365–6, and the reply in Strawson 1994: 51–2.

93 Russell 1927b: 101.

94 Ryle 1949: 186–9. The idea is an old one.

95 I argue for this in Strawson 1999b: section 10.

96 See, again, van Fraassen 1980: ch. 5.

97 As in the old quantum-theory model of the atom, *c*.1910–24. The standard way to convey the amount of empty space inside an atom is to say that if the nucleus is imagined to be as big as a 1 mm pinhead, then the nearest electrons – themselves much smaller than the nucleus – are 100 metres away.

98 See also Foster 1982: 67–72, and Harré and Madden 1975: ch. 9.

99 "In the modern theory of elementary particles known as the Standard Model, a theory that has been well verified experimentally, the fundamental components of nature are a few dozen kinds of field" (Weinberg 1997: 20). We continue to talk in terms of particles because the quantization of the field, whereby each different (normal) mode of vibration of the field is associated with a discrete ladder of energy levels, automatically gives rise to particle-like phenomena so far as observation is concerned.

100 Harvey Brown, personal communication; see Saunders and Brown 1991. Perhaps Descartes was right, deep down, in his theory of the plenum.

101 Compare Maxwell 1978: 399, James 1890: 1.147 n. It is arguable that Schopenhauer holds something close to this view.

102 Using this very wording, Kant (1781/7: B427–8) remarks that the "heterogeneity" of mind and body is merely assumed and not known.

103 Russell 1927a: 402.

104 Richard Price is consistently outclassed by Priestley in their *Free Discussion of the Doctrine of Materialism*, but he gets this point exactly right: "if . . . it comes out that [Priestley's] account of matter does not answer to the common idea of matter, [and] is not *solid* extension, but something *not solid* that exists in space, it agrees so far with spirit," or mind (Priestley and Price 1778: 54; Price held that spirit was not only located in space but might also be extended). This is a rather good description of how things have come out in physics. The account of matter given by current physics does not "answer to the common ideas of matter"; it does not take matter to be "solid extension," but rather "something not solid that exists in space." So far,

then, it agrees with our understanding of mind or consciousness, although the agreement can only be negative, given that we have no non-mathematical grasp of the non-structural nature of the non-Experiential being of matter – apart (perhaps) from our grasp of its spacetime structure.

105 Nagel 1998: 338; McGinn 1991: 1, 100; Greenfield, *Fundamentals*, BBC Radio 4, June 21 1997. In spite of these quotations I think that all three of these writers are fundamentally on the right track when it comes to the mind–body problem.

106 Russell 1927b: 78.

107 Lewis 1999: 5; qualia are Experiential phenomena.

108 Lewis 1994: 412.

109 Eddington 1928: 259; my emphasis on "silly."

110 They will find Russell's line of thought equally unnecessary as a way of reaching a conclusion they already fully accept: "having realised the abstractness of what physics has to say, we no longer have any difficulty in fitting the visual sensation into the causal series. It used to be thought 'mysterious' that purely physical [i.e., non-mental] phenomena should end in something mental. That was because people thought they knew a lot about physical phenomena, and were sure they differed in quality from mental phenomena. We now realise that we know nothing of the intrinsic quality of physical phenomena except when they happen to be sensations, and that therefore there is no reason to be surprised that some are sensations, or to suppose that the others are totally unlike sensations" (Russell 1927b: 117).

111 It seems considerably more implausible than Xenocrates' suggestion that the soul is a self-moving number (see Aristotle *c*.340 BC: 408b–409a).

112 In fact it is not clear that Descartes does make this mistake, although it is clear that some eliminativists do. Descartes was for a long time seen as a dangerous source of materialist views, and there are some reasons for thinking that his official dualism was motivated partly by the desire to stave off persecution by religious authorities.

113 Strawson 1989: 80; 1994: 93; compare McGinn 1989: 1.

114 Chalmers 1995: 200.

115 I don't know where these zombies come from – but they may be Australian. Some years ago philosophical zombies were far more plausible creatures: they were defined to be *outwardly* and *behaviorally* indistinguishable from human beings while having unknown (possibly non-biological) insides, and were, accordingly, of considerable interest to functionalists and behaviorists.

116 It is worth noting that a perfect physical duplicate of an actual human being would have to be governed by the same physical laws.

117 Compare Crane and Mellor 1990.

118 According to the view presented in Strawson 1994: 162–75, the latter pair suffice on their own.

119 As remarked in note 24, I do not think that there is any clear parallel argument for the claim that non-*mental* being must exist if Experience exists, because the term "mental" is too unclear for such an argument to be constructed. Nor do I know how someone like Bradley would respond to this argument.

120 I address this question in "What is the relation between an experience, the subject of the experience, and the content of the experience?".

121 Here at last, it seems, I may differ from Chomsky – but only, I think, on a point of

terminology. (Obviously "physical" can also be retained in so far as it is synonymous with "material.")

122 There is, in particular, no good reason to think that it is especially tied to the seventeenth-century conception of matter as something passive and inert. The conception of matter as essentially energy-involving, or at least as something to which motion is intrinsic, is already present in the work of Democritus and Epicurus.

123 I should add that I take it that spacetime itself is material (it is a disputed question). In quantum-field theory, reality consists of spacetime and a collection of fields defined on spacetime; what we think of as material objects are emergent (in the non-spooky sense) features of these fields. But spacetime is not a merely passive container, for according to general relativity the action–reaction principle applies as between spacetime and matter (this is the phenomenon Pullman (1998: 351) usefully calls "the vacuum-matter complementarity, or . . . the virtual-material duality of particles"). Moreover, the gravitational field, unlike the other fields, is not distinct from spacetime itself. Rather, the gravitational field, within a given region, just is the spacetime geometry of that region. The structural relations it involves are physical or material because they are spatiotemporal (they are not merely abstract-dimensional). Note also that energy can be stored and propagated within the gravitational field, and hence within the spacetime fabric itself, which again suggests that spacetime is substantial and hence – given materialism – material, in a way that Newtonian space and time, say, are not. (My thanks to Michael Lockwood and Harvey Brown for discussion of this matter.)

124 See, e.g., Seager 1995. Note that panpsychism does *not* require one to believe that tables and chairs have Experience.

125 There may be phenomena in the universe that cannot interact causally given their nature (rather than their position in spacetime), or that do so only on the first Thursday of every seventh century, in a highly peculiar way.

126 Sebastian Gardner proposed that I am a "?-ist" in 1990 (see Strawson 1994: 105). It is hard to find satisfactory names, and Grover Maxwell, who holds essentially the same position as I do, calls himself a "nonmaterialist physicalist" (1978: 365).

127 La Mettrie 1747: 39.

128 Eddington 1928: 276–7.

129 I am especially grateful to Noam Chomsky, Michael Lockwood and Undo Uus for the leads they have given me. I would also like to thank Lucy Allais, Harvey Brown, Jeremy Butterfield, Arthur Collins, Tim Crane, Mark Greenberg, Isaac Levi, Barry Loewer, Brian McLaughlin, Philip Pettit, Mark Sainsbury, Simon Saunders, Stephen Schiffer, Peter Unger, Bas van Fraassen, and audiences at the University of Birmingham, CUNY Graduate Center, and Columbia University in 1997 and 1998.

References

Algarotti, F. 1737/1739: *Sir Isaac Newton's Philosophy Explain'd For the Use of the Ladies*. Tr. Elizabeth Carter. London: E. Cave.

Aristotle *c*.340 BC/1936: *De anima*. Tr. W. S. Hett. Cambridge, Mass.: Harvard University Press.

Arnauld, A. 1641/1985: Fourth Set of Objections. In *The Philosophical Writings of Descartes*, vol. 2. Tr. J. Cottingham et al. Cambridge: Cambridge University Press.
Auden, W. H. 1940: *Another Time*. New York: Random House.
Bell, J. 1964: On the Einstein Podolsky Rosen Paradox. *Physics*, 1, 195–200.
Chalmers, D. 1995: Facing up to the Problem of Consciousness. *Journal of Consciousness Studies*, 2, 200–19.
Chalmers, D. 1996: *The Conscious Mind*. New York: Oxford University Press.
Chalmers, D. 1997: Moving Forward on the Problem of Consciousness. In J. Shear (ed.), *Explaining Consciousness: The Hard Problem*, Cambridge, Mass.: MIT Press.
Chomsky, N. 1968: *Language and Mind*. New York: Harcourt, Brace & World.
Chomsky, N. 1988: *Language and Problems of Knowledge*. Cambridge, Mass.: MIT Press.
Chomsky N. 1992: Language and Interpretation. In J. Earman (ed.), *Inference, Explanation and other Philosophical Frustrations*, Berkeley: University of California Press.
Chomsky, N. 1993: Explaining Language Use. *Philosophical Topics*, 20, 205–31.
Chomsky, N. 1994: Naturalism and Dualism in the Study of Language. *International Journal of Philosophical Studies*, 2, 181–209.
Chomsky, N. 1995: Language and Nature. *Mind*, 104, 1–60.
Chomsky, N. 1996: *Powers and Prospects*. London: Pluto.
Chomsky, N. 1998: Comments: Galen Strawson's *Mental Reality*. *Philosophy and Phenomenological Research*, 58, 437–41.
Churchland, P. 1995: *The Engine of Reason, the Seat of the Soul*. Cambridge, Mass.: MIT Press.
Crane, T. and Mellor, D. H. 1990: There is No Question of Physicalism. *Mind* 99, 185–206.
Demopoulos, W. and Friedman, M. 1985: Critical Notice: Bertrand Russell's *The Analysis of Matter* and its Historical Context and Contemporary Interest. *Philosophy of Science*, 52, 621–39.
Dennett, D. 1991: *Consciousness Explained*. Boston: Little, Brown.
Descartes, R. 1641/1985: *The Philosophical Writings of Descartes*, vols. 1 and 2. Tr. J. Cottingham et al. Cambridge: Cambridge University Press.
Deutsch, D. 1997: *The Fabric of Reality*. Harmondsworth: Penguin.
Eddington, A. 1928: *The Nature of the Physical World*. New York: Macmillan.
Evans, G. 1980/1986: Things Without the Mind. In Evans, *Collected Papers*, Oxford: Clarendon Press.
Foster, J. 1982: *The Case for Idealism*. London: Routledge.
Frege, G. 1918/1967: The Thought: A Logical Inquiry. In P. F. Strawson (ed.), *Philosophical Logic*, Oxford: Oxford University Press.
Harré, R. and Madden, E. H. 1975: *Causal Powers: A Theory of Natural Necessity*. Oxford: Blackwell.
Hirsch, E. 1986: Metaphysical Necessity and Conceptual Truth. *Midwest Studies*, 11, 243–56.
Hume, D. 1739/1978: *A Treatise of Human Nature*. Eds. L. A. Selby-Bigge and P. H. Nidditch. Oxford: Oxford University Press.
Isham, C. and Butterfield, J. 2000: Some Possible Roles for Topos Theory in Quantum Theory and Quantum Gravity. *Foundations of Physics*, 30, 1707–35.
James, W. 1890/1950: *The Principles of Psychology*, vol. 1. New York: Dover.
Kant, I. 1781–7/1933: *Critique of Pure Reason*. Tr. N. Kemp Smith. London: Macmillan.

Ladyman, J. 1998: What is Structural Realism? *Studies in the History and Philosophy of Science*, 29, 409–24.

La Mettrie, J. 1747/1996: Machine Man. In A. Thomson (ed. and tr.), *Machine Man and Other Writings*, Cambridge: Cambridge University Press.

Lange, (F.) A. 1865/1925: *The History of Materialism*. Tr. E. C. Thomas with an introduction by B. Russell. London: Routledge and Kegan Paul.

Leibniz, G. 1720/1965: *Monadology and Other Philosophical Essays.* Tr. P. and A. M. Schrecker. Indianapolis: Bobbs-Merrill.

Lewis, D. 1983: *Philosophical Papers*, vol. 1. Oxford: Oxford University Press.

Lewis, D. 1986: Introduction. In D. Lewis, *Philosophical Papers*, vol. 2. Oxford: Oxford University Press.

Lewis, D. 1994: Reduction of Mind. In S. Guttenplan (ed.), *A Companion to the Philosophy of Mind*, Oxford: Blackwell.

Lewis, D. 1999: Introduction. In D. Lewis, *Papers in Metaphysics and Epistemology*, Cambridge: Cambridge University Press.

Locke, J. 1690/1975: *An Essay Concerning Human Understanding*. Ed. P. Nidditch. Oxford: Clarendon Press.

Locke, J. 1696–9/1964: The Correspondence with Stillingfleet. In A. D. Woozley (ed. and abridged), *Concerning Human Understanding*, London: Collins.

Lockwood, M. 1981: What *Was* Russell's Neutral Monism? *Midwest Studies in Philosophy*, 4, Minneapolis: University of Minnesota Press.

Lockwood, M. 1989: *Mind, Brain, and the Quantum*. Oxford: Blackwell.

Lockwood, M. 1993: The Grain Problem. In H. Robinson (ed.), *Objections to Physicalism*, Oxford: Clarendon Press.

Lockwood, M. 1996: "Many Minds" Interpretation of Quantum Mechanics. *British Journal for the Philosophy of Science*, 47, 159–88.

Maxwell, G. 1978: Rigid Designators and Mind-Brain Identity. In C. Wade Savage (ed.), *Perception and Cognition: Issues in the Foundations of Psychology*, Minneapolis: University of Minnesota Press.

McGinn, C. 1983: *The Subjective View*. Oxford: Clarendon Press.

McGinn, C. 1989/1991: Can We Solve the Mind–Body Problem? In McGinn, *The Problem of Consciousness*, Oxford: Blackwell.

McGinn, C. 1991: *The Problem of Consciousness*. Oxford: Blackwell.

McGinn, C. 1995: Consciousness and Space. *Journal of Consciousness Studies*, 2, 221–30.

Moore, G. E. 1905–6/1922: The Nature and Reality of Objects of Perception. In G. E. Moore, *Philosophical Studies*, London: Routledge.

Nagel, T. 1998: Conceiving the Impossible and the Mind–Body Problem. *Philosophy*, 73, 337–52.

Priestley, J. 1777/1818: *Disquisitions Relating to Matter and Spirit*. In J. T. Rutt (ed.), *The Theological and Miscellaneous Works of Joseph Priestley*, vol. 3, London.

Priestley, J. 1777/1965: *Priestley's Writings on Philosophy, Science and Politics.* Ed. J. A. Passmore. New York: Collier.

Priestley, J. and Price, R. 1778/1819: *A Free Discussion of the Doctrines of Materialism, and Philosophical Necessity*. In J. T. Rutt (ed.), *The Theological and Miscellaneous Works of Joseph Priestley*, vol. 4, London.

Pullman, B. 1998: *The Atom in the History of Human Thought*. New York: Oxford University Press.

Russell, B. 1912/1959: *The Problems of Philosophy*. Oxford: Oxford University Press.
Russell, B. 1925: Introduction. In F. A. Lange, *The History of Materialism*, tr. E. C. Thomas, London: Routledge and Kegan Paul.
Russell, B. 1927a/1992a: *The Analysis of Matter*. London: Routledge.
Russell, B. 1927b/1992b: *An Outline of Philosophy*. London: Routledge.
Russell, B. 1948/1992c: *Human Knowledge: Its Scope and Limits*. London: Routledge.
Russell, B. 1956/1995: Mind and Matter. In Russell, *Portraits from Memory*, Nottingham: Spokesman.
Russell, B. 1967–9/1978: *Autobiography*. London: Allen and Unwin.
Ryle, G. 1949: *The Concept of Mind*. Harmondsworth: Penguin.
Saunders, S. and Brown, H. 1991: *The Philosophy of Vacuum.* Oxford: Clarendon Press.
Seager, W. 1995: Consciousness, Information, and Panpsychism. In J. Shear (ed.), *Explaining Consciousness: The Hard Problem*, Cambridge, Mass.: MIT Press.
Shoemaker, S. 1990/1996: First-Person Access. In Shoemaker, *The First-Person Perspective and Other Essays*, Cambridge: Cambridge University Press.
Strawson, G. 1989: *The Secret Connexion: Causation, Realism and David Hume*. Oxford: Clarendon Press.
Strawson, G. 1994: *Mental Reality*. Cambridge, Mass.: MIT Press.
Strawson, G. 1998: Précis of *Mental Reality* and Replies to Noam Chomsky, Pierre Jacob, Michael Smith, and Paul Snowdon, Symposium on *Mental Reality*. *Philosophy and Phenomenological Research*, 58, 433–5, 461–86.
Strawson, G. 1999a: Realistic Materialist Monism. In S. Hameroff, A. Kaszniak, and D. Chalmers (eds.), *Towards a Science of Consciousness III*, Cambridge, Mass.: MIT Press.
Strawson, G. 1999b: The Self and the SESMET. *Journal of Consciousness Studies*, 6, 99–135.
Strawson, G. 2002: Knowledge of the World. *Philosophical Issues*, 12.
Strawson, G. (forthcoming a), What is the Relation between an Experience, the Subject of the Experience, and the Content of the Experience? *Philosophical Issues*.
Strawson, G. (forthcoming b), Triangle and TRIANGLE.
Toland, J. 1704: *Letters to Serena*. London.
Uus, U. 1994: *Blindness of Modern Science.* Tartu: Tartu Observatory.
Van Fraassen, B. 1980: *The Scientific Image*. Oxford: Clarendon Press.
Weinberg, S. 1997: Before the Big Bang. *New York Review of Books*, June 12 (New York).
Wittgenstein, L. 1922/1961: *Tractatus Logico-Philosophicus.* Tr. B. McGuinness and D. F. Pears. London: Routledge.
Worrall, J. 1989: Structural Realism: The Best of Both Worlds? *Dialectica*, 43, 99–124.

4
Naturalistic Inquiry: Where does Mental Representation Fit in?

FRANCES EGAN

Three related themes run through Noam Chomsky's recent work: (1) a defense of *methodological naturalism* in the study of mind and language; (2) the idea that *internalist* theories of mental and linguistic phenomena will prove to be most explanatory; and (3) the suggestion that certain aspects of mind and language – notably, intentionality and language use – may lie outside the scope of naturalistic inquiry. In this paper I shall attempt to elucidate these three themes, especially as they bear on the attempt to provide a naturalistic, in particular a computational, characterization of the mind. I shall argue that by failing to recognize the role that representational content plays in computational models, Chomsky has underestimated the potential of naturalistic inquiry for explaining intentionality.

1 Methodological Naturalism

Chomsky distinguishes two forms of naturalism. *Metaphysical naturalism* is an ontological thesis which holds, roughly, that everything that exists is physical; there are no extra-physical entities, properties, or events. *Methodological naturalism* is not really a thesis at all. It is a commitment to apply scientific, empirical methods to the study of mental and linguistic phenomena, with the hope of eventually integrating our accounts of these phenomena with the "core" natural sciences (see Chomsky 1994, 1995). A naturalistic approach does not preclude other ways of apprehending the world. Rather, it seeks to provide a particular form of understanding, what we might call "theoretical understanding," as opposed to, say, the kind of understanding to be acquired from reading novels, or from the study of our shared concepts, an enterprise that Chomsky calls "ethnoscience" (Chomsky 1994: 195).

Most theorists of mind and language in our time espouse both forms of

naturalism. Chomsky, however, claims that metaphysical naturalism has had no coherent formulation since the demise of Descartes's corpuscular mechanics in the early eighteenth century. (As he puts it, "Newton eliminated the problem of 'the ghost in the machine' by exorcizing the machine; the ghost was unaffected" (ibid.: 188).) Accordingly, "[u]nless offered some new [post-Newtonian] notion of 'body' or 'material' or 'physical,' we have no concept of naturalism apart from methodological naturalism" (Chomsky 1995: 37). The term "physical," in its current usage, simply provides a loose way of referring to phenomena that we more or less understand (Chomsky 1994: 188).

Methodological naturalism, on the other hand, appears to be relatively uncontroversial.[1] But Chomsky notes that, despite the widespread acceptance of methodological naturalism, various explanatory theories of aspects of mind and language have been seriously challenged on non-naturalistic grounds. It has been claimed, for example, that computational accounts of cognitive capacities are defective because they fail to meet some adequacy condition specific to the explanation of mental phenomena, such as the requirement that any knowledge attributed to a subject must be consciously accessible, at least in principle (Searle's (1991) "connection principle"), or because they do not preserve and respect the classifications of common sense. Such critiques, according to Chomsky, should be seen for what they are – manifestations of *methodological dualism*, the idea that the study of mind and language, unlike scientific inquiry in other domains (which is allowed to be *self-policing*), should be held to independent, "philosophical" standards.

2 Internalism

Externalism holds that mental states are individuated by reference to features of the subject's environment or social context. *Internalism* denies that external properties have any individuative significance, holding that two subjects in the same internal (neural) states are in the same mental states, whatever their contexts. (Internalist individuation is said to be "narrow.")

Externalism is the more fashionable position.[2] Chomsky, however, claims that "though naturalism does not entail an internalist approach, it does seem to leave no realistic alternative" (Chomsky 1995: 49). Although our commonsense explanatory practices ("folk psychology") may presuppose externalist individuative schemes, such practices fall within the purview of ethnoscience and outside that of naturalistic inquiry. Within science itself, according to Chomsky, explanatory theories of mind and language are invariably internalist, individuating the states and processes they characterize narrowly, or independently of surrounding environment. Interpreters who have argued that scientific psychology is externalist have been misled by scientists' casual talk about their practice, failing to give proper attention to the practice itself.[3]

Let us turn specifically to the study of language. The notion of an *I-language* plays a central role in Chomsky's recent work. An I-language (for "internalized language") is a mechanism or procedure that generates linguistic expressions, or, more precisely, *structural descriptions* that characterize linguistic expressions.[4] An I-language is characterized by a *grammar*. *Universal grammar* is the theory of human I-languages, "a system of conditions deriving from the human biological endowment that identifies the I-languages that are humanly accessible under normal conditions" (Chomsky 1986: 23).

Importantly, an I-language is a narrowly described property of the brain, individuated independently of surrounding environment. Indeed, a given I-language would count as the same type of object even if it were embedded differently *within* the subject. Although an I-language is a component of the language faculty, along with various performance systems involved in language comprehension and speech production, "It is only by virtue of its integration into such performance systems that this brain state qualifies as a language. Some other organism might, in principle, have the same I-language (brain state) as Peter, but embedded in performance systems that use it for locomotion" (Chomsky 1992a: 213). It follows, then, that an I-language is an abstractly characterized mechanism or procedure that is not *essentially* part of the language faculty at all. We will examine the implications of this idea below. For now, I simply want to stress that Chomsky's conception of an I-language is an internalist conception, indeed a rather radically internalist conception that prescinds from both external and internal (intra-organism) environment.

Chomsky has been highly critical of the externalist trend in semantics, arguing that Putnam's (1975) Twin-Earth thought experiments upon which externalist claims about meaning are based make illegitimate appeal to speakers' intuitions about such notions as *reference*, *extension*, and *true of*, technical notions about which speakers can have no theory-neutral intuitions (Chomsky 1992a: 225; 1995: 42). The thought experiments trade on speakers' intuitions about whether a liquid phenomenally identical to water counts as the *same liquid*, but Chomsky notes that speakers' intuitions about such matters are sensitive to contextual factors. Change the circumstances or pragmatic presuppositions of the Twin-Earth story and judgments will vary accordingly (Chomsky 1992a: 225). As Chomsky puts it, "From the natural language and commonsense concepts of *reference* and the like, we can extract no relevant 'relation between our words and things in the world'. And when we begin to fill out the picture to approach actual usage and thought, the externalist conclusions are not sustained" (Chomsky 1995: 44). With respect to the "other prong" of externalist theories of language and thought, which stresses deference to experts and community norms (developed mainly in the work of Tyler Burge), Chomsky is characteristically blunt: "There is simply no way of making sense of this prong of the externalist theory of meaning and language, as far as I can see, or of any of the work in theory of meaning and philosophy of language that relies on it" (ibid.: 49). But Chomsky's

disagreement with externalist semantics runs deeper than a clash of intuitions about ordinary usage. Even supposing that externalist intuitions about Putnam's and Burge's thought experiments were correct, Chomsky dismisses the idea that this work has any relevance for naturalist inquiry into mind and language:

> Suppose we accept . . . [the] intuitions. What would this tell us about language, belief, and thought? At most, that sometimes we might attribute beliefs etc. to X in terms of other people's beliefs and intentions; but that is clear from simple and ordinary cases. Again, inquiry into the ways we attribute belief as circumstances vary is a legitimate topic of linguistic semantics and ethnoscience, but the study of how people attain cognitive states, interact, and so on, will proceed along its separate course. (ibid.: 47)

At the heart of Chomsky's disagreement with externalist semantics is his rejection of the notion of a "shared public language." The notion is central to the externalist case. The moral of the thought experiments is that, typically, nothing in an individual speaker's head is sufficient to fix the meanings of the terms in her language (nor of the contents of beliefs expressed by these terms;[5] they are fixed rather by community-wide reference fixing practices). Chomsky doubts that common sense is committed to anything like community-wide referential practices, or a shared public language, but even if it is, he claims that these notions serve no legitimate theoretical or explanatory purpose. Any explanatory interest purportedly served by the notion of a shared public language would be better served by talk about similarity or identity of I-languages across persons, in other words, by purely internalist notions (ibid.: 48–51).[6]

3 The Limits of Naturalistic Inquiry

According to Chomsky, "general issues of intentionality, including those of language use, cannot reasonably be assumed to fall within naturalistic inquiry" (ibid.: 27). Sometimes he puts the point in stronger terms, claiming that "[n]aturalistic inquiry will always fall short of intentionality" (Chomsky 1992a: 229).

We can discern in Chomsky's work several reasons why aspects of the study of intentionality and language use might be thought to lie outside the realm of naturalistic inquiry. In the first place, many of the concepts involved in our understanding of these phenomena, including *belief*, *desire*, and *meaning* (ibid.: 207), are part of our commonsense conception of things. Their study therefore belongs to the realm of ethnoscience: "intentional phenomena relate to people and what they do as viewed from the standpoint of human interests and unreflective thought, and thus will not (so viewed) fall within naturalistic theory, which seeks to set such factors aside" (ibid.: 208). We might wonder why the fact that intentional phenomena are involved in our commonsense understanding

of ourselves, and reflect our own perspective on the world, entails that they cannot be studied from the more detached (objective) perspective of science. The clear implication is that nothing in reality answers to the concepts involved in our commonsense conception of intentional phenomena, that there are no real phenomena here to be explained. Chomsky says the following:

> It is possible that natural language has only syntax and pragmatics; it has a "semantics" only in the sense of "the study of how this instrument, whose formal structure and potentialities of expression are the subject of syntactic investigation, is actually put to use in a speech community" . . . In this view, natural language consists of internalist computations and performance systems that access them . . . There will be no provision for what Scott Soames calls "the central semantic fact about language . . . that it is used to represent the world" [Soames 1989: 591], because it is not assumed that language is used to represent the world, in the intended sense . . . (Chomsky 1995: 26–7)

Elsewhere he puts the point more starkly: "The question, 'to what does the word X refer?' has no clear sense, whether posed for Peter [i.e., Peter's I-language], or (more mysteriously) for some 'common language'. In general, a word, even of the simplest kind, does not pick out an entity of the world, or of our 'belief space'" (Chomsky forthcoming: 14).

Chomsky does not deny the existence, or indeed the explanatory interest, of naturalistic semantic theories that purport to explain certain facts about language in terms of a relation R alleged to hold between linguistic expressions and some domain of extra-linguistic entities. According to such accounts, for example, the relation R holds between the word "London" (used in a certain way) and the city on the Thames. It is natural to interpret R as the reference relation, or perhaps as the representation relation. But doing so gains us no additional explanatory purchase, according to Chomsky, since these latter relations are obscure. It is better, he claims, to regard so-called "semantic" theories as characterizing a level of syntax, and the technical notions introduced by such theories (e.g., relation R) as denoting aspects of the structural descriptions postulated by the grammar (Chomsky 1992a: 223). In particular, semantic theories are to be regarded as "part of an interface level" (ibid.: 223), although as far as I know Chomsky does not develop this idea, or explain why "the information relevant to the . . . meaning [of an expression]," as he puts it (Chomsky 2000: 181), is part of the interface between the language faculty and other cognitive systems, including sensorimotor systems. I shall have something to say about this idea, in the context of computational theories, in the next section.

The idea that naturalistic (internalist) semantics is really a form of syntax is only partially clarified by Chomsky's discussion of concrete examples. He elaborates what is surely a "best case" for his thesis. The differential behavior of pronouns in various contexts is accounted for by theories of anaphora and binding (Chomsky 1992a: 223). Such theories posit a relation, call it R, that we

might be inclined to interpret as *coreference*, but the explanations that R supports do not presuppose any relation between linguistic expressions and some domain of extra-linguistic entities. But it is less clear how to understand Chomsky's general claim that internalist semantics is a form of syntax. Asking us to consider the sentence "John is painting the house brown," he says "a semantic property is that one of the final two words can be used to refer to certain kinds of things, and the other expresses a property of these" (ibid.: 219), but, given what Chomsky has said about "the reference relation" and the idea that language represents extra-linguistic entities, one wonders how to interpret his talk of a word "refer[ring] to . . . kinds of things" or "express[ing] a property." He notes that formal relations between lexical items can be expressed in terms of properties of things, as, for example, we can express the relation between *house* and *building* by saying that all houses are buildings, but then adds, rather obscurely, that "[c]ertain [entailment] relations happen to be interesting ones . . . because of the ways that I-languages are embedded in performance systems that use these instructions for various human activities." This remark is intended to cover a lot of ground.

Chomsky's critique of semantics, and of the fundamental idea of representing the world, cuts a wider swath than the study of language. Discussing David Marr's and Shimon Ullman's work in visual perception, he says:

> There is no meaningful question about the "content" of the internal representations of a person seeing a cube under the conditions of the experiments . . . or about the content of the frog's "representation of" a fly or of a moving dot in the standard experimental studies of frog vision. No notion like "content", or "representation of," figures within the theory, so there are no answers to be given as to their nature. The same is true when Marr writes that he is studying vision as "a mapping from one representation to another . . ." (Marr 1982, 31) – where "representation" is not to be understood relationally, as "representation of." (Chomsky 1995: 52–3)

Just as in the naturalist study of language, Chomsky maintains that talk of *representations* in the study of perceptual systems is to be interpreted as referring to postulated internal structures whose theoretically important properties are formal or syntactic. Questions as to what these structures *represent* receive no answers within these theories, the notion of *representing* (and its correlate *misrepresenting*) playing no role in the theories. Naturalistic theories are concerned solely with the processes by which these structures are derived, and the uses to which they are put by subsequent performance systems. If Chomsky is right, then the disputes among interpreters of computational vision theories concerning the "problems solved" by visual mechanisms, or whether the postulated structures have narrow or wide content, are seriously misguided.[7] Such disputes reflect the preoccupations of the philosophical community, not legitimate concerns of naturalistic inquiry.

Chomsky has suggested another reason why issues of intentionality may elude naturalistic inquiry. A full understanding of the mind (including how conscious-

ness arises from neural structures) may lie outside our biologically determined cognitive capacities, and hence is likely to remain a mystery for us (Chomsky 1991; 1995: 27). It should not be surprising, he claims, if we are unable to answer all questions that we are capable of posing (Chomsky 1992b: 123–4). In my view, however, it would be premature at this stage of inquiry to conclude (as does McGinn 1989, 1991) that apparent "explanatory gaps" in our understanding of crucial aspects of language and mind are due to inherent limitations in our cognitive capacities, and Chomsky does not do so. These "gaps" may be an artifact of unreasonable constraints on the explanation of mental phenomena engendered by methodological dualism. I shall have more to say on this possibility below.

In my view, Chomsky's characterization of naturalistic inquiry into mind and language is fundamentally correct, at least in the following respects. The study of the computational mechanisms underlying our cognitive capacities is, typically, internalist. Moreover, there is a general presumption in naturalistic inquiry in favor of internalist, or narrow, individuative schemes.[8] And finally, the characterization of these mechanisms in computational cognitive theories is, in an important respect, non-semantic. However, I shall argue, representational content does play an important role in computational models. By failing to recognize this role, Chomsky has underestimated the potential of naturalistic inquiry for explaining intentionality.

4 Computation and Content

It is widely held that intentionally characterized cognitive or rational capacities of agents require explanation by appeal to intentionally characterized internal states. I have argued in a series of papers that this idea is mistaken, at least in the following sense: the explanations provided by computational cognitive theories advert to states and processes that are not essentially intentional. In other words, computational states and processes are not *individuated* by reference to their semantic properties; a given computational state may, in some counterfactual circumstances, have a different semantic content, or no content at all, and nonetheless be the same computational state. However, the states and processes characterized by genuinely explanatory computational theories do have representational content, and – this is where I disagree with Chomsky – their content serves an important explanatory function in such theories.

Disputes about whether or not computational theories individuate the states they characterize in semantic terms turn on how the level of description that Marr called the "theory of the computation" should be interpreted. The theory of the computation provides a *canonical description* of the function(s) computed by a computational mechanism, what the device does.[9] By a "canonical description" I mean the characterization that is decisive for questions of individuation

or taxonomy. Interpreters of Marr's theory of vision have taken the canonical description of the function(s) computed by the various components of the visual system to be a semantic characterization, although they disagree about what the correct semantic characterization is, and whether this semantic characterization is externalist or not.[10]

Marr often describes the visual system in semantic terms. He speaks of it "detecting edges" and "representing objective aspects of the visual world," but, as Chomsky notes, we should be careful not to read individuative significance into everything a theorist says. I have argued (see Egan 1995) that the canonical description of the function computed by a computationally characterized mechanism is a *mathematical description*. For example, Marr describes a component of early visual processing responsible for the initial filtering of the image. Although there are many ways to informally describe what this filter does, Marr is careful to point out that the theoretically important characterization, from a computational point of view, is a mathematical characterization: the device computes the Laplacean convolved with a Gaussian (Marr 1982: 337). As it happens, it takes as input light intensity values at points in the retinal image and calculates the rate of change of intensity over the image. But as far as the computational characterization of the device is concerned, it does not matter that input values represent *light intensities* and output values the rate of change of *light intensity*. The computational theory characterizes the visual filter as a member of a well-understood class of mathematical devices that have nothing to do with the transduction of light.

My claim that the canonical description of a computational device is not a semantic characterization needs an obvious qualification. Given that the canonical description specifies the mathematical function computed by the device, it *is* a semantic characterization. But mathematical characterization is not what theorists typically have in mind when they talk about "the semantic interpretation of a device." The semantic interpretation of a visual mechanism assigns *visual* contents to the states it characterizes. For example, it may interpret some structures as representing visible *edges* in the scene. A parsing theory will assign appropriate linguistic contents. It will interpret some structures as noun phrases, others as verb phrases. The canonical characterization prescinds from these contents. It construes a computational mechanism as representing mathematical objects, not perceptible properties of the scene or linguistic objects.

Let me return to the visual filter described by Marr. Considered as a computational mechanism, the filter computes the mathematical function that it does whether it is part of a visual system or an auditory system, in other words, independently of the environment – even the *internal* environment – in which it is normally embedded. In fact, it is not implausible to suppose that each sensory modality has one of these same computational mechanisms, since it just computes a curve-smoothing function.

If I am right, then Chomsky's radical internalism is confirmed by computa-

tional practice. Computational mechanisms are individuated independently of environment, even intra-organism environment. On Chomsky's account, a given I-language – a mechanism that generates a particular class of structural descriptions – is not essentially a *linguistic* object. It counts as a language only in virtue of its integration into performance systems involved in language comprehension and speech production. Similarly, the filter described above is a visual mechanism only because it is embedded in systems involved in the transduction of light. It is not essentially a visual mechanism. The *same* mechanism may play an important role in auditory or tactile perception.

Similar considerations may seem to support Chomsky's claim that questions about the content of the structures postulated by computational models are of no theoretical interest. But we need to be careful here. From the fact that a computational theory is an account of the processes by which certain internal structures are derived, and the fact that representational content plays no individuative role within a computational theory, it does not follow that representational content plays no *explanatory* role within the theory. It is legitimate to ask *which internal structures* a computational vision theory, in particular, is concerned to describe and what constructing these structures does for the organism. These questions cannot be answered by appeal to purely syntactic or formal notions.

As noted above, a computationally characterized mechanism is not essentially a linguistic or a visual system. We can specify the *cognitive* (as opposed to the *mathematical*) function subserved by a computational mechanism only by considering how it is embedded in the surrounding environment (including the internal environment). The answer to the question *which internal structures is a computational vision theory concerned to describe?* is *structures that co-vary with tokenings of visible properties, such as changes in depth and surface orientation, in the immediate environment*. An organism that has a mechanism that can detect such properties by transducing light is an organism that can *see*.

The fact that a given computational mechanism constructs structures that track visible property tokenings is a purely contingent matter, from the point of view of its computational characterization. These structures would not track the same properties in all possible worlds. The structures constructed according to a well-defined algorithm specified by the computational theory, might, in different environments, track any number of different properties, not all of which are salient or useful for the organism to detect. Of course, we presume that computational mechanisms are *adaptations*; organisms have such mechanisms because they enhanced fitness in the ancestral environment. But being adapted to an environment is itself a contingent feature of a computational mechanism, so regarded. It is coherent to imagine the same (type of) computational mechanism being built by IBM or coalescing fully formed out of a swamp.

The semantic interpretation of a computational mechanism specifies which properties are tracked by the posited structures when the mechanism is functioning properly in its normal (internal and external) environment. An interpretation

of a computational system is given by an interpretation function that specifies a mapping between equivalence classes of physical states of the system and elements of some represented domain. To interpret a device as a parser is to specify a mapping between states of the device and syntactic items such as noun phrases or verb phrases; to interpret a device as a visual system is to specify a mapping between states of the device and tokenings of visible properties such as changes of depth in the scene. Precisely because the mechanism, as computationally characterized, would not track these properties in every environment, the semantic interpretation of the device is not an essential characterization, and cannot serve to individuate it. However, the semantic interpretation enables us to specify the cognitive function of the mechanism, to characterize it as computing depth from disparity, for example, or as computing the syntactic structure of a sentence. Without it we would be unable to see, or to say, what the device does, in any sense that is of interest to the theorist of cognition, as opposed to the mathematician. The semantic interpretation is required to explain how a formally characterized process, in a given context (its "normal environment"), constitutes the exercise of a cognitive capacity, such as detecting depth or parsing a sentence.

While computational mechanisms are individuated narrowly, independent of both the external and internal environment, the individuative conditions on content typically will depend on features of the environment. For computational models of perception, the content ascribed to the states and structures of the device will be straightforwardly externalist, specifying the distal properties tracked by the tokened internal structures. For example, the structures that Marr calls *edges* are tokened in the presence of a disjunctive distal property, namely, a change in depth, surface orientation, illumination, or reflectance. The content is (in part) externally determined; these structures cannot be expected to track, hence to represent, this property in every possible environment. In some weird counterfactual environments they may track no salient or easily characterizable property; in such circumstances they would represent no distal property.[11] A similar general point applies to computational linguistic mechanisms. Whether a particular structure constructed by a parser is correctly interpreted as a *noun phrase* will depend, in part, upon how the mechanism is embedded in the organism; it will depend, for example, upon the performance systems to which it is connected. A type-identical structure, serving as input to systems involved in locomotion, would not be correctly interpreted as a noun phrase. The general point is that the conditions on semantic individuation will be "wider" than those that determine computational individuation. The conditions on semantic interpretation will include aspects of the environment. Given that the meaning of a computationally characterized state or structure is not an essential property of it, the claim that computational theory, or scientific psychology more generally, is internalist is not thereby impugned.

Defenders of orthodox representationalism will object to the claim that content

is a non-essential (i.e. non-individuative) property of computational states and structures. But I can see no other way to square computationalism with intentional psychology, more specifically, with the view that cognitive, or intentionally characterized, capacities of agents require explanation by reference to internal states that have representational content. Computationally characterized structures are individuated non-semantically, their individuation conditions given by a *realization function* that maps equivalence classes of physical states to elements of a symbol system.[12] The computational processes that operate on these structures are not sensitive to whatever content the structures are assigned by the appropriate *interpretation function* – they are sensitive only to the non-semantic, or, broadly speaking, "physical" features of the structures specified by what I'm calling the *realization function*. This requirement is known as the *formality condition* (see Fodor 1980). It is precisely because computational psychology respects the formality condition that the computational model of mind has held such promise for materialistically minded philosophers of mind. By construing mental processes as formal (i.e., non-semantic) processes, computational models illustrate how mental states can have both causal and representational properties, how mental processes can be physically realizable and also respect canons of rationality. But then this invites the following question: if mental processes are not sensitive to semantic properties – if semantic properties are epiphenomenal in computational models – then what work are semantic properties really doing? Indeed, it invites eliminativism about mental content (see Stich 1983).

It doesn't invite eliminativism *tout court*. The fact that content is epiphenomenal in computational models does not undermine the claim that content plays an indispensable role in commonsense psychology, or that content is individuative of beliefs and desires, the states appealed to in commonsense explanations of behavior. But this should be of little comfort to an intentionalist who is also committed to naturalism, more precisely to the view that scientific investigation will eventually reveal more about the nature of the intentional states causally responsible for our behavior. The propositional attitudes characterized by folk psychology and the states characterized by a computational theory differ in the following respect: unlike propositional attitudes, computational states have an independent (i.e., non-semantic) characterization that serves as the basis for a precise specification of the role that these states play in the behavior and capacities of the system. Beliefs are subsumed by folk psychology's predictive and explanatory principles only by their contents. The worry is that there is a clear role for content only when there is no independent specification of intentional states. But an independent characterization of intentional states is exactly what the intentionalist who is also a naturalist is hoping science will eventually provide. It would be small consolation to the friend of intentional psychology if content has a role to play in psychology only in its folk or pre-scientific incarnation. This, of course, is precisely Chomsky's view of the matter.

But on the account sketched above, the existence of an independent specifica-

tion of psychological states does not make content ascription idle. Content plays an indispensable explanatory role in computational models, even though it plays neither an individuative nor a causal role. A semantic interpretation of a computational mechanism is necessary to explain how a formally characterized process, in a certain context (say, when connected to certain performance systems, or situated in a certain external environment) constitutes the exercise of a *cognitive* capacity, such as computing the depth of the scene, or the syntactic structure of an acoustic input. The explananda of cognitive theories are formulated in intentional terms – the device recovers 3D structure from 2D images, or it adds, or it parses input strings. To explain how it performs these intentionally characterized tasks some of the internal structures constructed by the device must be interpreted as representing visible distal properties, or addends and sums, or noun phrases. The semantic characterization forms a bridge between the explananda of the theory and the formal characterization of the device that constitutes the explanatory core of the theory.

5 Intentionality and Naturalistic Inquiry

Chomsky, however, would not be impressed by the alleged reconciliation between computationalism and the concerns of intentional psychology. Like Quine, he is eliminativist about mental content, holding that it has no place in legitimate science. Commenting on an earlier paper of mine on the role of content in computational psychological models, he says:

> We can say, if we like, that "where the constraints that normally enable an organism to compute a cognitive function are not satisfied, it will fail to represent its environment" (Egan 1995); but that "failure" is our way of describing some human end we impose for reasons unrelated to naturalistic inquiry . . . Nor is it relevant that consideration of "representation" in normal environments allows us to associate the system under analysis with the informally described function of vision. It's no task of science to conform to the categories of intuition, or to decide if it's still "vision" in abnormal environments . . . The study of perception naturally begins with informally presented "cognitive tasks," but cares little whether something similar to them is discovered as it progresses. (1995: 56)

I have claimed that a semantic characterization of a mechanism allows us to answer questions about the behavior and capacities of that mechanism (or the organism in which the mechanism is embedded) that we would be unable to answer with only a formal, computational characterization. For example, the fact that the visual system constructs structures that co-vary with changes in depth and surface orientation in the scene, and hence *represents* these distal properties, explains the organism's success in negotiating its environment. On occasions when it tokens a structure in the absence of the distal property to which the

structure is mapped in the appropriate interpretation, we can say that it misrepresents its environment, that it makes a *mistake*. Of course, this is a normative characterization of the device's behavior. As Chomsky notes, to describe the device as making a mistake is to impose our own interests and expectations on it. He concludes that such a characterization has no place in naturalistic inquiry. But the attribution (of a mistake) also helps to explain aspects of the organism's interaction with the environment, perhaps why it fails at certain tasks essential for its own survival. The fact that such notions as *representation*, *misrepresentation*, or *error* can be reconstructed within computational models is innocuous, from a naturalistic point of view. There are no unreduced normative elements, no unexplicated representation relations (such as *intrinsic intentionality*), in computational accounts of cognitive capacities. Content attribution is just interpretation, which is grounded in facts about how computational mechanisms are connected to their (internal and external) environment. Computational cognitive science has made significant progress towards understanding the place of such mental phenomena as representation, misrepresentation, and error in the natural world, that is, in the world described by the rest of science. To the extent that computational models are well confirmed, they provide naturalistic explanations of these phenomena. Computational theories do not assume, or take for granted, the intentionality of the mechanisms they characterize; rather, they aim to provide genuine explanations of intentionality. They give an essentially formal characterization of processes and capacities that we, pre-theoretically, describe in intentional terms. Computational theories explain how certain natural processes – those underlying the activities we pre-theoretically describe as seeing, adding, or parsing input strings – adhere to (or, occasionally, violate) rational strictures.

Chomsky disparages the idea that a scientific theory should concern itself with its informally described explananda. But it is a legitimate constraint on the acceptability of a scientific theory that it explain the phenomena in its domain at least as well as its competitors. This doesn't require that a theory preserve the categories of pre-theoretic intuition. A new theory may, in fact, imply that nothing in reality answers to the old categories, much as the Copernican-Keplerian model of the universe implied that there was no retrograde motion of the planets. But when it does imply that nothing answers to the old categories, the theory must at least provide the basis for an explanation of the appearances, as, for example, the C-K model did in explaining apparent retrograde motion as an effect of the relative motion of the earth.[13] Chomsky's view to the contrary notwithstanding, a computational theory that purported to provide an account of human vision would be justly criticized if it failed to explain our ability to recover the spatial properties of the scene before our eyes. I am claiming that a theory cannot meet this explanatory burden unless some of the states it postulates are interpreted as representations of those properties. To go further, to insist that the theory posit states that, arguably like the states posited by commonsense psychology, are essentially intentional, or states that would count as visual states

in every possible world, or states that have "intrinsic intentionality," is to manifest the attitude that Chomsky calls *methodological dualism*, the idea that the study of mind, unlike scientific inquiry in other domains, should be held to independent, "philosophical" standards.[14] I agree with Chomsky that such claims have no legitimate basis.

A final point about content: it would be a mistake to think that content attribution is to be viewed as a temporary expedient, to be dispensed with once we have a fully naturalistic account of the mind in hand. As long as we have interests in treating ourselves and others as rational agents, content will continue to be an indispensable element of our understanding of ourselves. Chomsky suggests (1995: 57) that these interests are the commonsense residue of a fundamentally dualistic picture of the world, and hence that they should be ignored by naturalistic inquiry. He would relegate our interests in understanding ourselves as rational agents to ethnoscience, and leave the satisfaction of these interests in the hands of novelists and folklorists. But whatever the origin of these interests, a commitment to naturalism does not preclude acknowledging them and attempting to satisfy them within science, as, I have argued, computational theorists are in the business of doing. It is not only the cognitive sciences that are grounded in the desire to understand ourselves and our place in the world – the biological sciences are grounded in such interests as well. From the detached perspective of fundamental physics, the difference between life and non-living matter is no less arbitrary than the difference between a rational process and a mistake. There is no reason why science should not aim to tell us about the features of our experience that interest us, even if it eventually tells us that we are fundamentally mistaken about these features. It is, after all, our science.

Chomsky's reply: pp. 268–74.

Notes

1 It is not entirely uncontroversial – Thomas Nagel, for example, does not endorse it. Nagel claims that naturalistic methodology, which strives to be objective, will fail to account for the most important feature of the mental, namely, the subject's own point of view. See Nagel 1986.

2 For a sample of the externalist literature see Millikan 1984; Burge 1986; Papineau 1987, 1993; Davies 1991; Shapiro 1993, 1997; Peacocke 1994; and Wilson 1994, 1995. See Egan 1999 for a critique of this trend.

3 Chomsky cites the interpretation of the work of vision theorists Marr and Ullman as a case in point (see Chomsky 1995: 52–5). Their informal exposition suggests an externalist perspective, but close examination of the theories themselves reveals

otherwise. See Egan 1995 for my defense of a similar claim. Of course, distilling a theory's individuative principles from the theorist's exposition is a non-trivial task.

4 The I-language is not to be identified with the set of linguistic expressions (structural descriptions) generated by the procedure. Chomsky notes that distinct I-languages might, in principle, generate the same set of linguistic expressions (Chomsky 1992a: 211).

5 Putnam (1975) does not draw this further conclusion, though Burge (1979) does, and it is generally accepted by externalists.

6 "Anything you can do, I can do better" is a common internalist refrain. See, for example, Egan 1992 and Patterson 1994. Of course, externalists have denied the claim. See the references in note 2, especially Burge 1986; Peacocke 1994; and Wilson 1995.

7 See, for example, Burge 1986; Segal 1989, 1991; and Shapiro 1993, 1997.

8 See Egan 1999 for an argument for this general claim.

9 The other two levels in Marr's explanatory hierarchy are the *algorithm* (which specifies a rule for computing the function characterized by the theory of the computation) and the *implementation* (which provides a physical description of the device).

10 See note 6.

11 They would still represent features of the retinal image. See Egan 1995 for elaboration of this point.

12 The claim that computational devices are individuated non-semantically applies also to connectionist devices, although they don't have symbols.

13 I say "basis for an explanation" because the theory may not contain the vocabulary required for the explanation, or the full explanation may require appeal to background assumptions or facts that are not part of the theory itself.

14 In my view, the distinction between "intrinsic intentionality" and merely "derived intentionality" (see Searle 1980 and Haugeland 1981) is an artifact of methodological dualism. Intrinsic intentionality is simply *unanalyzed* intentionality, and derived intentionality, the kind that computational mechanisms are said to have, is intentionality for which we have a naturalistic explanation. The distinction reveals a gap in our understanding, not a difference in the world. It is a gap that the success of computational theorizing will narrow.

References

Burge, T. 1979: Individualism and the Mental. *Midwest Studies in Philosophy*, 4, Minneapolis: University of Minnesota Press.

Burge, T. 1986: Individualism and Psychology. *The Philosophical Review*, 95, 3–45.

Chomsky, N. 1986: *Knowledge of Language*. New York: Praeger.

Chomsky, N. 1991: Linguistics and Cognitive Science: Problems and Mysteries. In A. Kasher (ed.), *The Chomskyan Turn*, Oxford: Blackwell.

Chomsky, N. 1992a: Explaining Language Use. *Philosophical Topics*, 20, 205–31.

Chomsky, N. 1992b: Language and Interpretation: Philosophical Reflections and Empirical

Inquiry. In J. Earman (ed.), *Inference, Explanation, and Other Frustrations*, Berkeley: University of California Press.
Chomsky, N. 1994: Naturalism and Dualism in the Study of Language and Mind. *International Journal of Philosophical Studies*, 2, 181–209.
Chomsky, N. 1995: Language and Nature. *Mind*, 104, 1–61.
Chomsky, N. 2000: Internalist Explorations. In Chomsky, *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press.
Davies, M. 1991: Individualism and Perceptual Content. *Mind*, 100, 461–84.
Egan, F. 1992: Individualism, Computation, and Perceptual Content. *Mind*, 101, 443–59.
Egan, F. 1995: Computation and Content. *The Philosophical Review*, 104, 181–203.
Egan, F. 1999: In Defense of Narrow Mindedness. *Mind and Language*, 14 (June), 177–94.
Fodor, J. A. 1980: Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology. *Behavioral and Brain Sciences*, 3, 63–73.
Haugeland, J. 1981: Semantic Engines: An Introduction to Mind Design. In J. Haugeland (ed.), *Mind Design*, Cambridge, Mass.: MIT Press.
Marr, D. 1982: *Vision*. New York: Freeman.
McGinn, C. 1989: Can We Solve the Mind–Body Problem? *Mind*, 98, 349–66.
McGinn, C. 1991: *The Problem of Consciousness*. Oxford: Blackwell.
Millikan, R. 1984: *Language, Thought, and Other Biological Categories*. Cambridge, Mass.: MIT Press.
Nagel, T. 1986: *The View from Nowhere*. Oxford: Oxford University Press.
Papineau, D. 1987: *Reality and Representation*. Oxford: Oxford University Press.
Papineau, D. 1993: *Philosophical Naturalism*. Oxford: Blackwell.
Patterson, S. 1994: Success-Orientation and Individualism in Marr's Theory of Vision. In K. Akins (ed.), *Perception*, New York: Oxford University Press.
Peacocke, C. 1994: Content, Computation, and Externalism. *Mind and Language*, 9, 303–35.
Putnam, H. 1975: The Meaning of Meaning. In *Mind, Language, and Reality*. Cambridge: Cambridge University Press.
Searle, J. 1980: Minds, Brains, and Programs. *Behavioral and Brain Sciences*, 3, 417–24.
Searle, J. 1991: *The Rediscovery of the Mind*. Cambridge, Mass.: MIT Press.
Segal, G. 1989: On Seeing What is Not There. *The Philosophical Review*, 98, 189–214.
Segal, G. 1991: Defense of a Reasonable Individualism. *Mind*, 100, 485–93.
Shapiro, L. 1993: Content, Kinds, and Individualism in Marr's Theory of Vision. *The Philosophical Review*, 102, 489–513.
Shapiro, L. 1997: A Clearer Vision. *Philosophy of Science*, 64, 131–53.
Soames, S. 1989: Semantics and Semantic Competence. *Philosophical Perspectives*, 3, 575–96.
Stich, S. P. 1983: *From Folk Psychology to Cognitive Science*. Cambridge, Mass.: MIT Press.
Wilson, R. 1994: Wide Computationalism. *Mind*, 103, 351–72.
Wilson, R. 1995: *Cartesian Psychology and Physical Minds*. Cambridge: Cambridge University Press.

5
Chomsky, Intentionality, and a CRTT

GEORGES REY

The appropriateness of the intentional idiom, at a certain level, is an obscure matter of fact.

Chomsky (1980b: 47)

1 Introduction

Chomsky's work has been a major inspiration for the cognitive revolution in psychology and related disciplines, whereby questions about the structure of mind have begun to be susceptible to serious scientific investigation. It is the source of some of the best ideas of this revolution, showing how a perfectly ordinary human competence like grammar can possess an enormously subtle and intricate structure; how this likely involves a similarly intricate, largely innate Language Acquisition Device (LAD) in the human brain; and, indeed, how underlying competencies of a system, not merely its superficial performance, are appropriate objects of theoretical inquiry. All of these Chomsky seems to me to have established beyond reasonable doubt, and I will take for granted in my discussion here.

One crucial idea that seems to underlie these developments has been the Computational Representational Theory of Thought (CRTT), or the theory that thought processes consist in computations defined over representations entokened in the brain. Chomsky has frequently presented his account of grammatical competence in what would appear to be precisely the terms of such a theory. Just how seriously he intended to do this, however, is a topic of considerable controversy, partly because it has not been at all clear what he has meant by the crucial terms "computation" and "representation." Psychologists and computer scientists have worried about how the computations Chomsky proposes correspond to actual computations that occur in space and time. Is he describing steps that are actually executed in parsing, production or acquisition, or in quiet reflection? Or is he merely *characterizing* a competence that might be *computed* in the brain in any number of ways, some of them as remote from his characterization as standard multiplication algorithms are from Peano's axioms?

Many philosophers have been concerned both with this issue and with an issue

closer to their traditional concerns, not so much the *computational* as the *representational* aspect of CRTT. Here the question might be said to be not merely the *reality* of the computational states or processes Chomsky describes, but, as it's often put, their *psychological* reality. Are the "computations" Chomsky describes really part of the *mental* life of, say, an ordinary child who might be quite incapable of understanding linguistic theory? Are the categories, and even the rules and principles,[1] of universal grammar really *represented*, *followed*, and even *known* innately by all normal human neonates, as Chomsky, resuscitating Rationalist doctrines, has often claimed? Or are the categories merely non-representational states of some internal system that only *satisfies* the principles, in the way that processes of digestion and metabolism satisfy principles and laws of chemistry and thermodynamics, presumably without representing anything?[2]

A remarkable fact about these issues is how they've failed to be resolved over the thirty or so years during which they've been disputed: to mention a representative sample, they were raised early on in Nagel 1969/1972, Quine 1972, and Harman 1974, and have been reiterated in various forms in Searle 1980, Matthews 1980, Rosenthal 1980, Sober 1980, Stich 1980, Stabler 1983, Soames 1984, Devitt and Sterelny 1987/forthcoming, and Wright 1989, and persist in Searle 1992 and Devitt (forthcoming). Chomsky has not been shy in replying (see, e.g., his 1975c, 1980b, 1986). However, I think it safe to say that few of these critics have felt that their worries have been adequately addressed. Speaking as someone quite sympathetic both to CRTT generally and to much of Chomsky's program in particular, I have often been puzzled by his replies. Although he has rightly noted a number of errors in many of the discussions, he seems to me not to have addressed some deeper issues that I think are reasonably bothering many of these writers.

In the short space here I want to focus on one specific issue that I think is at the heart of many of these debates, the issue of intentionality. This is, roughly speaking, the property by virtue of which representations, thoughts, ideas, and propositional attitudes (such as thinking, knowing, desiring, cognizing) have the *contents* they do, are *about*, e.g., objects, edges, contours, pitches, people, noun phrases, complementizers, traces; indeed, anything we can "think about."[3] Intentionality is a notoriously problematic property, evidently not needed in accounts of any part of nature other than minds and their products. I suspect that what's bothering many of Chomsky's critics is the unclarity about the ways in which his accounts of grammatical competence are actually committed to any sort of intentionality, for example, to grammatical principles being the represented *content* of any mental state of a normal speaker.

I will begin by considering Chomsky's presentation of his theory as an instance of a CRTT (section 2), proceeding to consider the prospects and problems of such a theory, and particularly what justifies taking on its representational burdens (section 3). Chomsky sometimes claims that the intentional idioms he employs should be understood technically, but in fact he hasn't provided any

explicit accounts of them, and natural candidates that he implicitly suggests are notoriously problematic (section 4). Sometimes he also seems to claim that intentional notions are inessential to his account, and seems even to retract his commitment to CRTT. This may be defensible, but it does fly in the face of the rich intentional descriptions he and other linguists provide of linguistic phenomena, and their evident commitment to a CRTT (section 5). He seems to want to have intentionality and a CRTT and to deny them too. I will conclude (section 6) by setting out the dilemma that seems to me to be posed by the passages of Chomsky's that I cite, and which I hope he may clarify in his replies.

A disclaimer from the start: I am not a linguist, and, aside from a few peripheral *psycho*linguistic issues that I will simply note in section 5, I do not expect to say anything that turns on details of linguistic theory. I am concerned solely with Chomsky's *meta*linguistics, in particular his appeals to representations, psychological reality, and a CRTT. In a way the point of my entire discussion is to see what real difference, if any, his metalinguistic claims make to his or anyone's substantive linguistic ones.

2 Chomsky's Commitment to CRTT

Chomsky has very often couched his theory expressly in terms of a CRTT (e.g., Chomsky 1969: 155–6; 1977: 3–6; 1980a: 54; 1986: 239; 1993: 40, 52; 1994: 153–4; 1995b: 225–35), and he is sometimes quite explicit about what he takes this to mean.[4] For example:

> I know of no other account that even attempts to deal with innumerable facts of [a standard linguistic nature], and I know of no proposed explanation for the fact that our judgments and behavior accord with certain rule systems other than the assumption that computation involving such rules and the representations they provide takes place in the mind – exactly the assumption one would make without hesitation in the study of perception and other topics. (Chomsky 1980a: 129–30)

He sometimes expands on this idea by making precisely the kind of comparison with *computers* that defenders of CRTT often have in mind. Not many pages after the above passage, he continues:

> Linguistics is the abstract study of certain mechanisms, their growth and maturation. We may impute existence to postulated structures at the initial, intermediate, and steady states in just the same sense as we impute existence to a program that we believe to be somehow represented in a computer or that we postulate to account for the mental representation of a three-dimensional object in the visual field. (ibid.: 188)

Even more specifically, he contrasts a "non-cognizing missile that acts in accord with laws governing the orbit of the moon" with a "cognizing missile that

computes its trajectory by reference to a representation of these laws," the latter (abstracting from its specific task-orientation) serving as a model for a normal speaker's understanding of language (Chomsky 1980a: 102–3, 275n. 61).

Now, many have wondered what it is about grammatical competence that requires CRTT. What makes the normal speaker like a cognizing rather than a non-cognizing missile? It's difficult to find passages where Chomsky directly addresses this issue. He and others frequently *describe* grammatical competence in a fashion that invites CRTT.[5] For example, Chomsky describes language acquisition as a kind of inferential process, specifically a kind of abductive hypothesis confirmation:[6]

> A child who is capable of language learning must have:
> (i) a technique for representing input signals
> (ii) a way of representing the structural information about these signals
> (iii) some initial delimitation of a class of possible hypotheses about language structure
> (iv) a method for determining what each such hypothesis implies with respect to each sentence
> (v) a method for selecting one of the (presumably, infinitely many) hypotheses that are allowed by (iii) and are compatible with the given primary linguistic data. (Chomsky 1965: 30)

However, when pressed about why we need to understand grammatical competence in such a robustly mentalistic fashion, he tends merely to reiterate the claims themselves. Thus, in response to worries about the psychological reality of grammar, he replies:[7] "What is 'psychological reality', as distinct from 'truth', in a certain domain? . . . I am not convinced that there is any such distinction, and see no reason not to take our theories tentatively to be true at the level of description at which we are working" (Chomsky 1980a: 107). It's as though he takes the question of *psychological* reality to be identical to the question of a *realistic* interpretation of his theory generally:

> The "realist interpretation" of linguistic theory is assumed throughout, and it is argued that the competence attained by the normal speaker-hearer is represented by a transformational generative grammar, which determines the *representation* of each sentence . . . These *representations* are then employed in the use and understanding of language. (Chomsky 1975c: 45; italics mine)

But, of course, someone could be a thoroughgoing realist (i.e., non-instrumentalist) about the categories of a grammar, and believe that the rules truly describe some system in the brain, much as they might be a realist about immunology describing some system in the body, without for a moment taking either the categories or the rules to be *psychologically* real – i.e., to be either *represented* or defined over *representations*, in addition to being "real"(ized). Why shouldn't, for example, a PP be a node in a perfectly real, but *unrepresented* tree structure in

the brain, like a node in the structure of some molecule? (We'll return to this possibility in section 5 (iv).)

At one point, Chomsky does recognize that merely appealing to the 'truth' of a theory isn't adequate. Harman rightly points out that:

> Given any theory we take to be true, we can always ask what aspects of the theory correspond to reality and what aspects are mere artifacts of our notation. Geography contains true statements locating mountains and rivers in terms of longitude and latitude without implying that the equator has the sort of physical reality the Mississippi River does. (Harman 1980: 21)

Chomsky agrees: "[Harman] correctly points out an error in my formulation: there is a question of physical (or psychological) reality apart from truth in a certain domain" (Chomsky 1980b: 45). Unfortunately, however, he goes on to consider only a very specific case of extensionally equivalent variants of his theory. Alternative formalizations of a grammar do raise subtle and important issues, and one should always ask which of their expressions refer to causally relevant phenomena. The question I want to raise here is specifically why we should regard any *representations* (*qua* representations) of a grammar as causally efficacious, i.e., why we should regard either its principles or its categories as *psychologically* real at all.[8] Why not regard any intentional talk here as on a par with the talk of latitude and longitude,[9] or treat all the talk of "representation" as metaphor or in some other way an "artifact of our notation"?

In order to get some grip on this issue, it will help to review one major motivation for a fully realistic version of CRTT.

3 Prospects and Problems of CRTT

(i) Computation and its prospects

The chief motivation for CRTT is that it offers a general framework in which to begin to systematize both ordinary and scientific insights about the mind. People and many animals exhibit all sorts of patterns of behavior that seem explicable only on the assumption that they are capable of perceiving, remembering, reasoning, planning, decision making, and, for some, of systematically expressing their thoughts in natural languages. There are two general properties of such thinking things: the transitions between their states are often *rational*, and they occur often by virtue of the *intentional content* of the states. Thus, people are sensitive to deductive, inductive and abductive relations among their thoughts, and to decision-theoretic patterns in much of their planning and behavior, most of which can only be expressed by reference to the truth-valuable contents of the constituent states (it's standardly truth-valuable contents that are at least one of

the relata of "*imply*," "*infer*," "*confirm*," or "*rationalize*"). Even much *ir*rational thought and behavior is understood intentionally, as in, e.g., the gambler fallacy, disregard of base rates, and discounting of future satisfactions.[10]

Until recently, it was widely thought that these properties weren't assimilable to a general physical theory of the world. Descartes (1637/1970: 117–18) famously claimed that "the rational soul . . . could not be in any way extracted from the power of matter . . . but must be expressly created," and Brentano (1874/1973: 97) that the intentional could never be understood physically (see note 17 below for Chomsky's quarrels with all this). However, many people have thought that recent advances in various formalizations of reasoning and computation suggest otherwise, and so have pursued CRTT. It postulates that there exists a medium of formal representation and a set of computational processes defined over it that could plausibly account for rational (and many irrational) phenomena. Since these computations and representations could be physically realized in the brain, it promises an answer at least to Descartes's challenge that *reason* couldn't be "extracted from the power of matter." Indeed, it provides a framework for understanding how a physical event (e.g., a retinal stimulation) could cause a thought (that there's an abyss underfoot) that would in turn cause a physical event (a rush of adrenalin): physical events cause tokenings of representations that, sometimes in rational, computable patterns, cause other tokenings that sometimes cause other purely physical events. The "psychological reality" of the ideas is established by the *causal efficacy* of the entokened representation.

I emphasize that CRTT "promises" such an account. As things stand, CRTT is not so much a theory as a research program. Dan Osherson once compared it to Boyle's postulation of "atoms" as constituents of chemical phenomena, without any clear idea of what the precise character of those atoms might be. It leads us to ask interesting and often empirically testable questions about, for example, the precise character of the medium of representation – its expressive power, the kind of information it needs to represent, whether it consists of sentence-like structures or of "mental images" and "models" – as well as about the character of the computations defined over or between those representations – whether they are serial, parallel, "connectionist" or "dynamic," to what extent they are modularized or "encapsulated" from one another. Discovering the subtle principles and algorithms by which we understand the world and adjust our behavior to it is, of course, not something to be expected in our grandchildren's lifetimes, if ever (Chomsky 1975c: 156–7, 1980a: 250–2). But CRTT does seem to be the only serious framework in which these questions can be posed.

In any case, CRTT has been fruitfully enlisted in a variety of psychological domains such as vision, memory, learning, reasoning, decision making, and natural language parsing. For example, many perceptual illusions can be accounted for by presuming the visual system is making inferences about the relations of surfaces in a three-dimensional space. In theories of decision making,

contradictory attitudes towards the same decision problem can be explained by different evaluative attitudes towards represented gains vs. losses. And standard accounts of natural language parsing seem to involve computing structural descriptions of acoustic input.[11]

Most such models, however, focus largely on the computational aspects of the processes. There has been some, but by no means equal, progress in understanding the notion of representation. Even though computations can be *specified* and studied without appeal to content, content is arguably *presupposed* by any computational theory: in vision, representations of, e.g., edges and surfaces; in decision making, of options, losses, and gains; in parsing, of nouns and verbs. Indeed, even the computations of Turing Machines are standardly defined over *numerals* that represent numbers. Yet the specific content of the representations, and their relations to what they represent, is simply intuitively – more often than not, haphazardly – taken for granted. In this respect, *pace* Chomsky (1988: 108; 1994: 154), CRTTs and "internally represented grammars" are *not* "by and large well understood."

(ii) Representation and its problems

Computational theorists can hardly be blamed. Although representation can seem a natural idea, it displays a number of peculiarities that have long resisted theoretical understanding. Here the problem is not just incorporation into a physical theory, but understanding the bases for intentional ascription *even in its own right*. Just to run quickly through some of the standard problems: Is representation a relation? If so, to what? (How can "Zeus" represent a non-existent god, or "trace" an inaudible syntactic element?) How to account for the difference in content between different representations of the same thing (e.g. "Hesperus" vs. "Phosphorus," "Adjective" vs. "[+N,+V]")? By differences in inferential role? In which case, how do we account for the wide divergences in inferential roles between individuals in states with the same intentional content (e.g., people with different theories of the planets, or different linguists with different theories of "+N")? What about inferentially impoverished creatures (can dogs ignorant of anatomy really represent *bones*? Or kids mystified by linguistics, *c-command*?)? How do we distinguish a disjunctive content that has been correctly applied from a simple one that has been misapplied (does a frog snapping at a beebee correctly represent it as a *fly or beebee* or misrepresent it as a *fly* – or is it some other content, say, *moving dot*?)? To what extent do our answers to these questions depend upon consideration of the normal environment of the system? (Mightn't the same structure that represents NPs in one environment represent some odd social group in another? Don't the "Socrates" thoughts of my Twin-Earth doppelgänger obviously represent someone other than mine do – viz., Socrates' Twin-Earth doppelgänger?[12])

All of these specific questions take on a special significance in the light of what seems to be a worrisome insularity and lack of discipline to intentional ascription. The insularity is manifest in Brentano's tempting claim that the intentional is irreducible to the physical: whatever the ultimate status of this claim, it is certainly true that intentionality seems peculiarly resistant to incorporation in the rest of science. It hasn't *begun* to be explained in anything like the way that, say, the explosiveness of bombs, the transparency of glass, or even the growth of organisms can be.[13] What non-intentional properties do two entities have to share in order to share an intentional one, e.g., to represent the same thing, say, Vienna, or Santa Claus, or the number 2 – or a noun phrase, or a trace? A particular molecular structure? A certain sort of internal organization? A disposition to behave? Some causal relation to the environment? The main part of the problem here doesn't seem to be one of making further empirical discoveries, but of figuring out which ones would be *relevant*. There's something not merely empirically, but *conceptually*, puzzling about the intentional and its relation to the rest of the world (which is why, *pace* Chomsky (1993: 29–30, 90–3; 1996: 18–19; 2000: 147), the question of whether machines can "think" is not as simple and terminological as the question of whether planes "fly").[14]

This insularity is, I think, partly responsible for the even more serious lack of discipline to the idiom. People traffic in intentional explanation by and large with the greatest of ease, often with an impressive measure of predictive success in a wide range of ordinary contexts, from plans and games, to governments, economies, wars. Moreover, as Chomsky (1959) pointed out, intentional explanation seems indispensable to our best scientific explanations of large amounts of human and animal behavior. The problem is that it seems often too *unconstrained*; people cheerfully apply intentional description to all manner of phenomena that don't genuinely satisfy it: hurricanes, comets, flowers, sometimes the universe as a whole (cf. Chomsky 2000: 146).

Particularly troublesome in the twentieth century has been the degree to which it has become commonplace to ascribe all manner of often unconscious psychological states to people on the slenderest of objective bases – often merely because the ascription "makes sense" to, say, the patient or the therapist. The relatively inarticulate world of the small child seems to be a particularly inviting canvas for intentional ascription: consider the extraordinary thoughts attributed to young children by a Freud or a Melanie Klein, on the most amazingly meager evidential base. Small wonder Chomsky encounters some wariness about his attribution to them of cognitive processes that are in many ways even more elaborate.

I rehearse all of these standard problems to emphasize how they arise quite immediately from taking *any* intentional idiom seriously: they can be raised about *any* idiom that involves "aboutness" – whether it be "reference," "thought," "representation," "information" – and so arise equally for our ordinary descriptions of dogs or frogs or human neonates, quite independently of any *technical* intentional notions that may have been introduced.[15]

Indeed, it's important to notice that the problems arise independently of the *specific* ways they arise for natural language. Bach (1987/1994: ch. 4) and Lormand (1996) rightly distinguish "linguo-" or *natural language* semantics, from a "psycho-"semantics. It may well be that the only way to understand the behavior of speechless rats is to suppose that they also perform computations over representations (see, e.g., Gallistel 1990). A theory of the content of those representations might have little to do with the semantics of any natural language, perhaps even in the human case (Chomsky 1993: 34–5).

I hasten to mention problems that are *not* problems for CRTT that need to be considered at this time: consciousness, qualia, and free will. Although these are serious problems, they are by and large not implicated in the wide range of phenomena (memory, reasoning, grammar) for which Chomsky and most other cognitive scientists have enlisted CRTT. CRTT could apply to a system, or portions of it, that, for whatever reason, were inaccessible to consciousness, lacked qualitative feel, and were quite incapable of voluntary, morally responsible action.[16] Indeed, an attraction of CRTT is that it could be realized on practically any system of things that could be organized in the right way, whether or not it is composed of the "right stuff" (Cartesian, Newtonian, quantum-mechanical or otherwise)[17] for these further problematic phenomena.

For the nonce, the problems with intentionality are problems enough. Some of them are so peculiar and unique that they strongly suggest that intentionality marks a major division, if not in nature, then at least in our theories about it, distinguishing the idiosyncratically "mental" from the "physical" along lines Brentano claimed. To a first approximation, a state or process, or a product of a process, is *mental* iff you can ask what it is "about."[18] At any rate, I shall assume this throughout the present discussion. It makes substantive sense of many of the debates about the mind from Descartes to Brentano, from Max Weber's defense of "Verstehen explanation" to Skinner's and Quine's defense of behaviorism. Determining the character and scope of intentional explanation is arguably one of the most important and perplexing problems facing contemporary science. In contrast, the use of mental vocabulary in a way that does not presuppose intentionality would indeed seem to be a mere terminological decision.[19]

4 Technical Notions?

(i) Chomsky's proposals

One reason Chomsky doesn't address many of these problems is that he sometimes seems to regard the notions I am calling intentional as "technical" ones that shouldn't be burdened with the difficulties of ordinary use. In response, for example, to philosophers' qualms about claiming that the principles of grammar are *known* by a child, he retreats to what he calls a "technical term," "cognize":

> To avoid terminological confusion, let me introduce a technical term devised for the purpose, namely "cognize," with the following properties. The particular things we know, we cognize. In the case of English, presented with the examples, "the candidates wanted each other to win" and "the candidates wanted me to vote for each other," we know that the former means that each wanted the other to win, and that the latter is not well-formed with the meaning that each wanted me to vote for the other. We therefore cognize these facts. Furthermore, we cognize the system of mentally-represented rules from which the facts follow. That is, we cognize the grammar that constitutes the current state of our language faculty and the rules of this system as well as the principles that govern their operation. And finally we cognize the innate schematism, along with its rules, principles and conditions. In fact I don't think "cognize" is very far from "know" . . . but this seems to me a relatively minor issue, similar to the question whether the terms "force" and "mass" in physics depart from their conventional sense (as they obviously do). (Chomsky 1980a: 69–70; see also 1975c: 164–5; 1986: 32–3, 265)

On the face of it, however, this is no introduction of a "technical" term: "cognize" is patently just another word for *know* – just less constrained. Indeed, given that a few pages later Chomsky thinks that "it is not at all clear that the ordinary concept of 'knowledge' is even coherent" (1980a: 92), one might wonder why "cognize" is any better off.[20] In any case, the same questions arise for "cognize" as for "know": if we cognize grammars, do we also cognize laws of gravitation, principles of fluid flow, synaptic transmission, or metabolism?[21] Where does the mind (cognizing, representation, intentionality) begin? Why?

Of course, a genuinely computational account, such as *CRTT*, might well afford a definition of "cognize" in terms of some causal/computational relation to representations entokened in the brain. But then we would need some reason for thinking that such a relation is actually implicated at some stage of natural language processing. It is this sort of reason (as well as any such definition) that Chomsky has persistently failed to provide, presumably regarding it as too much of a "performance" issue (1965: 9). The problem he faces is how to capture *cognize* or the like without any indication of the causal role it is meant to play, or whether and why it must involve "representations."

"Representation" and "content," themselves, Chomsky also regards as "technical" terms:

> Take, say, discussion about wide and narrow content, specification of mental representations, or individuation of thought and belief. If the inquiry falls into ethnoscience, we ask how people think and talk about such matters – recognizing, however, that the question cannot be raised directly for "content" and "mental representation," used here in a technical sense. (Chomsky 2000: 165)

But what are these technical senses? No explicit definitions of them are provided. Perhaps they should be regarded as merely "implicitly" defined by their role in his theory.[22] But what he says about that role is not helpful:

> The internalist study of language also speaks of "representations" of various kinds, including phonetic and semantic representations at the "interface" with other systems. But here too we need not ponder what is represented, seeking some objective construction from sounds to things. The representations are postulated entities, to be understood in the manner of a rotating cube, whether it be the result of tachistoscopic presentations of a real rotating cube or stimulation of the retina in some other way; or imagined, for that matter. Accessed by performance systems, the internal representations of language enter into interpretation, thought, and action, but there is no reason to seek any other relation to the world, as might be suggested by a well-known philosophical tradition and inappropriate analogies from informal usage. (ibid.: 160)

More specifically, with regard to linguistic representations:

> PHON(E) and SEM(E) are elements at the "phonetic level" and "semantic levels," respectively; they are phonetic and semantic "representations." These terms have their technical sense; there is nothing "represented" in the sense of representative theories of ideas, for example. These levels are the "interface" between LF [Logical Form] and other systems, providing the information used by the sensorimotor apparatus and other systems of language use. (ibid.: 173–4)

Again, no "technical" characterizations, explicit or implicit, are really on offer here. What are Chomsky's alternatives to common usage? Just what is representation in "the manner of a rotating cube" (what makes the content of some brain state *a rotating cube*)? How do PHON(E) and SEM(E) "provide information"? Doesn't that mean that they are *about*, e.g., phonemic and semantic properties? "Information," again, is an intentional notion,[23] and all the problems we raised about "content" can be raised in regard to it. Of course, the representation relation need not be an "objective construction from sounds to things" – this much we know from the existence of vacuous representations like "Zeus." But it had better be *some* objective construction or postulation. In particular, if we are to seek even merely an implicit definition of "represent" by its role in Chomsky's theory, we need to know what it is about the way internal representations of the structure of language enter into interpretation, thought and action that determines that they have one informational or representational content and not another.[24]

Although Chomsky provides no technical characterizations of "represent(ation)," the above and other passages in his writings do suggest various approaches to some. Different passages suggest at least *two*, an "externalist" and an "internalist" one, depending on whether the approach appeals to facts external or internal to the brain (which is why, despite his protestations, the corresponding issues of "wide and narrow content" *do* arise for his work). It's not clear how to integrate these passages with each other or with all of the things Chomsky has to say on the matter, but it's worth setting out the issues, if only to invite his further comments.

(ii) Externalism: indication semantics?

A natural suggestion that seems implicit in the above and other passages (e.g., Chomsky 1980a: 105) is that a state represents whatever phenomenon it *indicates*, or would *co-vary* with under suitable circumstances. This would seem to be the idea behind talk of "information," "feature detectors" (ibid.: 226) and "the manner of a rotating cube," and is the point of departure for "informational" theories of representation and intentionality of the sort that have been pursued by a number of recent philosophers.[25] And it affords a plausible basis for positing representations in the linguistic case.[26] In general, sensitivity to *non-local and/or non-physical* phenomena is a *prima facie* sufficient reason for the need of a CRTT: how else can we account for human sensitivity to, e.g., the past presence of dinosaurs, the authenticity of a Rembrandt, or the intentions of speakers than by supposing people are making inferences and confirming hypotheses about these matters? Even non-local gravitation doesn't seem enough.[27] It is hard to see how a system could discriminate, e.g., nouns and verbs without confirming hypotheses that represent them.

There are, however, a number of problems that an indication approach would need to confront. Just for a start:

(1) as Fodor (1990: 93) notes, *indication* is transitive, while *represents* is not (photoreceptors indicating light might thereby also indicate a certain form of heat; but they presumably *represent* only the former);
(2) it is hard to see how to begin to apply indication meaningfully beyond peripheral perceptual systems where the co-variance can be established relatively securely (do "quark" representations really co-vary with quarks, "god" ones with gods, "IP"s with IPs?);
(3) indeed, this approach would seem less available to Chomsky than to a standard vision theorist, given his (1996: 22; 2000: 160) repeated insistence that there actually are no objective phenomena corresponding to representations, most emphatically in the case of "I-language"! (Chomsky 1986: 19–24; 2000: 136) (a problem made particularly acute in the case of phonologically empty categories like *trace*: with what phenomenon does a trace marker co-vary? An absence?);
(4) it's not clear how to regard dispositions to co-vary with phenomena in environments very remote from the one in which the system is ordinarily situated (e.g., one where a state that co-varies with thin cracks here co-varies with thin shadows there (Burge 1986); or one where "NP" might co-vary with sectors of some weird social organization).

One can understand a temptation to retreat inside.

(iii) Internalism: implicit definitions?

At least with regard to *linguo*semantics, Chomsky often advocates an "internalist" semantics "of the Cartesian variety that doubts the existence of external things" (Chomsky 1993: 26; see also 1996: 38). It

> deals with certain postulated levels of mental representation, including representations of syntactic and lexical form and others called "models" or "pictures" or "discourse representations" or "situations" or the like. But the relation of these latter systems to the world of objects . . . is often intricate and remote. (Chomsky 1986: 44)

And some remarks suggest that he would advocate a similar approach for a *psycho*semantics as well:

> There is a rich and intriguing internalist semantics, really part of syntax, on a par with respect to phonology. Both systems provide "instructions" for performing systems, which use them in complex and largely pre-determined ways for articulation, interpretation, expression of thought, and various forms of human interaction. (Chomsky 1996: 53)

Thus, a typical human mind has thoughts not only about cats and dogs, love and hate, but also about VPs, IPs, traces and c-command – and all these contents are determined entirely by facts inside their skulls.

What sort of facts? Facts about the *stuff* we're made of? Or facts about how it's put together? It's tempting to think that it's just some "primitive" fact about our minds, or maybe our "biology," that it just automatically refers to things in the world (see note 33). However, Chomsky's allusions here and elsewhere to a broad notion of "syntax" (see also Chomsky 1993b: 52) suggests that he sometimes would like to take them to be facts about the mind's *computational organization*: the role the representations play in computational relations to one another and to inputs and outputs.

Any such view, however, faces two famous difficulties: (1) the Quinian challenge of specifying just *which* internal relations short of the entire system of them are to serve as the contents of specific representations,[28] and (2) the problem of multiple interpretation: a formalism does not in general fix its own interpretation.[29] Indeed, abstracting from *truth*, most symbols can mean most anything. Two computers could enjoy exactly the same internal physical and computational states and yet be representing different things, one a war, the other a game of chess; the whole internal recursive apparatus postulated by Chomsky for processing "grammar" in a normal human being could be taken to represent simply a set of recursive mathematical structures, or some other isomorphic domain (e.g., some weird society), talk of "nouns," "verbs," and "complementizers" be hanged.

However, while it's true that a formalism doesn't *generally* fix its own

interpretation, there is a (much disputed) possibility that there might be exceptions in the case of *certain* contents: e.g., those expressed by logical operators ("and," "all"), indexicals ("I," "now") and perhaps phenomenal representations ("looks red"), whose content arguably depends only upon their internal causal/conceptual role:[30] thus, a certain configuration arguably means "and" iff it both causes separate entokenings of its constituent states and is caused by the pair of them (mirroring (P&Q) ⊢P,Q and P,Q ⊢(P&Q)). What seems distinctive of such cases is that they seem to play extremely stable, idiosyncratic roles across a wide variety of cognitive systems: the roles of "and," "I," or "looks red" don't vary across people quite in the way that the roles of, say, "gold," "love," or "war" do. These stable roles seem to afford an explanatorily useful basis for content ascription and so blunt both the Quinian and multiple interpretation challenges. Perhaps such a case could be made for the categories of linguistics. Essentially, the linguist would let the grammar implicitly define all the categories peculiar to it and declare the resulting definitions to express the content of the representations manipulated by the LAD. Thus, a child would represent NP iff there was a symbol that played a specific computational role in her brain.

The success of such a proposal depends, however, on the prospects of responding to the Quinian challenge and actually delimiting stable, idiosyncratic roles for grammatical categories to play. Given the actual character of linguistic theory, it is by no means obvious that this would be possible. As it is standardly presented, the grammatical categories seem inextricably entwined with the intentional content of other systems, notably the "LF-C/I" (or "logical form-conceptual/intentional") system(s) whereby the outputs of the grammar receive their semantic interpretation. Conspicuous examples are the constraints on coreference imposed by binding theory, and "theta-marking," whereby certain classes of lexical items are marked for the kinds of roles that must accompany them (e.g., "agent," "patient," "theme," "goal," "location"). More generally, at least in the "minimalist" program, it is the LF-C/I system that provides the crucial "legibility conditions" on the grammar, "in the sense that other systems must be able to 'read' the expressions of the language and use them as 'instruction' for thought and action" (Chomsky 1995b: 2; 2000: 9). These connections would appear to be especially crucial for Chomsky insofar as the development of the LAD depends upon them. In his influential, largely Chomskian theory of the LAD, Pinker, for example, writes:[31]

> I have assumed that the language induction mechanisms exploit internally existing cognitive abilities to represent mentally many conceptual type-distinctions . . . Examples include objects, actions, attributes, manners, agents, causes, instruments, goals . . . person, number, humanness, animacy. (Pinker 1984: 362–3)

(He actually lists forty-one such categories, adding "and many others.") If the content of grammatical representations involves all *these* sorts of contents as well,

the prospects of implicitly defining their idiosyncratic content seem pretty daunting. One problem is that any such implicit definition entails that people who diverge on *any* one concept appearing in the definition *ipso facto* diverge on *all*, since each is defined in terms of all the others. Thus, people who differed about *person* or *animacy* couldn't share a conception of *noun*. The greater such entanglements the less likely different individuals would satisfy any of them. Of course, one would like to distinguish just those entanglements that are relevant to the *content* of a representation. But, as they proliferate into the central cognitive system, we are brought back to the Quinian challenge to do this in a principled way; and, as they come to involve increasingly referential concepts such as *object*, *animacy*, *number*, in a way that could possibly rule out unintended interpretations.

One can understand a temptation to return outside, e.g., to external accounts like those of Putnam (1975) and Burge (1986), or the various "informational" approaches that we've already discussed.[32]

(iv) Intentionality as primitive

"But," someone might protest, "surely it's unfair to ask Chomsky to solve all the problems of philosophy. He does terrific linguistics and you should let him alone. Newton had to take the notion of gravity for granted; why can't Chomsky simply provisionally assume an unanalyzed notion of represention?"[33] Certainly nothing said here should be taken to interfere in the least with Chomsky's *linguistics*. That's not at issue. What's at issue is why Chomsky insists *over and above his specific proposals about grammar* that these grammars are *represented* in the brains of ordinary speakers, without sufficient indication of what he means by this. If he is merely taking for granted *ordinary* intentional notions, that's fine; but then he should indicate as much – and not explicitly disavow those notions. If he is proposing "technical" notions instead, he should provide some indication of what he has in mind, clarifying the various externalist and internalist suggestions different passages suggest. In either case, he should indicate what about his *linguistics* requires what sort of *representation*, where and why, and what about his linguistics requires him to spurn discussions of, e.g., wide and narrow content that philosophers have provided (philosophers might leave him alone to do his terrific linguistics if he left *them* alone to do their intentionality – admittedly less terrific, but so far the best we seem to have to go on).

The comparison with gravitation is, however, instructive. It is not enough merely to claim that Newton appealed to a maybe "occult" force of gravitation: the point is that *Newton was nevertheless able to provide extremely impressive laws relating this occult force to mass, force and motion* (cf. Chomsky 1996: 6). So we need to ask: *what are anything like the general laws of "representation" or intentionality on which linguistic theory could similarly rely?* As we have just seen,

suggested candidates for such theories are notoriously problematic, and at least in their present form won't suffice to settle the controversial cases that Chomsky raises. What is there beyond these proposals, or simply the ordinary, unanalyzed notions of intentionality, that would help?

The lack of clarification of this issue in Chomsky's writings here can't but make one wonder whether his account would be any less explanatory were intentional idioms like "representation" dropped from it altogether. Perhaps all this talk of innately "representing" and "cognizing" principles, and "selecting a grammar that is compatible with the primary linguistic data," is unnecessary, more like Newtonian appeals not to "gravitation," but to "absolute space."[34] Talk of that was also thought to be necessary for understanding motion, but it turned out that physics could get along better without it. So we need to ask: *what would be the effect on Chomsky's theory if all intentional talk were deleted*?

5 Does Chomsky Need Intentionality?

Interestingly, in a number of passages Chomsky himself flirts with such an "eliminativist" possibility. Some passages suggest (i) an instrumentalism about content, others (ii) an eschewal of it based on a particular reading of the work of Marr (1982); still others seem to look to (iii) a brute reduction of linguistics to biology; but, most recently, to (iv) an elimination of any mental or intentional talk altogether. I'll discuss each in turn.

(i) Instrumentalism?

Relegation of intentionality to a merely instrumental status surfaces in a startling passage in which Chomsky seems actually to be *renouncing* CRTT:

> It has been common to try to relieve uneasiness about computational approaches by invoking computer models to show that we have robust, hard-headed instances of the kind: psychology then studies software problems. That is a dubious move. Artifacts pose all kinds of questions that do not arise in the case of natural objects. Whether some object is a key or a table or a computer depends upon designer's intent, standard use, mode of interpretation, and so on. The same considerations arise when we ask whether the device is malfunctioning, following a rule, etc. There is no natural kind of normal case . . . Such questions do not arise in the study of organic molecules, nematodes, the language faculty, or other natural objects, viewed . . . as what they are, not in a highly intricate and shifting space of human interests and concerns. (Chomsky 1993: 43–4; repeated in 1994: 154)

It's hard to reconcile this passage with our earlier quotations (section 2), in which linguistics structures were compared to "a program that we believe to be

somehow represented in a computer" or a "cognizing missile" (Chomsky 1980a: 102–3, 275 n. 61, 188). Isn't the "language faculty" (not to mention the visual system) a "computational" device? Aren't computations realized (and often tested) on computers?[35]

Chomsky is, of course, correct to note that the artifactual status of most existing computers permits much of their semantics to be whatever semantics the artifactor chooses. But this doesn't imply that a machince – either one produced artifactually, or one occurring naturally – mightn't have its *own* natural semantics, one determined by its computational organization and its relation to the external world.[36] People, after all, might be *used* by Martians as computers for determining Martian bank balances without in the least disturbing the fact that certain of their internal states may naturally represent contours – or NPs.

(ii) Marr's three levels

I think what is inclining Chomsky to such an uncharacteristic instrumentalism here is a possibility that has been discussed by Frances Egan in her discussion of Marr's theory of vision.[37] Egan (1992: 488) claims that, although the *exposition* of "computational" theories of vision like Marr's may make use of intentional idioms as "reference fixers," *the actual "computational" theory itself* does not, relying instead on purely physical-mathematical descriptions such as the Laplace-Gaussian convolution of intensity arrays (roughly, a rate of change in a distribution of them; see Marr 1982: 54–7, 337–9). Like Egan, Chomsky thinks this exempts Marr's theory from the externalist worries about intentionality raised by Burge (1986):

> It is, correspondingly, a misreading of informal talk to conclude that Marr's theory of vision attributes "intentional states that represent objective, physical properties" because "there is no other way to treat the visual system as solving the problem that the theory sees it as solving" (Burge [1986: 28–9]). The theory itself has no place for the concepts that enter into the informal presentation, intended for general motivation. (Chomsky 2000: 161)

Now, of course, at *some* level this is precisely what is to be expected from a "computational" theory, given that computations can be specified entirely formally, without reference to interpretations. The question is not whether there exists a theory of that sort, but whether it is one at an essentially *mental* level of description.

As is well known, Marr distinguished three "levels at which an information processing device must be understood before one can be said to have understood it completely" (Marr 1982: 24), the "computational," the "representation and algorithmic," and the "hardware implementational." He characterizes the first ("level-1") as: "the abstract computational theory of the device, in which the

performance of the device is characterized as a mapping of one kind of *information* to another, the abstract properties of this mapping are defined precisely, and its *appropriateness* and *adequacy for the task at hand* are demonstrated" (ibid.: 24; italics mine). He further glosses this level by examples of the kinds of questions it addresses: "What is the *goal* of the computation, why is it appropriate, and what is the *logic of the strategy* by which it can be carried out?" (ibid.: 25: italics mine). Of course, if *these* are indeed the questions to be addressed at the top level, calling it a "computational" level is extremely misleading.[38] These days "computational" often serves as a near synonym of "algorithmic." However, the above characterizations in terms of "information," "appropriateness and adequacy for the task at hand," "goal" and "logic of the strategy" have patently to do with the *interpretation* of the representations specified at the "representation and algorithmic" level, i.e., with the very questions that Chomsky thinks Burge inappropriately addresses. Indeed, level-1 would far more fittingly be called the "intentional" level, and, *pace* the above claim of Chomsky, Marr clearly – and surely quite reasonably – thinks it has a place in his theory.[39]

What does invite further confusion here is a puzzling remark of Marr's that Egan (1992) cites, that "from a computational point of view, [the Laplace-Gaussian] is a precise specification of what the retina does" (Marr 1982: 337). But surely the Laplace-Gaussian is closer to a "*level-2*," "algorithmic" characterization: although it doesn't specify an algorithm, it does provide the mathematical characterization of the output of that algorithm, one step closer than "edge" to a level-2 "choice of representation for the input and output" (ibid.: 24) (perhaps it should be regarded as a "level-1.8" characterization – a specific characterization of the "function in extension"). In any case, it is surely not a description of the "goal" or "appropriate[ness]" or "logic" of the "strategy" of the computation by which Marr characterizes his level-1.[40] For that level, one reverts to precisely the talk of "contours," "surfaces," "motion," "occlusions," and "2-1/2-D sketches" that Marr and most other vision theorists employ *throughout* their discussions (and not merely in the motivational glosses).[41]

Perhaps Chomsky simply wants to reject Marr's level-1 talk, and to restrict science to the level-2 algorithmic descriptions. But then we might ask why the algorithmic level is appropriately regarded as *mental*. After all, one might think that mental states have their contents *essentially*. Egan worries about this possibility, but argues that it is "plausible inasmuch as we have no other way to characterize the internal states posited by commonsense psychology . . . but it is not plausible for mental states construed as computational states" (Egan 1992: 450). However, although a *state* itself could, of course, be characterized in any number of ways, Egan needs to show that *qua mental* state it can be so characterized. What is it to *think* or *believe* or *cognize* a mere formula abstracted from its content? As I and many others have argued, intentionality seems to be the substantive mark of the mental: to settle for less is to render the issue merely terminological. But, now, why on earth should we deprive ourselves of precisely

the rich level-1, mentalistic explanations that Marr and others so admirably provide?

In any case, whether or not Egan's view is an adequate way to understand a theory of *vision*, it seems a far cry from the heavily intentional way in which Chomsky consistently presents his own theory of *grammar*. Here the discussion is presented in terms of *nouns*, *verbs*, and all the conceptual entanglements we mentioned earlier (section 4 (iii)): binding theory, theta roles, the legibility conditions imposed by the LF-C/I interface, and the semantic material (e.g., *agent*, *patient*, *person*, *cause*) that seems to be essential to the operation of the LAD. All of this patently involves categories of a Marrian level-1 *intentional* theory, not the physical/mathematical ones of a level-2 (or "1.8") *algorithmic* one. Chomsky himself compares what he does to "what Gunther Stent has called 'cerebral *hermeneutics*', referring to the abstract investigation of the ways in which the visual system constructs and *interprets* visual *experience*" (Chomsky 1986: 40; italics mine). Indeed, in the opening chapter of the same recent book in which he endorses Egan's view, he also writes:

> The cognitive perspective regards behavior and its products . . . as data that may provide evidence about the inner mechanisms of *mind* and the way these mechanisms operate in executing *actions* and *interpreting experience*. The properties and patterns that were the focus of . . . structural linguistics find their place, but as phenomena to be explained . . . The approach is *"mentalistic," but in what should be an uncontroversial sense*. It is concerned with "*mental aspects* of the world," which stand alongside its mechanical, chemical, optical and other aspects. (Chomsky 2000: 3, italics mine)

So the categories of linguistics would seem overwhelmingly like the visual categories of *edges* and *contours*, *not* like that of a Laplace-Gaussian (which I presume is not an *uncontroversial* bit of anyone's visual "hermeneutics"!). The linguistic categories may well be cognitively and socially constructed "fictions," but that doesn't render them any less the *intentional content* by which we *interpret experience*, and so the subject of standard linguistic theorizing in precisely the way that *edges* and *contours* are for Marr.

If the comparison to *Egan*'s reading of Marr is to be serious, then we need some physical-mathematical categories to *replace* all such intentional talk in the way that the Laplace-Gaussian can replace talk of "edge detectors." There are two ways Chomsky might effect such a replacement: reduction or elimination.[42] I myself find neither of these possibilities particularly plausible, but it is worth setting them out if only to chart regions of the logical space that various of Chomsky's remarks suggest.

(iii) Brute reduction?

Sometimes Chomsky seems to anticipate a straightforward reduction of linguistic to physical categories:

> Linguistics, conceived as the study of I-language . . . becomes part of psychology, ultimately biology. Linguistics will be incorporated within the natural sciences insofar as mechanisms are discovered that have the properties revealed in these more abstract studies . . . It could be that the brain sciences would show that each [principle of grammar] corresponds to some determinate complex of neural mechanisms. (Chomsky 1986: 27, 39)

At least one way of reading such claims is as suggesting brute, 'type–type' reductions of the linguistic to the biological, roughly on the model of Mendelian genetics to modern biochemistry. However, *that* model seems extremely implausible for linguistics, for precisely the reasons having to do with both Marr's correct insistence on level-1 theories of a domain, as well as the familiar plasticity of the brain and multiple-realizability of its structures. Surely there are any number of Marr-style level-2 algorithms that could *implement* Chomsky's (implementable) grammars, which, themselves, could be realized in any number of physically diverse materials. The question is whether Chomsky's theory is at the physical, algorithmic or intentional level. The various conceptual phenomena that we've already mentioned suggest that it's largely the last. Consequently, we have no reason to think that it could be rephrased so as to conform to Egan's reading of Marr. As Peacocke rightly remarks in reaction to the above passage, "one who appeals to determinate neural complexes in elucidating the psychological reality of grammars will need to rely at some point on informational notions which he aimed to replace" (Peacocke 1989: 121).

Indeed, as I mentioned earlier (section 2 (ii)), it is the very fact that linguistic categories are *not* reducible to physical categories that would seem to argue for the need of representations at least in characterizing a system's sensitivity to those categories. It's the human sensitivity to such apparently non-physical phenomena as nouns and verbs that seems to require positing *inferential* or *confirmatory* computations over *representations*. But such an account then needs to supply a reasonable notion of representation over and above computation over physically specified states, i.e., it must deal with some of the difficulties we have discussed.

(iv) Elimination?

At one point, Chomsky raises an interesting empirical issue on which much of this discussion might turn:

> Learning seems pretty much like what Peirce called "abduction," a process by which the mind forms hypotheses according to some rule and selects among them with reference to evidence and, presumably, other factors. It is convenient sometimes to think of language acquisition in these terms, as if a mind equipped with universal grammar generates alternative grammars that are tested against the data of experience with the most highly valued one selected; I have occasionally used this metaphor, but I don't think it should be taken seriously. If we take it partially seriously, then under this concept of learning as "abduction" or "self-design," the question whether language is learned or grows will depend upon whether the mind equipped with universal grammar presents a set of grammars as hypotheses to be selected on the basis of data and an evaluation metric, or whether the steady state grammar arises in another way – for example, by virtue of a hierarchy of accessibility . . . and a process of selection of the most accessible grammar compatible with given data. (Chomsky 1980a: 136)

If I understand *one* tendency of both the P&P and the more recent "minimalist" programs correctly, this latter possibility is being taken increasingly seriously (see Chomsky 1994: 161; 1995b: 5–9). "Rules" are being superseded by the less mentalistic "principles," and, indeed, Chomsky (2000: 184) recently replied to Searle's skepticism about "unconscious rules" by defending "specific structures and principles" with no mention of mentality at all. These structures are constrained by such principles of economy as "least effort," "shortest move," and "last resort" (Chomsky 1995b: 28, 138ff, 181–5), which seem to be principles regarding certain *topological* properties of trees, not their contribution to the probability of *truth*.

This is all further encouraged by Chomsky's (1986: 19–24; 2000: 136) denial of the role of an external "E-language" as an object of linguistic concern. If, unlike the case of vision, there is nothing to be right or wrong about, then there is no need of any *confirmation* relation between the input to the LAD and its settled state. Moreover, if the LAD is "modularized," or "informationally encapsulated," as many have claimed, then there would appear to be no need of any inferential, confirmational or other rational relations to any *other* contentful states.[43] To take a comparison Chomsky (2000: 65) himself likes to make: *immunity* is not a state that is *confirmed* by the antigens that may nevertheless be required to trigger it; why should language growth be?

But if confirmation is irrelevant, then perhaps we really don't need talk of *representations* either. As we noted in section 3 (i) above, it is the rational relations of, e.g., *evidence*, *inference*, and *confirmation*, involving truth-valuable *contents* that largely motivate CRTT. If a child is not confirming hypotheses, then perhaps there is no reason to regard the states as "structural descriptions" or "representations" of anything. Maybe they are *just structures* (e.g., tree structures) that are realized by the brain without being represented in it, in the way that certain complex structures are realized without being represented in, e.g., crystals, molecules, genes, plants, and antibodies.[44] As Egan (1992: n. 17) points out, "the

mathematical formula that describes [the Laplace-Gaussian] function is not assumed by Marr and Hildreth to be explicitly represented in the retina." The visual system *satisfies* the laws described by Marr just as it satisfies laws of gravitation and electromagnetism. Why shouldn't the same be said of the language faculty?

Of course, as we already noted with regard to Egan's proposal, *were* such an elimination of linguistic for physical categories to be effected, the result would cease having a claim on *psychological* reality. Again: a *mental* state *qua mental state* involves a relation to a *content*, not merely to some physical magnitude such as a Laplace-Gaussian. The metabolic or immune systems are not part of a *mind*. Just so: on this view the language faculty would not be part of the mind either. Of course it could interact with one. The "intuitions" on which linguists rely could result from running a non-mental grammatical sub-system "off line" and checking out its outputs, much as one might find out which foods are indigestible by trying to digest them.[45]

I myself am skeptical about the prospects of such a non-intentional rendition of linguistic theory. As I said, it is *one* tendency of minimalist approaches, but there is also the emphasis already noted on the "legibility conditions" on the grammar provided by the LF-C/I system (Chomsky 2000: 9–11), as well as the myriad ways we noted that psycholinguists like Pinker (1984) claim that the operation of the LAD depends upon a child's grasp of conceptual material. Moreover, even if there are *in fact* no such things as "E-languages," still, children might *think* (or "cognize") that there are, and that thought might well influence the course of their linguistic development. Indeed, what invites an *inferential*, as opposed to a merely causal, model of the LAD, is that many of the states attained by the child are characterized by quite *general* principles that are *true of what speakers classify* as the stimuli that caused them: specific "head first" stimuli give rise to a *generalized* "head first" parameter setting; kids seem to acquire French not just by any causal route, but by hearing *instances* of *what people classify as that language* (cf. Fodor 1975: 37).[46] Consequently, there would appear to be *content-sensitive* relations between the input and output of the LAD. If such facts are essential to an "explanatorily adequate" linguistic theory (Chomsky 1965: 25; 1996: 3), then CRTT would seem to be precisely the kind of theory Chomsky would need – just as he originally thought! – and the prospects of a non-intentional account of the LAD would seem quite dim.

Well, but maybe even confirmation and the conceptual material of the LF-C/I system is not to be understood intentionally or mentalistically either! This would appear to be the drift of some *total* eliminativist suggestions Chomsky has lately been entertaining:

> More generally, intentional phenomena relate to people and what they do as viewed from the standpoint of human interests and unreflective thought, and thus will not (so viewed) fall within naturalistic theory, which seeks to set such factors aside.

> Like falling bodies, or the heavens, or liquids, a "particular intentional phenomenon" may be associated with some amorphous region in a highly intricate and shifting space of human interests and concerns. But these are not appropriate concepts for naturalistic inquiry . . . If "cognitive science" is taken to be concerned with intentional attribution, it may turn out to be an interesting pursuit (as literature is), but is not likely to provide explanatory theory or to be integrated into the natural sciences. (Chomsky 2000: 22–3)

> We may ask whether the theory of people, meteors, and flowers should involve [intentional] notions. The current answer is "definitely not" in the case of flowers and meteors, and unknown in the case of people . . . (ibid.: 146–7)

> We can be reasonably confident that "mentalistic talk" will find no place in attempts to describe and explain the world . . . The notion "common store of thoughts" has no empirical status, and is unlikely to gain one even if the science of the future discovers a reason, unknown today, to postulate entities that resemble "what we think (believe, hope, expect, want, etc.)." [Frege's "principle" that there is a common store of thoughts] seems groundless at best, senseless at worst. (Chomsky 1996: 45–7; see also 1993: 18 and 1994: 165–6)

It's a little hard to know what to make of these remarks. Suffice it to say that, if they are to be taken seriously, then Chomsky's theory begins to look rather different than it originally appeared: if there is no common store of thoughts then, *a fortiori*, there is no common store of the *principles* of (universal) grammar *cognized* by normal (neonate) human beings; if mentalistic or intentional notions will find no place in serious attempts to describe or explain the world, then, *a fortiori*, they will not find a place in the description of the human language faculty.[47] But, fine: Chomsky's numerous critics (section 1) will be relieved to know that the language faculty can be happily characterized without any of his controversial mentalistic claims.

6 Chomsky's Dilemma

The upshot of my discussion can be put very simply: either Chomsky's account of human grammatical competence requires intentionality and a CRTT or it does not. If it does, then he owes us some indication of how he understands these phenomena, and particularly the many intentional idioms ("representations," "structural descriptions," "cognizing," "information") he employs, where he thinks they are required and why. If the plan is to implicitly define such talk, he should indicate the reasons for thinking it is plausible to do so in the light of familiar difficulties. If, however, CRTT and intentionality are *not* required, then he owes us some rendition of his theory without them – and a good reason why we should regard grammatical competence, so conceived, as genuinely part of the

mind, and not merely as some non-mental system that a mind might happen to exploit.

Alternatively, he could remain agnostic. He could regard linguistic theory as simply an abstract systematization of a domain to which the brain is obviously sensitive *in some manner or other*, much as someone might characterize, say, arithmetic, leaving the character of that manner open, as entirely the "mystery" that he sometimes claims it to be (Chomsky 1986: 39).[48]

Perhaps it involves representations, intentionality and a CRTT; perhaps entities and theories far less problematic. If the relation of mind and brain is such a mystery, then surely: *wovon man nicht sprechen kann, darüber muss man schweigen*.[49] Perhaps this is the view he should take in the long run – but not, of course, before clarifying his position and the puzzling passages I have cited.

Chomsky's reply: pp. 274–86.

Notes

1 For the most part, we won't need to distinguish between rules and principles, although since Chomsky 1981, the "Principles and Parameters" ("P&P") theory has increasingly stressed principles as primary (see Chomsky 1995b: 5–6, but also section 5 (iv) below for a potential difference to the discussion here).

2 Of course, many of the *words of a natural language* (e.g., "cat") may represent things (cats); and, of course, many of the *terms of a grammar* (e.g., "noun," "rule") may represent grammatical categories (nouns, rules). The question here is whether linguistic theory is committed to there being *states of a normal speaker* that represent the categories and rules represented by the grammar.

3 I will use "representation" as the generic for items that display one or other sort of intentionality: e.g., *pictures* that "depict things," *words* that are (or stand) "for things," *thoughts* that "concern things." These many intentional idioms exhibit a vexing ambiguity with regard to how we are to understand the "things represented," i.e., whatever goes in for "X" in, e.g., "representation of X," "word for X," "thought about X." Sometimes, X is understood to be some actual thing, as in "representation of Mozart's nephew" – it verges on contradiction to say "this represents Mozart's nephew, although he didn't have one." But in other uses, it is not understood in this way, but rather seems merely to express the "idea," "concept" or "content" of X, as in "image of a rotating cube" – there is no problem with ". . . image of a rotating cube, though there is none" (cf. Chomsky 2000: 160). Throughout this paper I will presume this latter, what might be called "purely intentional" usage of "represent" and other intentional idioms, and in talking about the "contents" of representations I will mean nothing more than *however we are to understand the substitutends for "X" in this peculiar usage*, remaining neutral between different theories of what they might be (thus X might be an abstract "intension," or "X" might be merely a way of classifying representations).

4 I'll consider passages in which Chomsky seems explicitly to reject CRTT in section 5.

5 It is really quite astonishing how effortlessly talk of representation slips into contemporary discussions. Even in a rare moment in which the issue is explicitly addressed, Higginbotham (1987: 125) writes, "I assume the grammar is represented in the mind; to endorse that proposition is just to identify language as a faculty."

6 See also the comparison of language acquisition to Peircean abduction in Chomsky 1968/1972. I'll discuss further reasons for a confirmation model – as well as Chomsky's more recent retraction of it – in section 5 (iv) below.

7 See also Chomsky 1975c: 37–8; 1980a: 108–9, 112, 189, 192, 196–7; 1986: 249–57; and 1993: 52. Sometimes it seems as if Chomsky thinks the only alternative to a claim of psychological reality is some form of *behaviorism* (see, e.g., Chomsky 1975c: 189; 1980a: 48, 102–3; 1986: 9–10, 43; 2000: 51, 59–60), a dilemma perhaps encouraged by the general instrumentalist and behaviorist inclinations of some of his opponents, e.g., Quine 1953b, 1972 and Dummett 1981. However, someone might think that grammars describe an internal competence in the brain that far exceeds behavior, but *one that still involves rules merely being satisfied, not represented*, as in the case of the digestive or metabolic systems (Stabler 1983 and Devitt and Sterelny 1987/forthcoming: section 8.9 explore possibilities of this sort).

8 Interestingly, one place Chomsky does raise the question of psychological reality in the way being raised here is in relation to possible world semantics:

> If one believes that possible world semantics is the way in which some domain of semantics works, then he has one of two choices. He can say: this relates to cognitive psychology in a broad sense, that is, it is about the ways in which all of this is represented in the mind. Or he could say: it has nothing to do with the way things are represented in the mind, it is just a kind of mathematics. If he takes the first approach, there are serious problems, because it is not at all clear how people can have access to calculations using possible worlds when they make their judgments. (Chomsky 1988: 93)

Of course, one might wonder what's wrong with *representations* of possible worlds: indeed, it's insofar as a system's behavior is sensitive to possibilities that representations arguably have a role to play (cf. Fodor 1975: ch. 2)! But the interesting issue here is that Chomsky finds the issue interesting at all.

9 A line that has in various ways been independently recommended by, e.g., Churchland 1979, Loar 1981, and Dennett 1987 for mentality generally. For reasons I sketch in the next section (and in Rey 1997: sections 5.2, 10.3), I myself reject this line, but think it deserves to be taken seriously – if only in view of the very skepticism about the "objects" of attitudes that Chomsky (1993: 18; 1994: 165–6; 1996: 47) himself endorses.

10 See, e.g., Kahnemann, Slovic, and Tversky 1982 and Ainslie 1992. This is why, *pace* Chomsky 2000: 166, thinking of people decision-theoretically is an interesting empirical research program.

11 Excellent surveys of recent CRTT-driven research can be found in the several volumes of Osherson 1995, e.g., of vision research in the contributions of Nakayama, He, and Shimojo, and of Biederman in vol. 2; and of decision making in that of Shafir and Tversky in vol. 3. Not to beg questions presently being raised about the

characterization of grammatical *competence*, one might look to the recent computational work on *parsing* that "balances" statistical and rule-based approaches, as in Klavans and Resnik 1996.

12 See Stich and Warfield 1994; Rey 1997: ch. 9; and Devitt and Sterelny 1987/forthcoming for recent surveys of the enormous literature on these and other problems, and some proposed strategies for dealing with them.

13 Which is not to suggest that all the work in these other sciences is nearly said and done. Levine (1983) has nicely put the point in terms of an "explanatory gap" between the physical and (the case that interests him) the phenomenal, a point that can be extended to the intentional and much of the mental generally (see Rey 1997: ch. 2): the problem is of imagining not only what *actually*, but what *possibly* could "upwardly necessitate" the mental in the way that it's perfectly imaginable that, e.g., patterns of cell replication could do for growth.

14 The *conceptual* nature of the puzzles is perhaps clearest in the cases of the "semantic paradoxes" such as those surrounding "This representation is not true": surely *here* the issue is not one of mere "decision" about how to use the semantic vocabulary! I submit that the other mentioned problems are more like this than like deciding whether planes "fly." Comparable puzzles in the history of science are those surrounding the notion of "limit" in calculus and "absolute space" in physics.

15 Chomsky (1995a: 42; 2000: 148) sometimes makes it seem as though we're not entitled to raise "intuitive" problems about a purely "technical" Fregean notion of "reference" (*Bedeutung*). Quite apart from the generality of the problems that I'm emphasizing here, it is surely appropriate to ask whether a technical notion will solve certain problems particularly since the Fregean notion may well turn out to be the one appropriate for science (Chomsky 1996: 46), perhaps even linguistics! Indeed, Chomsky (1988: 91) himself raises some of these very problems with regard to possible world semantics.

16 Some (e.g. Searle 1992: 155–6 and Strawson 1994: 166–7) claim that consciousness *is* conceptually essential to the understanding of any process as intentional, but this just seems dogmatic and flies in the face of proposals that go back not only to Freud and the Romantics, but to the Rationalists, the Scholastics and the Greeks (see Whyte 1978). Chomsky himself sometimes writes as if ordinary intentional explanation, at least understood causally, were hostage to the "mysteries" surrounding free will (Chomsky 1994: 162–3; 1996: 3, 10), but, of course, there is plenty of interesting intentional causation and explanation (what makes people mad, sad, laugh, cry) that can steer clear of them. (Moreover, mental states and competencies could be causes that "incline" without "compelling" (Chomsky 1980a: 7; 1986: 222–3) in the way dispositions generally do: my computer is "inclined" to *print* when I press F7, but of course it is only "compelled" to do so when conditions are *just right*. Voluntary action may well be merely a complex *interaction* effect, like the *actual* activity of a machine, or the precise path of a leaf in a lake, about which, indeed, general, systematic theorizing or prediction may be quite idle; cf. Rey 1997: section 8.6, but also Strawson 1987.)

17 Chomsky (1993: 38–9; 1994: 157; 1996: 6; 2000: 166–8) sometimes claims that mind–body problems can't be coherently posed in post-Newtonian science, what with its postulation of "action at a distance" and its abandonment of "Cartesian mechanism." But the problems of intentionality can be posed by simply asking whether any *non-*

intentional account could be provided for intentional phenomena – it's just that the physical as it's presently understood is the only serious *non*-intentional vocabulary we know! If not, then obviously we have no explanation of the intentional (cf. note 33). Interestingly, Chomsky (1993: 37–8) does think there's a sufficiently identifiable "mind–body problem" to claim that what many of us regard as the one advance we've made on it, the notion of Turing computability, doesn't solve it! I'm not clear precisely what problem he has in mind here, but note that Turing computation *could* be realized in a local mechanical way, whether or not there is also "occult" action at a distance.

18 Thus, e.g., thoughts are mental states, thinking a mental process, and utterances mental products.

19 Someone might offer a "paradigm case" argument that mentation is just *whatever* human beings do, e.g., when they use language, and that this could turn out not to involve intentionality. As with paradigm case arguments generally, besides rendering the issue boringly verbal, it would miss the substantial and interesting issues that are raised, in this case, by Chomsky's persistent employment of intentional idioms.

20 Chomsky adds the following on behalf of "cognize":

> If the person who cognized the grammar and its rules could miraculously become conscious of them, we would not hesitate to say that he knows the grammar and its rules, and that this conscious knowledge is what constitutes his knowledge of the language. Thus "cognizing" is tacit or implicit knowledge, a concept that seems to me unobjectionable . . . (Chomsky 1980a: 70)

Of course, should someone become miraculously conscious of the principles of digestion, we might well say she knew them – *then*. But of course that doesn't imply she knew – or cognized – them all *beforehand*. The skeptic here wants a reason for thinking that they are *represented* even when they are not (accessibly) conscious; otherwise they seem, as Strawson (1994: 166) compares them, like "sounds" on an unplayed CD. Searle (1980: 38) rightly notes that *causal efficacy* would be one reply (even if he himself ignores this reply in his later work), a reply that, as I've said, is usually a major part of CRTT. But it's not a reply that I've found Chomsky willing to defend.

21 Apparently not:

> [T]here would be little point to a concept of "cognizing" that did not distinguish "cognizing the rules of grammar" from the bicycle rider's knowing that he should push the pedals or lean into a curve . . . [W]e take bicycle riding to be a skill, whereas knowledge of language and its concomitant . . . is not a skill at all. The skill in question might, perhaps, be based on certain reflex systems, in which case it would be incorrect to attribute a cognitive structure in the sense of this discussion. (1980a: 102)

But the distinctions here are orthogonal: "skills" (e.g., to do mental math) and "reflexes" (e.g., parsing, as in Fodor 1983: 72) might well involve representations, and plenty of what are presumably *non*-"reflexive" *non*-skills (e.g., the immune system) might *not* involve representation. It doesn't help that Chomsky (2000: 52)

now *does* regard bicycle riding as "irreducibly cognitive"! It is difficult to see how he is relying on anything more than the pretty erratic "ordinary use" of intentional idioms (as at Chomsky 1980a: 92ff; 1986: 27, 266–9; 2000: 165–6). (See Devitt and Sterelny 1987/forthcoming: section 8.4 for discussion of the diversity of relevant uses of "know.")

22 Implicit definition consists in defining a set of terms entirely in terms of the role they play in a theory: e.g., a "quark" just *is* whatever satisfies some specific "quark" theory proposed by a physicist.

23 I presume Chomsky doesn't have in mind the genuinely technical Shannon–Weaver notion of information, which would be too pegged to performance for his purposes. There may, of course, be still other notions of "information" that might in interesting ways fall short of full intentionality, which, for that matter, may well itself come in varieties and degrees. (One promising possibility relevant here is of "non-conceptual content" of the sort explored in Evans 1981, Davies 1989, and Crane 1992.) All of this bears much discussion.

24 Chomsky also claims that

> Misperception raises no difficulties for this approach; it is a matter of how people assign interpretations to interactions they observe – to the reactions of a frog or a person in an experiment, a photo receptor that is "deceived," etc. – a fair topic for internalist inquiry into the psychology of the person who is deciding what to call a "misperception." (2000: 160)

However, *any* serious theory of representation entails a "misrepresentation" or "misperception" problem. What "information" a visual or grammatical system is representing determines what will count as a misrepresentation, and *vice versa*: a state's information is *fly* iff its being occasioned by flies is *not an error*; a state's information is NP iff its being occasioned by, e.g., "[PH REP The skaters waltz PH REP]" is *not an error*. The questions of representation and misrepresentation are inseparable, and have to be confronted by *the theorist who uses such intentional idioms*, and not merely by the "ethnoscientist" who studies that theorist.

25 E.g., Stampe 1977; Dretske 1981, 1988; Stalnaker 1984; and Fodor 1987, 1990. Note that this "indication" relation is a *dispositional* one and so can obtain without the indicated phenomenon actually existing with every tokening of the state, as when a photoreceptor fires randomly, or when mere tachistoscopic stimulations produce the appearance of a rotating cube (Chomsky 2000: 160). On such a view, these are not the suitable circumstances for defining the right indication relation. Such a dispositional account, of course, renders the distinction between "internalist" and "externalist" semantics less than vivid: is a discrimination ability that is never exercised an internal or external property of system? (Is solubility in the absence of water an internal or external property of salt?) There are, of course, other non-dispositional externalist theories – for example, causal/historical ones (e.g., Devitt 1981) and teleological ones (e.g., Millikan 1984) – but there are no passages in Chomsky that support them.

26 Note that phonemic and other motor representations might secure their content co-variationally, not from their co-variant *causes* (as plausibly in the case of perception), but from their standard co-variant *effects* (e.g., causing certain utterances). This, of course, deserves a lot more discussion.

27 I develop this argument in Rey 1997: section 4.3. Chomsky (1965: 58) himself suggests such an argument, as does Fodor (1986), who claims representation is needed for "non-projectible" properties (i.e., properties that don't appear in laws).

28 Quine (1953a, 1956) has questioned the viability of the distinction between claims constitutive of meaning and claims merely expressive of empirical belief, and has claimed that "the unit of meaning is the whole of science." This latter "semantic holism" has been defended by, e.g., Block 1986 and is critically discussed in Fodor and LePore 1992.

29 This is sometimes expressed in terms of the Löwenheim–Skolem theorem, according to which any consistent first-order theory about *anything* has an interpretation on the natural numbers (see Putnam 1980). However, the point is more trivial than that: the most one can hope of *any* formalism is a specification of *isomorphic* models – and this only if all of its theorems are taken to be true (hardly an assumption to be made of people). This is the (only) truth at the heart of Searle's "Chinese Room" argument, which, given his interest in internalist semantics, it's surprising to find Chomsky (1988: 12) also endorsing.

30 See, e.g., Harman 1982 and Peacocke 1992 for discussion of logical particles, Kaplan 1979 for indexicals, and Rey 1998 for phenomenal states.

31 See Pinker 1984: ch. 8 and Bloom 1994 for specific examples of how acquisition of lexical distinctions may be conceptual.

32 At an extreme there is Fodor and LePore's (1992) semantic "atomism." However, arguably some combination of internal and external approaches is required, as in "two factor" theories (see Field 1977; Loar 1981; Block 1986). But such proposals are still largely vague and programmatic. I discuss many of these issues in greater detail in Rey 1997 ch: 9.

33 Chomsky at one point actually seems prepared to accept intentionality as not only provisionally, but perhaps *ultimately*, primitive:

> It may seem offensive to common sense and sound thinking to suppose that certain matters (intentionality and aboutness and consciousness, behavior that is uncaused but appropriate, or whatever) are among "the ultimate and irreducible properties of things" that physicists seek to catalogue (Jerry Fodor's formulation). But the stipulation is not very helpful. Why these but not attraction and repulsion? (Chomsky 1996: 44)

Well, barring parapsychology, one difference is *the closure of physics*: every *individual event* (including every particular animal motion and linguistic utterance) is explicable in it, if explicable at all (see Weinberg 1992; Loewer 1994; Rey 1997: section 3.1.1). Other properties had better therefore "supervene" on the fundamental physical ones. However, this may be something of a red herring, since many of the worst problems about intentionality that are relevant to our discussion can be raised about it whether or not it is primitive (e.g., What are its laws? Where does it pop up and why?).

34 A more contrived comparison might be closer to the point: suppose that, to please the Church, Newton had prefixed all his postulations with "angel beloved": e.g., "the angel beloved force of angel beloved gravity . . ." Presumably merely *this* concession to theology, without any indication of just how these things come to be beloved of angels, would be empty. That is the worry here about the prefix "representation of."

35 Part of the problem here is that "computational" is used with exasperating freedom in these discussions, sometimes to refer to actual processes in spacetime, sometimes to a way of abstractly characterizing those processes, and sometimes to merely the input, output and "goals" of those processes, as in Marr 1982: 24–5, to be discussed shortly. One source of Chomsky's antagonism to "computer" models *per se* may be their early association with an excessively behavioral conception of their role, e.g., in getting something to pass the "Turing Test," or to facilitate machine translation. One can sympathize with Chomsky's (1975b: 39–40) antipathy to these projects without prejudice to the word "computer," or to the possibility of CRTT characterizing the structure and competencies of various subsystems of the mind.

36 Chomsky seems to concede as much in his own example of how we might investigate "a machine that fell from the sky" (1986: 239) (although it's not clear that he intends this investigation to include any semantics).

37 Chomsky (2000: 203 n. 9) refers to an unpublished ms. of Egan's, the relevant ideas of which are however set out in Egan 1992.

38 Egan (1992: 445 n. 3) also finds Marr's usage misleading, though for different reasons.

39 Chomsky (2000: 203 n. 9) complains that discussions of "'what Marr meant' are somewhat strange," since "what matters is what a scientist does, not what he may have had in mind." But Chomsky himself seems of two minds about how to characterize what Marr *did*. On the one hand, he characterizes it as "mostly concerned with mappings of retinal images to the visual cortex . . . no notion like 'content', or 'representation *of*' figures within the theory" (1980a: 21) but, on the other, as we have already noted (sections 1, 3), he describes it in terms of "representations *of* a three-dimensional object," "*feature* detection," "sensitiv[ity] *to* lines with certain orientations," and representations *of* a "rotating cube" (Chomsky 1980a: 105–6, 188, 226–7; 2000: 160; all italics mine; see also 1965: 8; 1980b: 52). Indeed, he is evidently quite prepared to think of it intentionally: "The shift from '*linguo*semantics' to '*psycho*-semantics' . . . is significant only so far as psychological explanation reaches. Quite far in some domains (e.g. visual perception)" (Chomsky 2000: 165).

40 I suspect Marr was himself misled by his use of "computational" for both the intentional and the function-in-extension computed at the algorithmic, what I call "level 1.8." (Soames (1984: 171) independently proposes three levels very like Marr's, but confines "computational" to the second, algorithmic one.) Note that, in the specific passage cited, Marr is at pains merely to distinguish himself not from an intentionalist, but from "a retinal physiologist" (Marr 1982: 337). Note, too, that, even if it's a Laplace-Gaussian that's being computed, still that seems to presuppose that the states are *representing the values of that function*! So are these then the real contents of visual perception? See also Marr's discussion of the complexities of the distinction between "level-1" and "level-2" theories in Marr 1981, as well as Peacocke's (1986, 1989) further insistence on the existence of a "level-1.5" that specifies the information to which the level-2 algorithm is sensitive (and which, of course, is also intentional).

41 Again, see the contributions of, e.g., Nakayama, He, and Shimojo; of Biederman; and of Spelke, Gutheil, and van de Walle in Osherson 1995, vol. 2.

42 I want to distinguish "brute" (what is sometimes called "type–type") reduction of, e.g., linguistic to physical categories from the kind of potentially very complex "implicit definitions" of them mentioned earlier, since these latter, unlike the former,

arguably preserve an important explanatory level (see Rey 1997: section 6.4 for discussion).

43 See, e.g., Stich 1978; Fodor 1983; Davies 1989. Note that, although Fodor (1983) did indeed neglect discussion of the relation of "input" to "output" systems, this doesn't entail, as Chomsky (1986: 14 n. 10) claims, that the LAD is part of a "central" cognitive system, as Fodor understands it. It's still true that (to take Stich's kind of example) beliefs acquired in a linguistics seminar might be systematically unable to combine with the states of one's language faculty.

44 Uriagereka (1998: ch. 1) compares minimalist structures to the Fibonacci series exhibited in, e.g., the growth of shells and plants. Substantially the same point as this last paragraph was set out very well by Matthews 1980, to which Chomsky (1980b: 47) replied with the epigraph of this chapter.

45 It may be possibilities such as this that led some psychologists to want harder "processing" evidence of represented grammars than intuitions alone provide, since, as merely accounts of intuitions, they seem every bit as plausible as the "deductive" account that Chomsky (1986: 270) occasionally sketches (whereby intuitions are "deduced" from grammars as "theorems" in a formal system, cf. also Chomsky 2000: 69). Thinking of grammars as non-intentional structures in the brain sharpens the issue of their causal efficacy in a way that too loose talk of "representation" seems to me to obscure.

46 Indeed, Crain and Thornton, discussing how children resolve ambiguities, observe that "The sentence-processing system attempts to access a linguistic analysis that makes the sentence true in the discourse context" (Crain and Thornton 1998: 52–3). So the LAD itself had better have access to the notion of truth (and so of other intentional notions such as representation and aboutness).

47 Of course, there is the question of how we are to deal with the rest of human behavior – particularly the confirmation and refutation of scientific theories! Chomsky (2000: 138) seems to regard even the importance of the eliminativist question as "unclear" (cf. notes 17 and 33 above). For the nonce perhaps we should read his eliminativism as confined merely to the theory of the grammatical faculty (give or take some part of the LF-C/I interface).

48 This was, of course, almost the position of his early (1965) work: "A generative grammar is not a model for a speaker or a hearer. It attempts to characterize in the most neutral possible terms the knowledge of the language that provides the basis for actual use of language" (Chomsky 1965: 9).

(This was reiterated in 1980a: 222.) Drop "the knowledge of" and any other intentional idioms and you have a perfectly reasonable agnostic view, without all the headache we have examined here.

Note that such an agnostic view would not entail the view that linguistic theory is only instrumental, or merely a systematization of a corpus of *sentences*, as in Quine 1953b; Chomsky 1957; Soames 1984; and Devitt and Sterelny 1987/forthcoming. One could still think one was characterizing some real "faculty" in the brain – one was just doing so in considerable abstraction from many of its details, e.g., from whether it was intentional or characterizable by a CRTT, or even from whether it was part of *a mind* at all (immunology, after all, is the study of a non-mental, bodily faculty, and not just a systematization of the "corpus" of antibodies; and a theory of aerodynamics might place constraints on wings, whether of birds or planes (cf. Marr

1982: 27)). Just how much of one's description would turn out to describe causally efficacious components of that faculty would depend upon precisely which facts (parsing, production, acquisition, reflective intuitions) the characterization was used to explain (cf. note 45, as well as the discussion of Harman's point at the end of section 2 above).

49 In somewhat more stilted English: "Whereof one cannot speak, thereof must one be silent" – the closing aphorism of Wittgenstein's *Tractatus*.

Acknowledgments

I'm grateful to Louise Antony for the invitation to write this material up, and for discussions with her, David Auerbach, Janet Fodor, and Brian Loar. Some of it emerged from fertile joint seminars with philosophers and linguists at the University of Maryland at College Park, and the joint Georgetown/Maryland semantics discussion group. I'm also indebted to Kent Bach, Ramon Das, Michael Devitt, Jerry Fodor, Diane Gottlieb, Janet Levin, Joe Levine, Paul Pietroski, Jesse Prinz, Juan Uriagereka and especially Michael Morreau for invaluable comments and advice.

References

Ainslie, G. 1992: *Pico-Economics: The Strategic Interaction of Successive Motivational States Within the Person.* Cambridge: Cambridge University Press.

Bach, K. 1987/1994: *Thought and Reference.* Oxford and New York: Oxford University Press.

Berwick, R. and Weinberg, A. 1984: *The Grammatical Basis of Linguistic Performance.* Cambridge, Mass.: MIT Press.

Block, N. 1986: Advertisement for a Semantics for Psychology. In P. French, T. Euhling, and H. Wettstein (eds.), *Studies in the Philosophy of Mind. Midwest Studies in Philosophy*, 10, Minneapolis: University of Minnesota Press.

Bloom, P. 1994: Syntax-Semantics Mappings as an Explanation for Some Transitions in Language Development. In Y. Levy, *Other Children, Other Languages*, Hillsdale, NJ: Lawrence Erlbaum Associates.

Brentano, F. 1874/1973: *Psychology from an Empirical Standpoint*. Tr. A. Rancurello, D. Terrell, and L. McAlister. London: Routledge and Kegan Paul.

Burge, T. 1986: Individualism and Psychology. *Philosophical Review*, 95 (1), 3–46.

Chomsky, N. 1957: *Syntactic Structures*: The Hague: Mouton.

Chomsky, N. 1959: Review of *Verbal Behavior* by B. F. Skinner. *Language*, 35 (1), 26–58.

Chomsky, N. 1965: *Aspects of the Theory of Syntax*. Cambridge, Mass.: MIT Press.

Chomsky, N. 1969: Comments on Harman's Reply. In S. Hook (ed.), *Language and Philosophy*, New York: New York University Press.

Chomsky, N. 1968/1972: *Language and Mind*. New York: Harcourt, Brace and World.

Chomsky, N. 1975a: *Logical Structure of Linguistic Theory*. New York: Plenum Press (material written *c*.1955).

Chomsky, N. 1975b: Introduction to *Logical Structure of Linguistic Theory*. New York: Plenum Press (written 1973).
Chomsky, N. 1975c: *Reflections on Language*. New York: Pantheon.
Chomsky, N. 1977: *Essays on Form and Interpretation.* Amsterdam: Elsevier North-Holland.
Chomsky, N. 1980a: *Rules and Representations.* New York: Columbia University Press.
Chomsky, N. 1980b: Précis of *Rules and Representations* with Replies to Commentators. *Behavioral and Brain Sciences*, 3, 1–61.
Chomsky, N. 1981: *Lectures on Government and Binding*. Dordrecht: Foris.
Chomsky, N. 1986: *Knowledge of Language*. New York: Praeger.
Chomsky, N. 1988: *The Generative Enterprise*, a discussion with Riny Huybregts and Henk van Riemsdijk. Dordrecht: Foris.
Chomsky, N. 1993: *Language and Thought*. London: Moyer Bell.
Chomsky, N. 1994: Entry on (and by) Chomsky. *Blackwell's Companion to Philosophy*. Oxford: Blackwell.
Chomsky, N. 1995a: Language and Nature. *Mind*, 104 (413), 1–61.
Chomsky, N. 1995b: *The Minimalist Program.* Cambridge, Mass.: MIT Press.
Chomsky, N. 1996: *Powers and Prospects.* Boston: South End Press.
Chomsky, N. 2000: *New Horizons in the Study of Language and Mind.* Cambridge, Mass.: MIT Press.
Churchland, P. 1979: *Scientific Realism and the Plasticity of Mind.* Cambridge: Cambridge University Press.
Crain, S. and Thornton, R. 1998: *Investigations in Universal Grammar: A Guide to Experiments on the Acquisition of Syntax and Semantics.* Cambridge, Mass.: MIT Press.
Crane, T. (ed.) 1992: *The Contents of Experience.* Cambridge: Cambridge University Press.
Davies, M. 1989: Tacit Knowledge and Sub-Doxastic States. In George 1989.
Dennett, D. 1987: *The Intentional Stance.* Cambridge, Mass.: MIT Press.
Descartes, R. (1637/1970) Discourse on the Method. In *The Philosophical Works of Descartes*, vol. 1, tr. E. S. Haldane and G. R. T. Ross, Cambridge: Cambridge University Press.
Devitt, M. 1981: *Designation*. New York: Columbia University Press.
Devitt, M. forthcoming: Ignorance of Language. In A. Barber (ed.), *Epistemology of Language*, Oxford: Oxford University Press.
Devitt, M. and Sterelny, K. 1987/forthcoming: *Language and Reality.* Cambridge, Mass.: MIT Press.
Devitt, M. and Sterelny, K. 1989: What's Wrong with the Right View. In *Philosophical Perspectives*, 3: *Philosophy of Mind and Action*. Atascadero, Calif.: Ridgeview Press.
Dretske, F. 1981: *Knowledge and the Flow of Information.* Cambridge, Mass.: MIT Press.
Dretske, F. 1988: *Explaining Behavior: Reasons in a World of Causes.* Cambridge, Mass.: MIT Press.
Dummett, M. 1981: Objections to Chomsky. *London Review of Books*, 3–16 September.
Egan, F. 1992: Individualism, Computation and Perceptual Content. *Mind*, 101 (403), 445–59.
Evans, G. 1981: Semantic Theory and Tacit Knowledge. In S. Holtzman and C. Leich (eds.), *Wittgenstein: To Follow a Rule*, London: Routledge and Kegan Paul.
Field, H. 1977: Logic, Meaning and Conceptual Role. *Journal of Philosophy*, 69, 379–408.
Fodor, J. 1975: *The Language of Thought*. New York: Crowell.
Fodor, J. 1983: *Modularity of Mind.* Cambridge, Mass.: MIT Press.

Fodor, J. 1986: Why Paramecia Don't Have Mental States. *Midwest Studies in Philosophy* 10, Minneapolis: University of Minnesota Press.
Fodor, J. 1987: *Psychosemantics*. Cambridge, Mass.: MIT Press.
Fodor, J. 1990: *A Theory of Content and Other Essays*. Cambridge, Mass.: MIT Press.
Fodor, J. and LePore, E. 1992: *Holism: A Shopper's Guide*. Oxford: Blackwell.
Gallistel, C. 1990: *The Organization of Learning*. Cambridge, Mass.: MIT Press.
Garrett, M. 1975: The Analysis of Sentence Production. In G. H. Bower (ed.), *The Psychology of Learning and Motivation*, San Diego: Academic Press.
George, A. 1989: *Reflections on Chomsky*. Oxford: Blackwell.
Harman, G. 1974: Review of *Language and Mind*. In G. Harman (ed.), *On Noam Chomsky: Critical Essays*, New York: Doubleday.
Harman, G. 1980: Two Quibbles about Analyticity and Psychological Reality. *Behavioral and Brain Sciences*, 3, 21–2.
Harman, G. 1982: Conceptual Role Semantics. *Notre Dame Journal of Formal Logic*, 23, 242–56.
Higginbotham, J. 1987: The Autonomy of Syntax and Semantics. In J. Garfield (ed.), *Modularity in Knowledge Representation and Natural Language Understanding*, Cambridge, Mass.: MIT Press.
Kahnemann, D., Slovic, P., and Tversky, A. 1982: *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.
Kaplan, D. 1979: The Logic of Demonstratives. In P. French, T. Euhling, and H. Wettstein (eds.), *Contemporary Perspectives in the Philosophy of Language*, Minneapolis: University of Minnesota Press.
Klavans, J. and Resnik, P. 1996: *The Balancing Act: Combining Symbolic and Statistical Approaches to Language*. Cambridge, Mass.: MIT Press.
Levine, J. 1983: Materialism and Qualia: The Explanatory Gap. *Pacific Philosophical Quarterly*, 64, 354–61.
Loar, B. 1981: *Mind and Meaning*. Cambridge: Cambridge University Press.
Loewer, B. 1994: An Argument for Strong Supervenience. In E. Savellos (ed.), *New Essays on Supervenience*, Cambridge: Cambridge University Press.
Lormand, E. 1996: How to be a Meaning Holist. *Journal of Philosophy*, 93(2), 51–73.
Marr, D. 1981: Artificial Intelligence: A Personal View. In J. Haugeland (ed.), *Mind Design*, Cambridge, Mass.: MIT Press.
Marr, D. 1982: *Vision*. Cambridge, Mass.: MIT Press.
Matthews, R. 1980: Language Learning *versus* Grammar Growth. *Behavioral and Brain Sciences*, 3, 25–6.
Millikan, R. 1984: *Language, Thought and Other Biological Categories*. Cambridge, Mass.: MIT Press.
Nagel, T. 1969/1972: Linguistics and Epistemology. In G. Harman (ed.), *On Noam Chomsky: Critical Essays*, New York: Doubleday.
Osherson, D. (ed.) 1995: *An Invitation to Cognitive Science* (4 vols.). Cambridge, Mass.: MIT Press.
Peacocke, C. 1986: Explanation in Computational Psychology: Language, Perception and Level 1.5. *Mind and Language*, 1, 101–23.
Peacocke, C. 1989: When is a Grammar Psychologically Real? In George 1989.
Peacocke, C. 1992: *A Study of Concepts*. Cambridge, Mass.: MIT Press.

Pinker, S. 1984: *Language Learnability and Language Development*. Cambridge, Mass.: Harvard University Press.
Putnam, H. 1975: The Meaning of "Meaning." In *Mind, Language and Reality* (Philosophical Papers, vol. 2), Cambridge: Cambridge University Press.
Putnam, H. 1980: Models and Reality. *Journal of Symbolic Logic*, 15, 464–82.
Quine, W. 1953a: Two Dogmas of Empiricism. In *From a Logical Point of View*, New York: Harper and Row.
Quine, W. 1953b: The Problem of Meaning in Linguistics. In *From a Logical Point of View*, New York: Harper and Row.
Quine, W. 1956: Carnap and Logical Truth. In *Ways of Paradox and Other Essays*, revised and enlarged edition 1976, Cambridge, Mass.: Harvard University Press.
Quine, W. 1972: Methodological Reflections on Current Linguistic Theory. In G. Harman and D. Davidson (eds.), *Semantics of Natural Language*, New York: Humanities Press.
Rey, G. 1997: *Contemporary Philosophy of Mind: A Contentiously Classical Approach*. Oxford: Blackwell.
Rey, G. 1998: A Narrow Representational Account of Qualitative Experience. In J. Tomberlin (ed.), *Philosophical Perspectives*, 12: *Language, Mind and Ontology*, Atascadero, Calif.: Ridgeview Press.
Rieber, R. and Voyat, G. 1983: Dialogue 1 (with Noam Chomsky). In *Dialogues on the Psychology of Thought and Language*, New York: Plenum Press.
Rosenthal, D. 1980: The Modularity and Maturation of Cognitive Processes. *Behavioral and Brain Sciences*, 3, 32–4.
Searle, J. 1980: Rules and Causation. *Behavioral and Brain Sciences*, 3, 37–8.
Searle, J. 1992: *The Rediscovery of the Mind*. Cambridge, Mass.: MIT Press.
Soames, S. 1984: Linguistics and Psychology. *Linguistics and Philosophy* 7(2).
Sober, E. 1980: Representation and Psychological Reality. *Behavioral and Brain Sciences*, 3, 38–9.
Stabler, E. 1983: How are Grammars Represented? *Behavioral and Brain Sciences*, 3, 291–421.
Stalnaker, R. 1984: *Inquiry*. Cambridge, Mass.: MIT Press.
Stampe, D. 1977: Toward a Causal Theory of Linguistic Representation. *Midwest Studies in Philosophy*, 2, Minneapolis: University of Minnesota Press.
Stich, S. 1978: Belief and Sub-doxastic States. *Philosophy of Science*, 45, 499–518.
Stich, S. 1980. What Every Speaker Cognizes. *Behavioral and Brain Sciences*, 3, 39–40.
Stich, S. and Warfield, T. 1994: *Mental Representation: A Reader*. Oxford: Blackwell.
Strawson, G. 1987: *Freedom and Belief*. Oxford: Oxford University Press.
Strawson, G. 1994: *Mental Reality*. Cambridge, Mass.: MIT Press.
Uriagereka, J. 1998: *Rhyme and Reason*. Cambridge, Mass.: MIT Press.
Weinberg, S. 1992: *Dreams of a Final Theory*. New York: Pantheon.
Whyte, L. 1978: *The Unconscious Before Freud*. New York: St. Martin's Press.
Wright, C. 1989: Wittgenstein's Rule-Following Considerations and the Central Project of Theoretical Linguistics. In George 1989.

6
Referential Semantics for I-languages?[1]

PETER LUDLOW

Introduction

It is widely assumed that Noam Chomsky is hostile towards semantics, and at least in the oral tradition, it is supposed that this hostility carries over to virtually everything that goes by the name "semantics," including lexical semantics and model-theoretic semantics. This latter supposition is almost certainly mistaken, and it appears to reflect a confusion about precisely what kinds of semantic theories Chomsky objects to. For reasons that we will look at later, Chomsky has no problem with lexical semantics or model-theoretic semantics (at least in principle). Even when we turn to the domain of "referential" semantics – broadly construed as theories which articulate language/world connections – there are some versions that Chomsky will find unobjectionable.

My task in this paper will be to get as clear as possible on where Chomsky takes exception (and why he does). Surprisingly, we will find that once the landscape is properly surveyed, Chomsky's views do not exclude the kind of referential semantics that many philosophers favor. Indeed, I think that where there is disagreement, it stems not from disputes about referential semantics itself, but rather from certain subsidiary assumptions which I will try to make explicit and, in certain cases, defend.

1 Some Important Distinctions

To get some traction here we need to lay out some of the notions we will be working with. First, we need to get clear on the kinds of semantic theories that are in play (including those that might be characterized as being "referential" in some sense or other). Then we will need to get a bit more clear on Chomsky's notion of I-language.

1.1 Three Kinds of Semantic Theories

The first thing we need to understand is that "reference" is not a pre-theoretical notion, but as Chomsky (1995) has stressed, is a theoretical term the meaning of which will depend upon the broader semantical theory. As a very crude taxonomy, we can distinguish three ways in which "reference" is employed. As we will see, this three-way distinction still underdetermines the range of positions in play, but it will help us to clarify the subsequent discussion. I'll distinguish the senses as R^0, R^1, and R^2.

Many philosophers would take R^0 to be a misuse of the term "reference." According to this notion, reference does not involve relations to the external world, but is primarily a relation to internal representations. This might involve "reference" to elements of the "domain D" of Chomsky (1981), or to some sort of data structure or representation in the mind.

The R^0 notion of reference is fairly commonplace in linguistic theory (e.g., in lexical semantics like those in Levin and Pinker 1992) as well as in artificial intelligence research. Philosophical theories which employ the R^0 notion of reference include the structural semantics of Katz and Fodor 1963, Katz and Postal 1964, and more recent work in this vein such as Jackendoff 1972, 1983.

The key idea behind all of these proposals has been the notion that providing a semantics for a natural language expression (or syntactic form) requires that one provide a mapping of that expression (form) onto some representation which in some sense encodes the meaning of the expression. So, for example, in Jackendoff 1972, Katz and Fodor 1963, and Katz and Postal 1964, the idea is to map a given natural language expression onto a "semantic marker," which is in turn a symbol in a particular representational language (following Lewis 1972, we might call this language "Semantic Markerese").[2] In turn, Semantic Markerese is designed so that ambiguous expressions of a given natural language are disambiguated in Markerese. In addition, entailment relations and synonymy relations between natural language expressions are supposed to follow by virtue of the forms of their Semantic Markerese counterparts.

I would argue that the R^0 notion of reference is also employed by the various versions of model-theoretic semantics grounded in the work of Montague (1974). The idea that model-theoretic semantics cannot deliver language/world connections is controversial, and a number of model-theoretic semanticists are under the impression that language/world connections are precisely what they are offering (see Dowty, Wall, and Peters 1981; Partee 1980), but as Higginbotham (1990) and Lepore (1983) have argued, this betrays a certain confusion about the nature of model-theoretic semantics. Their argument is that model-theoretic semantics fails in exactly the same way that structural semantics is supposed to fail. What we expect a semantic theory to deliver is a characterization of the speaker's knowledge about the connection between language and the world.

Structural semantics does not deliver this because it only gives us a mapping from one language into some other system of representation. Model-theoretic semantics likewise fails, because it never makes the connection between language and the world, but at best gives us a mapping from expressions of a language onto certain model-theoretic objects.[3] This paper is not the place to take up the debate about model-theory and its ability to deliver language/world connections. The key point is that model-theoretic semantics may well be a theory which employs the R^0 notion of reference.

Other kinds of semantic theories more clearly explore some form of language/world relations, but they accomplish this in different ways. Let's distinguish reference in sense R^1 from reference in sense R^2, where R^1 takes the reference of a term to simply be some direct (perhaps causal) relation between that term and the world, and R^2 takes the relation to be rather more complex, involving at a minimum a four-place relation that involves the speaker, the expression used, context, and aspects of the world. One way of thinking about the difference is that on R^1 the term itself refers, but on R^2 a speaker uses the term to refer to certain aspects of the world under certain contextual conditions.

I gather that Chomsky takes R^1 to be what most direct reference theorists have in mind when they talk about reference, and indeed at least some of them may be using "reference" in that sense. On the other hand, it is clear that not all of them are (consider, for example, the literature on demonstratives, which routinely includes argument places for speaker and context). In any case, I propose that we set aside the R^1 notion of reference. For purposes of discussion, by "referential semantics" I will mean some form of semantic theory which employs R^2. I'll let individual theorists decide whether they subscribe to R^1 or R^2, although I suspect that many philosophers take R^2 to be the operative notion in their work.

Of course, saying that a semantic theory employing R^2 can express language/world connections still doesn't make matters crystal clear, since we still need to know what we mean by "language" not to mention what we mean by "aspects of the world." As we will see, for Chomsky everything turns on our understanding of these two notions.

For example, consider what turns on how we are to understand the fourth place of the R^2 reference relation. Clearly there are a number of ways to take the phrase "aspects of the world." We could take it as being about phenomenological aspects of the world, or about Humean sense data, or about the posited objects of science, or about objects in the world (in some robust sense of 'object'). We will come back to this issue a bit later, but suffice it to say, for now, that some ways of filling in the fourth place will result in semantic theories where Chomsky would take exception.

First, however, we need to take a closer look at the nature of language, and in particular at the conception of linguistic inquiry that Chomsky has articulated. A number of philosophers working in the theory of reference have conceptions of language as a social object – conceptions which are completely antithetical to

Chomsky's views on the matter. If we are not on the same page as regards the nature of language, then any sort of debate about language/world relations is going to be at cross purposes.

1.2 I-languages vs. E-languages

Here I follow Chomsky's (1986) distinction between I-language and E-language. Strictly speaking it may not even count as an actual *distinction*, since on Chomsky's view there is I-language, and then there is a collection of poorly defined if not incoherent views which might be lumped together under the general title of "E-language theories."

An I-language is not a spoken or written language (whatever they might be), but is rather a state of an internal system which is part of our biological endowment. I might have I-language representations of English sentences, but those internal representations are not to be confused with spoken or written English sentences. They are rather data structures in a kind of internal computational system that humans are born with and which they have co-opted for communication among other purposes.

Crucially, we have to understand the I-language computational system *individualistically*. That means that the properties of the system can be specified completely independently of the environment that the agent is embedded in. Thus, it involves properties like the agent's rest mass and genetic make-up (and unlike relational properties like the agent's weight and visibility to hungry tigers).

From the E-language perspective, on the other hand, a natural language is a kind of social object the structure of which is purported to be established by convention (however "convention" is to be understood), and persons may acquire varying degrees of competence in their knowledge and use of that social object. I gather that on Chomsky's views such objects would be of little scientific interest if they did exist, but that in any case such objects *don't* exist.

Consider a possible candidate for an E-language. The French Academy is supposed to dictate that French is to have a certain set of properties which cover the French lexicon, pronunciation, etc. The only problem is that the French Academy doesn't dictate as much as it thinks it does. At best, it dictates a small range of superficial rules about French, all the while relying upon shared tacit knowledge of I-language which provides the substrate upon which those prescribed rules are parasitic. What they have are a collection of proclamations about what the structure of French *should* be, which in total are not even sufficient to constitute a respectable candidate E-language.

In short, on Chomsky's view, E-languages (insofar as we can make sense of such things) would be unsuitable objects for naturalistic inquiry, but the point is academic, since there are no such things. I-languages are the only serious candidates for the objects of linguistic inquiry.

2 Are Referential Semantics for I-languages Possible?

Given the notion of I-language sketched above, we can take up the following question: is it possible to have a referential (R^2) semantics for an I-language?

Most would suppose that Chomsky's answer to this question is in the negative, and on the face of it, that answer is the natural one. An I-language, after all, is supposed to be part of an agent's psychology, and ultimately part of the agent's biology. The properties of an I-language are therefore (on Chomsky's view) individualistic – as noted above, they are properties that hold of the agent in isolation (like having particular genetic make-up).[4] The properties of a referential semantics, on the other hand, appear to be anything but individualistic. They are not properties that an agent can have in isolation, since they express relations between linguistic representations and, among other things, aspects of the world external to the agent.

But it seems to me that there is room to hold that so long as we avoid the R^1 notion of reference, and adhere to the R^2 notion, a theory which traffics in language/world relations is completely reasonable. How is this possible, given that I-language is individualistic and reference, even in sense R^2, apparently involves language/world connections?

The general phenomenon of individualistic and relational sciences informing each other is fairly common. For example, it is routine in primate ecology to draw upon facts about primate physiology (which is individualistic) to support a particular claim about the relation of the primate to its environment. More generally, as Webster and Webster (1988) observe, anatomical structure (which is presumably individualistic) can place constraints on the types of (relational) environmental functions that are possible, and vice versa.[5]

Chomsky has written little on this topic, but what he has written seems to be consistent with this general point of view. For example, in Chomsky 1995, we have the following:

> Naturalistic study is of course not limited to such [internalist] bounds; internalist inquiry into a planet or an ant does not preempt or preclude the study of the solar system or an ant community. Non-internalist studies of humans can take many forms: as phases in an Oxygen-to-Carbon Dioxide cycle or gene transmission, as farmers or gourmets, as participants in associations and communities, with their power structures, doctrinal systems, cultural practices, and so on. Internalist studies are commonly presupposed in others with broader range, but it should be obvious that the legitimacy of one or another kind of inquiry does not arise. (Chomsky 1995: 28)

The key thing to see here is Chomsky's view that internalist investigations can be presupposed by relational (i.e., non-internalist) investigations (although, at

least in this passage, Chomsky is conspicuously silent on whether the converse is true).

Of course, even if we allow that relational and individualistic sciences can inform each other, due to familiar arguments[6] one cannot draw *direct* inferences from the structure of an organism to organism/environment relations, but that does not mean that no inferences are possible. Rather it means that caution is necessary, and that each inference must be supported by careful argumentation and sound evidence – in short, it is the usual sort of scientific reasoning wherein nothing comes for free.[7] But the key point here is that our knowledge of the environment can inform our knowledge of I-language. Indeed, there is no need for Chomsky to take exception to the idea that knowledge of the environment might shed light on the nature of I-language.

So, then, what is all the noise about Chomsky and referential semantics? What precisely is the real bone of contention between Chomsky and most philosophers who espouse some form of referential semantics?

It seems to me that there are two deal-breakers here. One turns on how we are to understand the fourth place of the R^2 reference relation: "aspects of the world." As noted above, there are a number of ways to take this phrase (e.g., as being about phenomenological aspects of the world, or about Humean sense data, or about the posited objects of science, etc.).

There is another deal-breaker we should consider first, however, since it will help to clarify our discussion. This deal-breaker (a hypothesis that I will call language/world isomorphism (LWI)), strikes me as being at the root of many of Chomsky's concerns about referential semantics.

3 Language/World Isomorphism

The LWI label is a little bit crude, since the key idea is that there is an isomorphism holding between logical forms and the world. The basic idea is not restricted to the analytic tradition in philosophy. It has been at the root of the "semantic tradition" (in the sense of Coffa 1991) from Kant, through Bolzano and the early Wittgenstein, up to the present. As early as Kant, for example, we find some suggestion that there is an isomorphism between a representation and the thing represented.

> [Representation] is that determination of the spirit (*Bestimmung der Seele*) that refers to other things. What I call referring (*Beziehen*) is when its features conform to those of the external things.[8]

> [The representation] is composed out of its component concepts in the same way in which the entire represented thing is composed out of its parts. Just as, for example, one can say that the notes of a musical piece are a representation of the

> harmonic connection of the notes, not because each note is similar to each tone but because the notes are connected to each other just as the tones themselves.[9]

The most celebrated version of this general idea is surely the picture theory of meaning advanced by Wittgenstein in the *Tractatus*. On the view articulated there, just as "[w]hat a picture must have in common with reality, in order to be able to depict it – correctly or incorrectly – in the way it does, is its pictorial form" (Wittgenstein 1949: 2.17), a proposition must be isomorphic in structure to a state of affairs in order to represent or be about that state of affairs.

Of course we are not talking about the "surface form" of a proposition here.[10] We are talking about the proposition "under analysis" – in other words, we are talking about its logical form. For Wittgenstein, if one got down to the complete analysis of proposition it would reflect the structure of the world.[11] The key idea here is that we do not suppose there is an isomorphism between the surface form of a sentence and the world, but rather that there is an isomorphism between the sentence under analysis (i.e., its logical form) and the world.[12]

Most of the following arguments from Chomsky are often construed as being arguments against the possibility of semantics. In fact, however, it seems to me that they are not (by themselves) arguments against semantic theories which employ reference in sense R^2. They are rather arguments against any semantic theory which traffics in language/world isomorphism. Ultimately, they are intended to count against certain conceptions of referential semantics, but seeing how they work against these conceptions will require that we first carefully walk through some of these arguments.

4 Chomsky's Arguments against LWI

In Chomsky's recent writings there are a number of arguments which are designed to show that if we adopted a referential semantics (using R^1 or, I submit, any R^2 theory employing the LWI hypothesis), the kind of ontology we would predict does not appear to track our intuitions about the kinds of things we are really talking about. (i) Such theories allegedly commit us to things that we would never acknowledge as existing; (ii) they allegedly commit us to types of things which are different from the types of things that we ordinarily suppose we are talking about; and (iii) the things we talk about are of such a character that they are allegedly too unruly for such theories to handle. To have labels for all these possibilities, let's call them the "argument from implausible commitments," the "type mismatch argument," and the "misbehaving object argument."

I'll review all three of Chomsky's arguments before offering what I take to be the most natural reply to them. As we will see, this reply will lead us directly into the teeth of Chomsky's concerns about "aspects of the world," a topic that will occupy the final section of this paper.

4.1 The Implausible Commitments Argument

Chomsky draws attention to the fact that a referential semantics (apparently in conjunction with the LWI hypothesis) commits us to some apparently implausible entities, flaws for example, and suggests that we really can't take seriously a theory which commits us to such entities.

> If I say "the flaw in the argument is obvious, but it escaped John's attention," I am not committed to the absurd view that among things in the world are flaws, one of them in the argument in question. Nevertheless, the NP *the flaw in the argument* behaves in all relevant respects in the manner of the truly referential expression *the coat in the closet*. (Chomsky 1981: 324)

Pursuing a similar line of attack, Hornstein (1984: 58) has drawn attention to constructions like (1).

(1) The average man is concerned about his weight

Hornstein contends that "no one wishes to argue that there are objects that are average men in any meaningful sense."

The general concern introduced by examples like "flaw in the argument" and "the average man" is that the doctrine that there must be a tight connection between semantics and ontology is suspect, for it commits us to apparently absurd views.

This argument needs to be hedged up a bit before it is fully functional. As it stands, the argument makes the assumption that the logical form of these constructions is transparent – e.g., that the expression "the flaw" really is nothing more than a simple NP containing a determiner and a noun. This is not such an innocent assumption. After all, the LWI hypothesis does not hold that there is an isomorphism between surface linguistic form and the world, but rather between some ultimate logical form and the world. As Higginbotham (1985) has argued, before we admit a commitment to objects like flaws and average men, we would have to look more closely at the underlying logical form of these constructions.

With respect to the case of "flaw in the argument," Higginbotham suggests that the construction is parallel to "bad singer." The semantics of "bad singer" is not such that we say someone is a bad singer iff they are bad and are a singer. Rather, we understand that someone is a singer and that their singing is bad. Likewise we should not say that "that is a flaw in the argument" is true iff that is a flaw and that is in the argument. Rather, we should say that the phrase is true iff something is an argument and it is flawed. Despite appearances, "flaw" may not be a nominal, but may be a modifier like "bad."

Higginbotham offers a similar analysis for the case of "the average man." First,

he notes that the expression is actually ambiguous between an individual who has typical properties, and the sense in which we say that the average family has 2.3 children. It is presumably the latter sense which is problematic here. In this latter case, Higginbotham suggests that "average," despite appearing to be an adjective, is functioning as a kind of adverbial. The construction is parallel to examples like (2), discussed by Haïk (1983).

(2) Let's have a quick cup of coffee

Clearly in (2) we are not suggesting that there are cups of coffee which are in some sense quick. Rather "quick" is behaving as an adverbial, modifying the activity. Returning to "the average man," Higginbotham proposes that in a construction like (3)

(3) The average family has 2.3 children

the analysis will be something along the following lines:

(3′) On average, a family has 2.3 children

Chomsky might reply in a number of ways here.[13] Consider the rather more complex (4):

(4) Your report on the average family fails to make it clear that it has 2.3 children[14]

Even if suitable glosses can be worked out, they will be technical spellouts, of course, and these spellouts may introduce difficulties of their own. But the glosses themselves are not necessarily innocent. As Chomsky (pers. com.) asks, are we to continue with the strategy for examples like (5)?

(5) That his income is falling bothers John Doe

That is, are we to take John Doe to be standing proxy for "an average man," which in (5) is to be further unpacked with "average" operating as a kind of adverbial? This begins to constitute a very strong and possibly counter-intuitive thesis about the logical form of these constructions.

Returning to the case of "flaw in the argument," there are likewise responses to Higginbotham available. For example, Chomsky (pers. com.) asks how could one treat "flaw" as a kind of predicate in an example like (6)?

(6) We fixed three of the flaws you found but the rest of them resisted our efforts

One possible answer is to say that (6) can be glossed, as in (6′), where it is the steps in the argument that are intuitively flawed.

(6′) We fixed three of the flawed steps you found but the rest of them resisted our efforts

Again glosses of this nature have to be justified at some point, and there is a heavy burden to show that the introduction of this proposed hidden structure comports well with the rest of what we know about the syntactic form of these constructions. Still more, it has to be shown that general rules are available and not just case-by-case fixes. Any attempt to pursue the Higginbotham course here will not be a trivial exercise.

Even if the "average man" and "flaw" cases *can* be analyzed away, it is perhaps doubtful that *all* such cases can be analyzed away. One can, of course, imagine a project which would attempt to analyze away all "implausible" objects, but this would amount to rejecting certain core assumptions that we have agreed to grant Chomsky here. Such a project would essentially be committed to the super-externalist view that linguistic form could only be determined after ontology was complete – in other words, that linguistic form was determined by the embedding circumstances of the language. And that project, assuming it is coherent, would radically depart from Chomsky's individualistic assumptions about the nature of I-language.

4.2 The "Type Mismatch" Argument

Even if we are comfortable with perhaps counter-intuitive entities like flaws, Chomsky (1995) notes that there is an apparent mismatch between the type individuation that objects and substances intuitively have, and the type individuation that a referential semantics will provide. To get clear on the issue a terminological distinction will be helpful. Suppose we distinguish an I-substance from a P-substance, where a P-substance (if there is such a thing) is the sort of stuff that would play a role in physical theory (H_2O for example), and an I-substance (if there is such) is the stuff that we are intuitively talking about when we use language (the intuitive referent of "water"). To put the point in a more theory-neutral way, the I-substance is what it appears we are talking about based upon our use of language. The alleged problem is that P-substances and I-substances just don't match up right.

If, following Putnam 1975, "water" refers to H_2O, then a referential semantics will assign a P-substance (H_2O) as the semantic value of "water." But the problem is that the stuff we are actually talking about when we use the term "water" – the I-substance – is something else altogether. To see this, consider the fact that what we find in the Hudson River is called "water," though it could

hardly be considered H_2O. Also problematic is the fact that there are substances like ice tea which chemically approximate H_2O much more closely than Hudson River water, yet we don't call them "water." According to Chomsky, the situation is even more problematic than this. If someone at the water company poured tea leaves into the system so that what came out of the tap was chemically identical to Lipton Ice Tea, we would still call it "water" – although we might complain about its impurity.

So, here is the situation. What we are talking about when we use the term "water" – the I-substance – depends upon the social setting in which we find that substance. But according to referential semantics, the meaning of the term is supposed to depend upon the chemical composition of the substance referred to – it's supposed to be a P-substance. Conclusion: referential semantics (if respecting the LWI hypothesis) will not track the intuitive notion of meaning.

One might think it possible to get off the hook if one appeals to social theories of external (referential) content (in the sense of Burge 1979). Rather than P-substances, we might posit S-substances, substances that are individuated according to community norms. So, while my concept of water might not accord with H_2O, it might still accord with a certain socially determined object, which has the property of being water when it comes from the faucet, but not when it is served at a restaurant. The problem here is that there are plenty of examples where such S-substances (if there could be such things) would not track our intuitions about the extension of terms – Burge's own examples about arthritis and brisket are cases in point. That is, there is a mismatch between I-substances and S-substances. We will come back to this point a bit later.

4.3 The Misbehaving Object Argument

There are a number of interesting features to the water/tea story, one of which can be broken out as a separate objection to referential semantics.

We have already seen in the type mismatch argument that I-substances don't track P-substances, but there is another problem. The water/tea story also seems to show that the I-substance we are talking about when we use the term "water" is a most ill-behaved sort of substance. Something may cease to be water even if no internal physical changes have taken place. For example, the same chemical compound is water when it comes from the tap, but ceases to be water when it is served at a restaurant.

If that's the intuitive character of I-substances, then there is really little hope that referential semantics can "give the reference" of what we talk about when we talk about "water" and "tea," since referential semantics is supposedly going to say that the content of these terms is H_2O in the first case and H_2O plus certain other elements in the latter. That is, I-substances are so unruly that it is wildly implausible to suppose that they could have any counterparts in the

physical world. Hence they have no counterparts that a referential semantics could utilize as their referents.

A related, if somewhat more general, point is made in Chomsky (1975: 203), where Chomsky notes that the very notion of whether we are talking about a single object or a collection of objects turns on any number of social and institutional factors:

> We do not regard a herd of cattle as a physical object, but rather as a collection, though there would be no logical incoherence in the notion of a scattered object, as Quine, Goodman, and others have made clear. But even spatiotemporal contiguity does not suffice as a general condition. One wing of an airplane is an object, but its left half, though equally continuous, is not . . . Furthermore, scattered entities can be taken to be single physical objects under some conditions: consider a picket fence with breaks, or a Calder mobile. The latter is a "thing," whereas a collection of leaves on a tree is not. The reason, apparently, is that the mobile is created by an act of human will. If this is correct, then beliefs about human will and action and intention play a crucial role in determining even the most simple and elementary of concepts.

Moving that discussion into the current debate, we might say that it is implausible for even very simple semantic concepts like object and collection to correspond in any interesting sense with P-substances.

4.4 Analysis of the Arguments: Some Replies

The conclusion that we can draw from examples like the above is that any referential semantics purporting to respect the LWI hypothesis is going to misfire badly, since it is bound to utilize P-substances as referents, and P-substances just don't track the intuitive meanings of natural language expressions.

One reply for the semanticist is to simply reject the LWI hypothesis. That is, one could say that yes, a semantic theory is concerned about the connection between language and the world (as part of a four-place relation which also involves speaker and context), but this says nothing about there being an isomorphic mapping between primitive linguistic expressions (under analysis) and things (or kinds of things) in the world. If there is a single lexical entry corresponding to "water," it does not map onto any single substance, but rather various things depending upon circumstances, discourse participants, etc. Such theories are no less referential than those which respect LWI. After all, they are still in the business of articulating language/world relations, albeit through a very complicated story, and despite the fact that these relations are often dynamic one/many relations. I am guessing that Chomsky would have no objection to such a move.

There is, however, another maneuver – indeed, a standard philosophical

maneuver – which I gather is the place where Chomsky parts company from Putnam and Burge and the like. According to this strategy, individualist claims about the individuating conditions of "water" and "arthritis," etc. have no force. In the case of Burge's (1979) example of Oscar (who claims that he has arthritis in his thigh), one can say that "arthritis" refers (R^2) exclusively to ailments of the joints, and reject the claim that for Oscar, "arthritis" might apply to a painful condition of the thigh. Thus, we preserve R^2, LWI, and the intuition that the world is furnished by P-substances (or at least S-substances).

But now it is not clear that we can retain the I-language thesis. If there is an isomorphic relation between the content of Oscar's utterance and the syntactic form of that utterance (in other words, if LWI holds), and if the content of Oscar's utterance depends upon the environment in which he utters it, then it looks like the syntactic form of Oscar's utterances is going to depend upon the (possibly social) environment in which the utterance takes place. But then the syntactic form of Oscar's utterances doesn't depend solely upon facts about Oscar in isolation. And this is just to reject the I-language thesis.

Is it possible to retain I-language, LWI, and externalism about content? Could one argue that the linguistic representations remain unchanged but that their contents shift as Oscar moves between Earth and Twin-Earth? This seems unlikely. Following the dialectic of a related argument due to Boghossian (1997), consider the case where Earth and Twin-Earth differ only in that on Twin-Earth there is no gold (substitute metals are used). Twin-Earthians have the concept of gold, but it is by virtue of certain experts who have postulated the possibility of gold on the basis of physical theory. Thus, the concept of gold on Twin-Earth is a complex one, having to do with numerous facts involving the periodic table, etc. On Earth the concept of gold is an unstructured concept about gold itself. Now consider the cases of Oscar and Twin-Oscar, both of whom utter "I want some gold." Assuming isomorphism between linguistic structure and conceptual structure, when Oscar utters the word "gold," he employs a singular referring expression. Meanwhile, when Twin-Oscar utters "gold," he employs a structured description.[15] But by hypothesis Oscar and Twin-Oscar are internally identical. So, if LWI is respected, the difference in the syntax of their utterances must depend exclusively upon the respective environments in which they are embedded. Hence I-language has to be given up. We are forced to a notion of E-language, which Chomsky has already rejected as incoherent.[16]

Summarizing the strategies covered thus far, one can retain the notion of I-language with R^2 reference to P(S)-substances, but only if one gives up LWI. If we try to hang onto LWI and P(S)-substances, then we have to give up I-language. Is there another way out?

Why not give up the P(S)-substances? That is, why not bite the metaphysical bullet and acknowledge that there really are I-substances – that is, things like flaws, etc. They are clearly not logically absurd entities, and it need not be conceded that they are particularly odd entities. In the case of flaws, at least, one

might say that they are altogether common in the arguments one runs across, and one might wonder why they should be considered any less real than, say, tables and chairs. Likewise, coats hanging in the closet need not have any particular ontological priority over average families and flaws.

But is this a metaphysical bullet that we can bite? Can our ontology admit such things? Here (finally) we come to the issue of "aspects of the world" and the kind of things that we can be realists about.

5 Aspects of the World

Operating throughout all three of Chomsky's arguments above is the assumption (shared with Putnam et al.) that P-substances are the kinds of substances that a referential semantics is going to favor. For example, Chomsky (1992: 208) remarks:

> To be an Intentional Realist, it would seem, is about as plausible as being a Desk- or Sound-of-Language- or Cat- or Matter-Realist; not that there are no such things as desks, etc., but that in the domain where questions of realism arise in a serious way, in the context of the search for laws of nature, objects are not conceived from the peculiar perspectives provided by the concepts of common sense.

But why should we make the assumption that "the domain where questions of realism arise in a serious way" is "in the context of the search for laws of nature"? One possible answer would be that physical theory gets to say what's real, and hence if we are to have a genuine referential semantics in which the referents are "real" existing entities, then we are stuck with the kinds of entities and substances posited by physical theory. But this answer makes a strong assumption about scientific realism – one which is controversial to say the least, and most likely false, in my view.

There is a great deal of literature in the philosophy of science (for example van Fraassen 1980) which holds that the entities which science posits do not exist in the same sense as mid-sized, earth-bound objects like tables and chairs. Pursuing this line of thinking, we might say that scientific theories, despite their great interest and utility, are not the arbiters of what is real.

If we set aside the exclusive claim of the physical sciences on our ontology (i.e., if we dismiss P-substances as our semantic values), then we may well find that I-substances are entirely plausible candidates for the referents of a semantic theory. So, for example, it may be that the semantic value of "water" just is water – that complicated I-substance which moves through pipes into our homes, etc. If we take that route, then it is far from clear that the LWI hypothesis must be surrendered.

But perhaps the appeal to I-substances is a cheat.[17] Or perhaps it is entirely

parasitic on the notion of I-language representations. What is an I-substance if not simply "whatever corresponds to a particular I-language representation"? Seen in this light, isn't the talk of I-substance vacuous? Or at best, isn't it a misleading way of talking about I-language?

Here we need to head off a confusion. Even if it should turn out that I-language representations are in some sense prior to I-substances (perhaps logically prior or metaphysically prior) it does not follow that our path as empirical investigators will begin with the linguistic representations. The idea is that the linguistic representations will indeed underwrite our metaphysical intuitions, but that because of this we can expect our metaphysical intuitions to shed some light on the nature of I-language. To take a concrete example, the lexical entry for "water" when fully fleshed out is bound to be a very complicated representation (or at least to interact in complex ways with other I-language representations). Our knowledge of this fact does not come from direct investigation of the lexicon, but rather is guided by metaphysical intuitions about water like those that Chomsky evinces in his articles – intuitions which take into account rich contextual information about whether the material in question is coming from a faucet, or is being served at a restaurant.

It will not do to argue that Chomsky's water intuitions are not metaphysical – that they are only about his "I-concept" of water, or about the phenomenology of water. If there is no world beyond these concepts or beyond the phenomenology, then once we have made the step from talking about linguistic representations to talk of concepts, or phenomena, or sense data, we have basically stepped into the world – or as far into it as we are ever going to get.[18]

Of course, this picture is as Kantian as it can be. We have metaphysical intuitions, and we want to know what underwrites those intuitions. The first departure from Kant lies in the answer given – not "the categories of reason" but rather the structure of I-language. The second departure lies in the fact that our approach need not be entirely transcendental. We do have substantial independent knowledge of the language faculty, and we can use that knowledge to gain insight into the nature of reality.

6 Conclusion

I've made two claims in this chapter. The first one is that Chomsky is not hostile to referential (R^2) semantics *per se*, but rather toward certain auxiliary hypotheses about language/world isomorphism and toward ways we might understand talk of "the world." The second, somewhat contentious, claim is that language/world isomorphism is not implausible and that we can retain the hypothesis if we are prepared to be less ontologically stingy. More contentious still, the hypothesis may have some value to actual empirical research into the nature of I-language.

Naturally, caution is necessary in any attempt to draw conclusions about the nature of I-language from our knowledge of I-substances. Our knowledge of such substances is certainly fragmentary (are there really flaws and average guys or not?), and the nature of the link between those substances and their linguistic representations remains to be explored. One plausible hypothesis, however, is that any grasp we have on metaphysics is by virtue of our having the linguistic representations that we do. That is, were it not for our I-language representations having certain properties, we would find ourselves in a rather different world of I-substances. If this view should turn out to be true, then firm intuitions about I-substances can certainly shed light on the nature of I-language. Matters work in the reverse direction as well. It is entirely possible that if semantic theory requires that we quantify over flaws and average guys, then that helps shed light on our imperfect intuitions about ontology.

In sum, if referential semantics is possible, then it provides a new source of evidence (intuitions about I-substances) which could shed light on the nature of I-language. Likewise, the study of I-language could shed light on the nature of metaphysics. These possibilities illustrate the great promise of developing referential semantics for I-languages, and they also explain part of the motivation that some of us have for pursuing such theories.

Chomsky's reply: pp. 287–95.

Notes

1 This paper has benefited much from an extended correspondence with Noam Chomsky on these issues. Earlier versions of this material were presented to the Riejka Branch of the Section for Analytical Philosophy, Croatian Philosophical Society (1995), at the 2nd Central European Summer School in Generative Linguistics, Olomouc, the Czech Republic (1995), at the Department of Cognitive Science at Johns Hopkins University (1995), and at the Conference on Naturalizing Semantics, Maribor Slovenia (1996). I am indebted to Nenad Miscevic, Boran Bercic, Snjezana Prijic, and George Wilson for helpful discussion, to Noam Chomsky, Michael Devitt, Norbert Hornstein, and Richard Larson for comments on various drafts of this paper, and to Susan Larson for the references on kangaroo rats, etc.

2 So far as I know, Fodor may still hold some view like this.

3 It might be objected that we can always specify an intended model, and that the actual world should be the intended model. But Lepore (1983) has observed that this move will not work. If there is to be an intended model, then we should like to know how to identify it. But, on the face of it, this is no easy task.

> How much about a world do we need to know before we can distinguish it from all other worlds? Presumably a lot. There presumably is a class of worlds in which the number of trees in Canada is even and one in which the number is odd. So far are

> we from being able to single out the actual world from all others that we do not even know which class it falls in. But do we need to distinguish the actual world from all others to understand our language? (Lepore 1983: 184)

4 Here I am granting Chomsky the assumption that cognitive science generally is individualistic. This assumption, while plausible, is certainly not universally held. See Burge 1986 for the anti-individualism view, and Chomsky 1995 for a response.

5 One interesting example comes from the study of the anatomy of the kangaroo rat, where significant progress has been made by making judicious inferences from physical structure to organism/environment relations, and vice versa. When kangaroo rats were first discovered, biologists observed that they had particularly large middle-ear cavities. According to Webster and Webster (1988), biologists have posited a number of hypotheses about the function of the cavity, and the hypotheses based on anatomy alone appeared to be the weakest. For example Hatt (1932) argued that the enlarged cavities shifted the weight of the head to the posterior and thereby assisted upright saltatorial locomotion.

An alternative, and rather natural assumption, however, was that the enlarged cavities served the function of improved hearing. Research undertaken in the 1950s incorporated both anatomical and environmental studies and concluded that the enlarged cavity assisted in hearing, and subsequent work by Webster and Webster (1971) showed that the function was to pick up low frequency sound waves typically generated by predators like owls and snakes.

6 In popular literature, this point has been stressed numerous times by Stephen Jay Gould.

7 As Webster and Webster (1988) summarize the situation:

> Implicit in many morphological studies is the idea that structure determines, and therefore can reveal, function. Unfortunately this is an oversimplification. A more realistic view is that structure – whether of an entire organism or its parts – places constraints on what functions are possible (Gans, 1985) and may, to the observant, suggest some that are plausible.

8 Kant 1910–83: vol. 15, 76–7. Translation is from Coffa 1991: 31.

9 Kant 1910–83: vol. 15, 78. Translation from Coffa 1991: 31.

10 The very idea of a surface form is problematic in any case, since sentences (or utterances) do not come with their forms on their sleeves. The form that they receive when written, for example, is hardly a surface form in any interesting sense. Even word and sentence boundaries presuppose a fairly rich theory about the underlying form of our utterances.

11 As Wittgenstein explains the matter in his 1915 notebooks:

> It is clear that the constituents of our statements can and should be analyzed by means of definitions, and must be, if we want to approximate to the real structure of the statement . . . The analysed proposition mentions more than the unanalysed. The analysis makes the proposition more complicated than it was, but it cannot and must not make it more complicated than its reference was from the first. When the proposition is just as complex as its reference, then it is *completely* analysed. (Wittgenstein 1961: 46)

12 The basic idea and its application to questions of ontology can be illustrated by means of a brief detour through the history of the philosophy of language. In medieval philosophy of language (even earlier, actually), it was observed that standard assumptions about the connection between language and the world lead to certain surprising conclusions. Consider a simple sentence like (i):

(i) A dog barked

Specifically, if we regard "a dog" as a referring expression, then it appears that there must be some object to which it refers. But what is the referred-to object in this case? Some vague dog? Such proposals were entertained by medieval philosophers of language, and the consequences were taken quite seriously (and still are in, for example, Fine 1985). Clearly, one operative assumption was that if one successfully expresses a proposition by uttering (i), then the components of (i) must refer to something.

Much later, Meinong (1904) argued from the assumption that all expressions refer, to the conclusion that there must be nonexistent objects. So, for example, if all expressions refer and sentence (ii) is true,

(ii) Smith seeks the Golden Mountain

then "the Golden Mountain" must refer, not to an existing object, but to a *nonexistent* object.

Frege is often credited with showing us how to steer clear of the imbroglios created by sentences like (i). If we use the expression "some dog" we are not referring to some object or other, but rather are saying that some object in the domain is a dog. The logical form of (i) is best represented as in (i′), which says that there is an x which is a dog and which barked.

(i′) $(\exists x)(\text{dog}(x) \,\&\, \text{barked}(x))$

Here there is at most reference to the properties *doghood* and *barking*. If the semantics of (i) is as in (i′) then (i) is properly understood as being about these properties rather than as being about an abstract vague dog.

The same general strategy was exploited by Russell (1905) in his theory of descriptions. Against Meinong, Russell argued that definite descriptions like "The present King of France" do not refer, but rather *denote*. Thus, a sentence like (iii) has the underlying logical form in (iii′).

(iii) The present King of France is bald

(iii′) $\exists x(\text{Present-King-of-France}(x) \,\&\, \forall y(\text{Present-King-of-France}(y) \rightarrow x = y) \,\&\, \text{Bald}(x))$

(iii′) is not about any particular object according to Russell, but is rather about the property baldness and the property of being the present King of France. (iii′) asserts that the world is such that it contains one object which has the property of presently being the King of France, and everything which has that property is bald. (See Neale 1990 for discussion.)

The examples I have just given are cases where assumptions about the nature of reality have influenced our theory of the logical form of language. If we cannot countenance vague objects or nonexistent objects, then perhaps we need to take a closer look at the underlying form of our language. Alternatively, we can also find cases where beliefs about the underlying form of our language have led to substantive metaphysical claims.

For example, Davidson (1968/9) argued for the existence of events on the basis of the logical form of action sentences. To illustrate, we can easily make the inference from (iv) to (v):

(iv) John ate the chips gracefully

(v) John ate the chips

Davidson proposed that this inference was a logical inference, and that the inference could be made formally once the underlying logical form of "John ate the chips gracefully" was revealed. Davidson suggested that its logical form should be as in (vi):

(vi) $(\exists e)$ [ate(John, the chips, e) & graceful(e)]

Roughly, (vi) can be understood as saying that there was an event e which was an eating of the chips by John, and e was graceful. The conclusion (v) follows by simple conjunction reduction. Davidson took this to be evidence for the existence of events.

While Chomsky would certainly be sympathetic to a number of these technical proposals, I suspect he would *not* accept the notion that there are metaphysical consequences to these proposals. He would doubtless reject Meinong's assertion that "The Golden Mountain" must refer to some sort of object, and he would not feel compelled to give some alternative analysis of the expression *à la* Russell's theory of descriptions (the theory of descriptions might, however, be motivated by other considerations). But the reason for rejecting these metaphysical consequences has little to do with a rejection of referential semantics (in sense R^2). Rather, I think the metaphysical entailments are rejected because they rest on the assumption (apparently held by everyone from Russell to Meinong to the early Wittgenstein to Davidson) that there is an isomorphism holding between language (under analysis) and the world.

13 In the following discussion I am indebted to personal communication with Chomsky.

14 One possible response for Higginbotham here is to render (4) along the lines of (4′):

(4′) Your report fails to make it clear that on average a family has 2.3 children

There are, however, potential pitfalls in the (4′) gambit. For a start, the report described in (4′) need not be *about* the average family at all; it could very well be a report on milk consumption in urban areas. Is (4″) a possibility?

(4″) Your report on the average family fails to make it clear that on average a family has 2.3 children

Obviously not, since this brings us back to the apparent quantification over the average family. But perhaps the upstairs NP can be analyzed away in another fashion, along the lines of (4‴):

(4‴) Your report on what, on average, the state of families is, fails to make it clear that on average a family has 2.3 children

But this might not do either, since (4) does not say that the report is on the general *state* of families.

15 For more on the distinction between structured denoting expressions and unstructured referring expressions, see Neale 1993.

16 Chomsky has made numerous independent arguments against content externalism which I am passing over here, since my primary interest is in ferreting out semantical theories that are compatible with the notion of I-language. The philosophical consequences of content externalism are of course vast, and it may well be that it leads to consequences which, if taken to their logical conclusion, would prove to be reductios of the doctrine. For example, there are extremely complex issues about the consequences of the doctrine for the theories of self-knowledge, epistemic warrant, and memory. For a survey, see Ludlow and Martin (forthcoming).

17 Both Paul Horwich and Noam Chomsky have suggested as much to me.

18 Nor does it help to argue that the concepts mediate between language and the world. In that case, the same considerations apply to the concept/world relation. If conceptual structure underwrites our metaphysical intuitions about I-substances, then our intuitions about I-substances will provide insights into our conceptual structure, and our conceptual structure will provide insights into the structure of I-language. For the record, I consider conceptual structure to be a dispensable "middle man" here.

References

Boghossian, P. 1997: What the Externalist Can Know A Priori. *Proceedings of the Aristotelian Society*, 97, 161–75.

Burge, T. 1979: Individualism and the Mental. *Midwest Studies in Philosophy*, 4, Minneapolis: University of Minnesota Press.

Burge, T. 1986: Individualism and Psychology. *The Philosophical Review*, 95, 3–45.

Chomsky, N. 1975: *Reflections on Language*. New York: Pantheon.

Chomsky, N. 1981: *Lectures on Government and Binding*. Dordrecht: Foris.

Chomsky, N. 1986: *Knowledge of Language*. New York: Praeger.

Chomsky, N. 1992: Explaining Language Use. *Philosophical Topics*, 20, 205–31.

Chomsky, N. 1995: Language and Nature. *Mind*, 104 (413), 1–61.

Coffa, J. A. 1991: *The Semantic Tradition from Kant to Carnap*. Cambridge: Cambridge University Press.

Davidson, D. 1980: The Logical Form of Action Sentences. In *Essays on Actions and Events*, Oxford: Oxford University Press.

Dowty, D., Wall, R., and Peters, S. 1981: *Introduction to Montague Semantics*. Dordrecht: D. Reidel.

Fine, K. 1985: *Reasoning with Arbitrary Objects.* Oxford: Blackwell.

Gans, C. 1985: Vertebrate Morphology: Tale of a Phoenix. *American Zoology*, 25, 689–94.

Haïk, I. 1983: Syncategormatic Adjectives. Manuscript. MIT, Cambridge, Mass.

Hatt, R. T. 1932: Vertebral Columns of Ricochetal Rodents. *Bulletin of the American Museum of Natural History*, 63, 599–738.

Higginbotham, J. 1985: On Semantics. *Linguistic Inquiry*, 16, 547–94.

Higginbotham, J. 1990: Contexts, Models, and Meanings: A Note on the Data of Semantics. In R. Kempson (ed.), *Mental Representations: The Interface Between Language and Reality*, Cambridge: Cambridge University Press.

Hornstein, N. 1984: *Logic as Grammar*. Cambridge, Mass.: MIT Press.

Jackendoff, R. 1972: *Semantic Interpretation in Generative Grammar.* Cambridge, Mass.: MIT Press.

Jackendoff, R. 1983: *Semantics and Cognition.* Cambridge, Mass.: MIT Press.

Kant, I. 1910–83: *Gesammelte Schriften*, vols. 1–29. Edition by Königliche Preussische Akademie der Wissenschaften und Deutsche Akademie der Wissenschaften zu Berlin. Berlin: De Gruyter & Reimer.

Katz, J. and Fodor, J. A. 1963: The Structure of a Semantic Theory. *Language*, 39, 170–210.

Katz, J. and Postal, P. 1964: *An Integrated Theory of Linguistic Description.* Cambridge, Mass.: MIT Press.

Lepore, E. 1983: What Model Theoretic Semantics Cannot Do. *Synthèse*, 54, 167–87.

Levin, B. and Pinker, S. 1992: *Lexical and Conceptual Semantics*. Cambridge, Mass.: Blackwell.

Lewis, D. 1972: General Semantics. In D. Davidson and G. Harman (eds.), *Semantics of Natural Language*, Dordrecht: D. Reidel.

Ludlow, P. and Martin, N. (eds.) forthcoming: *Externalism and Self-Knowledge*. Stanford: CSLI Publications.

Meinong, A. 1904: Über Gegenstandstheorie. Translated as "The Theory of Objects," in R. Chisolm (ed.), *Realism and the Background of Phenomenology*, New York: Free Press, 1960.

Montague, R. 1974: *Formal Philosophy*. Ed. Richmond Thomason. New Haven: Yale University Press.

Neale, S. 1990: *Descriptions.* Cambridge, Mass.: MIT Press/Bradford Books.

Neale, S. 1993: Terms Limits. In J. Tomberlin (ed.), *Philosophical Perspectives*, 7: *The Philosophy of Language and Logic*, Atascadero, Calif.: Ridgeview Press.

Partee, B. 1980: Montague Grammar, Mental Representation, and Reality. In P. French, T. Euhling, and H. Wettstein (eds.), *Midwest Studies in Philosophy*, 5, Minneapolis: University of Minnesota Press.

Putnam, H. 1975: The Meaning of "Meaning." In K. Gunderson (ed.), *Minnesota Studies in the Philosophy of Science*, vol. 7: *Language, Mind and Knowledge*, Minneapolis: University of Minnesota Press.

Russell, B. 1905: On Denoting. *Mind*, 14, 479–93.

Van Fraassen, B. 1980: *The Scientific Image*. Oxford: Oxford University Press.

Webster, D. and Webster, M. 1971: Adaptive Value of Hearing and Vision in Kangaroo Rat Predator Avoidance. *Brain, Behavior and Evolution*, 4, 310–22.

Webster, D. and Webster, M. 1988: Hypotheses Derived from Morphological Data: When and How They Are Useful. *American Zoologist*, 28, 231–6.
Wittgenstein, L. 1949: *Tractatus Logico-Philosophicus.* London: Routledge & Kegan Paul.
Wittgenstein, L. 1961: *Notebooks 1914–1916.* Ed. and tr. G. E. M. Anscombe. New York: Harper & Row.

7
Meaning and its Place in the Language Faculty

PAUL HORWICH

1 Methodological Background

Some projects are incoherent: their alleged aims self-contradictory or not well defined.[1] Others, though coherent, are impractical or unfeasible: we have no reason to think they can be properly carried through. And some projects, though both clear and do-able, are not worthwhile: their completion would yield results of too little value.

Studies of language – which vary enormously from one to another with respect to their subject matter, ambitions, methods, and potential products – well illustrate these three respects in which a project may be sound or suspect. Such studies can be concerned with language in general, or with some particular language; their focus may be historical, sociological, neurological, cultural, philosophical, literary, or psychological; and their goals can be descriptive, explanatory, normative, or pedagogical. Thus, dozens of ways of investigating language are imaginable; but few can be expected to satisfy our three desiderata.

One that does appear to satisfy them is the scientific investigation into which internal properties of a person are responsible for his or her linguistic activity, for the individual's capacity to speak and understand a language. This endeavor – which Chomsky has termed "I-linguistics" – is clearly coherent, for it is simply one part of the global scientific enterprise; moreover the prospects of arriving at knowledge in this domain are initially quite reasonable and have been bolstered by the development of increasingly successful theories; and the value of such results can be no less than that of reaching a scientific understanding of any other significant aspect of the natural world.

Of course, the legitimacy of I-linguistics – judged with regard to coherence, feasibility, and value – has no tendency to imply the illegitimacy of other studies

of language. But, if we are not to waste time and energy, it is important to focus attention on these desiderata and on how hard it is to satisfy them, and to concentrate our efforts on projects that do. So-called "philosophical" accounts – those that are self-consciously *non*-scientific – are particularly vulnerable in this regard, as Chomsky has frequently pointed out.[2]

2 Evidence

Since the aim of I-linguistics is to discover the basis of an individual's linguistic activity – his speech and understanding – those phenomena will provide the primary evidence for any theory within the field. This is not to draw an invidious distinction between what may and may not give support to an I-linguistic theory. Insofar as such a theory is *scientific*, there can be no such distinction: any fact in any domain *might* somehow prove relevant, and a great variety of facts *have* proved relevant. The point is merely that since the goal is to explain certain phenomena, then those phenomena must be amongst the facts that are accommodated.

More specifically, what must ultimately constrain our theory are vocal sounds that a person produces, the circumstances in which they are produced, and the characteristic reactions to sounds that are heard. Such facts are plainly relevant and uncontroversially observable, presupposing relatively few of the psychological hypotheses whose correctness might be at issue. Thus, although they by no means exhaust the data that might well be deployed in support of an I-linguistics theory, they will provide an essential epistemological foundation for it.

Notice, however, that although the fundamental data in I-linguistics will tend to be fairly behavioral, we cannot expect to be able to predict and explain *specific items* of linguistic behavior; for any such item will be the product of an unmanageably large array of interacting causal factors including not merely the individual's knowledge of his language, but also his beliefs, what he desires to express, and his levels of alertness, intoxication, etc.

Following Chomsky, it seems reasonable to postulate that a person's "knowledge of language" is embodied in a distinct mechanism or faculty and that the focus of I-linguistics should be on the character of this faculty. The more-or-less behavioral facts one might expect roughly to correlate with the properties of a person's language faculty are extremely crude, vague, qualified, approximate generalizations regarding the way that his or her words sound, and the tendency for them to be deployed in certain distinctive combinations in certain circumstances: for example,

Peter says things of the form "x seems to be sleeping," but not usually of the form "x seems sleeping"

Peter is normally disposed to accept "That's red" in the presence of red things

Peter is normally inclined to infer "Mary intended to take her medicine" from "Mary was persuaded to take her medicine"

Such crude behavioral patterns are presumed to result from the operation of the language faculty in "normal conditions." More specifically, it might be tentatively supposed that Peter's language faculty immediately engenders such facts as that

"x seems to be sleeping" is grammatical but "x seems sleeping" is not

"That's red" is true of red things

"Mary intended to take her medicine" follows from "Mary was persuaded to take her medicine"[3]

and that these in turn, given further, normal facts about Peter, account for his observable behavior patterns.

It is not unnatural to characterize the immediate output of the language faculty as items of *knowledge*: for example,

Peter knows that sentences of the form "x seems to be sleeping" are grammatical and that sentences of the form "x seems sleeping" are not

Peter knows that the word "red" applies to red things

Peter knows that "was persuaded" entails "intended"

and to suppose that what explains this knowledge – namely, the *fundamental* content of the language faculty – is also a body of knowledge. For insofar as Peter's language faculty is governed by certain law-like regularities, R, which result in certain facts, D, we may speak of Peter's *implicitly* knowing R and D. However, our describing the laws and immediate products of the language faculty as states of "knowledge" is more of a *façon de parler* than a substantive theoretical move. We are entitled to recharacterize the fact *that Peter's language faculty entails that "John is sleeping" is grammatical* as the fact *that Peter implicitly knows that "John is sleeping" is grammatical*. But the latter formulation makes no additional commitments, has no explanatory advantages, and does nothing to illuminate the nature of the fact "known."[4]

The immediate products of the language faculty (which can be regarded as

states of implicit knowledge) are the objects of deeper explanation: they are the "data" for a linguistic theory which characterizes the internal structure of that faculty. The virtue of focusing on such "data" rather than verbal behavior is that they are more sharply formulated and bear more directly on what we want to find out about – namely, the basic nature of that component of the mind/brain. However, this advantage is acquired at a certain cost, which should not be forgotten. For it is an uncertain theoretical hypothesis that the inclination to engage in some particular verbal behavior is the joint product of some (theoretically articulated) state of a "faculty of language," together with other factors that determine whether or how that state will be manifested on any given occasion. This is indeed a reasonable conjecture; for it is indeed plausible to suppose that a person's speech behavior – although affected by such factors as environmental conditions, beliefs, tiredness, desire to speak, honesty, etc. – is also an expression of his "knowledge of a language." But it should not be forgotten that both the faculty-of-language postulate, and specific theses about its immediate deliverances, are theoretical conjectures whose plausibility derives from the observation of behavioral tendencies together with assumptions about the interaction between the language faculty and other mental and physical factors.[5]

3 A Very Simple Picture

In developing this idea let us assume with Chomsky (a) that each human being indeed has a faculty of language, FL, a component of his mind/brain constituting the primary causal/explanatory basis of his linguistic activity; (b) that the possible states, L1, L2, . . ., of FL are, by definition, possible I-languages; (c) that each such state, L, is a computational procedure that generates infinitely many I-expressions, E1, E2, . . .; (d) that each such expression, E, is a pairing <PHON(E), SEM(E)> of phonetic and semantic objects, which, through their respective interaction with the perceptual/articulatory system (P/A) and the conceptual/intentional system (C/I), determine an association of a sound with a thought; (e) that these PHON–SEM pairs are constructed from lexical items, LI1, LI2, . . .; and (f) that these lexical items are stored in a lexicon which is accessed by the computational procedures that form I-expressions.[6]

Figure 7.1 shows a simple model of how things might work – a model which incorporates all these assumptions (but which nevertheless is not adopted by Chomsky, as we shall see).

In this very simple picture (which I'll call VSP) each lexical item consists in an I-sound paired with formal properties that indicate the "functional type" to which it belongs. Thus, some lexical items will be *schemata* containing "slots" into which items of specified functional types may be inserted to yield terms that may in turn be inserted into other schemata, and so on. A "construction"

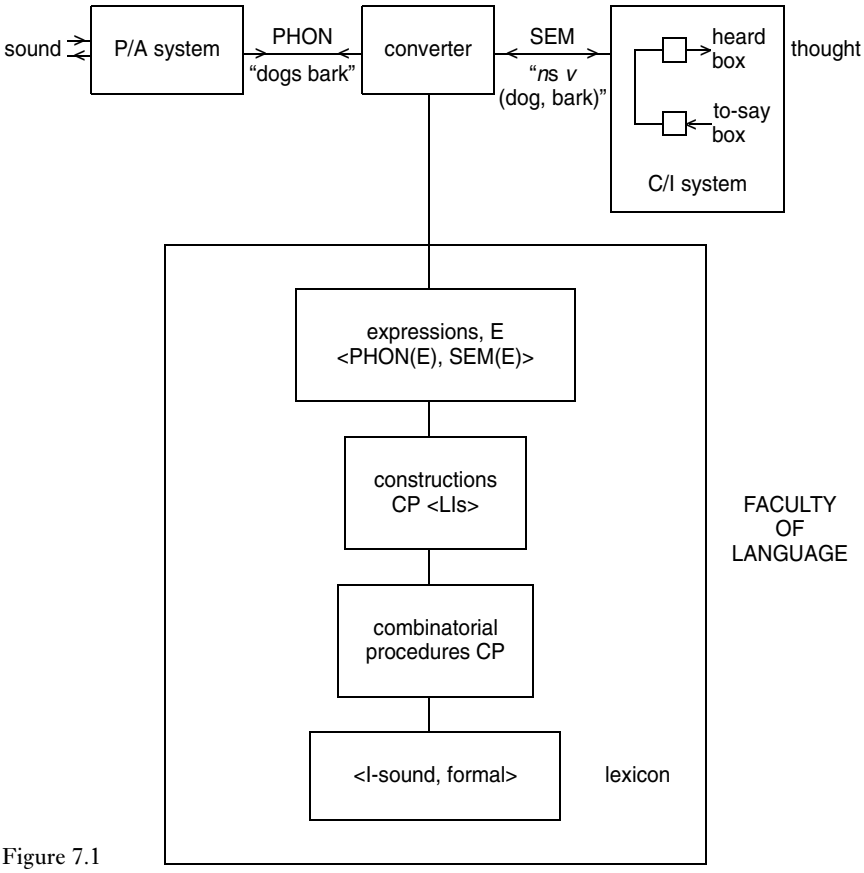


Figure 7.1

specifies certain lexical items to be combined in that way, and specifies in which order this is to be done – e.g.,

Apply LI(15) to the sequence <LI(3), the result of applying LI(7) to the sequence <LI(22), LI(1), LI(4)>>]

Thus, a construction is a pair consisting of an abstract combinatorial procedure, e.g.,

Apply u to <v, the result of applying w to <x,y,z>>

and a sequence of lexical items, e.g.

<LI(15), LI(3), LI(7), LI(22), LI(1), LI(4)>

on which the procedure is to be imposed. The expression E that is computed from such a construction is a pair consisting of a phonetic object, PHON(E), and a semantic object, SEM(E). A crucial difference between these two components of E is that in PHON(E) the information about how the LIs have been combined is largely deleted – all we have, roughly speaking, is a modified string of I-sounds. But in SEM(E) that information is retained – encoded by means of bracketing. The phonetic object, PHON(E), determines (via P/A) how the expression sounds. And vice versa. So when that sound occurs, the corresponding semantic object, SEM(E) – which consists in a combinatorial procedure imposed on a sequence of I-sounds – enters the "heard box" and passes into thought.[7] It is an expression of the person's language of thought. Each I-sound expresses (or embodies) the concept it does in virtue of its basic conceptual role – which consists in the explanatory priority of the fact that certain specified thought-formulations, which employ that I-sound, are maintained in certain circumstances.[8] Each SEM derives its content from the concepts expressed by the LIs from which it was constructed and from the combinatorial procedure imposed on those LIs. (See section 7 for further discussion of the compositionality of meaning.) Note that there may be an innate predisposition toward the instantiation of particular conceptual roles.

For example (and simplifying enormously for the sake of illustration), a person's lexicon might contain the lexical items <"dog," (f1, f5)> and <"bark," (f2, f4, f9)> and <"n̲s v̲," (f15)>. The formal features, f1, f2, . . ., fN, might be such that an expression can be formed by applying something with f15 to a pair of terms with f1 and f4. Thus, we have a permissible construction consisting of the procedure, "Apply the first item to the second and third items" imposed on the sequence of three LIs

<"n̲s v̲," (f15)>, <"dog," (f1, f5)>, <"bark," (f2, f4, f9)>

This construction determines a certain phonetic object, "dogs bark," and a certain semantic object, "n̲s v̲ (dog, bark)."[9] The first of these determines (given the character of the P/A system) a certain sound. The second passes into the C/I system: to think that dogs bark is for this object to have a particular "location" in that system (i.e., for it to be in the "belief box"). Its elements, namely "dog," "bark," and "n̲s v̲," express or embody the concepts they do in virtue of their conceptual roles – or, more specifically, in virtue of certain of their deployments (in postulates, inference patterns, etc.) being explanatorily basic.[10]

4 Mentalese

Let us now consider various objections and alternatives to this picture. One modification of VSP, suggested in the work of Jerry Fodor, involves the idea that

the language of thought is the same for everyone – a universal "Mentalese."[11] This idea can be represented by amending the above model slightly. We might suppose that each lexical item contains, not merely an I-sound paired with formal properties, but also a symbol ("word") of Mentalese. And we can suppose that the SEM part of an expression is made up of the Mentalese components (rather than the I-sounds) of the LIs from which it is constructed. Thus, one possible I-expression would be

<"dogs bark," "n* v ($, #)">

where "n* v," "$," and "#" are the Mentalese translations of "ns v," "dog," and "bark."

However, as Chomsky has pointed out, this sort of move has various disadvantages.[12] In the first place it is evidently less simple than VSP – postulating an additional inner language. In the second place it is not plausible to suppose that languages are so easily translated into a universal Mentalese. Experience with translating spoken languages into one another suggests that exact translation is rarely possible, because words in one language tend not to have exactly the same meanings as words in another – which they would have if they could each be translated into Mentalese. This argument breaks down if Mentalese is imagined to be so rich that it has separate terms for all the subtle variations of meaning exhibited in ordinary languages – but in that case the first objection (regarding the extra complexity of the Mentalese hypothesis) becomes even more telling.

Amongst the considerations that might nonetheless be thought to militate in favor of Mentalese are:

1 the explanatory value of a language in which to have the thought processes involved in learning one's first spoken language;
2 the need to integrate linguistically encoded information with representations from the visual system;
3 the presumed fact that an individual, growing up in isolation and never acquiring a spoken language, would nevertheless be capable of elementary reasoning.

However, these considerations may be accommodated by supposing that the terms of a normal person's language of thought include, not only I-sounds, but also a strictly limited number of universal Mentalese terms (expressing such very basic concepts as RED, OBJECT, LATER THAN, etc.). This small concession towards the Mentalese hypothesis is perfectly consistent with VSP. For we can continue to maintain that most of a person's language of thought is composed of the I-sounds of his or her spoken language. The association of the few Mentalese terms (e.g., "red*") with their spoken-language equivalents can be constituted in the belief system (e.g., by holding true "x is red* if and only if x is red"). Thus,

there is no need for translations into Mentalese to take place within the language faculty.[13]

5 Referentialism

A second way of departing from VSP is to suppose that each I-sound derives its content (i.e., embodies the thought constituents that it does) not in virtue of a fundamental, internal, conceptual role, but rather in virtue of standing in some *referential relation* to some aspect of the external world. This idea may be thought to be motivated by Davidson's contention that the compositionality of meaning (the fact that the meanings of sentences are explained by the meanings of their words) can be accommodated only by supposing that the meanings of sentences consist in truth conditions and that the meanings of words consist in their referents.[14] And the idea is further supported in the work of Fodor and others, suggesting how this reference relation might be "naturalized" – i.e., analyzed in non-semantic, scientifically respectable terms.[15] Thus, Fodor maintains that a predicate (in thought) has as its meaning the abstract concept (i.e., content) DOG, in virtue of the fact that tokenings of that predicate are nomologically correlated with the presence (to the speaker) of dogs. This alternative to the "internal conceptual role" account of concept-identity can be affixed either to my VSP (in which concepts are internally expressed by I-sounds) or to the Mentalese variant (in which concepts are expressed by Mentalese terms).

Either way, there are two objections. In the first place, the motivation for the referential approach, insofar as it comes from the need to account for the compositionality of meaning, is defective. For, as we shall see in section 7, it is very easy to explain compositionality in other terms.

And in the second place, the referential approach to concept-identity does not provide a framework in which the characteristic uses of words (sounds) can be explained. For in order to do that we must be able to explain the use in relation to one another of the concepts expressed by those words – e.g., Peter's very strong conviction that bachelors are unmarried, that if John is taller than Bill then Bill is not taller than John, that the sentence "dogs bark" is true if and only if dogs bark, that if electrons exist they are negatively charged, etc. It is very hard to see how these convictions could be explained merely in virtue of referential relations that are constituted by the tendency to think, "That's a bachelor, a man, bigger, married, true, an electron, charge, negative, atom, etc. . . ." in the presence, respectively, of bachelors, men, instances of "bigger than," etc.

Nor can the referentialist afford to retreat to the claim that his view is not *attempting* to address the above explanatory questions, but merely to provide an account of concept-identity. For this would be to concede that the view contributes nothing at all to one of the important goals of I-linguistics; and it is

unclear how any other goal of I-linguistics would be facilitated either. Moreover, it is plausible to suppose that theories of concept constitution *must* play a vital role in such explanations; for it is plausible that the event of *having* a certain thought (e.g., Peter's maintaining, on a particular occasion, that it is raining) is due in part to the *nature* of that type of thought (together with auxiliary facts regarding sensory input, background theories, etc.). In other words, we really must suppose that the facts that identify which concept is expressed (or embodied) by a given internal term (I-sound) are facts that play a fundamental role in explaining the overall deployment of that term within C/I. The referential properties of a term will not be able to perform this explanatory work.

6 Definitions

A third alternative to VSP involves the idea that the meanings of I-sounds are explicitly represented in the faculty of language. They are not, as in VSP, merely constituted by their behavior in the belief system. Rather the C/I system is capable of "interpreting" the specifications, given in the language faculty, of each term's meaning, thereby ensuring that the term is deployed appropriately. This picture is favored by Chomsky, who implements it by supposing (a) that lexical items consist, not merely of I-sounds (and their formal properties), but of I-sounds paired with I-meanings (and formal properties); and (b) that the input to (and output of) the C/I system – namely SEM(E) – is constructed, *not* from I-sounds (as in VSP), but from I-meanings. Thus we have something like figure 7.2.[16]

This picture – whereby meanings are represented in the faculty of language – may be developed in various alternative directions, suggested by the alternative answers one might give to the following pair of questions:

(Q1) What is the form of an I-meaning? Does it look like a classical definition, in which a word's meaning is articulated by providing a synonymous expression composed from a certain stock of primitives? Is it rather the representation of a regularity of use? Or is there perhaps some further possibility?

(Q2) What is the relation between the I-meanings and the constituents of thought? Is it that the I-meanings themselves are the constituents of beliefs and are manipulated within the conceptual system? Or is it rather that the I-meanings determine the use of the terms that are deployed within the conceptual system – terms such as I-sounds or words of Mentalese?

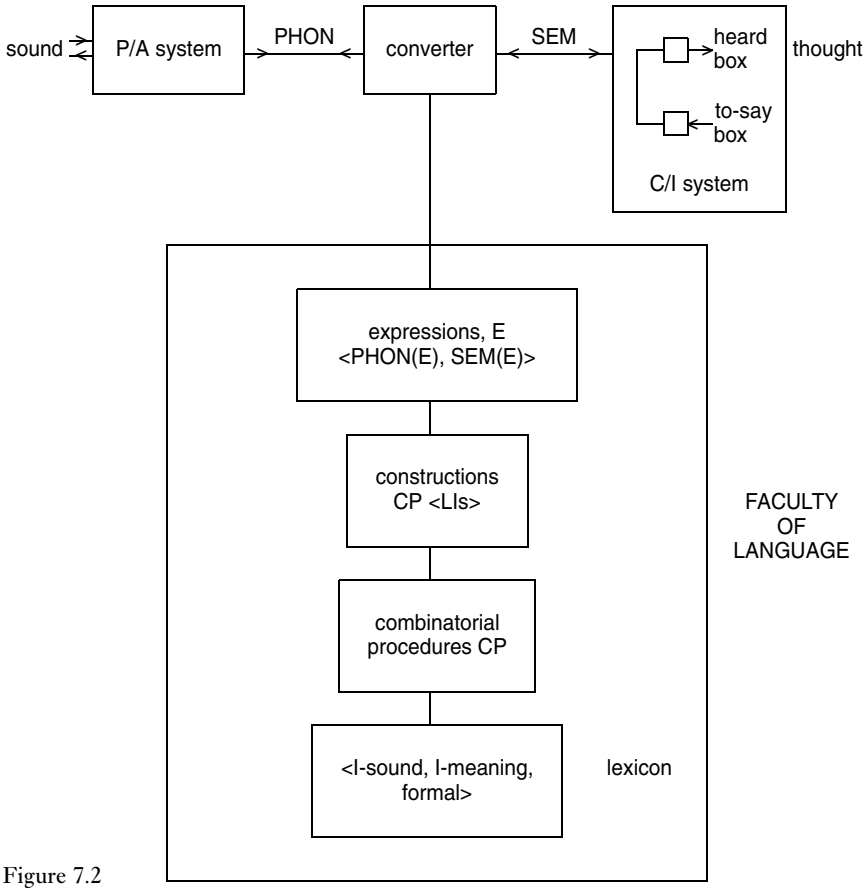


Figure 7.2

Regarding Q1, it seems more plausible to suppose that, in general, I-meanings take the relatively liberal, flexible form of the specification of a regularity of use, rather than the relatively constrained form of an orthodox explicit definition. For it is notoriously difficult to provide satisfactory, proper definitions of just about anything. Therefore, it seems more realistic to imagine that each lexical item pairs an I-sound with the representation of a basic regularity of use – which will not in general (though it may sometimes) take the form of an explicit definition.[17]

Turning to Q2, the issue is whether we should suppose that the terms in which we think are literally composed of the specifications of their conceptual roles, or whether, rather, these instructions – the I-meanings – are acted on by C/I to result in terms that are used accordingly. A merely apparent advantage of the former view – that mental terms are literally *composed* of their definitions – is that wherever the term is, the C/I system knows what to do with it; for

the C/I system will react to a term depending on its intrinsic characteristics. But on reflection it is obvious that there is no need for this sort of mechanism. Suppose the conceptual role of a mental term "f" consists in the underived acceptance within C/I of certain sentences, "#f," containing it. Then that fact alone (together with other general principles about the system) will constrain and influence the deployment of "f." It is not as if each token of "f" exhibits some characteristic, meaning-fixing behavior – so that it is plausible to locate the cause of this behavior within each token. It is rather that the variety of facts about "f"s deployment are best explained by the properties of *some* of its tokens – namely that there is a proclivity towards certain beliefs containing them. Thus it is gratuitous to suppose that each token carries its definition around with it.

Assuming these particular answers to Q1 and Q2, it remains to compare my VSP with the alternative model which incorporates them. We have, on the one hand, the picture presented initially, in which the basic internal conceptual roles of I-sounds are merely exemplified in C/I; and, on the other hand, a Chomskian model in which *in addition* these conceptual roles are explicitly represented in FL as I-meanings. Clearly, the postulation of I-meanings involves an extra layer of complexity, and therefore must be justified in terms of explanatory advantages. But it is unclear what these advantages could be.

One possibility is that considerations of learnability and innateness would favor the existence of I-meanings. The claim, more specifically, might be that since we learn our language so quickly and on the basis of such little evidence, it must be that we have available to us only a very limited stock of psychologically possible I-meanings, which are waiting innately in the language faculty for particular I-sounds to be associated with them. It might be held, for example, that the tendency to employ "red" as a response to red light emerges so quickly that there must be an innate proclivity toward this pattern of use, arising from a characterization of it being innately in the faculty of language. But clearly we can distinguish the hypothesis of there being an innate proclivity toward a certain restricted set of conceptual roles, from the further hypothesis that these are spelled out in the faculty of language, rather than merely instantiated in the belief system. All the explanatory work is being done by the innateness hypothesis: the postulation of I-meanings appears to be otiose.

Another idea might be that in order to explain why a term has a certain conceptual role we should suppose that there exists an explicit characterization of that role – instructions to which the term can be made to conform. But it is hard to see any merit in this explanatory strategy. In the first place, the existence of the explicit characterization would itself call for explanation. In the second place, the conceptual role of a term will often consist in nothing more complex than the presence of certain thought-formulations which contain that term; and this can well be something much simpler than the explicit characterization of that role; so there is no explanatory rationale for supposing that the exemplifica-

tion of role stems from some specification of it. We can suppose rather that a person, simply by holding true what he hears, provides such terms with their conceptual roles. And in the third place, it is possible (indeed plausible) that certain conceptual roles are innate. In such a case an innate Mentalese term might well possess that role. Again there is no need for anything in the faculty of language to represent it.[18]

Finally, it might be suspected that *the symmetry of sound and meaning* should dictate that since I-sounds are represented in FL, I-meanings must also be represented there. However, although it can be valuable to treat sound and meaning similarly, the analogy must of course break down somewhere. And one point at which it appears to fail is this. Vocal sounds are short events correlated with the P/A system temporarily entering into certain characteristic states. Meanings on the other hand – ways of understanding words – are long-lasting background conditions of the C/I system. Thus, although we can speak of understanding a word at just the same time as it is being heard, what is happening in C/I is an event caused by the sound (i.e., by the P/A activity) within a salient background state of C/I – a state that constitutes knowing what such sounds mean and which is responsible for providing the event with distinctive consequences (e.g., beliefs) within C/I. Thus we have, at bottom, an association between a sound-*type* and a background state of C/I. There is no need for something correlated with each *instance* of a given sound-type to give characteristic meaning-instructions to C/I. Rather, when the sound appears C/I is already in the state required to understand it.[19]

Thus, the asymmetry between the ways that sound and meaning relate to FL stem from the radical difference in nature between sounds and thoughts. A sound event is a distinctive pattern of air vibrations. What is correlated with it – i.e., a thought constituent – is an entity whose identity as the thought constituent it is depends, *not* on its intrinsic properties, but on its role in the conceptual system. For example, in the language of thought model, a thought constituent consists not merely of a certain mental term, but depends on the basic role of that term – i.e., the fact that certain sentences containing the term are (non-inferentially) "in the belief box." Consequently, in order to produce a particular sound element, the articulatory system must be sent the appropriate "instruction" – the appropriate PHON. But in order to produce a certain thought constituent it does not matter what element ("symbol") is generated in C/I – what is crucial is the *role* of that element. That conceptual role does not need to be specified every time the symbol possessing it is produced. Once the conceptual role of a mental symbol has been established, then, insofar as an element of SEM keys that symbol, it will key the appropriate thought constituent. This being so, there is no reason why the elements of SEM should not be I-sounds. Indeed, for the sake of economy, the elements of thought might also be assumed to be I-sounds. These elements will qualify as particular thought elements in virtue of having particular conceptual roles. And each such role, having once been learned, is tied

to an I-sound. There is no need for these ties to be articulated in the faculty of language.

Thus the postulation of I-meanings offers no explanatory advantages. Therefore, VSP is preferable to any of the pictures that involve I-meanings – whether they articulate orthodox definitions or conceptual roles, and whether they themselves are the terms of thought, or merely determine the deployment of those terms.

7 Compositionality

A fourth point of conflict between VSP and certain alternatives has to do with the compositionality of meaning. What explains the fact that we know the meanings of an unlimited number of sentences? Under the influence of Chomsky and Davidson some philosophers and linguists hypothesize that the language faculty contains explicit inferential processes in which conclusions about the meanings of complex expressions are inferred from premisses about the meanings of their parts and from further premisses about the ways those parts have been combined.[20] In addition, it is often supposed (following Davidson) that such inferences are possible only if the conclusions regarding the meanings of sentences are articulated as claims about the *truth conditions* of these sentences, and only if the premisses about the meanings of the words are articulated as claims about the referents of those words. For, in that case, Tarski-style demonstrations of the truth conditions of sentences may be converted into derivations of their meanings.

However, there are a number of good reasons to reject this idea. First, what needs to be computed is SEM(E) – a semantically structured correlate of PHON(E). But this is done simply by combining I-meanings (in Chomsky's picture) or I-sounds (in VSP) in a certain order as legislated by some combinatorial procedure (under the constraint of their formal properties). Second, there is no explanatory need for any conclusion of the form

"PHON(E) means that p"
or: "PHON(E) is true iff p"
or: "SEM(E) means that p"
or: "SEM(E) is true iff p"

to be *explicitly* drawn. Not that the speaker cannot be said to know such things. But this knowledge is merely implicit in the fact that he deploys SEM(E) as he does in his conceptual-intentional-belief system. Third, one might wonder how it can be that the content of SEM(E) is determined by the contents of its parts – i.e., why there is no possibility that the underlying property in virtue of which a complex mental expression has its meaning fails to square with the properties in

virtue of which the lexical items have their meanings (given the way these items have been combined). But this can be explained trivially – in a way that has nothing to do with truth conditions. It suffices to suppose that the content-property of a complex mental expression, SEM, is *constituted by* – one might even say *identical to* – its property of being constructed as it is from LIs with certain meanings.

Suppose for example that the term "%" embodies the concept FIDO in virtue of having underlying property P1, and that the term "#" embodies the concept BARKS in virtue of having the underlying property P2. Then how can we be sure that the SEM "#(%)" will possess that underlying property P3 which would constitute its embodying the complex concept FIDO BARKS? How can we be sure that "%"s having P1 and "#"s having P2 will together guarantee that "#(%)" has P3? Don't we need, as the Davidsonians maintain, to suppose that these content-constituting characteristics, P1, P2, and P3, are referential and truth conditional properties? I am claiming that we don't. For it suffices to suppose that P3 is the property

x results from applying a term with P2 to a term with P1

And similarly in other cases. We can suppose that what *constitutes* the content property of a SEM is its "construction property" – its property of resulting from the imposition of a certain syntactic/semantic structure (or combinatorial procedure) on lexical items whose own contents are provided by specified content constituting characteristics. Thus, compositionality is accommodated without making any assumptions whatsoever about what sort of property of a primitive is responsible for its embodying the concept it does.[21]

8 Conclusion

For over thirty years Chomsky's framework has been subjected to a stream of spurious philosophical criticism deriving for the most part from the failure to appreciate that when familiar words are deployed in scientific theories they may acquire new and technical meanings. Thus, he is accused of not discussing *language* properly so-called, of contradicting himself in referring to *unconscious rule-following*, of confusion about *representation* and *knowledge*, of blindness to the *externalist* character of meaning, and so on. It is not surprising that he has emerged unscathed from these criticisms, which are based on elementary confusion, and do little but bring philosophy into bad repute. This, of course, is not to suggest that Chomsky's framework for I-linguistics is correct. It is to say that its assessment must be *a posteriori*: it must concern the question whether or not we can find any simpler explanation of the phenomena within its domain. In that spirit I have tentatively proposed a model, VSP, in which the association of

sounds with meanings is achieved by virtue of the conceptual roles of those sounds. I have argued that VSP compares favorably to various alternatives, including those suggested by Fodor, the Davidsonians, the referentialists, and Chomsky himself.[22]

Chomsky's reply: pp. 295–304.

Notes

1 This paper was stimulated by Noam Chomsky's "Internalist Explorations," in *Festschrift for Burge*, ed. M. Hahn and B. Ramberg (Cambridge, Mass.: MIT Press, 1999). However, the Chomskian point of view that I discuss here is also articulated in various other works of his, including *Knowledge of Language* (New York: Praeger, 1986); *Language and Thought* (Wakefield: Moyer Bell, 1993); and "On the Nature, Use and Acquisition of Language," in *Mind and Cognition*: *A Reader*, ed. W. Lycan (Oxford: Blackwell, 1990).

2 There is a striking convergence of opinion between Chomsky and later Wittgenstein regarding the dubious legitimacy of philosophical theories of language.

3 These formulations cannot be quite right. For the facts determined by the language faculty will be *theoretical* – i.e., not exactly expressible using ordinary language expressions such as "follows from" and "the sentence 'x seems to be sleeping.'"

4 As Chomsky puts it, "In English usage, having a language is called 'knowing a language' . . ." ("Internalist Explorations," section 1). Similarly, we can say that having a faculty governed by the law that R(FL) can be described as "knowing that R(FL)."

5 In stressing the crucial evidential role of behavioral facts I do not mean to impugn the value of other kinds of data. Of particular importance is an individual's intuitive judgments regarding the grammaticality (or degree of grammaticality) of sentences presented to him. Note, however, that the utility of such information presupposes the theoretical assumption that we each have introspective access to the products of our own language faculty. This assumption can be reasonable only to the extent that the linguistic theories to which it leads can also accommodate the behavioral data.

6 See Chomsky, "Internalist Explorations," sections 2 and 3.

7 One could perhaps simplify further by getting rid of the "converter." But it is worth explicitly noting that it is one thing for FL to deliver a list of pairs of PHONs and SEMs, and another thing to have a device which, with access to that information, takes a given PHON as input and outputs the appropriate SEM (and vice versa).

8 When I speak of a term (e.g., an I-sound) "having a conceptual role" or "being governed by a basic regularity," I have in mind a fact about it to the effect that the acceptance of certain postulates or inference rules involving the term are explanatorily basic with respect to its overall deployment. For example (very roughly):

> Peter's acceptance of instances of "<p> is true iff p" is not the result of his inferring them from other assumptions about truth and is the explanatory basis of his overall use of "true."

> Peter's acceptance of inferences from "p" and "q" to "p and q," and vice versa, does not stem from his acceptance of deeper facts involving "and," and is the explanatory basis of his overall use of it.

The idea is that the meanings of I-sounds are constituted by such "regularities," which are exemplified in C/I. Moreover, according to VSP, these don't need to be explicitly articulated in FL (or anywhere else). This version of the "use theory of meaning" is developed in my *Meaning* (Oxford: Clarendon Press, 1998).

9 One might wonder why we should suppose that in order to get "dogs bark" we need to recognize as a lexical item, not merely "dog" and "bark," but also "ns v." The answer is that the combinatorial procedures can then be articulated in purely argument-functional terms, as just indicated. Moreover, there don't have to be any principles that specify what these procedures are allowed to be. Any way of combining any lexical items will be permitted as long as it squares with their formal properties. That is the essence of syntactic "Minimalism."

10 An utterance will seem ill formed to the extent that the hearer's P/A system responds to it by outputting a phonetic representation that does not sufficiently coincide with the PHON of any of the expressions constructed in his FL.

11 See, for example, J. Fodor, *The Language of Thought* (New York: Thomas Y. Crowell, 1975).

12 See Chomsky, "Internalist Explorations," section 4.

13 Some further considerations that might be thought to favor the Mentalese hypothesis are: (a) the presumed fact that thoughts, unlike sounds, are not ambiguous (but a given I-sound might have two distinct conceptual roles, and its tokens indexed to keep track of them: e.g., some beliefs may be formulated with "bank(1)" and some with "bank(2)." In that case, the process of understanding an instance of "I went to the bank" will involve, first, FL's assigning to it the unique SEM determined by that PHON; and, second, C/I's determining (on contextual/pragmatic grounds) the right conceptual role: i.e., guessing which of the two groups of sentences containing "bank" that are already in the belief box are sentences to which the present instance of "bank" is inferentially relevant, and then indexing it accordingly); (b) the possibility of more-or-less approximate translations (but could not these be established merely on the basis of similarities of basic conceptual role?); (c) the existence of thoughts that we are able to express in public only by means of metaphors and other evocative language (but is there any reason to assume that such "thoughts" can be literally expressed in Mentalese?).

For more on the pros and cons of Mentalese, see N. Smith and I. Tsimpli, *The Mind of a Savant* (Oxford: Blackwell, 1993).

14 D. Davidson, "Truth and Meaning," in *Truth and Interpretation* (Oxford: Clarendon Press, 1984).

15 See, for example, J. Fodor, *Psychosemantics* (Cambridge, Mass.: MIT Press, 1987); F. I. Dretske, *Knowledge and the Flow of Information* (Cambridge, Mass.: MIT Press, 1981); D. W. Stampe, "Toward a Causal Theory of Linguistic Representation," *Midwest Studies in Philosophy*, 2, 1977; D. Papineau, *Reality and Representation* (Oxford: Blackwell, 1987); P. Jacob, *What Minds Can Do* (Cambridge: Cambridge University Press, 1997).

16 Chomsky does not explicitly present this model: it is my reconstruction of what he suggests in sections 2, 3, and 4 of "Internalist Explorations."

17 This question has implications for whether or not I-meanings are, in a certain sense, "holistic" or "atomistic." The issue is whether the existence of a given I-meaning in Peter's lexicon can or cannot entail the existence of a certain other I-meaning in his lexicon. To the extent that I-meanings are explicit definitions, then the existence of each is independent of the other, and they are "atomistic." But to the extent that I-meanings specify regularities of use, then they require the deployment of concepts other than the one they directly determine. Note that this "holism" is in no way implausible or objectionable. In particular, it should not be confused with another doctrine, sometimes given the same name, according to which any change, no matter how slight, in what is believed by a person, alters to some degree the meaning of every one of his terms. The latter idea – endorsed by Davidson (*Truth and Interpretation*) and by Ned Block ("An Argument for Holism," *Proceedings of the Aristotelian Society*, 95 (1994–5), 151–69) – is indeed unattractive; but it is quite different from the view that the meanings of certain terms are inextricably interdependent.

18 There are at least two ways for *a priori* belief to arise from within the VSP perspective. One way is through the innateness of certain conceptual roles. For example, it may be an innate feature of C/I that there is a Mentalese token "%" such that "p%q" is inferred from "p" and "q," and vice versa; and in that case the belief that if dogs bark and pigs fly, then pigs fly, is *a priori*. Another kind of *a priori* commitment arises by virtue of the pragmatic decision to hold true a certain sentence (so that it is not held true for empirical reasons). For example, "Bachelors are unmarried" is maintained for the sake of a socially useful distinction, not in order to accommodate experience.

19 One might be tempted to try a further simplification – that is, to eliminate, not merely I-meanings, but I-sounds as well! But this is clearly not feasible. For there must be information of some sort passed between P/A and C/I; so there have to be terms in which that information is articulated. Moreover, there must be a system capable of taking the output of P/A – roughly, a string of sounds – and determining its compositional structure. Thus, neither I-sounds nor FL can be dispensed with.

20 I don't mean to suggest that either Chomsky or Davidson holds this view – only that their remarks on compositionality have led others to adopt it. See, for example, R. Larson and G. Segal, *Knowledge of Meaning* (Cambridge, Mass.: MIT Press, 1995).

21 This approach to compositionality is developed in my "The Composition of Meanings," *Philosophical Review*, vol. 106, no. 4 (1997), 503–31; reprinted as chapter 7 of *Meaning* and in my "Deflating Compositionality," *Ratio*, 14 (December 2001).

22 I am very grateful to Ned Block, Robyn Carsten, Tim Crane, Michael Harnish, Rita Manzini, Alec Marantz, Barry C. Smith, Neil Smith, and especially to Noam Chomsky, for their helpful comments on an earlier draft of this paper.

8

Small Verbs, Complex Events: Analyticity without Synonymy

PAUL M. PIETROSKI

I was introduced to linguistics and philosophy of language at the same time, in a setting which encouraged the thought that these enterprises can be fully continuous and mutually supporting. Noam Chomsky provided my initial exposure to this field of inquiry. For this, I am deeply grateful.[1]

In this chapter, I combine an aspect of Chomsky's (1995a) minimalist program with a version of Davidson's (1967) event analysis. This may seem odd, given Chomsky's recent claims about truth-conditional semantics. (See Chomsky 1995b: 23–7.) But perhaps what follows is an exercise in what Chomsky calls "internalist" semantics – a "form of syntax" that does not contribute to "an account of communication in terms of shared entities" or Fregean *Bedeutungen*. Then again, perhaps we really do need to say that certain verbs are semantically associated with events, in order to explain certain facts about human linguistic competence. In any case, minimalism and the event analysis are made for each other, especially with respect to sentences like

(1) John boiled the water.
(2) The water boiled.

In my view, the inference from (1) to (2) is valid; (2) follows from (1), much as (2) follows from

(3) John did something, and the water boiled.

Correlatively, I think that

(4) If John boiled the water, then the water boiled.

is analytic – knowable by virtue of knowing a language. This view has been criticized, most recently by Fodor and Lepore (1998, 1999). But one can respond, by combining a Davidsonian treatment of action sentences with Chomsky's claim that transitive constructions like (1) result from merging intransitive verb phrases with "small" verbs – hidden lexical elements that resemble the causative morphemes posited by many theorists. This lets one maintain that the inference from (1) to (2) is an instance of a valid form; although specifying the logical form of (1), and saying *why* (2) follows from (1), requires both semantic and syntactic theory. The proposed account introduces a motivated twist on traditional causative analyses of (1), thereby avoiding various criticisms of such analyses. And analyticities like (4) do not rely on synonomy, meaning postulates, or any other semantic relations eschewed by Quineans. (This is fully in keeping with Chomsky's own defense of analyticity; see Chomsky 1992: 112–16.)

1 Hidden Tautologies

Any talk of analytic truths is likely to arouse suspicions. So let me begin by saying how such talk is to be understood for these purposes. It should be relatively uncontroversial that there are analyticities of the sort I have in mind. The controversial claim will be that (4) is an example.

1.1

All analytic truths are necessary truths, but the converse does not hold. Consider

(5) If John heated the water, then John raised the mean molecular energy of the water.

Following Kripke (1980), I assume that there is no possible situation in which John heated the water without raising its mean molecular energy. Yet (5) is not analytic. Competent speakers of English can fail to know that this conditional is true. Correspondingly, the argument "John heated the water, so John raised the mean molecular energy of the water" strikes us as invalid: failure to draw this inference is not a failure to exhibit logical competence, even though there is no possible situation in which the premiss (as interpreted) is true and the conclusion (as interpreted) is false. If we construe the premiss and conclusion as syntactic objects, though, there are possible interpretations of (the nonlogical elements of) the premiss and conclusion that make the former true and the latter false. Similar remarks apply to "if John boiled some H_2O, then John boiled some water" and "if Hesperus is big, then Phosphorus is big." By contrast,

(6) If John did something and the water boiled, then the water boiled.

is true on any interpretation of its nonlogical lexical items. And the inference from (3) to (2) is valid. As Kripke showed, we must distinguish modal relations between propositions from model-theoretic relations between syntactically individuated linguistic entities. (See also Etchemendy 1990; Hanson 1997.)

As I use the term, (6) is analytic. Intuitively, one can know that (6) is true, by virtue of knowing English – and a little logic, if this is not already part of knowing a language with logical connectives. I do not say that (6) is true by virtue of its meaning, or that (6) is true by convention. If a sentence is true by virtue of anything, it is true by virtue of how the world is. A sentence is true when the world is the way the sentence says it is. The world *is* such that: if John did something and the water boiled, then the water boiled. It takes only rudimentary logic to appreciate that the world is this way (since it must be this way); and any competent speaker of English knows that (6) is true, if the world is this way. (In general, if a sentence has the same truth value on any interpretation of its nonlogical elements, then its truth value is a function of its form and its logical elements; and such a sentence expresses a noncontingent proposition.)

One might complain that my usage of "analytic" is overly inclusive. So let me note that Quine himself distinguished two classes of alleged analytic statements:

> Those of the first class, which may be called *logically true*, are typified by
>
> (1) No unmarried man is married.
>
> The relevant feature of this example is that it is not merely true as it stands, but remains true under any and all reinterpretations of "man" and "married" . . . a logical truth is a statement which is true and remains true under all reinterpretations of its components other than the logical particles. But there is also a second class of analytic statements, typified by
>
> (2) No bachelor is married.
>
> The characteristic of such a statement is that it can be turned into a logical truth by putting synonyms for synonyms . . . We still lack a proper characterization of this second class of analytic statements, and therewith of analyticity generally, inasmuch as we have had in the above description to lean on a notion of "synonymy" which is no less in need of clarification than analyticity itself . . . Our problem, however, is analyticity; and here the major difficulty lies **not in the first class of analytic statements, the logical truths, but rather in the second class, which depends on the notion of synonymy**. (Quine 1953: 22–4, his italics and my bold)

For present purposes, I grant that: there are no second-class analyticities, or at least none that can be characterized without vicious circularity; and this loss need

not be mourned, since no theoretically interesting goals are served by positing these local semantic relations between distinct lexical items.

My interest lies with first-class analyticities, which Quine characterized in model-theoretic terms. I assume that "Someone ran, if John ran" is analytic, although its surface appearance does not make its analyticity manifest. An inference can be truth-preserving on any interpretation of its (nonlogical) lexical elements, even if determining that (and why) this is so requires theoretically interesting work. That is, first-class analyticities can be more or less "hidden," depending on their surface appearances. My claim is that instances of "y boiled, if x boiled y" are hidden first-class analyticities.[2]

1.2

An important motivation for this claim is that "boiled" appears to be one of many verbs "Φ," such that inferences of the form "x Φ y, so y Φ" are compelling. Consider: "melted," "froze," "broke," "opened," "closed," "sank," etc. Not every verb appears in both transitive and intransitive constructions. And if Nora counted the marbles, then Nora counted, while the marbles did not. Still, many instances of "x Φ y, so y Φ" are about as compelling as instances of "P and Q, so Q." When many compelling inferences exhibit a pattern, one suspects that the inferences share a (perhaps hidden) valid form.

Nonetheless, it remains a hypothesis (not a datum) that instances of "x boiled y, so y boiled" are valid. Validity requires that the *same* verb appear in both the transitive and intransitive constructions; else there would be no guarantee that every interpretation (of the nonlogical components) that makes "x boiled y" true also makes "y boiled" true. That is, the sound of "boil" must not be ambiguous as between: "boil_2," which takes two arguments, as in "John boiled_2 the water"; and "boil_1," which takes one argument, as in "the water boiled_1." (On this view, the pattern alluded to above is a surface pattern due to the homophony of distinct verbs.) Yet perhaps "boil" is ambiguous in just this way. Indeed, if the verb in the transitive construction takes more arguments than the verb in the intransitive construction, how can it be the *same* verb in both constructions? Presenting and defending a nonambiguity hypothesis will require the rest of this essay. Let me say briefly, though, why ambiguity hypotheses are unattractive.

If one denies that the same verb appears twice in "x boiled y, so y boiled," one must still account for the compellingness of such inferences. One might reply that "boil_2" and "boil_1" are related by a meaning postulate: if x boiled_2 y, then y boiled_1. But this is to posit a second-class analyticity. Moreover, even those unimpressed by Quine's arguments should be wary of positing so *many* meaning postulates – if x melted_2 y, then y melted_1, etc. – to account for what appears to be a widespread (and arguably productive) linguistic phenomenon. So if the

"hidden first-class analyticity" hypothesis can be defended, it is preferable to the "ambiguity plus meaning postulate" hypothesis.

In reply, one might advance an "ambiguity plus background knowledge" hypothesis. An invalid inference can be compelling, even when the corresponding conditional expresses a necessary truth. Any competent speaker of English who knows about the nature of heat will infer from "John heated the water" to "John raised the mean molecular energy of the water." Of course, competent speakers of English can fail to know facts discovered by scientific investigation. But competent speakers presumably share *some* background knowledge; and perhaps it is a piece of (extralinguistic) common sense that y boiled$_1$ if x boiled$_2$ y. (The parenthetical qualifier is needed to rule out truisms like "Q, if P and Q.")

On this view, the two verbs ("boiled$_2$" and "boiled$_1$") are homophones, but no more semantically related than "smoked" and "burned." Indeed, on this view, the following conditionals are on a par:

(7) If John burned$_2$ the wood, then the wood burned$_1$.
(8) If John burned$_2$ the wood, then the wood smoked.

Herein lies the difficulty. Let us stipulate a third meaning for "burned": y burned$_{1A}$ iff y smoked. Then

(9) If John burned$_2$ the wood, then the wood burned$_{1A}$.

has the same meaning as (8). But *prima facie*, (7) and (9) are not on a par. In (9), the homophony is unrelated to the compellingness of the inference; yet it is hard to believe that this is also true of (7). Correlatively, even if it is metaphysically impossible to burn$_2$ wood that does not burn$_{1A}$, this seems epistemically possible. We can imagine burning$_2$ wood that does not smoke. Yet it seems epistemically impossible to burn$_2$ wood that does not burn$_1$. Moreover, the fact that so many verbs fit this pattern calls for explanation. Why is homophony correlated with epistemic impossibility of the sort just noted? And why is background knowledge correlated with homophony?

Thus, there is considerable motivation for the claim that the same verb ("boil," "burn," "melt," etc.) appears in both transitive and intransitive constructions. But perhaps nonambiguity hypotheses will face even greater difficulties. Moreover, one might claim that the *same* verb appears *in different forms* in transitive and intransitive constructions. Perhaps "boil" can appear as a transitive or intransitive verb, taking two arguments or just one, and inferences of the form "x boiled$_T$ y, so y boiled$_I$" are not valid (absent meaning postulates). For reasons that will emerge below, I think this is an ambiguity hypothesis in disguise. Nonetheless, if one rejects such hypotheses, one owes another explanation for why "boil" can *appear as* a transitive or intransitive verb. Most importantly, those of us who say that

(4) If John boiled the water, then the water boiled.

is valid owe an account of *why* this is so. And serious objections await the attempt to discharge this debt. The question of whether (4) is analytic turns, in large part, on whether these objections can be avoided.[3]

A caveat before proceeding: the view to be defended does *not* entail that the number of lexical primitives is much smaller than the number of surface form words. But I assume that the relevant surface pattern, in which the transitive and intransitive constructions exhibit verbal homophony, is the manifestation of a regularity some of whose instances do *not* exhibit this pattern. Consider

(10) If John raised the pistol, then the pistol rose.

I assume that (10) is an instance of "if x Φ y, then y Φ." There are reasons for treating the phonological difference between "raised" and "rose" as an irregularity, not a sign of distinct verbs. And while

(11) If John killed Abe, then Abe died.

leaves more room for dispute, I treat it on a par with (10), in part because the Old English words for "kill" and "die" were "cwell" and "cwel." (Evidently, "cwel" was replaced with the Scandinavian "die"; Traugott 1972: 75.) But assimilating such cases to the general pattern requires independent argumentation (e.g., appeal to crosslinguistic data); and I am not wedded to any particular examples.

In particular, I do not say that (12–14) are instances of "if x Φ y, then y Φ":

(12) If Nora persuaded Nick to buy a car, then Nick intended to buy a car.
(13) If Nora tripped Nick, then Nick stumbled.
(14) If Nora saw a cat, then Nora saw a feline animal.

If such inferences are valid, the lexicon may contain far fewer primitive items than one might have thought. (Instead, it may contain resources for generating a plethora of surface form words from a significantly smaller stock of primitive items; see Jackendoff 1990, Pustajevsky 1995.) But defending this thesis is not my agenda. The phenonemon that interests me may well be restricted to instances that involve homophony, along with a handful of cases like (10–11) independently motivated on a case-by-case basis. This will seem like a reduction of the lexicon, only to the extent that one expected pairs like "boiled"/"boiled" ("raised"/"rose," "killed"/"died") to be examples of distinct lexical items – i.e., distinct primitive linguistic expressions.

2 Causatives without "Cause"

A traditional idea is that "x boiled y" means that x caused y to boil. And for many purposes,

(15) John caused the water to boil.

is an acceptable paraphrase of

(1) John boiled the water.

One might, however, advance a bolder hypothesis: the two surface forms have the same underlying form – perhaps because (1) is somehow derived from (15); and the inference from (1) to

(2) The water boiled.

is valid, because the inference from (15) to (2) is valid. (See, e.g., McCawley 1968.) In this section, I review some reasons for rejecting this bolder hypothesis in favor of a Davidsonian variation on it.

2.1

Fodor (1970) offers three reasons for not deriving "kill" from "cause to die." These reasons apply, as Fodor and Lepore (1998, 1999) note, to any suggestion that (15) reflects the logical form of (1).

The first objection is that (15) is not a perfect paraphrase of (1). If John sets a house on fire, and as a result some water in the house boils, it seems wrong to say that John boiled the water; but John caused the water to boil. Or suppose John is not an arsonist, but a chef who orders an underling to boil some water. Arguably, John caused the water to boil, but *he* did not boil it. Even if one insists that (1) does follow from (15), an asymmetry remains: why does it *seem* that (15) can be true, while (1) is false? A familiar reply is to introduce a technical term "cause*," and propose that (1) has the same meaning as

(15*) John caused* the water to boil.

Perhaps "cause*" means: caused in a normal way; whereas in the examples above, John's causal relation to the subsequent boiling is abnormal. But for reasons discussed below, I think this proposal is mistaken.

The second objection is that (1) and (15) differ, in the scope of temporal modifiers. Consider

(16) On Tuesday, John boiled the water.
(17) On Tuesday, John caused the water to boil.

Suppose John turned a dial on his stove, creating a flame beneath a pot of water, at 11.59 p.m. on Tuesday; but the water did not boil until after midnight. Then intuitively, (17) is true, while (16) is false. And even if one rejects this intuition, why do (16) and (17) seem to differ, if (15) and (1) are relevantly similar?

The third objection is that (15) creates a potential for ambiguity that (1) does not. Consider

(18) John boiled the water on Tuesday.
(19) John caused the water to boil on Tuesday.

While (18) has a single reading, (19) is ambiguous. (Cf. "John wanted the water to boil on Tuesday.") On one reading, (19) implies that the water boiled on Tuesday; on this reading, (19) would be false in the scenario described above. On the other reading, (19) implies that John's action occurred on Tuesday; on this reading, (19) would be true in the scenario described above. A similar contrast is exhibited by

(20) John boiled the soup without getting burnt.
(21) John caused the soup to boil without getting burnt.

In (21), "without getting burnt" can modify John's action (implying that *he* didn't get burnt) or the subsequent boiling of the soup (implying that *it* didn't get burnt); (20) has only the first reading.[4]

These objections point to important facts about the meaning of (1). But such objections should not convince us that instances of "x boiled y, so y boiled" are invalid. Claiming that (15) reflects the logical form of (1) is *not* the only way to show that instances of "x boiled y, so y boiled" are valid. Indeed, this may be a poor strategy. For it leaves us asking why instances of "x caused y to boil, so y boiled" are valid. Such inferences are truth-preserving, and they may well be valid. But their validity is not manifest. (See note 4 below. The logical form of "x *caused y to boil*" is not obvious; cf. "x turned the dial and y boiled, so y boiled.") In this respect, shifting from (1) to (15) may not help. But comparison with (15) is instructive. Perhaps the underlying syntax of (1), ignoring tense, is: [x [**v** [boil y]]]; where "**v**" is a hidden verb whose meaning has something to do with causation, and "boil" is an intransitive verb that can appear in transitive constructions (along with a hidden verb that takes a second argument). If the underlying syntax of "y boiled" is as it appears, modulo word order, one might

look for a semantics according to which: every interpretation that makes [x [**v** [boil y]]] true also makes [boil y] true.

Moreover, talk of causation suggests that talk of *events* may well be relevant. And as we shall see in the rest of this section, an independently motivated eventish semantics shows why inferences from [x [**v** [boil y]]] to [boil y] would preserve truth. This semantics also provides a plausible interpretation for the hypothesized hidden verb, with the consequence that two of Fodor's three objections are avoided. The more serious third objection will remain, until the eventish semantics is combined with the Chomskian syntax, which provides independent motivation for appeal to hidden verbs.

2.2

At this point, let me briefly review the main arguments for Davidson's (1967) event analysis, and the implications for the kinds of events over which we (tacitly) quantify. In section 2.3, I return to so-called causative constructions.

Inferences like the following appear to be valid:

(22) Booth shot Lincoln with a pistol; so (23) Booth shot Lincoln.
(24) Booth pulled the trigger with his finger; so (25) Booth pulled the trigger.

Similarly, if Booth shot Lincoln with a pistol on April 13, Booth shot Lincoln with a pistol. More generally, if x shot y with z, then x shot y; if x shot y with z on w, then x shot y with z; etc. According to the event analysis, such inferences are valid, because: action sentences are interpreted as existentially quantified constructions; and adjunct phrases (like "with a pistol" and "on April 13") serve as *conjuncts of* an event description. Modifying Davidson's notation slightly, the proposal is that the logical form of (22) is

(22D) $\exists$e[Shoot(e, Booth, Lincoln) & With-a-Pistol(e)]; so (23D) $\exists$e[Shoot(e, Booth, Lincoln)].

where "Φ(e, x, y)" means that e is a Φ-ing of x by y, and compositionality *within* adjunct phrases is ignored for simplicity. The inference from (22) to (23) is said to be an instance of "$\exists$e[Φ(e, x, y) & Ψ(e)], so $\exists$e[Φ(e, x, y)]," instances of which are truth-preserving on any interpretation of their lexical elements. (The logical form of "x Φ y with α on β" is said to be: $\exists$e[Φ(e, x, y) & With-α(e) & On-β(e)].)

Evidence for the event analysis is not confined to entailment patterns. As Parsons (1990) notes,

(26) After the shooting, Booth fled.

is true, iff an event of Booth's fleeing occurred after the shooting. Moreover, consider

(27) Nora heard Fido bark.

which differs from the propositional attitude report "Nora heard that Fido barked": in (27), "bark" is untensed, and substituting coreferential expressions for "Fido" preserves truth. Following Higginbotham (1983) and Vlach (1983), one can render (27) as "there is a hearing by Nora of a barking by Fido":

(27a) ∃e∃f[Hear(e, Nora, f) & Bark(f, Fido)].

This also explains the ambiguity of "Nora heard Fido bark in her apartment," if "in her apartment" can modify the e-position event (the hearing) or the f-position event (the barking).[5]

I assume, therefore, that Davidson's basic proposal is correct. Still, I think it is incomplete. If x shot y, it *follows* that there was a shooting, and that x did something. So notation like "∃e[Shoot(e, x, y)]" may mask further structure, as Castañeda (1967) suggested:

(22a) ∃e[Agent(e, Booth) & Shooting(e) & Theme(e, Lincoln) & With-a-pistol(e)]
(23a) ∃e[Agent(e, Booth) & Shooting(e) & Theme(e, Lincoln)].

The inference from (22a) to (23a) is valid. But on this view, a verb with n arguments is not true of $(n + 1)$-tuples. Instead an event is represented as being of a certain *sort*, and as standing in (binary) thematic relations to each of the verb's arguments. For present purposes, it will suffice to think of Agents as salient initiators of events, and Themes as things saliently affected in the course of events. Parsons (1990) provides motivation for elaborating the event analysis in this way. (For example, there is often no *contradiction* in saying that a thematic role went unfilled. Consider: in my dream, there was a shooting, and I was shot; but no one shot me.) Schein (1993) shows that a plausible semantics of plurals requires, not just the event analysis, but also the "separation" of arguments at logical form; for further arguments, see Herburger (2000), Pietroski (2000b, 2001, forthcoming). So henceforth, I will take this version of the event analysis as given.

The more important point, for present purposes, concerns the kinds of events over which the quantifier ranges. As we shall see, this bears on two of Fodor's three objections. Assume that: Booth shot Lincoln exactly once; he shot Lincoln with a pistol; he pulled the trigger exactly once; and he pulled the trigger with his finger. Then a unique event is the truth-maker for (22a–23a), and similarly for

(24a) ∃e[Agent(e, Booth) & Pulling(e) & Theme(e, the trigger) & With-his-finger(e)].
(25a) ∃e[Agent(e, Booth) & Pulling(e) & Theme(e, the trigger)].

One might think that the shooting *was* the pulling; and Davidson (1971) encourages this thought. But the shooting and the pulling must be distinct events. For consider

(28) Booth shot Lincoln with his finger. (29) Booth pulled the trigger with a pistol.

These sentences are false, and their logical forms are:

(28a) ∃e[Agent(e, Booth) & Shooting(e) & Theme(e, Lincoln) & With-his-finger(e)].
(29a) ∃e[Agent(e, Booth) & Pulling(e) & Theme(e, the trigger) & With-a-pistol(e)].

Let α be the event of Booth's shooting of Lincoln; let β be the event of Booth's pulling the trigger. Since α satisfies the first three conjuncts of (28a), and (28a) is false, α does *not* satisfy "With-his-finger." But β does satisfy "With-his-finger," since (24a) is true. Similarly, β satisfies the first three conjuncts of (29a), which is false; hence, β does *not* satisfy "With-a-pistol." But α does satisfy "With-a-pistol," since (22a) is true. So $\alpha \neq \beta$. The shooting is With-a-pistol, but not With-his-finger; while the pulling is With-his-finger, but not With-a-pistol.

Taylor (1984) and Parsons (1990) mention other examples of this sort in the context of the event analysis; see also Francken and Lombard (1992). Thomson (1971), Thalberg (1972), and others have made similar points, by focusing on the spatiotemporal properties of events. The pulling and the shooting seem to end in different places at different times. The intuition is even clearer with regard to killings. Booth shot Lincoln on April 13, 1865; but Lincoln died on April 14. So the shooting and the killing must be distinct events, since the latter event could not have ended before Lincoln died.

This presents a puzzle, if one wants to say that Booth's action of pulling the trigger *was* his action of shooting Lincoln. How can there be one action, if the pulling and the shooting are distinct events? Elsewhere (Pietroski 1998), I propose an answer. But here, I just want to note that these examples prompt two related thoughts: an action usually has a concertina of effects (see Anscombe 1957; Feinberg 1965); and some events, like wars, have parts that are causally related (see Thomson 1977). For plausibly, one of Booth's actions caused (i) the event of the trigger going back, and thus (ii) the event of the bullet entering Lincoln. Intuitively, the pulling ended with (i), while the shooting ended with (ii). So a

natural thought is that events like pullings and shootings have actions *and certain effects of actions* as parts.

Let us say that event D *grounds* event E, iff: D and E both occur; D is a (perhaps improper) part of E; and D causes every event that is a proper part of E but is not a part of D.[6] Every event grounds itself. But more interestingly, the toppling of one domino can ground the complex event that is the toppling of ten dominoes; and if ten dominoes topple, the toppling of the first domino also grounds the (complex event that is the) toppling of the first five dominoes. Thus, a single event can be the grounder of distinct complex events. If pullings and shootings are complex events, the event of Booth's pulling the trigger can be distinguished from the event of Booth's shooting Lincoln. The latter event had parts that began after the trigger went back. Let us say that event F *terminates* event E, iff: F and E occur; F is a (perhaps improper) part of E; and F is an effect of every event that is a proper part of E but is not a part of F. Then for each n, the toppling of the nth domino terminates the toppling of the first n dominoes. And crucially, *different* events terminate the shooting and the pulling.

Appeal to complex events lets us see how (22) and (28) can differ in truth value. The event of Booth's shooting Lincoln is *longer* than the event of Booth's pulling the trigger; and only the longer event is With-a-pistol. Similarly, (24) and (29) can differ in truth value, if only the shorter event is With-his-finger. Thus, I claim that action sentences quantify over events grounded by actions, as indicated in figure 8.1. (For these purposes, I take no stand on whether paradigmatic actions are: bodily motions; causes of bodily motions; or complex events with bodily motions as parts. See Pietroski 1998, 2000a.)

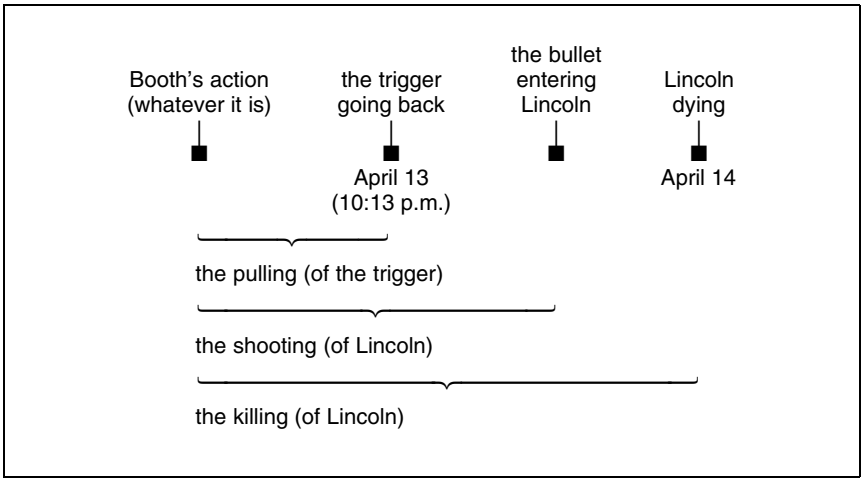


Figure 8.1

2.3

Let me now return to

(1) John boiled the water; so (2) the water boiled.

The event analysis does not rule out ambiguity hypotheses. One can say the logical forms of (1–2) are: $\exists$e[Agent(e, John) & Boiling_2(e) & Theme(e, the water)]; and $\exists$e[Boiling_1(e) & Theme(e, the water)].[7] One can say that "boil_2" is true of certain complex events grounded by actions, that "boil_1" is true of certain events in which a liquid undergoes a change of state, and that (as a matter of fact) a boiling_2 always ends in a boiling_1. But on this view, the inference above is invalid, absent a meaning postulate: $\forall$y{$\exists$e[Boiling_2(e) & Theme(e, y)] $\rightarrow$ $\exists$e[Boiling_1(e) & Theme(e, y)]}. And the earlier objections to ambiguity hypotheses remain. On the other hand, if "x boiled y" contains a hidden verb – i.e., [x [**v** [boil y]]] – then perhaps the Davidsonian logical forms of (1–2) are:

(1a) $\exists$e$\exists$f[Agent(e, John) & R(e, f) & Boiling(f) & Theme(f, the water)]
(2a) $\exists$f[Boiling(f) & Theme(f, the water)]

where "R" is a metalanguage predicate whose interpretation is yet to be given. The idea is that the hidden verb expresses a binary relation between events. And *whatever* relation this is, the inference from (1a) to (2a) is an instance of the valid form "$\exists$e$\exists$f[Φ(e, x) & Ψ(e, f) & Ω(f) & Π(f, y)], so $\exists$f[Ω(f) & Π(f, y)]."

That is, the interpretation of the hypothesized hidden verb is irrelevant to the validity of the inference. Nonetheless, the meaning of (1) tells us that relation R has something to do with causation. So the simplest hypothesis would be that R just *is* the relation of causation. This is Parsons's (1990) view, according to which (1) is true, iff:

(1-P) $\exists$e$\exists$f[Agent(e, John) & Cause(e, f) & Boiling(f) & Theme(f, the water)].

But (1-P) is true, while (1) is false, if John's act of arson caused the boiling of the water; and similarly for other examples of so-called deviant causal chains. Or consider

(16) On Tuesday, John boiled the water.
(16-P) $\exists$e$\exists$f[On-Tuesday(e) & Agent(e, John) & Cause(e, f) & Boiling(f) & Theme(f, the water)].

If the water did not boil until Wednesday, then (16-P) can be true while (16) is false.

These objections can be avoided, however, by saying that R is the relation *terminates in*: R(e, f), iff f terminates e. Recall that f terminates e, iff: f and e occur; f is a part of e; and f is an effect of every event that is a proper part of e but is not a part of f. Replacing (1-P) with

(1a) ∃e∃f[Agent(e, John) & Terminates-in(e, f) & Boiling(f) & Theme(f, the water)].

lets us say that the e-position events are *complex* events of which the f-position events are parts. In which case, e-position events do not cause f-position events. Rather, (1) is true, iff John is the Agent of a *complex event that terminates in* a boiling of the water. John's act of arson may not make him the Agent of any such complex event. For there may not *be* any event (grounded by John's action) that has as parts both John's act of arson *and* the subsequent boiling of the water. The mere fact that C caused E does not ensure the existence of a complex event C* that has C and E as parts. And if there is no event (grounded by John's action) that has the boiling of the water as a part, then John did not boil the water, even though John's action caused the boiling of the water.[8] Similarly, if the water did not boil until Wednesday, (16) is false. For there was no complex event e, such that: e occurred (entirely) on Tuesday, John was the agent of e, and e terminated in a boiling of the water.

This minor revision of Parson's semantics is independently motivated, since the event analysis demands complex events (which terminate in other events) anyway. So Fodor's first two objections do not refute the Davidsonian analysis that treats "x boiled y" as *importantly like* "x caused y to boil." Still, the third objection remains. Moreover, there are two bothersome features of (1a): John and the water are not represented as Agent and Theme of the *same* event; and it is not clear how these truth conditions can be determined *compositionally*. To appreciate the force of this latter point, consider two possible implementations of the current proposal, focusing on the interpretations of the verbal components (figure 8.2).

In each case, "boil" is an event sortal; the hidden verb is true of ordered pairs of events; and the complex verb phrase [**v** [boil y]] is true of events that terminate

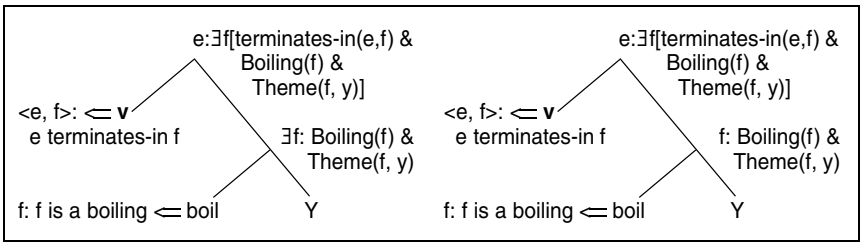


Figure 8.2

in a boiling of y. The difference concerns the interpretation of [boil y]: is it sentential (as indicated on the left), like "the water boiled"; or is it subsentential, like "boiled on Tuesday"? The event analysis treatment of "x heard y boil" would seem to require the former hypothesis, which is also supported by Fregean considerations about the relation of truth to function/argument structure (and unsaturated/saturated components of thought). But if truth values are the semantic values of sentences, the interpretation on the left will not satisfy a familiar compositionality constraint: the semantic value of each node is to be a function of the semantic values of its immediate constituents. For no suitably general function will map *truth values* and the extension of "terminates-in" onto events that terminate in a boiling of y. (Compare the interpretation on the right, which is compositional.) Perhaps another plausible choice for sentential semantic values will avoid this problem; though I doubt it. Or perhaps we should weaken the compositionality constraint, requiring only that the semantic value of each node be a function of its constituents; but this should be a last resort. Moreover, one would like to preserve an idea suggested by other eventish analyses: the interpretation of subsentential structure is as simple as conjunction; see Pietroski (2002). So revision is needed somewhere.

3 Minimal Syntax

Thus far, my perspective has been overtly semantic. I have spoken of inferences, truth conditions, and intuitions of possibility. In this section, I focus on some syntactic details, in responding to Fodor's third objection (and the worry about compositionality just voiced). As Fodor notes, one expects constructions with two verbs to be potentially ambiguous when adjuncts are employed. Recall

(19) John caused the water to boil on Tuesday.

Or consider the verb phrase "heard Fido bark in her apartment," which has the readings shown in figure 8.3. But "x boiled y on Tuesday" is *not* ambiguous as

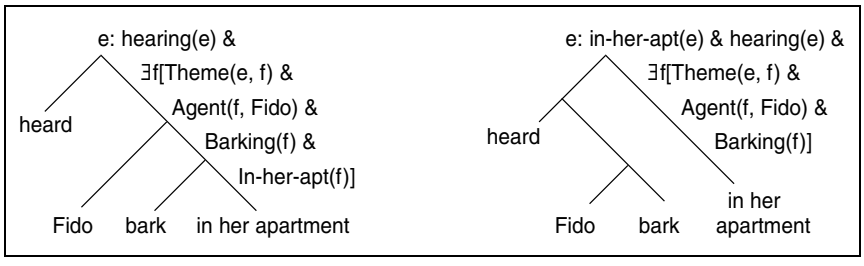


Figure 8.3

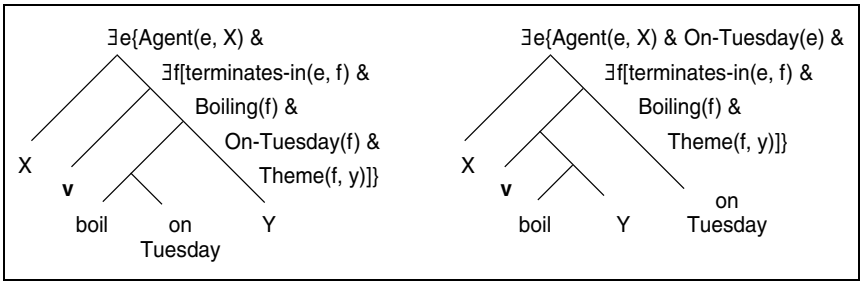


Figure 8.4

between the readings in figure 8.4. The adjunct phrase must modify the e-position event, as indicated on the right. As we shall see, one can provide principled reasons for excluding the downstairs adjunction structure indicated on the left. Still, if the underlying syntax of "x boiled y" is either [x [boil y]] or [x [**v** [boil y]]], the nonambiguity of

(18) John boiled the water on Tuesday.

might be a point in favor of the former analysis, which also preserves the appearance that (18) contains a transitive verb that takes two arguments. But Chomsky (1995a) gives us reason for preferring the hidden-verb analysis, once we ask where transitive structure comes from, and why it has the character it does.

3.1

One might claim that the ambiguity of (19) is irrelevant, once one abandons the idea of deriving (18) from (19). For one might say that the overt verb "cause" differs from the hidden verb, in that the latter provides a *mandatory* target for adjuncts. But as it stands, this is not a satisfactory reply to Fodor's objection; it merely describes the facts as they must be described, *if* (18) contains a hidden verb.

A more promising line of thought is developed by Baker (1988) and others. The tense marker appears on the matrix verb in "x caused y to boil" and "x heard y bark." So perhaps "x boiled y" is the result of a *transformation on* the basic structure: x [**v** [boil y]]. Suppose that "boil" moves and combines with the hidden verb, thereby forming a *unit* to which a tense marker can be attached: x [[**v**-boil$_i$][t$_i$y]]; where [**v**-boil$_i$] is treated *as* a unit – in effect, a terminal node of the tree. (This would satisfy a familiar principle governing transformations: a trace must be dominated by the node immediately dominating the antecedent of the trace.) Baker provides extensive support for the idea that one semantically

significant piece of language can come to be "inside" another. Indeed, he argues that verb incorporation is overt in many languages. Consider his example from Chichewa (pp. 10–11, agreement and aspect markers ignored):

Mtsuko	u-na-gwa-a			
The waterpot	past fall			(The waterpot fell)
Mtsikana	u-na-u-gw-ets-a		mtsuko	
The girl	past	fall causative	the waterpot	(The girl made the waterpot fall)

The latter construction results from the transformation indicated in figure 8.5, where small "v" indicates a functional item that combines with an ordinary lexical item (like "fall") to form a new verb. (I assume that the new unit keeps the label of the item adjoined to.) This construction can be interpreted as: the girl was the agent of an event that terminated in a falling of the waterpot.

If "boil" incorporates with a hidden causative-forming verb ("#"), then the structures to consider with respect to (18) are indicated in figure 8.6 (ignoring the order of "#" and "boil"). And the incorporation hypothesis suggests a potential reason for why the structure on the left does *not* reflect a possible reading of (18): perhaps the "downstairs adjunct" would block the movement of "boil"; for other potential reasons and related discussion, see Pietroski and Uriagereka (2002), Pietroski (forthcoming).

It is a familiar idea in linguistics that too much structure can block movement. And it is not implausible that the overt verb can move to the hidden verb only in the absence of intervening nodes – or alternatively, that there cannot be nodes between the incorporated verb and the overt verb's trace. But finding *independent* evidence for this suggestion will be hard, since Fodor and Lepore's alternative claim is that morphemes matter. They will object that the Baker-inspired proposal *presupposes* that English is relevantly like languages with overt causative elements: the downstairs adjunct (allegedly) blocks a transformation that is

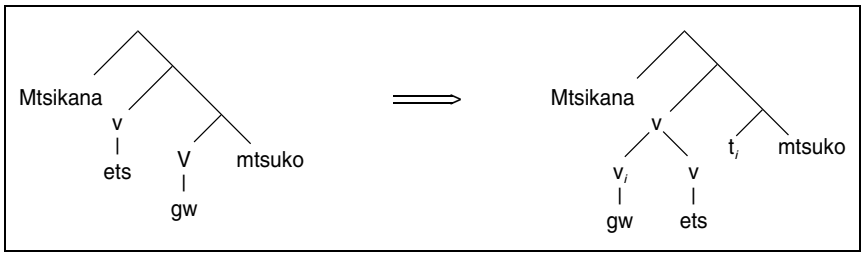


Figure 8.5

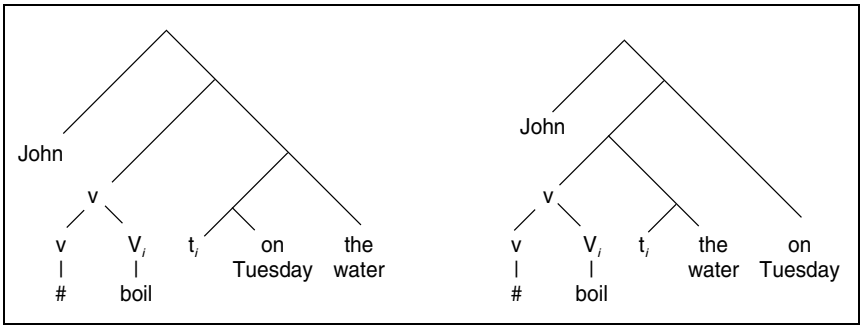


Figure 8.6

(allegedly) required by the (alleged) presence of a hidden verb. Nonetheless, we can now see at least one way in which syntactic constraints might rule out the nonexistent reading of (18), even if the hidden-verb hypothesis is correct; hence, the nonambiguity of (18) does not refute this hypothesis. Fodor and Lepore suggest that the nonambiguity of (18) still tells against the presence of a hidden verb, absent an independently confirmed account of the nonambiguity. But even if one grants this methodological point, one must also consider the *prima facie* implausibilities of treating "boil" as ambiguous. Moreover, crosslinguistic data can bear on the issue indirectly.

For example, Baker (1988, 1997) defends the Uniformity of Thematic Assignment Hypothesis (UTAH), according to which grammatical relations determine thematic roles (and vice versa); and Pestesky (1995) argues, in defending a slightly weaker thesis (viz., that Agents are always represented in a "higher" grammatical position than Themes), that English has an unpronounced causative morpheme. A potential problem for UTAH is that "be angry" appears to permit representations of the same event participants in different grammatical positions, as in:

(30) Bill was angry at the article.
(31) The article angered Bill.

But perhaps "Bill" is the subject of "be angry" in each sentence; while "the article" is an object in (30), and the subject of a causative verb (to which "angry" adjoins) in (31). The idea is that the underlying syntax is as follows: [Bill [be-angry [(at) [the article]]]]; and [the article [[make-angry$_i$][Bill t$_i$]]].

Pesetsky notes (citing Kuroda, in Akatsuka 1976) that the hypothesized transformation seems explicit in Japanese, as indicated below (where "ga" and "o" are markers for accusative and nominative case):

Tanaka-ga	sono	sirase-o	yorokon-da	(Tanaka was pleased at that news)
Tanaka	that	news	be pleased-past	
Sono sirase-ga	Tanaka-o		yoronkob-ase-ta	(That news pleased Tanaka)
That news	Tanaka		be pleased-causative-past	

Pesetsky then shows that a hidden English causative morpheme would explain a range of facts, including the distribution of nominalizations like "annoyance." And he draws an analogy with Latinate roots like "ceive" and "fer" that occur only with certain prefixes – like "re," "in," "per," and "pre" – suggesting that the causative morpheme can attach to stems that do not occur alone as pronounced verbs.

Yet even if appeal to hidden verbs and incorporation in English is not ad hoc, there is something unsatisfying about saying that downstairs adjuncts would block movement of overt verbs. It can seem that this "explanation" merely describes a salient feature of the structure representing the reading that (18) does not have – but should have, according to Fodor and Lepore, on a hidden-verb hypothesis. So even if we discover a correlation between nonexistent readings and barriers to movement (that overt English verbs would need to make given hidden verbs), Fodor and Lepore can deny that appeal to hidden verbs has been motivated on syntactic grounds. And even if the semantic motivation suffices, given ways of ruling out structures that would carry nonexistent readings, one would like to do better.

3.2

If downstairs adjunction blocks movement, one wants to know *why* it does so. Moreover, I have assumed that incorporation is *required*, given a causative-forming verb. But what is wrong with the structure: x [**v** [boiled y]]? (The tense consideration is only suggestive; perhaps "x caused y to boil" and "x heard y bark" differ, precisely because the matrix verbs are overt.) Most importantly, is there any reason to think that natural languages do *not* generate structures like: [x [boil y]]; where "boil" is a *primitive* transitive verb? At this stage of the dialectic, Chomsky (1995a) has much to offer.

The minimalist program has many facets; and I will not try to summarize it. But let me stress that Chomsky's agenda is *not* to abandon the "principles and parameters" view, according to which: the child faces a poverty of stimulus problem that is solved because the child knows (or cognizes) a set of linguistic principles that are universal (for humans) but parameterized; the input the child receives, in combination with universal grammar, suffices to determine the parameter settings; and a language (i.e., an I-language) is a possible setting of the various parameters, in conjunction with a lexicon (see Chomsky 1981, 1986).

Minimalists endorse this basic account, but subject its many posits (e.g., S-structure) to a kind of methodological scrutiny akin to Occam's razor. As an ideal, Chomsky aims for the simplest possible theory that posits only what is required (by "virtual conceptual necessity"), given a certain cluster of linguistic phenomena whose existence is beyond reasonable doubt.

The basic linguistic phenomena include: that linguistic expressions can have (arbitrarily many) constituents; that linguistic structures undergo transformations, in which constituents are displaced from their original positions; and that the resulting structures somehow interface with cognitive systems not exclusively concerned with language. With regard to this last point, each sentence must be able to interface with at least – and so, given minimalist ideals, at most – an articulatory/perceptual (A/P) system that relates linguistic structures to sounds (or signs); and a conceptual/intentional (C/I) system that relates linguistic structures to meanings. Chomsky thus endorses the ancient idea that sentences are structures by virtue of which sounds are paired with meanings.

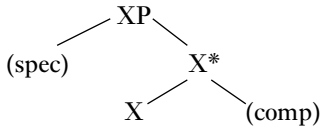
Indeed, a sentence is said to be a pair of instructions – called "PF" and "LF" – for the A/P and C/I systems. If these instructions are to be usable by the extralinguistic systems, PFs and LFs may have to respect constraints that would be arbitrary from a purely linguistic perspective. Chomsky speculates that herein lies the rationale for transformations: basic linguistic structures are being altered to meet demands imposed from outside. (Since we know little about the C/I system, there is little to be said about how it might constrain LFs; though one might view work on the relation between grammatical and thematic categories in this light. See Baker 1997 for discussion.) But whatever the rationale for transformations, minimalists encode the fact that transformations occur as follows: linguistic items are said to have formal (i.e., uninterpretable) features, which must be deleted before interfacing with other systems; and deletion requires displacement to a site at which the feature is "checked off."

This is a compellingly simple conception of natural language. For these purposes, I assume that Chomsky and others have rendered it empirically respectable. So if the minimalist conception favors the hidden-verb hypothesis (as supplemented by Baker), this is important support for that hypothesis – and the corresponding claims about hidden (first-class) analyticities. Indeed, Chomsky provides a framework that lets us see analyticities as spandrels of natural language architecture. There are two related lines of thought here, discussed in sections 3.3 and 3.4 below: a minimalist treatment of constituency suggests that transitive structure is *not* the manifestation of a single verb that takes two arguments; and a minimalist account of movement lets us say why the overt verb must move, in a way that downstairs adjunction would block, to a place that allows for a compositional eventish semantics.

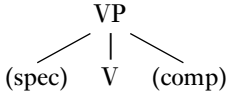
3.3

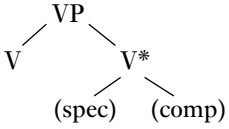
Perhaps the most basic linguistic phenomenon is that one expression can be *part* of another. Chomsky thus assumes a lexicon and an operation called "Merge" for combining expressions. For at a minimum, it must be possible to combine a determiner like "the" with a noun like "water," combine the result with a verb like "boil," etc. Moreover, complex expressions are interpreted (at the interfaces) as being of certain types; "the water that the man boiled" is interpreted as a nominative, although it contains a verb. So each result of Merge must be labelled, presumably taking its label from one of the Merged items. One can think of Merge as a recursive process for establishing a transitive relation C(ommand): C<"the,""water">, and this instance of C is labeled "DP"; C<"boil," DP>, which is labeled "VP"; etc. Or in more familiar notation: $[_{VP}\ [_{V}$boil$][_{DP}\ [_{D}$ the$][_{N}$ water$]]]$. Merge plays a significant role in the minimalist program, which thereby takes on a derivational flavor. Instead of talking about bad-making features of trees and barriers to movement, one emphasizes constraints on derivations, which are said to "crash" if they do not satisfy certain constraints (imposed by lexical items, formal features, and so-called economy considerations that I will not discuss here).

Appealing to Merge also lets one eliminate any separate theory of the structures in which lexical items can appear. It has long been common to think of lexical items, and especially verbs, as projecting structures of the following form (see Jackendoff 1977):



where the specifier and complement positions can be filled by projections of other lexical items, and the intermediate category X* is the site for adjunction. Such structures may be plausible for transitive verb phrases, and clauses like "the student of physics." But it seems gratuitous to suppose that *all* lexical items are surrounded by two further positions. (Correspondingly, the status of X* has always been less clear than that of lexical and phrasal categories.) Moreover, if verbs project sentential frames, one wants to know *why* these frames have the character indicated above, as opposed to:





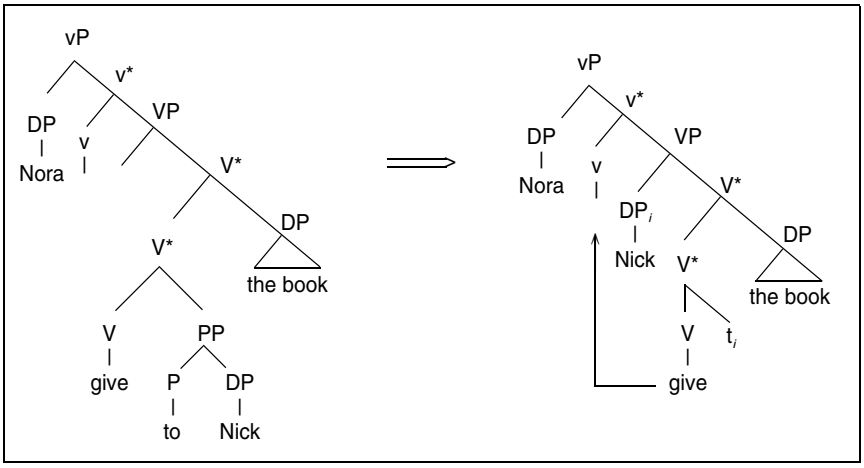


Figure 8.7

These structures reflect different command hierarchies, and no asymmetry between the so-called external (spec) and internal (comp) argument positions. But there are such asymmetries (see, e.g., Williams 1995); and there are reasons for thinking of the external argument as the argument of the unit formed by combining the verb with its internal argument (see Marantz 1984). One would like explanations for these facts, as opposed to the mere stipulation that this is how natural language structure is.

Worse, the posited frames do not provide enough room for all the arguments of a verb like "give." And Chomsky uses Larson's (1988) treatment of so-called double-object constructions as a model for how the structure associated with an (n + 1)-place verb can emerge from the structure associated with an n-place verb. According to Larson, "Nora gave Nick the book" results from a transformation based on a double verb-phrase structure in which: "give" is the *lower* verb; the higher verb position is *empty*; and "Nick" is the object of a prepositional phrase (see figure 8.7). The idea is that "give" moves into the higher verb position, while "Nick" raises to the specifier position of the lower verb-phrase, and the preposition is absorbed in the process. Larson's account has a host of welcome empirical consequences, in addition to the theoretical virtue of preserving binary branching. But let me stress: on this view, the higher verb position is not occupied by a hidden verb that takes "Nora" as its argument; the claim is rather that, given the constraints imposed by a certain theory of phrase structure, the outer vP shell is required to accommodate the third argument associated with "give." (I say "associated with," since Larson offers reasons – citing Marantz 1984 – for treating "the book" as the argument of the complex unit V*, and *not* literally as an argument of "give.")

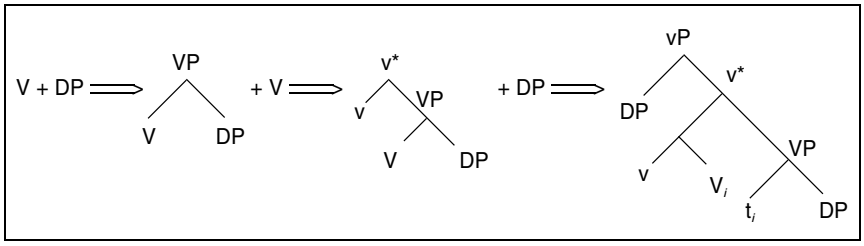


Figure 8.8

From a minimalist perspective, however, one does not take any theory of phrase structure as given. The structure on the left must arise through the operation Merge; so the higher verbal position must be occupied by a (hidden) lexical verb, which presumably does take an argument. Instead of saying that a verb which takes three arguments moves into an empty position, one is led to say that a verb which takes two arguments adjoins to (and incorporates with) something akin to an overt causative morpheme. But if Larsonian shells offer an independently motivated way of seeing how ternary predicate structure can arise from a syntactic process to which no ternary predicates are inputs, then as Chomsky notes, a natural suggestion is that transitive structure emerges the same way (see figure 8.8).

On this view, every *lexical* verb takes one argument; and the argument of a small verb "v" is a verb phrase, the head of which adjoins to "v." Causative constructions provide a model for this process, in which complex structure emerges from a simple operation and simple lexical items. But the claim is *not* that all overt verbs can be decomposed into small verbs and overt (nonpassive) one-place verbs. Recall Pestesky's analogy to the Latinate roots "ceive" and "fer." So we may have to use surface form transitive verbs, in order to give the meaning of some lexical verbs. Perhaps "hit" results from combining a small verb with a lexical verb that is true of events f, such that if x is the agent of an event that terminates in f, then x is the agent of a hitting (cf. Dowty 1979; Lombard 1986).

Chomsky does not explicitly apply his generalization of Larson's account to double-object constructions. But presumably, they would arise from two rounds of merger and incorporation. A verb-phrase headed by a one-place verb ("G") merges with a small verb ("%"), forming a unit that takes a second argument, followed by merger with the small verb "#" to form a unit that takes a third argument (see figure 8.9). One wants to know what "%" and "G" *mean*. But a plausible thought is that "G" is true of transfers (of an appropriate sort). So suppose that $[_{VP}$ G $[_{DP}$ Nick]] is true, iff: $\exists$g[Transfer(g) & Goal(g, Nick)]. The argument position made available by "%" is that of Theme. So the truth-conditional contribution of "%" is presumably some relation that holds between events (of which the book is the Theme) and transfers (of which Nick is the Goal). Perhaps the relation is *constitution*: $[_{v*}$ $[_{v*}$ %-$G_i[_{VP}t_i$Nick]] $[_{DP}$ the book]] is

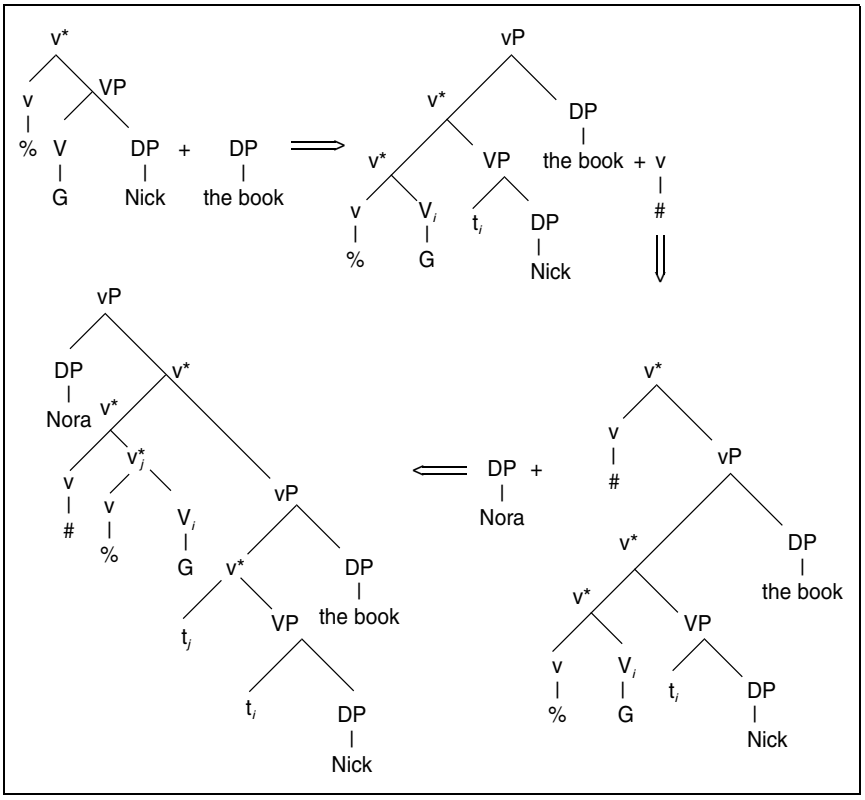


Figure 8.9

true of events that constitute a transfer to Nick of the book. If this is correct, Nora gave Nick the book iff: Nora was the Agent of an event e that terminated in an event f, such that the Theme of f was the book and f constituted an appropriate transfer of its Theme to Nick.[9]

One great advantage of this proposal is that it keeps the lexicon and the operation of Merge pleasingly simple. If one denies that there is a one-argument correlate of "hit," one must suppose that *from a purely syntactic perspective* there are at least two different kinds of verbs: those that merge with an argument, forming a unit (V*) that can – and perhaps must – merge with a second argument; and those that merge with an argument, forming a unit (VP) that cannot merge with another argument. But not only does this diverge from minimalist ideals, by positing linguistic distinctions not needed to explain the phenomena, it effectively reintroduces the idea that transitive verbs create transitive sentential frames. If one allows this, one has no principled reason for not positing a third kind of verb: it merges with an argument, forming a unit (V**) that merges with a second

argument, thereby forming a unit (V*) that merges with a third argument. Thus, one cannot link the phenomena explained by Larson's account of double-object constructions to the basic architecture of human languages. This is a theoretically regressive line of thought that minimalists avoid.

Another, less theory-internal motivation is that Chomsky's proposal suggests an explanation of Burzio's generalization (see Burzio 1986). It seems to be a general fact about human languages that verbs without external arguments fail to assign accusative case; correlatively, verbs that do not assign accusative case do not assign an external thematic role. Consider the Italian verbs "telefonare" and "arrivare," both of which appear with a single argument (like "telephone" and "arrive" in English). Only "telefonare" assigns accusative case, and only with "arrivare" is the single argument the *complement* of the verb. This is evidenced by a variety of linguistic phenomena, like the distributions of auxiliary verbs and the clitic "ne". But for these purposes, I simply assume that Burzio's generalization is correct, and follow Raposo and Uriagereka (1996) in noting that: this generalization presumably follows from more basic features of natural language; and minimalism provides an attractive explanation, if accusative case assignment depends on the presence of a small verb. This requires that we treat some (surface) intransitive constructions as the products of underlying structure involving a second verb. But this is a small price for a unified treatment of causatives and Burzio's generalization.

Moreover, it is a familiar point that many causative verbs appear in a second entailment pattern involving an adjectival form. If Nora opened the door, the door opened, and so (at some point in time) the door was open. (If x Φ_T y, then y Φ_I, and so y was Φ_{Adj}.) Similarly, if Tom broke the vase, the vase broke, and so the vase was broken; if Booth killed Lincoln, Lincoln died, and so Lincoln was dead; etc. So an attractive suggestion is that "Nora opened the door" results from the derivation shown in figure 8.10. If we assume that the adjective "open"

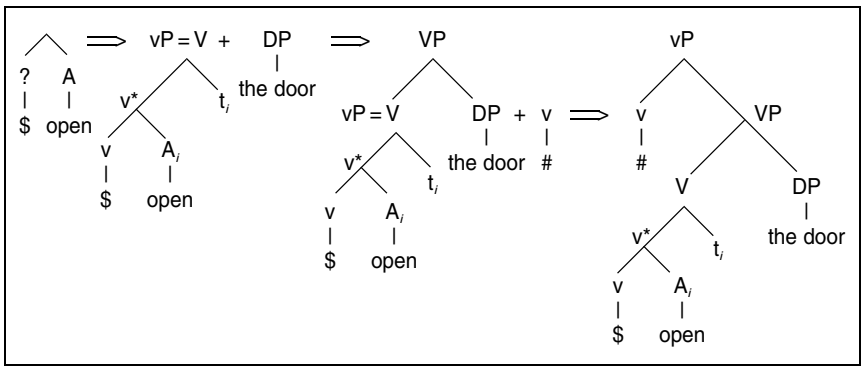


Figure 8.10

is true of *states*, perhaps there is a hidden lexical item "$" that expresses a function from states to *changes* of states (i.e., events). If some (surface form) intransitive verbs have the structure indicated above, they would be true of events in which something came to be open. The corresponding (surface form) transitive verbs would be true of events that terminate in events in which something comes to be open. And note that verbs like "hit" and "shoot" do have adjectival forms – as in "I am $\text{shot}_{\text{Adj}}/\text{hit}_{\text{Adj}}$," which is distinct from the passive "I was shot/hit by someone."[10]

3.4

I have been stressing a line of minimalist thought, based on the treatment of constituency, that supports appeal to hidden verbs. Let me turn briefly to a second line of minimalist thought, based on the treatment of transformations, that may explain why the overt verb in a causative structure must adjoin to the hidden verb (and why downstairs adjunction is incompatible with this).

Recall that minimalists encode the existence of transformations by saying that some lexical items have features that are "attracted" by (or perhaps to) other lexical items. But the deeper explanation for movement is said to lie with constraints imposed by the auditory/perceptual and conceptual/intentional systems. Perhaps the former system is relevant in the case at hand.[11] But we have already seen a sense in which [**v** [boil y]] may be semantically uninterpretable. At the end of section 2, I noted that this structure could not be interpreted in a (strongly) compositional fashion, given a (standard) sentential semantic value for [y boil]. So perhaps [**v** [boil y]] must be transformed, because the C/I system interprets linguistic structures in a compositional manner. The Bakeresque structure below (figure 8.11) *can* be interpreted compositionally, in accordance with a simple eventish semantics, as indicated. I have indicated no interpretation for the trace of "boil." For on a minimalist view, traces simply reflect derivational

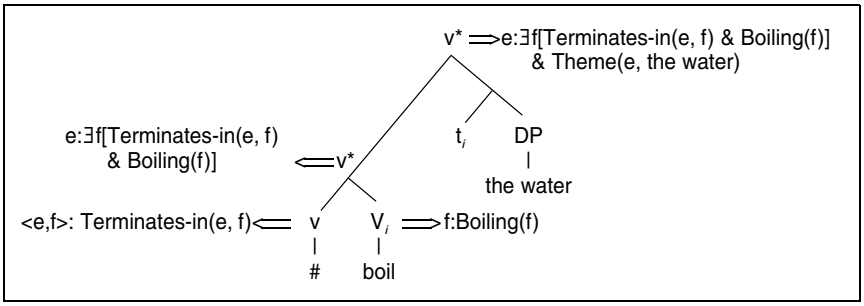


Figure 8.11

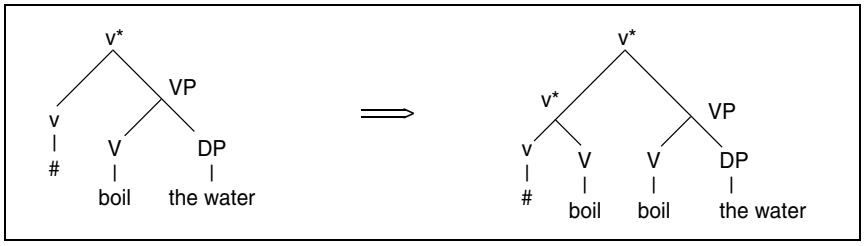


Figure 8.12

history. If the rationale for transformations is to ensure interpretability at the interfaces, then expressions are to be interpreted *where they end up*. (See Hornstein 1995 for defense.) One should *not* think of pre-transformation results of Merge as "deep structures" that determine the interpretation of sentences, with displacement occurring for semantically extraneous reasons. And if the trace of "boil" is ignored, "the water" is in the familiar theme (comp) position of the incorporated verb "#-boil."

This idea of "ignoring" traces is best understood in the context of the minimalist mechanism for transformation: Copy and Delete. The first operation is required unless expressions literally *move* expressions from one site to another; and literal movement, if intelligible, would be very complex. So in addition to Merge, minimalists posit an operation – Copy – that can modify a tree as in figure 8.12. But leaving expressions in their original places is unlikely to improve interpretability at the interfaces. On the contrary, introducing second copies may create new problems of interpretability. We know that both copies are not pronounced. So minimalists posit an operation – Delete – that has the desired effect (see figure 8.13). The subscript and trace notation reflect the history of the derivation. Originally, "boil" was merged with "the water." But the structure on the right does not contain a shadow of "boil" in the embedded VP. Indeed, the label "VP" ceases to have meaning (except as a reflection of derivational history), since its only constituent is the determiner phrase. And if

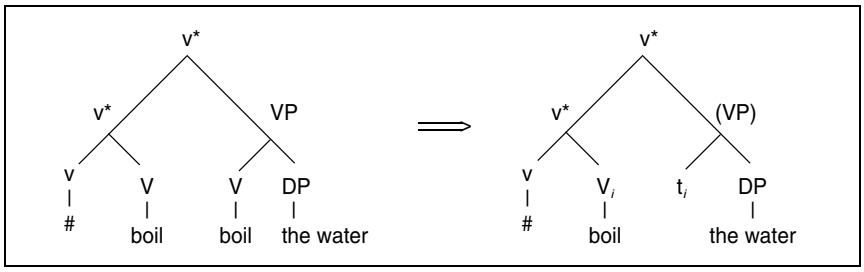


Figure 8.13

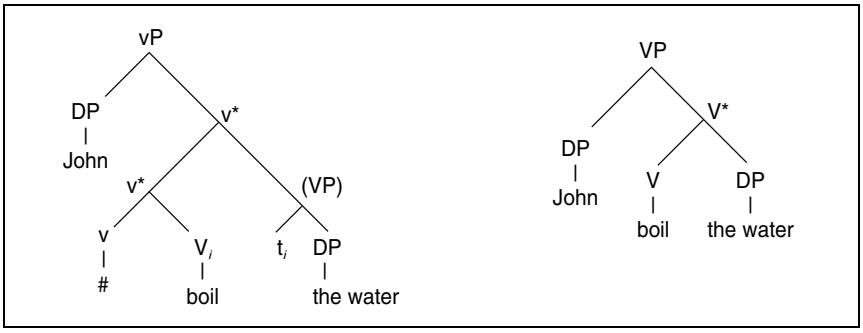


Figure 8.14

the incorporated verb is treated as a terminal node of the tree – see Uriagereka 1999 on the relation between adjuncts and "frozen units" in a minimal syntax – the result of transformation will mimic a transitive sentential frame (see figure 8.14).

So it is no surprise, on a minimalist account, that "x boiled y" has a transitive character. Indeed, it is not wrong to say that "boil" has a transitive form; and one can write "x boiled$_T$ y" and "y boiled$_I$." But the intransitive form is basic, in that "boiled$_T$ y" is short for: $[_{v^*}\ [_{v^*}\ \#\text{-boil}_i][_{(VP)}\ t_i[_{DP}\ y]]]$; where "boil" (i.e., "boil$_I$") takes a single argument. Thus, it would be a mistake not to see "boil" in "boil$_T$ y," just as it would be a mistake not to see "boil" in "boil on Tuesday." One can represent "y boiled" and "y boiled on Tuesday" as "βy" and "$\beta^* y$." But then one does not represent the inference from the latter to the former as valid; so one will be tempted to add the meaning postulate "$\forall y(\beta^* y \rightarrow \beta y)$." But this merely reports a fact, instead of explaining it, which suggests that "βy" and "$\beta^* y$" are theoretically inadequate representations of the natural language sentences. Similarly, I suggest, for "x boiled$_T$ y" and "y boiled$_I$."

A minimalist account of transformation also suggests that there would be something wrong with figure 8.15. One might wonder whether the original token of "boil" *can* delete. (The structure $[_{V^*}\ [\ ][_{AP}$ on Tuesday]] seems to depend on its presence.) But if it does delete, the embedded VP is still left with two major constituents, "on Tuesday" and "the water"; hence, the VP node cannot be ignored, even for purposes of interpretation. So "the water" is *not* in the complement position of the incorporated verb. This is a problem, on the assumption that every argument (DP) must be assigned a thematic role. For as Chomsky notes, a plausible thought is that thematic roles are assigned by virtue of structural relations that obtain between verbs and arguments: head-comp for Theme; head-spec for Agent. For this reason, one might suggest that "boil" originally merges with "the water," and that thematic relations are assigned in merger. In that case, the "downstairs adjunction" structure would be as in figure

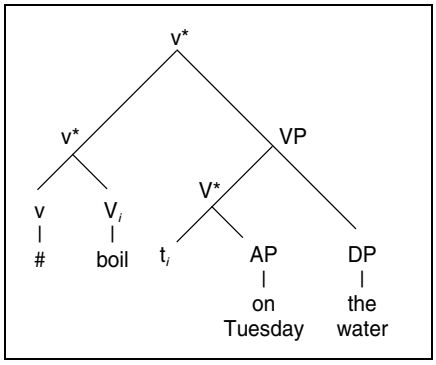


Figure 8.15

8.16. But even if "the water" is treated as a Theme, the meaning of the whole structure is not yet determined.

If Fodor's nonambiguity objection is sound, the higher "v*" should be true of events e, such that: ∃f[Terminates-in(e, f) & Boiling(f) & Theme(f, the water) & On-Tuesday(f)]. For this is the (relevant) nonexistent reading of "boiled the water on Tuesday." But why suppose this interpretation, given the indicated semantic values for the lower "v*" and "VP"? If any eventish interpretation is assigned to the higher "v*," one would expect it to be characterized by conjunction (or intersection). That is, one would expect the structure above to be true of those events e, such that: ∃f[Terminates-in(e, f) & Boiling(f)] & Theme(e, the water) & On-Tuesday(e). So even if this structure is generable – by Merge, Copy, and Delete – there is no reason to think it can bear an interpretation that would make "John boiled the water on Tuesday" ambiguous. Moreover, Chomsky (1995a: 329–34) argues against the possibility of the structure above on independent grounds; and Larson (1988) argues that adjuncts sit

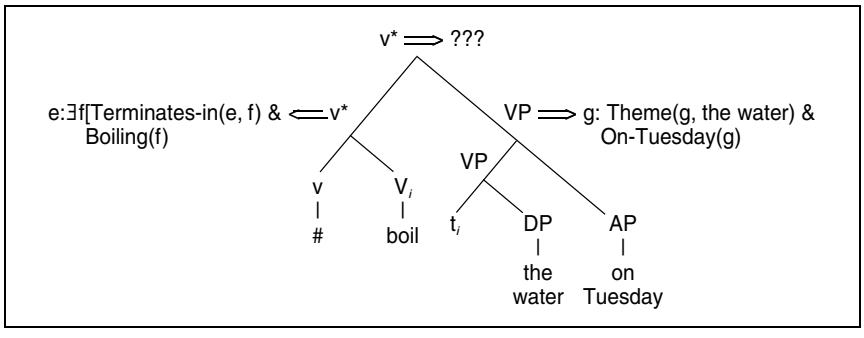


Figure 8.16

closer to verbs than arguments. (This last point coheres with the idea that "VP" reflects *saturation* of the verb – and so a *sentential* semantic value, absent subsequent movement.)

3.5

If the water is represented as Theme of the e-position event, which terminates in a boiling, one might wonder where that leaves us with respect to the validity of inferences like:

(1) John boiled the water; so (2) the water boiled.

For consider: ∃e{Agent(e, John) & ∃f[Terminates-in(e, f) & Boiling(f)] & Theme(e, the water)}; so ∃f[Boiling(f) & Theme(f, the water)]. This inference is not valid. First-order logic does not guarantee that the Theme of an event that ends in a boiling is the theme of that very boiling. But this does not show that the *natural* language inference is invalid. It shows that the representation in our standard predicate logic fails to capture a semantically significant aspect of the *transformation* that is part of the *derivation* (via Merge, Copy, and Delete) of the transitive construction "John boiled the water." It is a familiar point that the inference "John ran, so someone ran" does not look valid if rendered as "P, so Q." We know that the propositional calculus is an impoverished tool for reflecting the semantically relevant structure of natural language inferences; and we are used to the idea that (first-order) predicate logic fails to capture the validity of at least some inferences.[12]

Correspondingly, the model-theoretic notion of validity is not tied to the resources of any particular formal system. It can be applied to structures like those in figure 8.17. If the proposal developed here is correct, linguistic structures

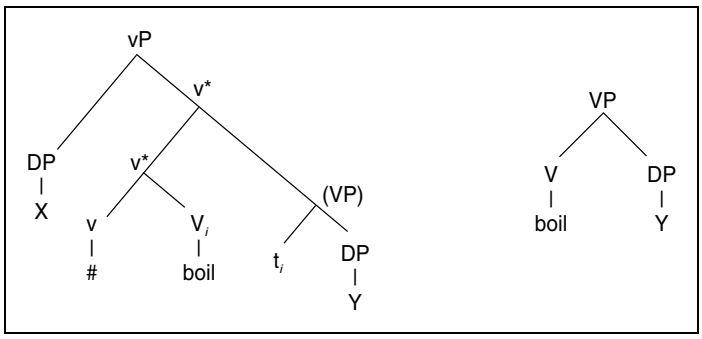


Figure 8.17

are assigned interpretations (in accordance with an eventish semantics), so that any interpretation that makes the structure on the left true also makes the structure on the right true. Given the semantic contributions of the various syntactic modes of combination – i.e., given the natural language equivalents of concatenation in predicate logic – the interpretation of the higher "v*" on the left is (compositionally) determined by the interpretations of: "#," "boil," and "Y." Given the semantic contribution of merging a verb with an argument, the interpretation of "VP" on the right is (compositionally) determined by the interpretations of: "boil" and "Y." In this sense, any interpretation of the structure on the left determines the interpretation of the structure on the right. For the set of lexical inputs to the process that generates the structure on the left – what Chomsky calls the "numeration" of the structure – *is a subset of* the set of lexical inputs to the process that generates the structure on the right. Given the semantic architecture of natural languages, this determination is also truth-preserving. And this is so, even though the incorporated verb is treated as a unit (and thus *like* a lexical item) *after* displacement of the overt verb. So *pace* Fodor and Lepore, the transitive character of "x boiled y" does not give us reason to treat "boil" as lexically ambiguous, as between "boil_{T}" and "boil_{I}" – at least not if the lexicon consists of semantically primitive items that (together with the semantically relevant syntactic modes of combination) determine the interpretations of all linguistic structures. On the contrary, if minimalist thinking is correct, the transitive character of "x boiled y" is best explained on the hypothesis that "boil" (like all verbs) takes a single argument.

This raises larger issues, about the relation of logical form to linguistic structures, that I leave for another occasion; see Pietroski 2002. But I hope to have illustrated that, in conjunction with the event analysis, minimalism offers an attractive framework for thinking about the relation of form to meaning. If all goes well, this program may even lead to productive hypotheses about the relation of human language to human thought. Indeed, Chomsky's talk of an LF-C/I interface suggests a way of pursuing the traditional analytic project in philosophy – that of investigating thought by investigating language – in a naturalistic mode that does not divorce philosophy from the rest of human knowledge. For perhaps we can use our best theories of syntax to frame, or at least constrain, hypotheses about what LFs interface with. (For example, one might speculate that the LF-C/I interface is trivial, and that expressions in the "language of thought" are structurally isomorphic to the LFs of natural language.) From this perspective, alleged analyticities will play their traditional role, as potential sources of evidence for claims about the structure of thought. Thus, the minimalist program offers the intriguing prospect (long hinted at by Chomsky) of using linguistic theory as a tool for investigating various aspects of human cognition.[13]

Chomsky's reply: pp. 304–7.

Notes

1 This paper, written in 1998, has descendants that have preceded it (in particular, Pietroski 2002); and Pietroski (forthcoming), while maintaining the basic line of thought, modifies some of the details. But for obvious reasons, I have not revised the paper, apart from updating some references.

2 At the end of his paper, Quine suggests that even logical truths occupy no special place in our web of belief: there is no principled distinction between (5) and (6); we can rationally revise our attitude toward either conditional, since both are confirmed in the same (global and empirical) way. But I know of no good argument for this claim. As Rey (1998) discusses, one must take care to avoid fallacious inferences from: fallibilism to the revisability of logic; our capacity to revise how we express thoughts, to the revisability of what we think; and the claim that (beliefs represented by) logical truths can be justified by holistic/empirical methods, to the claim that logical truths can be justified *only* by such methods.

3 It is an empirical question whether (4) is analytic, and the corresponding knowledge is *a priori*. This should be unsurprising, absent the dogma that analyticity is transparent – i.e., that competent speakers can just "tell" that analyticities are such. As used here, "analytic" is a technical term introduced in the course of explaining phenomena that appear to be manifestations of linguistic competence. Compare "grammatical," which is introduced in the course of explaining why speakers judge certain strings to be (un)acceptable.

4 The latter example is closer to Fodor's: "x killed y by swallowing his tongue," which lacks a reading had by "x caused y to die by swallowing his tongue." Such cases are complicated by the empty subject of the "by"-clause: x caused y to die by *e* swallowing his tongue; perhaps "*e*" cannot be coreferential with "y," even if the hidden-verb hypothesis is true. So I will focus on (18–19). Or consider "John caused the water to boil rapidly/twice," which seems to be ambiguous in a way that "John boiled the water rapidly/twice" is not.

5 Note that "x caused y to boil" differs from "x heard y bark": *"x caused y boil"; *"x heard y to bark." If "y to boil" specifies a fact, and not an event, then accounting for the validity of "x caused y to boil, so y boiled" may be a complicated matter. (Cf. "x correctly believed y to be a thief, so y was a thief.")

6 Let me stress: D grounds E, only if D and E are existing (occurring) events; and it is not my aim to provide an independent metaphysics of events. In particular, I am *not* saying that whenever one event causes another, there exists a third event that has the first two events as parts. On the contrary, I am inclined to reject such fusion theses. (See note 8 below.) In Pietroski (2002), I speak of one event *initiating* another, with "Initiater" and "Terminater" being treated as labels for relations between events (while "Agent" and "Theme" are labels for relations between events and participants).

7 I assume that the water is the theme of the boiling, since it is saliently affected (and not an actor). This is even more plausible, if "y boiled" results from movement, and the argument is the verb's logical *object*. (See Burzio 1986; Baker 1988: 46–7.) But one wants to know why "There boiled some water" and "There sank three ships" are less acceptable than similar expletive constructions like "There seems to be a mistake" and "There arrived three men." (Cf. Belletti 1988; Haegeman 1994: 331–7.) In this

chapter, I do not discuss views according to which the meaning of the intransitive verb is specified in terms of the transitive verb; see, e.g., Levin and Rappaport (1995).

8 By hypothesis, the alleged complex event would not be a *boiling*; and if it would not satisfy any ordinary event sortal, one cannot simply assume that the alleged event exists. Nor can one assume that if E and F are events, then the fusion of E and F is an event. Like Hornsby (1986), I am inclined to reject this thesis. (Is there an event whose parts include the falling of a tree and subsequent changes in the auditory system of some squirrel? Cf. Thomson 1977.) But perhaps fusion theorists can restrict my proposed thesis about Agents, along the following lines: α is the agent of E, if α performs an action that grounds E, *and* E satisfies some nonfusion condition for being a "genuine" event. If some fusions of events are not (genuine) events, one might wonder why not. But it is not the job of a *semantic* theory to tell us which particulars exist (occur); and "terminates-in" expresses a relation between existing events. Note that appeal to complex events accounts for the one-way entailment, from "x boiled y" to "y boiled," without replacing "cause" with a notion whose extension is a *subset* of the extension of "cause" (i.e., "cause in the right way"). Correspondingly, one need not take on the (futile) task of specifying some condition M, such that x boiled y, iff: x performed an action that caused a boiling of y, and M. Semantic decomposition does not require analysis in this sense, *pace* Fodor and Lepore.

9 See Larson 1988 and Baker 1997 on the "thematic hierarchy" <Agent <Theme <Goal>>>. One might wonder if small verbs have *any* substantial meaning of their own, or whether they simply provide a mechanism by which thematic roles are assigned in a minimalist syntax. I do not address this issue.

10 Let me mention one of Hale and Keyser's (1993) cases, addressed by Fodor and Lepore (1999). If Nora saddled the horse, Nora is the Agent of a complex event that terminated in a certain change of state – namely, a change in which the saddle came to be (appropriately) on the horse. So one might appeal to a functional item that transforms nouns into adjectives: $[_{AP}\ [_{A}[_{?}$ (en)$][_{N}$ saddle$][_{DP}$ the horse$]]]$, with the verb phrase being derived as described in the text. (This suggestion differs, in the details, from Hale and Keyser's own.) Verbs like "count," however, pose a difficulty for minimalists: if Nora counted the children, then *Nora* counted; and Nora is the *Agent* of the counting. Chomsky endorses Hale and Keyser's suggestion that an apparently intransitive verb is really a hidden transitive, if its overt argument represents the role of Agent. One might take this as endorsement of the following analysis of "X counted": $X[_{V^*}\ [_{v^*}$ #-count$_i][_{VP}\ t_i$ (Y)$]]$; where "(Y)" is unvoiced. But this structure would presumably be interpreted as meaning that X counted *something*; and one can count without counting anything (cf. "sing" and "dream"). So perhaps "(Y)" is deleted prior to interpretation at the interface. Or maybe the underlying form of "X counted" is: $X[_{v^*}\ [_{v^*}$ #-count$_i][_{V}\ t_i]]$; where this is the result of originally merging "count" with the small verb "#" (which is associated with the role of Agent) and it is an idiosyncratic property of "count" that it can merge directly with "#." In any case, a small verb analysis is motivated by the minimalist explanation of Burzio's generalization, if the subject in "Nora counted" is assigned accusative case. (Compare "There arrived/*counted three men.")

11 In languages with overt causative morphemes, raising may be required (at PF) for pronounceability; even in English, the difference between "raised" and "rose" is suggestive. On a minimalist view, LFs and PFs share some derivational history; and

they may diverge *after* the overt verb has moved, with such movement reflected at LF, even though the movement was for the benefit of the AP system.

12 See Pietroski (forthcoming) for related discussion and defense, via Tenny (1994), of the following claim: Theme(e, x) & Terminates(e, f) → Theme(f, x).

13 For helpful comments and discussion, my thanks to: Mark Baker, Susan Dwyer, Jerry Fodor, Norbert Hornstein, Ernie Lepore, Jim McGilvray, Juan Uriagereka, audiences at Carleton University and the University of Maryland.

References

Akatsuka, N. 1976: Reflexivization: A Transformational Approach. In M. Shibatani (ed.), *Syntax and Semantics*, 6: *The Grammar of Causative Constructions*, New York: Academic Press.

Anscombe, G. E. M. 1957: *Intention*. Oxford: Blackwell.

Baker, M. 1988: *Incorporation*. Chicago: University of Chicago Press.

Baker, M. 1997: Thematic Roles and Syntactic Structure. In L. Haegeman (ed.), *Elements of Grammar*, Dordrecht: Kluwer.

Belletti, A. 1988: The Case of Unaccusatives. *Linguistic Inquiry*, 19, 1–34.

Burzio, L. 1986: *Italian Syntax*. Dordrecht: Reidel.

Castañeda, H. 1967: Comments. In N. Rescher (ed.), *The Logic of Decision and Action*, Pittsburgh: University of Pittsburgh Press.

Chomsky, N. 1972: *Studies on Semantics in Generative Grammar*. The Hague: Mouton.

Chomsky, N. 1981: *Lectures on Government and Binding*. Dordrecht: Foris.

Chomsky, N. 1992: Language and Interpretation. In J. Earman (ed.), *Explanations and Other Frustrations: Essays in the Philosophy of Science*, Berkeley: University of California Press.

Chomsky, N. 1995a: *The Minimalist Program*. Cambridge, Mass.: MIT Press.

Chomsky, N. 1995b: Language and Nature. *Mind*, 104, 1–61.

Davidson, D. 1967: The Logical Form of Action Sentences. In N. Rescher (ed.), *The Logic of Decision and Action*, Pittsburgh: University of Pittsburgh Press.

Davidson, D. 1971: Agency. In R. Binkley et al. (eds.), *Agent, Action, and Reason*, Dordrecht: Kluwer.

Dowty, D. 1979: *Word Meaning and Montague Grammar*. Boston: Reidel.

Etchemendy, J. 1990: *The Concept of Logical Consequence*. Cambridge, Mass.: Harvard University Press.

Feinberg, H. 1965: Action and Responsibility. In M. Black (ed.), *Philosophy in America*, Ithaca, NY: Cornell University Press.

Fodor, J. 1970: Three Reasons for not Deriving "Kill" from "Cause to Die." *Linguistic Inquiry*, 1, 429–38.

Fodor, J. and Lepore, E. 1998: The Emptiness of the Lexicon. *Linguistic Inquiry*, 29, 269–88.

Fodor, J. and Lepore, E. 1999: Impossible Words? *Linguistic Inquiry*, 30, 445–53.

Francken, P. and Lombard, L. 1992: How not to Flip the Switch with the Floodlight: Causative-Inchoatives, The Instrumental "With," and the Identity of Actions. *Pacific Philosophical Quarterly*, 73, 31–43.

Haegeman, L. 1994: *Introduction to Government and Binding Theory*, 2nd edn. Cambridge, Mass.: Blackwell.

Hale, K. and Keyser, S. J. 1993: On Argument Structure and the Lexical Expression of Syntactic Relations. In K. Hale and S. J. Keyser (eds.), *The View From Building 20*, Cambridge, Mass.: MIT Press.

Hanson, W. 1997: The Concept of Logical Consequence. *Philosophical Review*, 106, 309–64.

Hart, H. and Honoré, A. 1959: *Causation and the Law*. Oxford: Oxford University Press.

Herburger, E. 2000: *What Counts.* Cambridge, Mass.: MIT Press.

Higginbotham, J. 1983: The Logical Form of Perceptual Reports. *Journal of Philosophy*, 80, 100–27.

Hornsby, J. 1980: *Actions.* London: Routledge & Kegan Paul.

Hornsby, J. 1986: Physicalism, Events, and Part-Whole Relations. In Lepore and McLaughlin 1986.

Hornstein, N. 1995: *Logical Form: From GB to Minimalism.* Oxford: Blackwell.

Jackendoff, R. 1977: *X-bar Syntax*. Cambridge, Mass.: MIT Press.

Jackendoff, R. 1990: *Semantic Structures.* Cambridge, Mass.: MIT Press.

Kripke, S. 1980: *Naming and Necessity*. Cambridge, Mass.: Harvard University Press.

Larson, R. 1988: On the Double Object Construction. *Linguistic Inquiry*, 19, 335–91.

Lepore, E. and McLaughlin, B. (eds.) 1986: *Actions and Events: Perspectives on the Philosophy of Donald Davidson.* Oxford: Blackwell.

Levin, B. and Rappaport, M. 1995: *Unaccusativity: At the Syntax-Semantics Interface.* Cambridge, Mass.: MIT Press.

Lombard, L. 1986: How not to Flip the Prowler. In Lepore and McLaughlin 1986.

Marantz, A. 1984: *On the Nature of Grammatical Relations.* Cambridge, Mass.: MIT Press.

McCawley, J. 1968: Lexical Insertion in a Transformational Grammar without Deep Structure. *Papers from the Fourth Regional Meeting, Chicago Linguistics Society*. Linguistics Department, University of Chicago, 71–80.

O'Shaughnessy, B. 1974: Trying as the Mental "Pineal Gland." *Journal of Philosophy*, 71, 365–86.

Parsons, T. 1990: *Events in the Semantics of English.* Cambridge, Mass.: MIT Press.

Pesetsky, D. 1995: *Zero Syntax.* Cambridge, Mass.: MIT Press.

Pietroski, P. 1994: A Defense of Derangement. *Canadian Journal of Philosophy*, 24, 95–118.

Pietroski, P. 1998: Actions, Adverbs, and Agency. *Mind*, 107, 73–112.

Pietroski, P. 2000a: *Causing Actions.* Oxford: Oxford University Press.

Pietroski, P. 2000b: On Explaining That. *Journal of Philosophy*, 97, 655–62.

Pietroski, P. 2002: Function and Concatenation. In G. Preyer and G. Peters (eds.), *Logical Form and Language*, Oxford: Oxford University Press.

Pietroski, P. forthcoming: *Events and Semantic Architecture*. Oxford: Oxford University Press.

Pietroski, P. and Uriagereka, J. 2002: Dimensions of Natural Language. In Juan Uriagereka, *Derivations: Exploring the Dynamics of Syntax*, London: Routledge.

Pustajevsky, J. 1995: *The Generative Lexicon*. Cambridge, Mass.: MIT Press.

Quine, W. V. O. 1953: Two Dogmas of Empiricism. In *From a Logical Point of View*, Cambridge, Mass.: Harvard University Press.

Raposo, E. and Uriagareka, J. 1996: Indefinite *Se*. *Natural Language and Linguistic Theory*, 14, 749–810.
Rey, G. 1998: A Naturalistic A Priori. *Philosophical Studies*, 92, 25–43.
Schein, B. 1993: *Plurals.* Cambridge, Mass.: MIT Press.
Taylor, B. 1984: *Modes of Occurrence*. Oxford: Blackwell.
Tenny, C. 1994: *Aspectual Roles and the Syntax-Semantics Interface*. Dordrecht: Kluwer.
Thalberg, I. 1972: *Enigmas of Agency: Studies in the Philosophy of Human Action.* New York: Humanities Press.
Thomson, J. 1971: The Time of a Killing. *Journal of Philosophy*, 68, 115–32.
Thomson, J. 1977: *Actions and Other Events.* Ithaca, NY: Cornell University Press.
Traugott, E. 1972: *A History of English Syntax*. New York: Holt, Rinehart, and Winston.
Uriagereka, J. 1999: Multiple Spell-Out. In S. Epstein and N. Hornstein (eds.), *Working Minimalism*. Cambridge, Mass.: MIT Press.
Vendler, Z. 1967: *Linguistics in Philosophy*. Ithaca, NY: Cornell University Press.
Vlach, F. 1983: On Situation Semantics for Perception. *Synthèse*, 54, 129–52.
Williams, E. 1995: *Thematic Structure in Syntax*. Cambridge, Mass.: MIT Press.

9
In Defense of Public Language

RUTH GARRETT MILLIKAN

a notion of "common, public language" that remains mysterious . . . useless for any form of theoretical explanation . . . There is simply no way of making sense of this prong of the externalist theory of meaning and language, as far as I can see, or of any of the work in theory of meaning and philosophy of language that relies on such notions, a statement that is intended to cut rather a large swath.

Chomsky 1995: 48–9

It is a striking fact that despite the constant reliance on some notion of "community language" or "abstract language," there is virtually no attempt to explain what it might be.

Chomsky 1993: 39

either we must deprive the notion communication of all significance, or else we must reject the view that the purpose of language is communication . . . It is difficult to say what "the purpose" of language is, except, perhaps, the expression of thought, a rather empty formulation. The functions of language are various.

Chomsky 1980: 230

I have yet to see a formulation that makes any sense of the position that "the essence of language is communication."

Chomsky 1980: 80 (see also 1992b: 215)

At frequent intervals over the years, Professor Chomsky has inveighed against both commonsense and technical notions of public language or "externalized language," claiming that they are confused, ill-defined, or of no scientific interest. As a scientist, he would be interested in public language only if it were a "real object of the real world" (Chomsky 1993: 39) rather than an "artifactual" and "arbitrary" notion (Chomsky 1985: 26). I propose to articulate such a notion of public language for him. Chomsky has also denounced the notion that the purpose of language is communication. I will argue, on the contrary, that a primary function of the human language faculty is to support linguistic conventions, and that these have an essentially communicative function.

Despite this bold display of sounded disagreements with Chomsky, however, I agree with his objections to common views of public language completely: "People who live near the Dutch border can communicate quite well with people living on the German side, but they speak different languages in accordance with the sense of the term Dummett argues is 'fundamental' . . ." (Chomsky 1992a: 101). Public language is a sprawling mass of crisscrossing, overlapping conventions, some known to some people, others to others. It does not divide into discrete portions, German versus Dutch versus French.

> What we say is that the child or foreigner has a "partial knowledge of English," or is "on his or her way" toward acquiring knowledge of English, and that if they reach the goal, they will then know English. Whether or not a coherent account can be given of this aspect of the common sense terminology, it does not seem to be one that has any role in an eventual science of language. (Chomsky 1985: 16)

> If Peter is improving his Italian or Gianni is learning his . . . [w]e gain no insight into what they are doing by supposing there is a fixed entity that they are approaching, even if some sense can be made of this mysterious notion. (Chomsky 1992b: 16–17)

Languages are not governed by discrete sets of public norms or rules of mysterious origin. There are no "fixed entities" which a child slowly approaches when learning language. On both the above points I entirely agree with Chomsky and entirely disagree, for example, with Dummett.

> This "externalized language" that Jones and Smith share must be an abstract object of some sort, a property of the community, perhaps . . . Suppose that Smith and Jones have more or less the same shape; we do not conclude that there is a shape that they partially share, and the interactions between Smith and Jones give us no more grounds to suppose that there is a language that they share. (Chomsky 1993: 39–40)

Further, public language is not just some property shared by Smith and Jones. It is not discovered, for instance, by averaging over the idiolects of people in "the community." I agree with all these points against the existence of what some conceive as "public language." I also agree with Chomsky's arguments for the importance of the study of I-language and the language faculty. And of course I agree that language is put to many different purposes. What, then, am I complaining about?

I will argue that there remains a legitimate way of looking at language as a public object, and also a legitimate way of looking at language function, that Chomsky has not taken into account. Learning language is not merely acquiring an "I-language." It is not just achieving a relatively steady state of the language faculty. Learning language is essentially coming to know various *public* conven-

tions and, with trivial exceptions, these conventions are around to learn only because they have functions.

What has brought Chomsky to deny these rather mundane truths, hence to embrace an unnecessarily extreme position on public language, may have been, in part, bad philosophical company. In particular, there are, I believe, two dominant but mistaken traditions in philosophy that probably played a role. One of these concerns the nature of language conventions. The other concerns the nature of language functions.

> A "convention" in Lewis's sense, is a regularity "in action or in action and belief" sustained by the belief that others conform to the regularity. Note that this is a rather restricted sense of the term "convention." There are, no doubt, conventional aspects of language: for example, the fact that one says "Hello" in answering a telephone or calls a chair "a chair." But . . . regularities in action and belief are quite restricted, at least if we insist that "regularities" have detectable probabilities; there is little reason to suppose that aspects of language that are commonly called "conventional" involve detectable regularities. (Chomsky 1980: 81, referring to Lewis 1969)

Chomsky is completely right about this. The conventions of language are not regularities, either *de facto* or *de jure*. They are not expressed in actual uniform behaviors, nor in people's beliefs about uniform behaviors, nor in rules or norms prescribing uniform behaviors. But it is exactly because conventions have been taken, mistakenly, to be regularities (Searle 1969; Lewis 1969, 1975; Schiffer 1972; Bach and Harnish 1979; Gilbert 1983, 1989/1992; Recanati 1987)[1] – and because language is so obviously conventional – that it has been thought that languages must belong to circumscribed social groups. You can't have a regularity unless there is a reference class for the regularity to be in. So there can't be a convention unless it is a convention for somebody – for the Jews but perhaps not for the Christians, for the Germans but perhaps not the French, for you but perhaps not for me. Hence the notion of a "'common, public language' that remains [so] mysterious," of a "'community language' or 'abstract language'," a "shared" language, "a property of the community." Chomsky sees the error in this way of thinking. But the philosophical tradition has supplied no alternative to this way of understanding conventions, hence no alternative reason to think the study of public conventions might be important in the study of language. I will try to remedy this situation by outlining a theory of language conventions that does not take them to be regularities, and then showing why this sort of public conventionality is of the very essence of human language. I know nothing about the language faculties of angels or Martians. They may come with all necessary language wired in so that they do not need to have language conventions. But in humans, I will argue, a pivotal job of the language faculty is to make language conventions possible, and the functions of language conventions are communicative functions.

The second philosophical tradition to which I will propose an alternative concerns the nature of language functions.

> What does it mean to say that language has [communication as] an "essential purpose"? Suppose that in the quiet of my study I think about a problem, using language, and even write down what I think. Suppose that someone speaks honestly, merely out of a sense of integrity, fully aware that his audience will refuse to comprehend or even consider what he is saying. Consider informal conversation conducted for the sole purpose of maintaining casual friendly relations, with no particular concern as to its content. Are these examples of "communication"? If so, what do we mean by "communication" in the absence of an audience, or with an audience assumed to be completely unresponsive, or with no intention to convey information or modify belief or attitude? (Chomsky 1980: 130)

Chomsky follows the dominant philosophical tradition in assuming that language functions would have to reduce to or be derived from speaker intentions. In fairness, the only alternative that mainstream philosophy has offered is to derive them somehow from language conventions, language conventions being understood, as above, to involve some obscure kind of regularity within some nebulous group. Chomsky is right to reject this latter alternative, but there is no need to accept the former. I will offer a third alternative that I hope he will find more palatable. But given this third alternative, it will become clear that the functions of the overwhelming majority of conventional language forms is to enable various kinds of communication.

The argument will be, then, that a central function of the language faculty in humans is to make language conventions possible, and that the functions of conventions are to make communication possible. In this case, I take it, "has as its function to" is a transitive relation. At the same time, I will argue that the web of conventions that forms the mass that is public language is not an abstract object but a concrete set of speaker–hearer interactions forming lineages roughly in the biological sense. These lineages and their interactions with one another are worthy of scientific study. Nor are their properties derivative merely from the properties of I-languages.

Both the thesis on conventions and the thesis on functions to be applied here have been fully stated and defended in other places.[2] Here I can only sketch results and try to ward off the most obvious objections. First, then, some words about conventions.

As Chomsky remarks, it is conventional to say "Hello" when you answer the phone. Moreover, nearly everyone, say, in America, does so. But it is conventional in exactly the same way to call a spigot "a spigot" and also to call it "a faucet" and also "a tap," and clearly there can be no group in which everyone does all three of these things regularly. And of course that is not what David Lewis had in mind when he claimed that language conventions exist only where there are regular behaviors. He didn't mean that spigots are always called

"spigots." What he meant was that if someone calls something "a spigot," then it is regular that that thing is a spigot, which looks quite a lot more plausible. Troubles with regularity theories show up not here but, first, with the problem how to delimit the relevant groups within which the regularities supposedly occur. Within what designatable group is the convention to hand out cigars on the birth of a boy regularly followed other than the group that in fact hands out cigars on the birth of a boy? Within what group is the convention that by saying "break a leg" one wishes another luck regularly followed other than the group that in fact uses "break a leg" to wish people luck? Troubles show up with the regularity theory, second, where people use language in ways that do not accord with convention, for example, when they use metaphors, sarcasm and other figures of speech, and when they lie, or make false promises, or demand things they know they won't get, or as Chomsky suggested (though in another context), when they sincerely assert what they know won't be believed, or when they talk only to themselves. There is a voluminous literature on such cases in speech act theory, which tries vainly to solve an insoluble problem. It tries to describe regularities involved over all such uses – very subtle regularities to be sure, usually taken to occur on the level of multiply embedded beliefs that speakers supposedly invariably intend to impart. Since Chomsky apparently agrees that this sort of problem is insoluble, I won't argue the case here.[3] I will simply start in and describe a kind of convention that has no connection with regularities, that is not a convention "for" any particular group. And I will try to show that this is a way in which natural human language is conventional.[4]

What is conventional in this sense is an activity or a pattern of activities. Exhibition of the pattern may require only one participant or it may involve more. Conventional ballroom dance steps, for example, are patterns involving two participants whereas the pattern that is playing ring a ring o' roses generally requires more. To become conventional, an activity or pattern of activity must, first, be reproduced, hence proliferated. (Please do not blanch. "Reproduced" will be used here such that language forms can be reproduced by a language faculty imposing a universal grammar.) Further, it must be proliferated due in part to the weight of tradition, rather than due, for example, to its intrinsically superior capacity to perform some function, or due to ignorance of any superior way to perform it. I will discuss each of these characteristics of conventions in turn.

First let me explain "reproduced." I will use this term in a limited way. A reproduction must be such that had the original been different in specifiable respects the reproduction would have differed accordingly. Inherited traits and behaviors are not "reproduced" in this sense. I have blue eyes not directly because my mother and/or father had, but because of my genes, which were copied from their genes, which were not, however, copied from their eyes. Had my mother's or father's eyes been yellowed from jaundice, that would not have caused my eyes to be yellow.

A reproduction is always a reproduction only in certain respects. The repro-

duction that comes out of a monochrome copying machine is a reproduction only with respect to pattern of light and dark. The background color and the paper texture are not reproduced, as these depend on the color and texture of the paper put in the paper feeder, not on the color and texture of the original. The color of the pattern is not reproduced because it depends on the color of the ink in the machine, not on the color of the original. Any object can be reproduced in an indefinite variety of ways, since any object has an indefinite number of properties. Nothing is a reproduction of anything else in all respects. Because properties that are reproduced can be relational properties as well as intrinsic properties, and since everything has an infinite number of relational properties, it is not possible to exhaust the variety of reproductions that might, logically, be made of an object. Reproductions that copy relations may be unlike their originals in very striking respects. For example, a painter who reproduces "the style" of an earlier painter may paint quite different subject matters and use quite different media, while another painter may reproduce this same painter's "style" by painting similar subject matters only. The products of both painters will be "reproductions" as I am using that term. Further, since an object can have many parts and many aspects or properties, it can be a reproduction of a number of different things all at once, borrowing some features from here, others from there, and so forth.

For one thing to be "reproduced" from another, all that is required is that there be a mechanism that produces the second on the model of the first, such that, had the first been different in specifiable respects, that would have caused the second to differ accordingly. Under this description, reproducing can of course occur unconsciously. For example, there can be conventional ways of talking or of moving the body, differing for men and women, or specific distances from one another at which it is conventional to stand when talking, that are handed down by being unconsciously reproduced. Also, reproducing need not be direct. If Mother tells Johnny to shake hands when being introduced, rather than, say, telling him to kiss on both cheeks or to sniff noses, granted that she tells him this only because other folk behave or have behaved in this way, then Johnny's behavior, when he does what he has been told, is indirectly reproduced. Had certain other people been greeting one another in other ways, Johnny's way of greeting would have differed accordingly as a result. Also, if Jim taps a nut to fit Jon's bolt, then Jan makes a new bolt to fit Jim's nut, on this definition, the thread on Jan's bolt has been reproduced from Jon's bolt. I will call this last kind of reproduction "nuts and bolts reproduction." Nuts and bolts reproduction probably figures in the propagation of certain kinds of conventional language patterns.

Now a preliminary word about the relation of function to convention. Many conventions have no apparent function. It is conventional for a bride to wear a blue garter, and to throw her bouquet to the bridesmaids. It is conventional to dress girls in pink and boys in blue, and to put a wreath on the door at

Christmas. But some conventional patterns obviously do serve functions. For example, the convention of driving on the right in the US serves a function. Where a reproduced pattern does serve a function, it is not a conventional pattern unless it is one that would have no particular reason to emerge again, rather than some alternative pattern, if once forgotten. This is because its intrinsic nature makes it no more able than other known or equally knowable patterns to serve its function. Conventional patterns are exemplified rather than other patterns owing only to historical accident, but having occurred, they cause their own recurrence.

A convention of this kind is not tied by definition to any particular group, nor is there anyone in particular who must know of it or follow it. Obviously, there have to be some people who follow or have followed it or the pattern is not reproduced hence not conventional, but it is not necessary that any particular people should do so, nor that a large number of people should do so. When conventions are associated with specific groups, this is because it happens to be the people in these groups who know the conventions and happen, for whatever reasons, to follow them. Which independently designated "group" knows that it is conventional to drink green beer on St. Patrick's day? How many who know accord? What else do these people have in common that would form them into a social group or "community"?

A convention does not prescribe that everybody in some predesignated group should follow it. To know a convention is not to know what to do categorically, but only to know what to do if you wish to follow the convention. For example, the conventions, the rules, of chess do not tell you what to do, but only what to do if you wish to play chess. You can get these conventions wrong, of course. You can fail to reproduce the conventional chess patterns faithfully even though you intend to. But the standard that has then been violated was set by your own intention. No public prescriptive rule will have been violated.

The appearance that conventions involve prescriptive rules may result partly from the fact that many conventions are conditional. The convention is to do something in a certain kind of context, or at a certain time, or if you are a certain kind of person, or in a certain situation. The convention is to put up red and green decorations at Christmas time, or to say "Hello" when you pick up the phone, or to wear a ring on a certain finger of your left hand if you are engaged or married. But that the conventional pattern includes a conventional setting does not mean that it mandates itself. A convention is not, as such, a rule that is required to be followed. A convention is not something having magic times or places or situations where it is mysteriously "in force." If there is some kind of mandate or compulsion to follow a convention (legal, moral, a norm of etiquette) or to follow it under certain conditions, that is a separate thing entirely, and not what makes it a convention, at least not in the sense of "convention" that applies to natural language forms and patterns.

Nor – and this will later emerge as crucial to the case of language conventions

– does a convention mandate that its pattern must be finished once begun. More generally, it does not mandate that the whole of the pattern must be reproduced and not its parts separately. One can quit a chess game in the middle, or set up just an end game or a middle game. Although there is a convention to have one's soup first, then one's salad, main course, and last one's dessert, one can follow the convention part way but not have dessert. For all this kind of convention is, basically, is a pattern of activity that gets reproduced.

Soon I will discuss conventions that serve to coordinate actions and thoughts between people, for these are very important to an understanding of language. And I will argue, contra Lewis in *Convention*, that like more simple conventions, coordinating conventions too do not, in general, involve regularities. First, however, we should examine some simpler aspects of public language in the light of its conventionality. We can begin with words.

Words, many philosophers have thought, are typed or individuated in accordance with physical form. My guess is that there are no legitimate theoretical purposes for which this way of typing is relevant, but I will not argue that here. Rather, I will examine some ways of typing tokens into "words" on which we commonly rely in ordinary contexts, and which reveal our ordinary sensitivities to the conventional aspects of language as I have described them. According to one way of typing, there are parts of Tennessee, for example, where the word "pen" sounds exactly the way my word "pin" sounds. It's exactly the same word, but it sounds different in Tennessee than Connecticut. "They say many of their words differently in Tennessee." Clearly the same word again need not be made of the same acoustic sounds again. Linguists may say that it must be made of the same phonological segments again, the same phonological segment sometimes taking a variety of forms, there being, for example, different allophones of it. On a similar way of typing, the same word must be made of the same letters again, but the very same letter has quite different shapes when different people write it, while different letters sometimes have the same shape, for example Greek P (rho) and Roman P. Typed in these ways, word types and their elements, phonological segments and letters, are like species.[5] In biology, what makes a dog a dog is, in the first instance, that it was born of a dog, not that it has some particular shape. Similarly, what makes a shape or sound into a token of a particular word on this way of reckoning is its lineage, what it was reproduced from, on what prior word tokens it was modeled, or alternatively (and this can be different), on what combination of phonological segments or letters, these being typed by their histories, their lineages. The lisping child who pronounces the word "red" like "wed" does not pronounce the word "wed" but mispronounces the word "red." If I repeat after a Scotsman, I say the same words, but many of them sound quite different. The same word is spelled "color" in America and "colour" in England. Conversely, the Twin-Earth word "water" is as much a different word than ours as the stuff to which it applies is a different liquid than ours. It is different the way Greek P (rho) is different from Roman P. And, of course, comments about

the river bank and about the bank and trust contain different words "bank," copied from completely different lineages. These ways of typing word tokens depends on the conventional nature of words, on their reproductive nature, in much the same way that the typing of individual animals depends on the reproductive nature of animals.[6] And it runs into similar problems. How much can a species change over time without becoming a different species? Is it really so that the word "moan" in "He moaned and groaned" is the same word as "mean" in "He didn't mean to" but a different word than "mean" in "no mean city," as the etymologist is liable to say?

Similarly, the difference between two tokens of a word being used in two different senses and being used once in a literal and once in an extended way, is a question of lineage. Word tokens used in extended ways are reproduced from words used literally. Tokens of the same word that have taken on different senses are words with a common lineage some distance back, but whose lineages have now separated – a matter which can, of course, be one of degree. For example, the question whether the word "red" in the phrase "red hair" has a different sense than the "red" in "red dress" depends on the ancestry of its tokens. Are its tokens currently being reproduced only from other tokens of "red hair"? Or do people individually and independently each stretch the word "red" as used in contexts like "red dress" to cover also red hair? Considered in this light, the conventionality, hence public nature, of words is certainly a real and interesting phenomenon, though often hidden from direct observation in its details.[7]

Besides words, grammatical forms are conventional patterns in the sense we are explicating. Remembering that a reproduction always copies only certain aspects of its model, clearly children reproduce syntax on the model of what they hear. Nor do I mean, of course, that grammar is a matter of patterns in sound. It is the linguist's job to discover exactly what kinds of patterns are materials for the conventions of grammar, to discover what forms of what aspects of speech are the reproduced ones in conventional grammars. The hypothesis of a universal grammar faculty that imposes restrictions on human grammars concerns this matter. Whatever the details, a UG, looked at one way, acts as a more or less complicated filter governing which aspects of what is heard will be reproduced, which aspects will vary depending on models heard and which will not. If there are certain aspects of grammar that are never reproduced at all, but always resupplied by the language module *de novo*, these act like the color of the ink cartridge in the monochrome copying machine. They are not conventional aspects of grammars. A huge question for linguists, of course, concerns which aspects of grammars are and which are not conventional, in exactly this sense of "conventional."

It is important to notice the similarity here between grammatical conventions and others. The perpetuation of any convention requires, categorically, that participants attend to the same aspects of the convention's embodiments and not to others, so that they generalize to new cases in the same manner. Thus, the

convention is to hang a wreath on the door at Christmas, not a sweet-smelling object, and not to do it at relative-visiting times, or at present-giving times, or on cold Tuesdays. In the case of every convention, there has to be some mechanism, whether accidental, or whether designed by nature or man, that operates to keep copies and copies of copies somewhat uniform. Otherwise there is, of course, no convention at all.

A UG is a mechanism effecting uniform reproduction of syntactic forms, hence the maintenance of syntactic conventions. Maybe UG got there accidentally, as Chomsky has sometimes suggested, or maybe it got there by natural selection; it doesn't matter. My present point is that, without something like it, there could be no grammatical conventions. Similarly, if there are inborn mechanisms, as there appear to be, that efficiently accomplish mastery of the phonological structures of languages that the young child hears, these function as another kind of filter that serves to narrowly channel linguistic reproduction, hence to aid the proliferation of linguistic conventions. Phonological mastery of a language yields a generalized sameness–difference schema for the language, dictating what is to count as another correctly executed linguistic utterance of the same type, along one relevant dimension of linguistic reproduction. Alvin Liberman has argued that phonological structure is the fundamental *sine qua non* allowing for the practical possibility of language innovation (Liberman forthcoming). Without it we could at best be stuck with a limited inborn vocabulary, that had slowly and painfully accrued during the course of genetic evolution. A great deal of attention has been paid to the kind of productivity made possible by a grammar that allows embeddings, so that an infinite number of sentences can be generated with a limited vocabulary. But this kind of productivity would have minimum utility if free to operate only on a tiny vocabulary. The capacity of the language faculty as guardian of phonological structures, thus allowing rapid vocabulary growth not just in the child but also in the public language, is productivity with a significance at least as profound.

> This idea [that new speech forms that a speaker has not heard are produced on analogy with those he has heard] is not wrong but rather is vacuous until the concept of analogy is spelled out in a way that explains why certain "analogies" are somehow valid whereas others are not . . . We can give substance to the proposal by explaining "analogy" in terms of I-language, a system of rules and principles that assigns representations of form and meaning to linguistic expressions . . . but . . . with this necessary revision in the proposal, it becomes clear that "analogy" is simply an inappropriate concept in the first place. (Chomsky 1985: 32)

"Analogy," in this context, could mean just sameness in abstract or relational form. Producing new speech forms on analogy with old could mean just some sort of reproduction, as I am using the latter term. But what people more usually have in mind when they speak of "analogy" in this context is a fairly free sort of reproduction, that might pick up one aspect of the original or might pick up

another. That language learning and use involve analogy in this free sense is something that both Chomsky and I want firmly to deny.

On the other hand, the present perspective on public convention gives us little reason to suppose that "I-language" should consist in something appropriately described as "a system of rules and principles," at least not if "system" implies much systematicity. To view the young child's language faculty as a filter through which language conventions are to be transmitted is to view it not as aiming toward some steady state as the child matures, to view it not, for example, as a process of acquiring permanent parameter settings, but as a faculty engaged in the accumulation of a larger and larger repertoire of conventional patterns it can recognize and reproduce on demand. If it reaches a steady state, that will be only if it runs out of local conventions to learn. It would be likely to reach a steady state only if there really were such a thing as The German Language, The French Language and so forth, to be learned. But there is quite a mass of conventions out there to be learned. These conventions are complex, and not particularly systematic, and sometimes crisscrossing or contradictory, getting in one another's way. Thus, with Italian words you use Italian grammar, with English words English grammar, except that in some cases people do quite a bit of mixing, even in conventional ways. With "expect" you use the infinitive but with "anticipate" the gerund. With most English verbs you add an "-ed," but not with "run" or "swim" – people just aren't doing it that way here these days. And the conventions cross at "Is the missionary ready to eat?" and (when it is spoken) "Gladly, the cross-eyed bear." Because linguistic conventions can be more or less compatible in various ways, and because they are often built on one another, exactly like genes in gene pools, then tend to get together in stable clusters. There are innumerable plant species that hybridize quite readily, but left to themselves in a relatively uniform environment, the genes fall back again into stable clusters, and separate relatively homogenous species emerge again. In the same fashion, German and French and other languages are formed out of compatible strands of convention, and if different peoples are isolated from one another for very long, clear demarcations soon emerge between the language conventions they follow. But the idea that there is at the center of each such language some univocal "system of rules and principles" seems as unlikely as that members of an animal species should be genotypically identical or that the whole gene pool should contain no incompatible genes. Possibly there is no real disagreement with Chomsky here. Maybe it is just that I am much more impressed – overwhelmed? – by the huge number of idiosyncrasies and idiomatic elements in any natural language. Perhaps whether "a language" is best idealized as a monolithic structure, or as a relatively loose texture of interlocking crisscrossing conventions, is only a matter of what you are interested in. Surely either way, idealization is involved. Chomsky and I agree, after all, that there actually is no such thing as "a language."

In any event, the phenomenon of public language emerges, I believe, not as a

set of abstract objects, but as a real sort of stuff in the real world, neither abstract nor arbitrarily constructed by the theorist. It consists of actual utterances and scripts, forming crisscrossing lineages. What language forms one is using, from the standpoint of public language, depends not only on the settings in one's language module but on what public conventions one is following. Moreover, public language has definite form, granted that it is passed on in a uniform way by people harboring I-languages that are definite in intension, as Chomsky requires.

If you take the gene's eye view, Dawkins tells us (1976/1989), the organism is just a gene's way of making another gene, but, of course, if you take the organism's viewpoint, the gene is just an organism's way of making another organism. Similarly, if you are interested in individual psychology, public language is merely a stimulus to transition from the initial state of the language faculty S_o to a more steady state S_s. If you are interested in public language forms, on the other hand, the language faculty is merely how public language forms reproduce themselves. A public language is interesting in its own right, I will argue, because it has certain functions that are all its own, that are not merely abstractions gleaned by averaging over speakers' intentions. The study of the functions of public language is a separate discipline, independent of the study of individual psychology.

To show this will require making the second move that I mentioned at the start of this paper, introducing an appropriate theory of language functions. The functions of language conventions are, for the most part, coordinating functions. So I must also show how the existence of conventions, given the way I have described them, can produce coordinations. Coordinations effected through language conventions typically involve communication. The second main argument of this paper will then be complete: a primary function of the human language faculty is to support linguistic conventions, and these have an essentially communicative function. Therefore, a primary function of the human language faculty is to support communication.

I will be using "function" here as short for "direct proper function" as defined in Millikan 1984: ch. 1. This notion is a relative of the notion of function that biologists use when they distinguish between functions and mere effects of an organ or trait or activity. Roughly, the function or functions of a conventional pattern are those effects of it that account for its continued reproduction. More accurately, the pattern is proliferated due in part to a correlation between it and certain of its effects. It is selected for reproduction, in accordance with conscious or unconscious intent, owing to its being coincident with these effects enough of the time. Correlations can, of course, be either very high or very low. Many biological items and traits have functions that they perform very seldom, yet just frequently enough to keep the genes responsible for them from drifting to extinction. Similarly, conventional activities and patterns of activity can have functions that they perform only once in a while, yet perform just often enough to keep them from becoming extinct.

Many conventions seem to have no functions. They seem to proliferate only because people are creatures of habit, or unthinking conformists, or because they venerate tradition, and so forth. Similarly, most patterns of activity that are reproduced due to their effects are not conventions, but rather are handed-down skills. Conventional *ways* of performing certain tasks, such as conventional ways of holding eating utensils in various cultures, or of sitting at meals, are proliferated, in a sense, because they serve certain functions. But it is only certain details of these activities that are conventional, namely, details not required to effect the functions that sustain the activities. Other kinds of details would do as well instead. On the other hand, some conventions have functions *as* conventions. The clearest and most interesting functions that conventions can have are coordination functions. I will call conventions with these functions, "coordination conventions."

Coordination conventions consist of patterns of activity (1) involving more than one participant, (2) proliferated because they serve a purpose had in common by the participants, (3) where the contribution to the joint pattern that each participant must make in order to reach the common goal depends crucially upon the contribution made by the other(s), and (4) where a variety of equally viable alternative joint patterns would achieve the same goal as well. Coordination conventions are thus patterns which might be said to be suitable to solving "coordination problems," though the sense of the latter phrase would not be quite David Lewis's (Lewis 1969).

Some conventional coordination patterns require partners to do the same as one another, while others require them to do complementary things. Examples of patterns requiring the same include driving on the right, shaking hands with the right hand with an up-and-down motion, and standing at a conventional social distance when conversing. Examples requiring complementary actions include conventional positions for partners of opposite sexes while dancing, the pattern *original caller calls back, original receiver waits* when a phone conversation is disconnected (Lewis 1969), the conventional pattern *driver on the right crosses, driver on the left waits* when arriving simultaneously at a four-way stop sign, and the pattern *patron puts the flag up when putting mail in, mailman takes the flag down when taking mail out* when a letter is mailed in a rural mailbox. The view that conventions require regularities has sometimes been supported in part by the assumption that all parties in a coordination pattern do the same thing. For example, Lewis describes patterns such as the telephone convention (above) as though both persons did the same, each following the same conditional rule, *if you are the original caller, call back, if not, wait*. But this is a vacuous move. Any pattern whatever involving more than one person can be described as though these people all did the same. Even the executioner and the executed do the same thing: *if you are the executioner, chop, if you are the executed, relinquish your head.*

For coordination conventions, it is important to consider whether in order actually to effect coordinations these conventions require regularities in the sense

that all or most in some group should accord with them. The answer depends directly on how observable one partner's contribution to the pattern is at the time when the other or others must make theirs. Consider, first, the telephone convention. Here, each partner must make his or her own contribution while completely blind to what the other is doing. For this reason, the coordination will not be achieved unless each partner produces their part of the pattern and before knowing whether the other will produce theirs. In this kind of case, use of the convention will be more or less effective depending on the frequency with which people who interact with one another conform to it, universal conformity producing the best results for everyone. Alternatively, each partner must have prior knowledge which of the various persons with whom they interact is likely to abide by the convention regularly, and if no one abides by it regularly, coordination cannot be achieved. Such a conventional pattern will produce coordinations often enough to encourage its own reproduction only if followed some critical proportion of the time.

The convention of driving on the right is almost as blind as the telephone convention, though not quite. One can see whether a car approaching ahead is driving on the right when the road is straight, but not, of course, around curves. Similarly, the mailman has no way of knowing whether the rural patron has performed his part of the conventional flag-raising pattern without looking in the mailbox to see if there is outgoing mail there. But the point of the convention is, of course, to avoid his having to look if the flag is not up, at the same time securing that all the outgoing mail gets collected. Again, these conventions will serve their functions better the more people follow them.

Contrast these cases of "blind" conventions with the case of assuming conventional postures for social dancing followed by the joint execution of various conventional dance steps. There are a number of alternative conventional postures that a couple may assume, and many alternative combinations of conventional steps can be taken to a given piece of music. But here the conventional pattern initiated by one partner is immediately known to the other: the conventions are not blind but open. There is no need then for different dancers all to conform to the same postures and ordering of conventional steps. All that is needed is for one partner to lead and the other to follow. What will not work, of course, is for both to lead. And there is a convention that the man leads, but, again, not one that everyone has to follow. Conventionalized social dancing thus involves "open" conventions rather than "blind" ones.

Warming up to the case of coordinating language conventions, here are four more cases of open conventions: (1) I stand behind your car and wave my hands this way and that in a conventional way and you complete the conventional cooperative patterns by backing your car to suit my gestures. (2) While biking, I hold my left arm out and you stay out of my way as I make a left turn. (3) I bid four clubs and my partner bids his strongest suit (the "gerber convention"). (4) I want you to pass the bread whereupon I say, "please pass the bread," and you

pass the bread, or, I want you to pass the bread whereupon I say, "Brot, bitte," and you pass the bread.

With "open" coordination conventions, there is no need for regular adherence to any one convention among others for achieving a given kind of coordination. Often many alternative open conventions coexist quite compatibly. Recognizing that a leader has initiated a particular conventional pattern and completing it is like chiming in after the first line of a familiar song. Nor is it necessary that the partners who follow should be regular in their responses to conventional leads. Conventional patterns, as noted before, often persist even though quite regularly broken or interrupted in execution, partners being unwilling, or unable, or having other plans in mind, and so forth. If the coordination effected by completion of a conventional pattern even occasionally has enough value to the partners involved, the pattern may be able to survive even though more usually fractured than not. Perhaps the best contribution that Skinner made to psychology was his demonstrations of the effectiveness of random reinforcement schedules even when reinforcement is, on average, very infrequent. Lead portions of conventional patterns also are often turned to secondary purposes that do not accord with their functions *as* conventions, that is, with the reasons for survival of the conventional pattern. A dancer might lead into conventional dance patterns that he knows his partner can't follow in order to embarrass his partner or to show off. Or the one standing behind the car might jokingly signal a turn that both know leads into a brick wall.

I now want to make plausible that the functions of language conventions are primarily coordinating functions. It is primarily for the service of coordination between speakers and hearers that language patterns are selected to be proliferated as conventions. That is, were it not for the fact that employing its conventions sometimes serves purposes common to both speaker and hearer, language as we know it would shrivel and die. Indeed, for emphasis I might make a stronger claim. Putting completely to one side the evolution of the language faculty itself, were it not for their roles in the achievement of communicative coordinations, there is every reason to suppose that the individual language faculties of individual humans would atrophy, just as with unfortunate children who are not exposed to human language forms at all. Imagine, for example, a child who hears a perfectly normal assortment of English sentences, but for whom no coordinations are ever achieved through use of conventions involving these. Such a child, I am suggesting, would develop no language at all. That, of course, would be an empirical hypothesis, but I will offer reasons to entertain it seriously.

On the other hand, it is possible that there exist some language conventions that have no functions at all. Saying "uhh" at intervals may be such a convention. The corresponding German convention is to say "also," and I am told that in Hungarian one says "öö." Possibly the use of expletives has no coordinating function. The function of expletives, granted they have functions, may

be simply to relieve oneself of, or to objectify for oneself, one's emotion. Thus, people use expletives just as often, perhaps more often, when alone as with other people. It is worth noting, however, that a different part of the brain is involved in the production of expletives than for other speech forms (Pinker 1994: 334).

The best argument that few if any language conventions lack coordinating functions is to discuss some that obviously do have coordinating functions. It will then become clear, I think, how ubiquitous such coordinating functions must be.

We can begin with functions of the grammatical moods. In English, at least, the syntactic forms identified with the indicative mood are proliferated in the service of a number of different coordinating conventions having different functions: conveying information or reminding someone of it ("The Athenians had slaves too"), giving orders ("You will report to the CO at 6 a.m. sharp"), conveying norms ("Johnny, we don't eat peas with our fingers"), and making declarations ("The meeting is adjourned") are four common ones, and undoubtedly there are more. The English indicative mood is like a homonym in this respect. Better, it is like a word that has several distinct though related meanings, instances of each use normally being modeled on prior instances of that same use, where all but one such use can still be heard as a dead metaphor. The use for conveying information is heard as the originating or most "literal" use of the indicative, so let us look first at that pattern.

The pattern begins with a speaker S believing some proposition p and accordingly speaking an indicative mood sentence that expresses p, given the truth-conditional part of the semantics of the language. It concludes when a hearer H, following the truth-conditional semantics for the language, translates the sentence into the thought that p, and accordingly believes that p. The pattern produces a coordination between speaker and hearer under the following assumptions: (1) the speaker is interested to convey information on the subject that p to the hearer, (2) the hearer is interested to gain information on that subject, and (3) the proposition p is true. The end that S and H have in common is that H should become informed about p. For the entire pattern to have attainment of this end as its *own* or *proper* function, (1) it must be a reproduced pattern or (the normal case) it must be composed of reproduced patterns (the words, the syntactic form(s)), (2) it and/or its reproduced elements must each continue to be reproduced only because they sometimes make a certain definite contribution in the service of coordinations, and (3) the sum of these contributions, if made in this case also, would effect that H would become informed about p. For the indicative mood form itself to have the transfer of information from speaker to hearer as a function, it must be that the contribution it makes to the whole is to utilize the functions of other reproduced parts of sentences exemplifying it to that end.

"Reproduction" in this case, and in the general case for coordinating functions

of language, is mainly by nuts and bolts reproduction (see above, p. 220). The speaker parts of the patterns are reproduced due to the effects they sometimes have in the presence of hearers who complete them, and the hearer parts are reproduced due to the effects they sometimes have in the presence of speakers who have initiated them. But the whole pattern, involving speaker and hearer, is arbitrary in relation to its function, that is, it is conventional. H responds to the indicative sentence by translating it into belief in accordance with certain semantic rules because, in H's experience, responding selectively to indicative sentences in this way has often enough resulted in the appropriation of useful information. S translates S's belief into an indicative mood sentence in accordance with these same semantic rules because, in S's experience, often enough hearers respond to such sentences by forming beliefs accordingly. Read the "becauses" here not as indicating conscious reasons, but as indicating causes. That is, turning the coin over, had H not lived where speakers often enough expressed true beliefs using the indicative pattern with these truth-conditional semantic rules, H would not translate from indicative sentences into belief in this way, and had S not lived where hearers often enough translated from indicative sentences into belief in this way, S would not speak in this way when interested to convey information. Thus, speakers and hearers collectively learn from each other how to speak and how to respond to speech in ways that serve purposes for both, each leaning on the settled dispositions of the other. The parallels with the evolution and fixation of symbiotic relations between animal species and with the evolution of animal signal systems should be apparent.

It should be clear from the above description that successful coordinations achieved in this way through use of the English indicative mood help not only to proliferate the indicative mood, but also to proliferate use of and reliance on semantic conventions applying to other elements of indicative mood sentences. That understanding the truth-conditional semantics this way is enabling me often to collect useful information reinforces my use of these conventions for interpreting the semantics, and that my hearers often enough seem to believe what I intend them to believe reinforces my use of these semantic conventions when speaking. If I never had either of these experiences when using or interpreting indicative mood sentences, and if, further, this use of the indicative mood was the sole convention in my language community involving intentional attitudes, hence the only one employing the truth-conditional part of its semantics, I could not possibly learn to understand either this function of the indicative mood or any of the truth-conditional semantic conventions of my language. Of course, the use of indicative forms to convey information is not the only convention employing the truth-conditional part of the semantics of any natural language. But the underlying principle here is crucial. It is only through various communicative uses of language and through the conventions that make these possible that the truth-conditional semantics of a language, which is also conventional, of course, is learnable. I will reinforce this point in a moment by discussing other

kinds of conventions that rely on the truth-conditional semantic conventions of languages, *and vice versa*.

A surprise on this analysis of the conventional nature of the information-transferring function of the indicative is that believing what you hear said in the indicative turns out to be a conventional act, something one does in accordance with convention. Compare: standing, while conversing, at what happens to be the conventional social distance in one's culture, is something one does in accordance with convention. And one learns to stand at the correct social distance in very much the same way, by unconsciously learning to fit with others who are already standing at that distance. The difference is only that in the case of social distance, the convention requires that partners do the same thing rather than different things. On the other hand, of course, the fact that H's believing p is a conventional outcome of S's telling H that p in no way mandates that H will or should believe p. The fact that a given outcome would be the conventional one in no way implies or mandates that it will be the actual one. As emphasized earlier, pieces of conventional patterns often occur independently. The speaker who lies also illustrates this point. He uses conventional structures in an attempt to induce partial completion of a conventional pattern, but both the beginning and the end of the pattern are missing. He does not translate from a genuine belief into words as required by convention, nor does the hearer end with a true belief as required.

Similar analyses can be applied to the functions of numerous other language forms, each of which continues to be reproduced due to its occurrence, often enough, in a pattern ending in a certain conventional hearer response. The root function of the imperative mood, for example, is to produce a corresponding action by the hearer, where the speaker is interested in having that action performed, and the hearer is interested in completing the conventional pattern, perhaps because he has a further interest in common with the speaker (hearers often want direction from speakers), or because conforming is sanctioned in one way or another. Performance of this imperative function is accomplished through the hearer's first forming an intention to perform the designated action, an intention formed following the truth-conditional semantic conventions of the language, so that reinforcement of the hearer's compliance tends to reinforce his observance of these semantic conventions as well. The imperative function also proliferates reproductions of indicative mood forms, especially in the armed services. Here the indicative mood functions, as I have said, rather like a dead metaphor. The root function of the interrogative is to get information of a designated kind conveyed to the speaker, and its conventional use when successful also reinforces any truth-conditional semantic conventions used with it. And so forth. On the other hand, elsewhere I have discussed the functions of a number of other language devices (Millikan 1984), claiming that the functions of sentences in which these forms occur do not directly implicate intentional attitudes on the part of either speaker or hearer. These sentences have as their functions to do

other things, as it were, to hearer's heads. Similarly, Strawson (1972) claimed that what an identity sentence *does* is to merge two information files in the hearer's head (see also Millikan 1984, 1993, 1997), and Wilfrid Sellars (e.g., 1963: essays 4, 5, 6, 8, 10) claimed that what the *"X" means "Y"* rubric *does* is to prompt the hearer to use "X" in the way he already knows to use "Y."

It is crucial that the functions of language forms are not the same as the uses to which they are usually put. Effective use of a set of language conventions requires the acquisition of a mass of skills and a good deal of inventiveness well beyond mere grasp of those conventions. Hearers can by no means be counted on to complete the conventional patterns that speakers initiate. They will do so only under special conditions. For example, they must trust the speaker, understand the subject matter, and have interests compatible with the coordination projected by the particular convention used. Moreover, the fact that a conventional linguistic form has a certain function does not prevent a speaker from using it to serve quite different ends. In general, there is no need that a device having a certain function be used to serve that function. A hammer can be used as a weapon, a human hair can be used as a cross hair on an instrument, and the eye blink reflex can be used by the psychologist to demonstrate classical conditioning. Similarly, language forms are often used to serve functions that are not their own. Sometimes these extrinsic ends are ends that also interest the hearer and sometimes they are not. Uncooperative uses of conventional forms include lying to the hearer, embarrassing the hearer, insulting the hearer, purposely putting the hearer in an awkward position, and so forth. Cooperative uses include pretend uses (acting, joking) and the whole hodgepodge of Gricean implicatures (Grice 1968).

Gricean implicatures are of particular interest because they involve uses of conventional forms to produce nonconventional coordinations. Unlike cases in which, say, the speaker lies or the hearer refuses to comply, so that one frustrates the purpose of the other, a successful Gricean implicature achieves exactly the same sort of coordination that proliferates conventional language patterns. For this reason, Gricean implicatures easily become conventionalized. Speakers soon come to reproduce the relevant language forms directly on the model of previous cases of successful implicature, and hearers also reproduce the intended response that way. Hackneyed examples are "Nice going!," "Where's the fire?," and "can you . . .?," "could you . . .?" (as contrasted with "are you able to?") as used to make requests. The effect is exactly the same as when metaphors enter "the language" (sic!) and become literal. The conventionalized use may continue to be associated with the original use more or less strongly, the strength of association differing also from person to person. It was only recently, for example, that I came to associate the idiom "going haywire" with tangled hay wire, and perhaps most people, excepting farmers, never do.

For both dying metaphors and conventionalized Gricean implicatures, entrance into "the language" is thus a matter of degree. How widely spread a conventional

usage is is always a matter of degree. Language conventions can develop between just two people, as frequently occurs with identical twins, or develop and proliferate widely within just one generation, as in the case of certain creoles. It does not follow that the distinction between the conventional functions of language and individual uses of language is the least bit arbitrary or trivial. Whether or not a person is dead can also, occasionally, be a matter of degree. All of the conventional functions of public language are as much its meanings as are its satisfaction conditions. The distinction between the meanings of linguistic forms and the meanings of speakers who use them is entirely real and important.

Davidson (1986) claims that there is no boundary around the information on which a hearer may need to draw in order to interpret a speaker, hence that there is no use for the notion that a language serves as "a portable interpreting machine set to grind out the meaning of an arbitrary utterance." He concludes from this that not only must we "abandon . . . the ordinary notion of a language, but we have erased the boundary between knowing a language and knowing our way around the world generally." Chomsky is right (1992a, 1992b) that Davidson has produced here no reason to abandon the distinction between the internal systems that know language and the systems that use this knowledge. On the other hand, the systems that know language must grasp or "represent" (as Chomsky rather puzzlingly puts it) more than just phonetic form (PF) and logical form (LF) if the latter is understood narrowly. They must also grasp conventional function. If one draws the semantics/pragmatics distinction with the purpose of cutting between meaning and use, between linguistic or "grammatical competence" and "pragmatic competence . . . relating intentions and purposes to the linguistic means at hand" (Chomsky 1980: 224), then the functions of all linguistic conventions fall on the semantics side of the dichotomy.

On another point Chomsky agrees with Davidson: we must "give up the idea of a clearly defined shared structure which language users acquire and then apply to cases." If we substitute "structures" for "structure," however, Davidson would surely be wrong. Language users acquire various shared structures, knowledge of public conventions, knowledge of the functions of various real-world lineages of idiom and form, without which they could not use language to communicate, nor for any of its myriad alternative uses that ride piggyback on public truth-conditional semantic conventions and conventions of "function" in a more classical sense. There is a very simple way to grasp this point. In order to communicate with language, I must be able to predict how other people will react to my language. But no such predictions could possibly be made were it not for the possibility of conventions of use and response. These are merely repetitions of prior usage and response, including both truth-functional semantic conventions and conventions of "function" in a more classical sense. The human language faculty plays a critical role in making it possible readily to proliferate such conventions. Whatever its history, its currently central functions are profoundly involved both with external language and with communication.

Returning to Chomsky's own views, I am not sure whether I am really one of "his critics." I have turned things a somewhat different way and introduced a somewhat different vocabulary to make my points. What I mean by "public language" is not the same as Chomsky's target when he decries "public language," and I have put a different gloss on language "purpose" or "function." I have used tools he himself has fashioned at the center of my argument, claims concerning the unique capacities of the human language faculty. There are very few points here on which I feel confident of either his agreement or his disagreement. I am therefore eager to learn his response and grateful to the editors of this volume for making it possible for me to speak with him at this length so directly.[8]

Chomsky's reply: pp. 308–15.

Notes

1 Recanati (1987) takes it that conventional language devices "indicate" or, using the linguists' term, "mark" uses of language, that is, conventions mandate that these devices shall be used only for those purposes. This is not explicitly stated, but see, for example, section 22.

2 The thesis on conventions is in Millikan 1998a; the thesis on language functions is in Millikan 1984, 1996, 1998b, 2001.

3 See, however, Millikan 1998b.

4 For more detail see Millikan 1998a.

5 I realize that this is not how contemporary linguists *define* phonological segments, but I am suggesting this as a theory concerning an aspect of their real nature. Similarly, biologists before Darwin thought that all members of the same species had an inner nature or form in common. Modern biologists don't think so.

6 This point about the typing of words is put in a broader context in Millikan 1984: ch. 4. Much the same point about words is also argued in Kaplan 1990.

7 There is more on all the above matters in Millikan 1984: chs. 3 and 4.

8 I am enormously grateful to Louise Antony for so patiently, thoroughly, and insistently misunderstanding an earlier version of this essay, forcing me, kicking and screaming, into actually saying what I meant in many places. Indeed, in many places she even forced me into knowing what I meant. Without question, she is the most brilliantly dense reader I have ever had and I hope and think that the essay was immeasurably improved by her interference.

References

Bach, K. and Harnish, R. M. 1979: *Linguistic Communication and Speech Acts*. Cambridge, Mass.: MIT Press.

Chomsky, N. 1980: *Rules and Representations.* New York: Columbia University Press.

Chomsky, N. 1986: *Knowledge of Language: Its Nature, Origin, and Use.* New York: Praeger.

Chomsky, N. 1992a: Language and Interpretation. In J. Earman (ed.), *Inference, Explanation and Other Frustrations*, Berkeley: University of California Press.

Chomsky, N. 1992b: Explaining Language Use. *Philosophical Topics*, 20, 205–31.

Chomsky, N. 1993: Mental Constructions and Social Reality. In E. Reuland and W. Abraham (eds.), *Knowledge and Language*, vol. 1, *From Orwell's Problem to Plato's Problem*, Amsterdam: Kluwer Academic Publishers.

Chomsky, N. 1995: Language and Nature. *Mind*, 104, 1–61.

Davidson, D. 1986: A Nice Derangement of Epitaphs. In E. Lepore (ed.), *Truth and Interpretation*, Oxford: Blackwell.

Dawkins, R. 1976/1989: *The Selfish Gene*. Oxford: Oxford University Press.

Gilbert, M. 1983: Agreements, Conventions, and Language. *Synthèse*, 54, 375–407.

Gilbert, M. 1989/1992: *On Social Facts.* Princeton: Princeton University Press (1992 edition, reprinted from Routledge 1989 edition).

Grice, H. P. 1968: Logic and Conversation. In H. P. Grice, *Studies in the Way of Words*, Cambridge, Mass.: Harvard University Press.

Kaplan, D. 1990: Words. *Aristotelian Society Proceedings, Supplementary Volume*, 103–19.

Lewis, D. 1969: *Convention: A Philosophical Study*. Cambridge, Mass.: Harvard University Press.

Lewis, D. 1975: Languages and Language. In K. Gunderson (ed.), *Language, Mind and Knowledge*, Minneapolis: University of Minnesota Press. Reprinted in H. Geirsson and M. Losonsky (eds.), *Readings in Language and Mind*, Oxford: Blackwell, 1996.

Liberman, A. M. (forthcoming): The Reading Researcher and the Reading Teacher Need the Right Theory of Speech. *Scientific Studies of Reading*.

Millikan, R. G. 1984: *Language, Thought, and Other Biological Categories.* Cambridge, Mass.: MIT Press.

Millikan, R. G. 1993: *White Queen Psychology and Other Essays for Alice*. Cambridge, Mass.: MIT Press.

Millikan, R. G. 1996: Pushmi-pullyu Representations. In J. Tomberlin (ed.), *Philosophical Perspectives*, vol. 9, Atascadero, Calif.: Ridgeview Publishing. Reprinted in L. May and M. Friedman (eds.), *Mind and Morals*, Cambridge, Mass.: MIT Press, 1996.

Millikan, R. G. 1997: Images of Identity: In Search of Modes of Presentation. *Mind*, 106 (423), 499–519.

Millikan, R. G. 1998a: Language Conventions Made Simple. *Journal of Philosophy*, 95 (4), 161–80.

Millikan, R. G. 1998b: Proper Function and Convention in Speech Acts. In L. E. Hahn (ed.), *The Philosophy of Peter F. Strawson*. The Library of Living Philosophers, LaSalle, Ill.: Open Court.

Millikan, R. G. 1998c: A Common Structure for Concepts of Individuals, Stuffs, and Basic Kinds: More Mama, More Milk and More Mouse. *Behavioral and Brain Sciences*,

22.1 (February 1998), 55–65. Reprinted in E. Margolis and S. Laurence (eds.), *Concepts: Core Readings*, Cambridge, Mass.: MIT Press, 1999.

Millikan, R. G. 2001: The Language–Thought Partnership: A Bird's Eye View. In H. J. Glock (ed.), *Language and Communication*, special issue on Language and Intentionality, 157–68.

Pinker, S. 1994: *The Language Instinct*. New York: William Morrow and Co.

Recanati, F. 1987: *Meaning and Force: The Pragmatics of Performative Utterances*. Cambridge: Cambridge University Press.

Schiffer, S. 1972: *Meaning*. Oxford: Clarendon Press.

Searle, J. 1969: *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

Sellars, W. 1963: *Science, Perception and Reality*. New York: Humanities Press.

Strawson, P. F. 1974: *Subject and Predicate in Logic and Grammar*. London: Methuen.

10
The Theory Theory as an Alternative to the Innateness Hypothesis

ALISON GOPNIK

One of the deepest and most ancient problems in philosophy is what we might call the problem of knowledge. There seems to be an unbridgeable gap between our abstract, complex, highly structured knowledge of the world, and the concrete, limited, and confused information provided by our senses. Since the *Meno*, there have been two basic ways of approaching this problem: rationalism and empiricism. Each era seems to have its matched pair of advocates of each view, making their way through the centuries like couples in some eternal philosophical gavotte: Plato and Aristotle; Descartes and Locke; Leibniz and Berkeley; Kant and Mill. The rationalist approach says that although it looks as if we learn about the world from our experience, we don't really. Actually, we know about it all along. The most important things we know were there to begin with, planted innately in our minds by God or evolution (or chance). The empiricist approach says that although it looks as if our knowledge is far removed from our experience, it isn't really. If we rearrange the elements of our experience in particular ways, by associating ideas, or putting together stimuli and responses, we'll end up with our knowledge of the world. There is both a tension and a kind of complementarity between these two ideas, and philosophy and psychology often seem to alternate between one view and the other. Rationalists can explain the abstract, complex nature of our knowledge quite well, but they can't explain, and so deny, the fact that we learn. Empiricists can explain the fact that we learn, but they can't explain, and so deny, the fact that our knowledge is so far removed from our experience.

Chomsky made two very important contributions to this debate. The first was to cast it as an empirical, scientific, psychological question, rather than simply as a philosophical question. Chomsky has argued that we can develop a scientific account of how knowledge develops in individual human minds, and that such an account will provide an answer to the problem of knowledge. Moreover, he

suggested that such an account would involve representations of the outside world and rules for manipulating those representations. These representations and rules characterize the cognitive capacities of human minds; they are the theoretical entities of cognitive psychology. That idea is at the heart of cognitive science, and underlies its success. It has been very widely influential. In fact, perhaps the greatest successes in cognitive science have been in domains that Chomsky didn't himself investigate, particularly vision, using methods – psychological experimentation, detailed computational modeling, and, most recently, neuroscience – that he has eschewed. There were other earlier sources for this idea, too, including Piaget and the Gestalt psychologists. But none of that alters the fact that, in its modern incarnation, this foundational idea of cognitive science is largely due to Chomsky.

We might call this part of Chomsky's view "cognitive naturalism," the idea that knowledge can be understood by scientific investigation of the mind. It assumes, of course, some version or other of scientific realism. If you think scientific investigation is the right course to find the truth of these questions, you must think that scientific investigation is the right course to find the truth, in general.

The second part of Chomsky's contribution was to present a particular, rationalist hypothesis, the innateness hypothesis, as the empirical answer to the problem of knowledge. Chomsky's thesis is that human minds are highly constrained innately. We can only formulate a very small set of possible representations and rules. Information from the outside world may trigger the development of those representations, and may narrow the set of possibilities even further by processes like parameter-setting, but the constraints remain unchanged throughout life. Representations and rules are not inferred or derived from the input. This rationalist answer was, of course, a radical, indeed revolutionary, departure from the empiricist asumptions of earlier psychology, particularly behaviorism. It has also been very widely influential. It has dominated Chomsky's own field, syntax, and has influenced theories of semantics (Pinker 1989), and phonology (Eimas 1975), and of our knowledge of everyday physics (Spelke et al. 1992), biology (Atran 1990), psychology (Leslie and Roth 1993), and even ethics (Barkow, Cosmides, and Tooby 1992).

There are a range of empirical phenomena, particularly developmental phenomena, that could be adduced to support the innateness hypothesis. But it is fair to say that these phenomena have played little part in Chomsky's own thinking and argumentation. Chomsky's arguments for innateness don't come from studying the development of language and thought in children but from considering the characteristics of adult language and thought. Chomsky's most important argument for rationalism is the same argument that Socrates originally formulated in the *Meno*; it has come to be called the poverty of the stimulus argument. The learning mechanisms we know about are too weak to derive the kind of knowledge we have from the kinds of information we get from the outside world.

The central argument I want to make in this paper is that these two important Chomskian theses – cognitive naturalism on the one hand and the poverty of the stimulus argument on the other – are at odds with one another. Suppose we apply the program of cognitive naturalism to scientific knowledge. If we do, and if we are scientific realists, we must believe that there are learning mechanisms that allow human minds to derive abstract, complex, highly structured, veridical representations and rules, namely theories, from limited input, namely evidence. They are just the mechanisms we use in science. But if we believe this, then it is at least logically possible that those same learning mechanisms are involved in our development of other kinds of knowledge, such as everyday physical, biological, psychological, and even linguistic knowledge.

Moreover, recent empirical developmental research suggests that this is precisely the case. Since the early 1990s the idea that there are deep similarities between scientific theory formation and cognitive development, an idea we have called "the theory theory," has become, at the least, a serious developmental hypothesis. The cognitive abilities involved in science do seem to also be involved in everyday cognitive development. Of course, this still leaves open the question of whether this is the correct account of development in any particular domain. But it also means that the poverty of the stimulus argument no longer applies. To discover whether any particular kind of knowledge is due to innate Chomskian devices or to theory formation we need to do detailed empirical studies, and, in particular, detailed developmental studies. We can't make those arguments simply by looking at the adult system.

The Theory Theory

First, I'll present a very brief sketch of the theory theory as it has emerged in cognitive development (for more detailed accounts see Gopnik and Wellman 1994; Wellman and Gelman 1997; Gopnik and Meltzoff 1997). The basic idea is that children develop their everyday knowledge of the world by using the same cognitive devices that adults use in science. In particular, children develop abstract, coherent systems of entities and rules, particularly causal entities and rules. That is, they develop theories. These theories enable children to make predictions about new evidence, to interpret evidence, and to explain evidence. Children actively experiment with and explore the world, testing the predictions of the theory and gathering relevant evidence. Some counter-evidence to the theory is simply reinterpreted in terms of the theory. Eventually, however, when many predictions of the theory are falsified, the child begins to seek alternative theories. If the alternative does a better job of predicting and explaining the evidence it replaces the existing theory.

This account has been successfully applied to explain children's developing understanding of the physical world (Smith, Carey, and Wiser 1985; Gopnik

1988), the biological world (Carey 1985; Keil 1989; Gelman and Wellman 1991), and the psychological world (Wellman 1990; Perner 1991; Gopnik 1993). It has been applied to adults (Murphy and Medin 1985), school-age children (Carey 1985; Keil 1989), preschoolers (Gelman and Wellman 1991), and infants (Gopnik 1988; Gopnik and Meltzoff 1997). We have suggested, in particular, that infants are born with initial innate theories, and that they begin revising those theories even in infancy itself. There are, of course, debates about which domains and which ages the theory applies to, and the theory is in competition with other accounts, including Chomskian innateness theories.

While the theory theory has primarily been articulated in cognitive development, it also helps to explain science itself. On this view, the success of science depends on the fact that we adults retain natural learning capacities that are most clearly evident in children (and were probably designed by evolution to be used by children). Science simply puts these universal and natural capacities to work in a socially organized and institutionalized way. Moreover, in science, we typically tackle just the problems that we have not solved in childhood, particularly problems where evidence is scarce and difficult to come by. Many of the social institutions and technological inventions of science are adapted to the special demands of these problems. Nevertheless, these institutions and inventions would be useless without the foundation of a universal set of cognitive capacities, capacities to form, test, and revise theories. These capacities were designed to give us a generally veridical view of the world, just as, for example, perceptual capacities give us a generally veridical view of the world. The theory theory, then, is both a theory of cognitive development and a scientific, naturalistic, psychological account of scientific knowledge (see Kitcher 1988; Gopnik 1996; see also some essays in Giere 1992).

We can translate the idea of the theory theory into the Chomskian language of rules and representations, in a way that is consistent with Chomsky's cognitive naturalism. A theory is a particular kind of system that assigns representations to inputs, in the way that the perceptual system assigns representations to visual input or the syntactic system assigns representations to phonological input. We can capture the distinctive structural features of these representations by talking about the specific abstract, coherent, causal entities and laws of the theory. The representations are operated on by rules that lead to new representations, for example, the theory generates predictions.

In a theory, however, unlike a Chomskian grammar, the very patterns of representation that occur can alter the nature of the representational system itself. This is what makes theory formation a kind of learning. The patterns of representation that occur can alter the nature of the relations between inputs and representations. As we get new inputs, and so new representations, the very rules that connect inputs and representations change. Eventually, we may end up with a system with a completely new set of representations and a completely different set of relations between inputs and representations than the system we started

out with. This differentiates theories from other kinds of representational systems.

To borrow Neurath's philosophical metaphor, the theory-formation view sees knowledge as a boat that we perpetually rebuild as we sail in it. At each point in our journey there may be only a limited and constrained set of alterations we can make to the boat to keep it seaworthy. In the end, however, we may end up with not a single plank or rivet from the original structure, and the process may go on indefinitely.

This kind of system may sound so open-ended as to be uninteresting. But, in fact, the theory-formation view proposes that the representational system will change in relatively orderly, predictable, and constrained ways. Moreover, there may still be some overall constraints on the kinds of representations that are generated. Not every logically possible theory will be formulated or tested by human beings. These constraints reflect the basic presuppositions of scientific inquiry, for example, that the world has a causal structure that can be discovered. However, the range of theories that can be generated is very much wider than the limited set of representations that are possible on innateness views. It includes, for example, the theories of relativistic physics.

The Theory Theory vs. Other Empiricist Alternatives

I want to argue, then, in favor of the theory theory as an empiricist alternative to the innateness hypothesis. However, my argument is different in important ways from earlier empiricist arguments. One such argument is that unless we can specify just which learning mechanisms we are talking about, and we know for sure that there aren't any other possible learning mechanisms, we can't be certain that some capacity is innate. The poverty of the stimulus argument doesn't go through. There may be general-purpose learning mechanisms that could account for the development of a particular type of knowledge. Putnam made this sort of argument many years ago when Chomsky first formulated the innateness hypothesis. Similarly, Piagetians argue that vague, possible, general-purpose processes of "assimilation" and "accommodation" can account for cognitive development. Chomsky is right in thinking that these are weak arguments. It seems unsatisfactory to prefer an unspecified and vague general-purpose learning mechanism, a mechanism that may or may not exist, to a specific proposal for an innate structure.

Of course, there have also been more specific proposals about the kinds of learning mechanisms that could be opposed to innateness. These arguments typically have the following form. A particular mechanism explains some particular type of learning. In this type of learning, inputs are transformed into representations that are at least somewhat more abstract, complex, and highly structured than the inputs themselves. The actual representations such mechanisms generate

are not nearly as abstract, complex, and highly structured as the representations we ultimately want to explain, the representations either of language or of everyday thought. But, the argument goes, if we just scale the system up, make the input richer, add more connections, we will get there eventually. Trust us.

This was, of course, precisely the argument of behaviorism. Operant reinforcement could get a pigeon from a series of specific experiences of lights and shocks to a representation that would allow it to perform a complex behavior (though, of course, behaviorists didn't like to talk about representations). Therefore, if the reinforcers and the stimuli and responses were complex enough, operant conditioning could get us to language and everyday cognition. It was also the argument of associationism. We could get from two specific simple experiences to a more complex representation (even on Hume's view, for example, a causal representation) by associating the simple experiences. Therefore, we could get to all representations if we made the system sufficiently complex.

A corollary of this sort of argument, of course, was that, in fact, the representations were not actually as abstract, complex, or highly structured as they appeared to be on the surface. If you looked more carefully you would see that they really were far more concrete, specific, context-dependent, and generally similar to the input. The two strategies go hand in hand as ways of bridging the gap between input and representation.

The most recent version of this argument invokes connectionist learning systems. These systems generate often surprisingly abstract and complex representations from repeated patterns of input. The learning mechanisms these systems employ are more powerful than the mechanisms of simple association or conditioning. Nevertheless, the representations these systems generate still look very different from the kinds of representations we want to explain. In particular, they lack the componential and inferential character of the representations of ordinary language and thought. Even the strongest exponents of these views still have to resort to a scaling-up argument (see, e.g., Bates and Elman 1992; Elman et al. 1996).

The theory formation alternative, in contrast, presents a very different kind of argument. If anything, it's actually a kind of scaling-down argument. No one would deny that scientific theories are very abstract, complex, highly structured kinds of knowledge; indeed, if anything, they seem more abstract, complex, and highly structured than everyday knowledge. Nor would anyone deny that theory change is based on evidence, and so, ultimately, on sensory input. And very few people, no one except perhaps Jerry Fodor, would accept that scientific theories are innately determined and are simply triggered by evidence. Scientific theory formation is a kind of demonstration proof that there are learning mechanisms in the universe that are powerful enough to generate the kinds of representations we want from the kinds of input we know we have. More than that, these learning mechanisms can be and have been instantiated in human brains right here on earth.

Theory formation isn't a vague, general-purpose learning mechanism that may or may not exist. Nor is it a specific learning mechanism that may or may not be able to generate sufficiently abstract, complex, highly structured representations with enough scaling-up. It's a mechanism that does exist and that does generate the right kind of knowledge from the right kinds of input.

Innate Theories and Starting-state Nativism

There is another aspect of the theory theory that makes it different from other empiricist alternatives. The classical empiricist views, as well as contemporary connectionist views, assume that all the representational structure comes from the nature of the input and the learning mechanisms themselves. Representational structure can be reduced to input. In contrast, scientific theory change never seems to involve raw, uninterpreted data. Instead, theory change is possible because theories themselves build on, revise or replace earlier theories. The earlier theories enabled us to select and interpret the evidence that eventually led to the new theories. In fact, as early as the 1950s, in Carnap's late work, for example, this view of science as a succession of changing conceptions of the world replaced the earlier logical positivist attempt to reduce science to "sense data."

In accounts of science, the origin of the first theory in this succession of changing theories is obscure. But if we think about theory formation as a mechanism of cognitive development, the origin of the first theories is quite simple. They are the theories we are, literally, born with. We learn by modifying, revising and eventually replacing those earlier theories with later ones. The empiricism of the theory theory is an empiricism of revision, not reduction. The idea that we are born with substantive theories of particular domains is also consistent with the empirical evidence of the last thirty years of infancy research. This idea, however, is not part of most empiricist theories in psychology. Notably, Piaget himself was only willing to grant infants reflex actions and accommodation and assimiliation mechanisms.

The reader may be reeling slightly at this point. I've been arguing that theory formation is an empiricist alternative to innateness and here I am attributing rich innate knowledge to newborn babies. There is, however, a very important difference between this kind of innateness and the Chomskian innateness hypothesis. On my view, which I have called "starting-state" nativism, these first theories are subject to radical revision and are indeed radically revised in the light of further evidence (Gopnik and Meltzoff 1977). By the time we reach adulthood the theories may look almost nothing like the initial theories we had at birth, and the further revision of theories in organized science may lead to still further radical changes. These changes are not arbitrary, though; they are inferred from the evidence of the experience we accumulate in the course of development. This is very different from Chomsky's innateness hypothesis, and from classical

rationalism. On that view, innateness is a claim about the constraints on the final state of the system, not about the state the system starts out with.

Again, though, while other philosophers might object to the idea of innate knowledge in principle, Chomsky obviously should not. In fact, Chomsky has made quite convincing arguments in favor of the idea that innate knowledge is possible, though his own proposals about the character of such knowledge are unlike mine.

The theory theory, then, like other empiricist theories, sees the development of knowledge as a kind of learning. The representations and rules of new theories are inferred from evidence and they change, often quite radically, as a result of our experience of the outside world. These changes lead our theories to become increasingly veridical. But the theory theory doesn't face the same gap between input and knowledge as classical empiricist theories. By granting that there are innate theories, representations that are already abstract, complex, and highly structured, the role of input becomes quite different. Input transforms existing abstract, complex, and highly structured representations into new abstract, complex, and highly structured representations, and this process can go on indefinitely. Most importantly, we have independent evidence that human beings can engage in this kind of learning: this is the kind of learning that goes on in science.

It has to be said right off that we still don't know, in detail, how these theory-formation mechanisms work, either in science or in cognitive development. We do, though, have a general outline, which is certainly more specific than the vague general-purpose mechanisms that Putnam or Piaget invoked, and at least as specific as most nativist accounts. Moreover, while there is a good argument for rejecting vague alternative possibilities, there is not an equally good argument for rejecting vague alternative certainties. It's not that we don't know what the learning mechanism is like or even if it exists: we don't know what it is like (in detail) but we do know that it exists, at least in scientific learning.

Of course, we might want to argue that our current inability to specify the nature of the learning mechanisms in science in detail is, in fact, a proof that they don't exist. Perhaps it's just an illusion that scientific theories provide us with veridical representations of the world based on evidence. Perhaps they are just arbitrary social constructions imposed by more powerful scientists on less powerful scientists. Just such a claim underpins the postmodern arguments against scientific realism. If the postmodernists were right, the theory theory would be doomed. But Chomsky can't accept that argument and still be a scientific realist, and realism underpins his whole enterprise (and mine and that of every other scientist who is not accepting money under false pretences). Chomsky's whole project of cognitive naturalism depends on the idea that science allows us to learn about the world by looking at evidence.

An alternative position that Chomsky might take, and sometimes indeed does seem to take, is that while science is veridical it is also intrinsically incomprehen-

sible. Chomsky is usually both a realist about science and a naturalist about cognition. He believes that knowledge can be understood using the techniques of empirical science. But he sometimes seems to think that this is only true of some kinds of knowledge. He certainly believes that we can understand syntactic knowledge empirically, but there are times when he seems to say that we can't understand such things as scientific knowledge (or ethical knowledge, etc.). This view is consistent with the innateness hypothesis. If the innateness hypothesis is generally correct, then some things are simply incomprehensible, just contingently beyond the limits of our innate capacities, and understanding how we do science might be one of them.

There seems to be very little good reason for this *a priori* pessimism, particularly since there has been hardly any empirical research on the question. In an age in which we humans have developed theories of quantum mechanics and relativistic physics, arguments about our innate inability to understand certain things are bound to be unconvincing. But even if this pessimism were accurate, the poverty of the stimulus argument would still be undermined in the same way. Given a choice between a specific proposal and an unspecified possibility, we might do well to at least tentatively adopt the specific proposal. But given a choice between two mechanisms that we have good reason to think exist, we can't adopt one because there is some contingent reason that we find it difficult, even impossible, to understand the other in detail. For a realist, this would be like arguing that because we can't know about events outside the light cone, we should believe that everything happens inside the light cone. It would be an unhappy outcome for psychologists if we had to accept that we could know nothing about scientific learning, but even that would not in itself be a reason for denying that scientific learning exists or accepting the innateness hypothesis.

Phenomenological and Social Objections

So far, I've been arguing that theory formation is at least a possibility as an answer to the poverty of the stimulus argument, and that it is not subject to the same objections as other empiricist alternatives. Indeed, Chomsky ought to accept that it is a possibility, given his cognitive naturalism and scientific realism. But, of course, merely saying that it is a possible answer doesn't make it a plausible answer, let alone the correct answer. The best argument that it is a plausible and correct answer is the detailed empirical developmental research I referred to earlier. Still, to many philosophers the very idea that children could be employing the same learning mechanisms as scientists generates a reaction of shocked, even indignant, incredulity. (For some reason, I've found this initial reaction to be stronger in philosophers than in psychologists or, especially, practicing scientists, who seem to find the idea appealing and even complimentary. It's odd since you'd have thought philosophers, in particular, would, like the Red Queen, have

had extensive practice in believing at least six impossible things before breakfast.) There might be specific objections to the theory theory, even if it escapes the usual objections to empiricism.

Some of these objections, though, at least shouldn't be Chomsky's objections. In fact, the objections that philosophers raise to the theory theory are often just the same objections that were raised against Chomsky's cognitive naturalism, and Chomsky's counter-arguments to those objections apply just as effectively. In fact, the arguments of theory theory advocates sound very much like Chomsky in his cognitive naturalist mode. (This is hardly surprising since most advocates of the view, including me, come from a strongly Chomskian intellectual tradition; indeed, sociologically and historically, you might think of the theory theory as a new species resulting from the cross-fertilization of that tradition and the Piagetian tradition in cognitive development.)

Two of the principal objections to the theory theory are, in fact, reminiscent of similar objections to Chomsky's account of linguistic knowledge as rules, and I would give replies that are very similar to his. Scientific theory formation, like the canonical kind of rule (say a traffic rule), is supposed to be typically conscious and even self-conscious; scientists consciously and reflectively consider how to gather evidence and how evidence tells on their specific hypotheses. The kinds of theory formation we see in children, the kinds that lead to everyday knowledge, do not, on the face of it, seem to be consciously accessible in the same way. In particular, children may not consciously assess evidence and consider its impact on theories. In the same way, of course, linguistic rules are not consciously accessible in the same way as traffic rules.

But, as Chomsky himself argues, it is difficult to see why conscious phenomenology of a particular kind would play an essential role in finding things out about the world. A characteristic lesson of the cognitive revolution in general, and Chomskian cognitive naturalism in particular, is that human beings (or, for that matter, machines) can perform extremely complex feats of representation without any phenomenology at all. It is rather characteristic of human cognition that it is largely inaccessible to conscious reflection. Why should this be different in the case of scientific knowledge?

In fact, one might be rather skeptical about how much real scientists, as opposed to the idealized scientists of philosophy, have conscious access to their theory-formation mechanisms. But even if scientists are, in fact, sometimes reflective about their theory-formation processes, it seems, at least, much too strong to say that this is a necessary condition for theory formation in science. It seems unlikely that it is the reflective phenomenology itself that gives scientists their theory-formation capacities or that gives those theories their epistemological force. Rather, it seems that some set of abstract rules and representations must underlie the flow of phenomenal experience in science, as they do in language.

A second objection, again reminiscent of the objections to Chomsky's account

of language, is that science is a socially constructed enterprise rather than an individual psychological phenomenon. One might have thought, in fact, that this objection applied more strongly to language. Again, though, Chomsky's own arguments seem quite applicable. Even if science, like language, is part of a rich social context, it nevertheless ultimately depends on individual cognitive processes in individual human minds. The social character of science at some level must boil down to individual scientists making individual decisions about which theories to formulate and to accept. Moreover, the socially oriented view of philosophy of science has always had a difficult time explaining how science gets it right. It has been difficult to reconcile with scientific realism. Just as it is hard to see how phenomenology, by itself, could lead to veridicality, it is hard to see how a particular social structure, by itself, could do so. Again, pursuing Chomskian cognitive naturalism, I would argue that some abstract, individual set of psychological rules and representations underpins the social interactions of science.

So Chomsky, at least, ought to be unpersuaded by objections based on the apparently conscious, social character of science, in contrast to the largely unconscious, and individual, character of cognitive development in children. The same arguments he applies to language can be applied to science. In both cases we can postulate that, underlying the conscious, social phenomena, there is a cognitive system, a set of rules and representations that accomplish certain ends.

Universality, Uniformity, and Learning

Another difference we might point to between children and scientists is that children converge on roughly similar representations at roughly similar times, both in language development and also in their understanding of the everyday world. It might be objected that scientists do not always show this sort of uniform development, and that this weighs against the theory-formation view. This would be a more serious objection from a Chomskian point of view, given that Chomsky has emphasized the universality and uniformity of language development as evidence for innateness.

The theory theory proposes that there are powerful cognitive processes that revise existing theories in response to evidence. If cognitive agents began with the same initial theory, tried to solve the same problems, and were presented with similar patterns of evidence over the same period of time, they should, precisely, converge on the same theories at about the same time. These assumptions are very likely to be true for children developing ordinary everyday knowledge. Children will certainly start with the same initial theory and the same theory-formation capacities. Moreover, the evidence is ubiquitous and is likely to be very similar for all children.

Notice, however, that, for scientists, these basic assumptions are not usually

going to be true. In science, the relevant evidence, far from being ubiquitous, is rare and difficult to come by, and often must be taken on trust from others. The social mechanisms of deference, authority, and trust, are, like all social mechanisms, highly variable. Moreover, different scientists also often begin with different theories, and quite typically approach different problems.

In fact, when the assumption of common initial theories and common patterns of evidence, presented in the same sequence, does hold, scientists, like children, do converge on a common account of the world. Indeed, even the timing of scientific discoveries is often strikingly similar, given independent labs working on the same problem with a similar initial theory and similar access to evidence (hence all those shared Nobel prizes). This convergence to the truth itself is the best reason for thinking that some general cognitive structures are at work in scientific theory change. Scientists working independently converge on similar accounts at similar times, not because evolutionary theory or the calculus or the structure of DNA (to take some famous examples) are innate, but because similar minds approaching similar problems are presented with similar patterns of evidence. The theory theory proposes that the cognitive processes that lead to this convergence in science are also operating in children.

This example may be generalized to make another point. The uniformity and universality of development of a cognitive ability does not in itself speak to whether it is a result of an innate device or of learning. By definition, a learning mechanism with the same initial state given uniform evidence will show as much convergence, universality, and uniformity as an innate device.

In fact, we have good reason to believe that this point is not lost on nature. Apparently, evolution doesn't much care whether convergence is the result of an innate device or of learning mechanisms, provided that the results are good, good because they are either functional or veridical. Closely related species of animals may use very different mixes of innate structure and learning to acquire very similar abilities. To take a very simple example, there is a great deal of variation among different songbirds in the acquisition of song. Some birds appear to have particular songs innately programmed, others rely on a tutor in a critical period, others don't have a critical period, for some the tutor must be of the same species, others will learn a "foreign" song if the tutor interacts with them in the correct way, and so on.

The difference between a learning account and an innateness account is not that an innate mechanism can lead to uniform outcomes and a learning mechanism can't. It is that a learning mechanism can lead to either uniform outcomes when the evidence is uniform, or varied outcomes when the evidence is varied. The innate mechanism should always lead to uniform outcomes. In the case of the theory theory, the learning mechanism leads to generally uniform outcomes in children, where the evidence is uniform, and progressively less uniform outcomes in adults and in organized science where the evidence is much less uniform.

This is once again part of why the poverty of the stimulus argument is so central for an innateness argument. If we could point to a sufficiently powerful learning mechanism, like scientific theory formation, and also show that the input to that mechanism was relatively uniform in many cases, we could predict that in those cases the development would be as uniform and universal. The argument for innateness only works if we think there are no such mechanisms.

Theory Formation and Language

Chomsky's innateness hypothesis is intended to apply quite widely, and it has been applied quite widely. Nevertheless, Chomsky himself clearly applies it centrally to language and, particularly, to syntax. In contrast, the theory theory has primarily been applied to our everyday knowledge of the world, our everyday understanding of biology, physics, psychology, and so forth. How does the theory theory relate to language acquisition?

First, the theory theory clearly can be, and has been, applied to understanding semantic and lexical capacities. In particular, there are a number of "theory theory" accounts of our understanding of many types of words. For instance, some accounts suggest that many nouns encode "natural kinds" which are further defined by particular theories (Carey 1982; Murphy and Medin 1985; Keil 1989; Gelman and Coley 1991). We have shown that the acquisition of early words is, empirically, closely related to the child's developing theoretical understanding of actions, appearances, and kinds (Gopnik 1984; Gopnik and Meltzoff 1986, 1987). Bartsch and Wellman (1995) have shown that children's acquisition and use of mental-state verbs tracks their development of successive theories of mind. Children seem to understand the meaning of the words they hear in terms of the theories they have; they treat the words of natural language the way that scientists treat theoretical terms. Moreover, rather than reflecting some fixed set of semantic primitives, children's understanding of words changes in parallel with their changing theoretical understanding of the world. Finally, language itself seems to play an important role in theory formation. We have also shown empirically that the words children hear influence the development of their theories (Gopnik, Choi, and Baumberger 1996).

Since semantics, by definition, relates linguistic expressions to our understanding of the world, and I have argued that our everyday understanding of the world is theory-like, this is not surprising. Moreover, insofar as semantics provides a foundation for syntax, theory formation also may play a role in syntactic development. We seem to use theory formation to develop an understanding of the meaning of words and sentences, and, as many people have argued before, that understanding might itself play an important role in developing more strictly syntactic abilities.

No one so far has applied theory-formation ideas to the development of syntax

itself. It is possible that syntax is simply not the sort of thing you can have a theory about. In particular, theory formation depends on a kind of realism, it depends on the idea that there is something out there that you can have a theory about, and that you can get a continuous stream of information from the world about that something out there. Interestingly, syntactic development is not like this. There is nothing out there that syntactic representations are representations of. Knowing a syntactic structure and having a syntactic structure are just the same thing. Unlike other kinds of knowledge, adult syntactic knowledge isn't defeasible even in principle. The phenomenon of creolization, which is often invoked in support of innateness, is a good example of this (Bickerton 1981). In creoles children don't develop an incorrect account of the pidgin language, which is subject to correction; they simply develop a new language.

But even if theory formation can't explain syntax for this reason, this outcome should not be an altogether happy one for Chomskian theory. For the same arguments that say that syntactic knowledge is not a kind of theory call into question whether it is really a kind of knowledge either. Chomsky rightly argues that philosophical attempts to define knowledge in terms of necessary and sufficient conditions are otiose; he thinks that knowledge is whatever a science of knowledge says it is. But if we take the view of cognitive naturalism we can at least say that the cognitive natural kinds that emerge from the recent research, the varieties of knowledge, look very different from the Chomskian picture.

There is a kind of knowledge that looks very much like the classical philosophical view of knowledge, a kind of knowledge that involves defeasible representations of an independent reality. The theory theory provides a naturalized account of such knowledge. There is also a kind of "knowledge" that is very different. It may be that we use representations and rules to guide our actions and, in particular, to coordinate our actions with the action of others, but not to get to a veridical account of the outside world. In this respect syntactic abilities look much more like certain kinds of social abilities, or perhaps motor abilities or musical abilities, than like scientific knowledge or our everyday understanding of the world. This doesn't mean that we can't reach a naturalistic understanding of these abilities, or that they don't involve rules and representations, or that we can't call them knowledge if we want to. It does mean that they are very unlike other kinds of knowledge.

The innateness hypothesis may apply to these kinds of knowledge. But there also might be learning mechanisms that lead to these representations, though they might be quite different from theory formation. The answer to the poverty of the stimulus argument here would have to be a rather different one. It would depend on the richness of our everyday capacities to develop certain kinds of actions in coordination with the actions of others. In fact, those who propose alternatives to innateness for syntax depend on just such arguments (e.g., Slobin 1985). If researchers could point to specific learning mechanisms that allowed us to coordinate our actions with those of others, given certain types of interactions,

they could oppose those mechanisms to the innateness hypothesis, just as I have opposed theory formation to innateness (for an example of such an argument for a learning mechanism for phonological knowledge see Kuhl 1994).

In any case, cognitive naturalism and the innateness hypothesis were exciting because they were supposed to explain language and knowledge in general, not just syntax. Even if Chomsky is right that syntactic structure is innate, syntax comes to look like a relatively isolated phenomenon using very different mechanisms than the rest of cognition, and even of semantic and lexical knowledge. In fact, sociologically speaking, at least, something like this seems to have come to pass. Where twenty or thirty years ago every cognitive psychologist knew the latest theories of syntax and employed them as models for cognition in general, now hardly any psychologists know even the names of the theories.

If we take the innateness hypothesis as a broad claim about the nature of language and mind, then it is clearly seriously challenged by the theory theory. If we take it as a narrow hypothesis about a syntactic capacity that is quite different from other linguistic and cognitive capacities, then it may survive, but it seems a rather pyrrhic victory.

On the other hand, if we think that Chomsky's real contribution was his cognitive naturalism, then the influence of Chomsky's ideas is both broad and important. They have allowed us to make real progress in solving the problem of how we come to understand the world around us. Cognitive naturalism has helped us to understand the myriad devices we use to get to the truth about the world around us, from human vision to science itself. It has helped us to begin to solve the ancient problem of knowledge.

Chomsky's reply: pp. 316–25.

References

Atran, S. 1990: *Cognitive Foundations of Natural History: Towards an Anthropology of Science*. New York: Cambridge University Press.

Barkow, J. H., Cosmides, L., and Tooby, J. 1992: *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. New York: Oxford University Press.

Bartsch, K. and Wellman, H. M. 1995: *Children Talk about the Mind*. New York: Oxford University Press.

Bates, E. and Elman, J. L. 1992: Connectionism and the Study of Change. In M. H. Johnson (ed.), *Brain Development and Cognition: A Reader*, Oxford: Blackwell.

Bickerton, D. 1981: *Roots of Language*. Ann Arbor: Karoma Publishers.

Carey, S. 1982: Semantic Development: The State of the Art. In E. Wanner and L. R. Gleitman (eds.), *Language Acquisition: The State of the Art*, New York: Cambridge University Press.

Carey, S. 1985: *Conceptual Change in Childhood*. Cambridge, Mass.: MIT Press.

Eimas, P. D. 1975: Speech Perception in Early Infancy. In L. B. Cohen and P. Salapatek (eds.), *Infant Perception*, vol. 2, *From Sensation to Cognition*, New York: Academic Press.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., et al. 1996: *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, Mass.: MIT Press.

Gelman, S. A. and Coley, J. D. 1991: Language and Categorization: The Acquisition of Natural Kind Terms. In S. A. Gelman and J. P. Byrnes (eds.), *Perspectives on Language and Thought: Interrelations in Development*, Cambridge: Cambridge University Press.

Gelman, S. A. and Wellman, H. M. 1991: Insides and Essence: Early Understandings of the Non-obvious. *Cognition*, 38 (3), 213–44.

Giere, R. N. 1992: *Cognitive Models of Science*. Minnesota Studies in the Philosophy of Science, vol. 15. Minneapolis: University of Minnesota Press.

Gopnik, A. 1984: Conceptual and Semantic Change in Scientists and Children: Why There are not Semantic Universals. *Linguistics*, 20, 163–79.

Gopnik, A. 1988: Conceptual and Semantic Development as Theory Change. *Mind and Language*, 3, 163–79.

Gopnik, A. 1993: How We Know Our Minds: The Illusion of First-person Knowledge of Intentionality. *Behavioral and Brain Sciences*, 16 (1), 29–113.

Gopnik, A. 1996: The Child as Scientist. *Philosophy of Science*, 63 (4), 485–514.

Gopnik, A., Choi, S., and Baumberger, T. 1996: Cross-linguistic Differences in Early Semantic and Cognitive Development. *Cognitive Development*, 11 (2), 197–227.

Gopnik, A. and Meltzoff, A. N. 1986: Relations Between Semantic and Cognitive Development in the One-word Stage: The Specificity Hypothesis. *Child Development*, 57 (4), 1,040–53.

Gopnik, A. and Meltzoff, A. N. 1987: The Development of Categorization in the Second Year and its Relation to Other Cognitive and Linguistic Developments. *Child Development*, 58 (6), 1,523–31.

Gopnik, A. and Meltzoff, A. N. 1997: *Words, Thoughts and Theories.* Cambridge, Mass.: Bradford Books/MIT Press.

Gopnik, A. and Wellman, H. M. 1994: The Theory Theory. In L. Hirschfield and S. Gelman (eds.), *Mapping the Mind: Domain Specificity in Cognition and Culture*, New York: Cambridge University Press.

Keil, F. C. 1989: *Concepts, Kinds, and Cognitive Development*. Learning, Development, and Conceptual Change Series. Cambridge, Mass.: MIT Press.

Kitcher, P. 1988: The Child as Parent of the Scientist. *Mind and Language*, 3, 217–28.

Kuhl, P. K. 1994: Learning and Representation in Speech and Language. *Current Opinion in Neurobiology*, 4 (6), 812–22.

Leslie, A. and Roth, D. 1993: What Autism Teaches us about Metarepresentation. In S. Baron-Cohen, H. Tager-Flusberg and D. Cohen (eds.), *Understanding Other Minds: Perspectives from Autism*, Oxford: Oxford University Press.

Murphy, G. L. and Medin, D. L. 1985: The Role of Theories in Conceptual Coherence. *Psychological Review*, 92 (3), 289–316.

Perner, J. 1991: *Understanding the Representational Mind*. Learning, Development, and Conceptual Change Series. Cambridge, Mass.: MIT Press.

Pinker, S. 1989: *Learnability and Cognition: The Acquisition of Argument Structure*. Learning, Development, and Conceptual Change Series. Cambridge, Mass.: MIT Press.

Slobin, D. I. 1985: The Cross-linguistic Study of the Language-making Capacity. In D. I.

Slobin (ed.), *The Cross-Linguistic Study of Language Acquisition*, Hillsdale, NJ: Erlbaum.
Smith, C., Carey, S., and Wiser, M. 1985: On Differentiation: A Case Study of the Development of the Concepts of Size, Weight and Density. *Cognition*, 21, 177–237.
Spelke, E. S., Breinlinger, K., Macomber, J., and Jacobson, K. 1992: Origins of Knowledge. *Psychological Review*, 99 (4), 605–32.
Wellman, H. 1990: *The Child's Theory of Mind*. Cambridge, Mass.: MIT Press.
Wellman, H. and Gelman, S. 1997: Knowledge Acquisition in Foundational Domains. In D. Kuhn and R. Siegler (eds.), *Handbook of Child Psychology* (5th edn), New York: Wiley.

11
Replies

NOAM CHOMSKY

Reply to Lycan

A few words of background, before turning to the important issues raised in William Lycan's essay.

Like Lycan, I take it for granted that the natural sciences (sometimes called "physics") seek to discover basic truths about the world. On that assumption, one can assess their accomplishments in various ways.

One view is that the core natural sciences have come close enough to achieving that goal so that any future changes will not affect the prospects for incorporating within them questions at the outer reaches of inquiry; even questions that are considered too remote to pose seriously as topics for investigation, including those that have been of primary concern for centuries in reflection on mental aspects of the world, or even their apparent analogues for insects; notably, questions of choice of action. A less confident stand is that the scientific enterprise remains open and evolving, and that surprises may lie ahead with unanticipated consequences, as in the past.

A different conception is that science does not even aim that high: we can at most try to increase our knowledge while recognizing that "the secrets of nature, of things-in-themselves, are forever hidden from us," Richard Popkin's (1979) description of a significant reaction to the Cartesian skeptical crisis, becoming more firmly entrenched after Newton undermined the hopes of the early scientific revolution, thus (as Hume saw it) restoring nature's "ultimate secrets to that obscurity in which they ever did and ever will remain." On this view, we revert to the "mitigated scepticism" of even pre-Newtonian English science, acknowledging "the impossibility of finding 'the first springs of natural motions'" (Joseph Glanvill, Henry 1986), and with more modest expectations generally than had once been hoped: we have "practical means for determining truths adequate for

human purposes" though no way to "overcome complete theoretical doubt" (Popkin 1979: 213). In part for different reasons, that becomes a guiding theme of post-Newtonian science.

A qualified version of this view is the Russellian conception that Strawson refines in his essay: that science can aim to give "a merely structure-specifying description," without telling us "what the physical actual is" beyond its formal structure, perhaps somewhat more.

Here I should, perhaps, correct a misunderstanding about Lycan's epigraph, a quote that appears to dismiss ontological pursuits as generally pointless, a form of harassment. As Lycan notes, it is taken out of context (and is not quite accurate). The context is relevant (Chomsky 2000: 77). After citing Popkin's observations on limits of science in the post-Cartesian conception, I wrote that:

> It may well be of interest to proceed beyond [these limits], but if so, the place to look for answers is where they are likely to be found: in the hard sciences, where richness and depth of understanding provides some hope of gaining insight into the questions. To raise them with regard to inquiries barely attempting to gain a foothold is pointless, scarcely more than a form of harassment of emerging disciplines.

The import is rather different.

In its Kantian adaptation, the conception Popkin describes influenced modern ethology, notably in Konrad Lorenz's notion of the "biological a priori," the "hereditary dispositions to think in certain forms," which he thought might be traced to "the innate working hypotheses present in subhuman organisms" (see Chomsky 1968 for quotes and discussion) – the term "hypothesis," of course, is used by Lorenz metaphorically, not in the literal sense of the theory theory (see Gopnik). As discussed elsewhere, I think it makes good sense to relate these ideas to those of Peirce on abduction and to trace them to earlier rationalist conceptions of the origin of our modes of interpreting and comprehending the world. From this perspective, we regard science as a human enterprise, making use of innate "cognoscitive powers" – perhaps a special science-forming faculty, a component of mind guided by certain "regulative ideals" (see Poland and reply). The reach of these powers is a matter of fact, not *a priori* stipulation.

Whatever one's stand on these matters, we should at least accept one conclusion that was recognized centuries ago, and has become a commonplace in the history of science: "a purely materialistic or mechanistic physics . . . is impossible" (Koyr 1957: 210), and however we may have "accustomed ourselves" to notions "hovering in a mystic obscurity between abstraction and concrete comprehension," the doctrines of "genuine Materialists" were abandoned once Newton's discoveries were assimilated and the scientific imagination was freed to explore even more "mystic" properties of the world with no regard to common-sense understanding (Lange 1925).

Even if Lycan's confidence in the present state of fundamental physics proves to be justified, the implications for "physicalism" in any serious sense of the term seem to me slight. We will have found, as a matter of empirical fact, that linkages exist among bodies of doctrine that seem equally remote from commonsense understanding. And in the background remains the question of whether it even makes sense to think that the human enterprise of science seeks to unravel the true "secrets of nature" or, more modestly, is an effort to make the best sense we can of selected aspects of our experience, maintaining a posture of "mitigated scepticism." See Poland and Strawson for relevant discussion.

Lycan offers several reasons for believing that the natural sciences as they now stand suffice for accounting for mental phenomena, but though the conclusion might turn out to be correct, I do not find the reasons he offers compelling.

The first is based on the fact that people are made up of molecules, which are "ordinary matter." But from that it does not follow that future discoveries about "ordinary matter" will not affect how it behaves in particular configurations and circumstances. Magnets, bees, and humans are made up of ordinary matter, but have quite different properties.

Lycan poses his "second qualm . . . in the form of a dilemma." One horn is "loony": that if physics comes to "incorporate entities and principles of hitherto unrecognized character," these "will be localized" in central nervous systems. As he interprets this thesis, it is doubtless "loony." But there is another interpretation in which everyone assumes it for current physics as well: namely, if we rephrase it to state that *the effects of* these entities and principles will be localized; perhaps in CNSs, in the case in question. Uncontroversially, whatever they are, they function quite differently for magnets, bees, and humans; that is, when differently localized. Possible new "entities and principles" will not "shy[ly] . . . hide in brains," except in the sense that such localization is taken for granted generally. Open-mindedness about the future of science does not lead us to maintain that "the mind involves new entities and principles that are confined within central nervous systems and not found in computers and auto engines," the conclusion that Lycan's "materialist claim rejects"; only that new entities and principles might be discovered that hold for all of nature, and yield different effects in the CNS, computers, auto engines, magnets, insects, . . . as do those currently assumed.

Hence, I am not convinced that these considerations "lend some support to the idea that reduction of mind to matter, if possible at all, does not in fact wait upon any expansion of physics." They seem to leave the situation where it was.

Furthermore, I have not been concerned with the question of "reduction of mind to matter," and do not even understand what the question is. As Lycan notes, I use the term "mind" with no ontological import: rather, as an informal way of referring to the "study of the body – specifically the brain – conducted at a certain level of abstraction" (see Lycan's note 7). More narrowly, referring to that part of the study of (mostly) the brain concerned with mental phenomena,

which are just things in the world – part of "whatever there is" alongside of other phenomena (optical, electrical, chemical, . . .), with delimitations that are not sharp or principled and may be modified as understanding progresses.

I also see no reason to question the general conclusion reached long ago that thought is "a little agitation of the brain" (Hume), or "a secretion of the brain" that should be considered no "more wonderful than gravity, a property of matter" (Darwin). In the standard contemporary version, "Things mental, indeed minds, are emergent properties of brain" (Mountcastle 1998), a close paraphrase of Joseph Priestley's conclusion two centuries earlier that "the powers of sensation or perception and thought" are properties of "a certain organized system of matter"; properties "termed mental" are "the result [of the] organical structure" of the brain and "the human nervous system" generally. Priestley, of course, could not account for the emergence, and at the conclusion of the Decade of the Brain (1990–9), little has changed in that regard: "these emergences are not regarded as irreducible but are produced by principles that control the interactions between lower level events – principles we do not yet understand" (Mountcastle 1998).

From this point of view, there is no place for Lycan's problems about "reduction of mind to matter." Mental phenomena are properties of matter, a term with no definite positive content in advance of understanding, but just the general term used to refer to the components of what Strawson calls "the whole of concrete reality." And I agree with Strawson on the pointlessness of most debate about these issues.

Consider Lycan's concerns about "the important distinction to be drawn before the sorts of things Descartes or Hume thought minds were and the sorts of things twentieth-century materialists think they are." Current conceptions of the nature of matter and its aspects (including mental phenomena) are radically different from those of earlier periods. Talk of "what minds are" makes some sense for Descartes, but not in the post-Newtonian tradition (with earlier antecedents) just illustrated. I'm putting aside here matters of textual interpretation, on which I don't entirely agree.

Lycan thinks I "overstate the case considerably in saying that the mind–body distinction has no metaphysical import," arguing that the import is substantial. What I wrote, however, was a little different; his interpretation may be based in part on misunderstanding of the epigraph, already noted. Thus, he believes that I am using "physical" as an epistemic rather than ontological term, basing this conclusion on a quote from an unpublished lecture. The context here is discussion of the collapse of a substantive "mind–body" problem after Newton "exorcised the machine," leaving the ghost intact. We are then left with the world, with its various aspects, including the mental aspects. We can call it "the physical world" if we like, adding nothing, taking "physical" to be what Strawson calls "the ultimate natural-kind term." Or, as quoted in Lycan's (e) from the same lecture of mine, "we can continue to distinguish 'physical' or 'material'

from 'mental,' but recognizing that the usage is only a descriptive convenience, with no metaphysical import." If we adopt this usage (NB: this terminology), then in practice "the domain of the 'physical' is nothing other than what we come more or less to understand, and hope to assimilate to the core natural sciences in some way" – the comment that Lycan interprets as insisting on an "entirely epistemic" characterization of "the physical." In print I've described my own usage (Chomsky 2000: 75, 106): "I will be using the terms 'mind' and 'mental' here with no metaphysical import," apart from whatever "metaphysical import" there is to the distinctions among various aspects of the world: mechanical, electromagnetic, optical, organic, mental, etc., not identical, but also not fundamental divisions of nature.

Lycan argues that a lot more is at stake, but I am not convinced. I also don't quite understand his examples. He holds that the materialist "distinguishes between the physical things that are also mental and those that are merely physical." This is not my terminology, but I don't understand his. What does "merely physical" mean? Are "mental things" the only things that are not "merely physical"? How about magnets? Stable molecules? Insects? What is the import of the word "things"? Are the phenomena of nature generally "things"? Whatever the answers, what are some of the "mental things" that are not "merely physical," according to Lycan's materialist?

Lycan's formulation of a materialist thesis, just quoted, requires a positive, more or less determinate account of the "merely physical." It won't do to say "current physics, or something near enough like it as not to matter." If the empirical assumption of (essential) completeness is correct, then "mental things" are "merely physical"; if the assumption is not correct, nothing else is merely physical either, strictly speaking. I know of no other proposal, since Newton exorcized the machine.

Lycan thinks there is a still "deeper" distinction that does not "presuppose the truth of materialism" (which I take to be the thesis that the world of concrete reality is the way it is; see Strawson). To illustrate, he cites various traditional concepts of mind that have no place in post-Newtonian science: Descartes's thesis that the soul is not "an organization or arrangement of matter," or the thesis that the soul is "indestructible, immortal." I do not see the relevance of these examples. Descartes was proposing a scientific theory; it did not survive Newton's undermining of the concept of body, which left no empirical basis for postulation of a second substance to account for what appeared to go beyond the bounds of mechanism. Within the context of scientific inquiry (Descartes's major concern, I think), the "metaphysical import" is comparable to the postulation of phlogiston, Newton's "subtle aether," or Lavoisier's "caloric." And if postulation of an immortal soul is held to be relevant in the present context (I do not see why), then why not the alleged demonstration that the properties of motion must be of divine origin (Robert Boyle and others)?

Turning to functionalism, Lycan objects to my statement that it is inappropriate

to adopt the slogan that "the mental is the neurophysiological at a higher level" (M-N) as a characterization or definition of "the mental." One reason I mentioned is that in many cases "the most credible assumptions about what the brain is and what it does are those of the computational theories"; the assumption that there is also some unknown neurophysiological account is, for the moment, "a leap of faith," so that what we should say, on naturalistic grounds, is that the neurophysiological may be "the mental at a 'lower level'" (N-M). Lycan glosses N-M as the assertion (his (f)) that "we should reject or at least not accept" M-N. That's incorrect. The passage states that we can accept it, restated as N-M, as an empirical hypothesis. He says that it is not obvious how I mean "to establish (f)." I don't mean to do so, because I don't accept (f). He proposes a possible argument, but it is not mine. Rather, my argument is as stated: for language, insect navigation, and many other phenomena of the world, the conclusions about the brain offered by the computational theories are currently more credible than those provided by neurophysiological theories (which do not, in fact, reach this far). The phenomena under consideration are "mental," in the sense in which I am using the word; Lycan too, if I understand him correctly. That neurophysiological theories are currently lacking is commonly recognized (see comments on Poland for a few examples).

Lycan's lucid review of the history of functionalism brings out the steps at which I've parted company. One has to do with "species chauvinism." Here, functionalism departs sharply, I believe, from the natural sciences, which "do not abstract in principle from the realization of postulated mechanisms, and are open to change of theory if something relevant is learned about them . . ., and do not care about tests and criteria for some category of performance except insofar as they lead toward understanding of the real systems under investigation" (Chomsky 1997, discussing formulations of "the computer model of the mind" and the pointless problems that would arise if analogues were adopted in the natural sciences). If we drop the condition of abstraction in principle that is crucial for the formulations quoted, then questions of species chauvinism do not arise any more than they did for chemistry throughout its modern history, when it postulated properties of unknown entities that could be satisfied by entities of many kinds, a matter taken to be of no significance. The problem is to find how things work in this world.

For this reason alone I don't find Lycan's "letter of recommendation for Functionalism" convincing. But there are others. Descartes's dualism was not abandoned because he could not solve "the interaction problem," but because his problem could not be posed; Functionalism did not repair a meaningful flaw in the Identity Theory (by addressing species chauvinism); and I feel rather skeptical about that theory too. Computational theories of language, insect navigation, etc., require no Identity Theory. They are theories about the nature of the organism (mostly its brain) that have to be judged on their merits as explanatory theories, like others. Prior to unification with core physics, chemistry needed no Identity

Theory, surely not one that linked it to the physics of the day, which had to be radically revised to be unified with chemistry; and the claim of leading scientists that chemistry (atoms, etc.) should be regarded only as a kind of calculating device is now understood to have been completely in error. I've suggested elsewhere (Chomsky 2000: 108ff) that these lessons apply to the study of mental aspects of the world, and should be taken quite seriously. There are all sorts of differences, of course, but enough similarities, I think, to suggest that neither the Identity Theory nor the concern for species chauvinism, understood in the framework of the natural sciences, will survive the comparison.

For similar reasons, I do not find Lycan's "direct argument" very convincing. For naturalistic inquiry, there is no interest in taking "mental types" to be non-biological, any more than there would be in defining "chemical" or "optical" types that share some properties of chemical and optical aspects of the world – except, irrelevantly here, to develop abstract models simple enough to yield insight into the real objects of concern. The computer analogy can be useful as a stimulus to the imagination, much as mechanical automata were for seventeenth- and eighteenth-century scientists. But they would have gone down the wrong path if they had avoided species chauvinism by establishing a non-biological "digestion type" that included de Vaucanson's duck, which simulated the process. Rather, they viewed these constructions as models, which might give some insight into the real objects that concerned them, not going on to "metaphysical speculations" about non-biological "types."

In the same connection, Lycan mistakes my comments on computers as a "refreshing" endorsement of "biological teleology." They offer no endorsement, but rather disclaim any teleological import, and are therefore not threatened by the fact that "natural teleology is hardly unproblematic" (as I have always assumed; see Millikan and reply).

Lycan also mistakes my comments on eliminative materialism. I do not say (his (h)) that it is "methodologically just silly" (though some do; see Strawson). Rather, that we cannot evaluate the thesis until crucial gaps are filled, those indicated in the comments he quotes. His response does not deal with my actual objections, so I will skip them.

Another letter of recommendation for Functionalism that I don't find persuasive is Lycan's observation that it shares with eliminative materialism "the claim that mental expressions are the theoretical terms of a folk or commonsense theory." There is every reason to undertake serious study of such conceptions in various cultures (see replies to Egan and Millikan; whether "theory" is the right notion is another question (see Gopnik and reply)). But there is no reason to expect that the commonsense conceptions subjected to investigation will survive serious inquiry into mental phenomena beyond its earliest stages, and they do not. The same is true, uncontroversially, of folk or commonsense physics. Lycan observes that if it is discovered that "ordinary folk mental concepts are ultimately unexemplified" (in the sense in which ordinary folk physical concepts are), it will

follow that "the Functionalist type-identifications are empty." That outcome, which would hardly be surprising, seems to me just another indication that Functionalism has taken the wrong course; in this case, miscasting ethnoscience as the natural science of the mind. There are no implications for computational theories that are within the latter category.

Let's turn next to Lycan's "solid objection" to my "claim" (observation, I think) that eliminativism remains obscure "until some account is given of the bounds of the material." He argues that the eliminativist can avoid the problem by adopting the "official thesis" that "no mental ascription has ever been true because nothing has ever been in a mental state." Adopting Lycan's account of the shared principle of eliminativism and Functionalism, the official thesis takes "mental state" to be a theoretical term of folk theory. The status of the official thesis is then comparable to that of the thesis that no "physical ascription" has ever been true because nothing has ever been in a physical state (as such states are construed in one or another kind of folk physics). Both theses are presumably true, for uninteresting reasons: natural science quickly departs from folk theories, and it is presumably on to something when it does so. Lycan argues that my criticism "just goes wide" of the official thesis. I'm happy to accept that objection. I did not have in mind a version of eliminativism and Functionalism based on the thesis that folk science will not long endure within naturalistic inquiry. If that is the official thesis, it seems too uninteresting to merit discussion.

On "problems and mysteries," Lycan's interpretation of what I meant (note 24) is accurate: I had in mind "serious *empirical* inquiry." By that standard, I don't think that the traditional questions have been addressed, and we can only speculate about whether they fall in the domain of "mysteries-for-humans." That category might be quite far-reaching. I've cited Hume's reaction to Newton's discoveries, which reflected Newton's own views and subsequent practice: the "hard problems" of earlier years were not so much solved as bypassed (see reply to Strawson). The traditional issues of will and choice remain off the agenda of empirical inquiry. There are valuable studies of organization of motor activity, but not on why an animal does this or that (beyond the most elementary cases). The observations that motivated the problem of other minds for the Cartesians remain unexplained (Chomsky 1966, 1968), and in fact are not empirically investigated. Helmholtz observed that we can shift attention "at will," yielding varied perceptual experiences of the same visual field. There is important recent work on the mechanisms involved (Kanwisher and Downing 1998), but not into what is concealed by the phrase "at will." In another domain, as far as empirical understanding is concerned, we are as much in the dark as ever about what it is that I know that Russell's blind physicist, who knows all of physics, does not (Russell 1927; see Strawson's valuable discussion of this too-much neglected work).

These seem to me fair conclusions, even in the light of the very instructive and illuminating work on these topics in recent years (including Lycan's fine

study (1987), which in the light of his note 2, I can now reread not only for enlightenment but even with a touch of pride).

Reply to Poland

I find myself in close agreement with the main thrust of Jeffrey Poland's "methodological physicalism." A few brief reactions:

The methodological physicalist is committed to scientific rationality and to unification, recognizing that this is at most a hope that might not be realized, either because nature really is not unified in the intended sense, or because human cognitive capacities are not capable of discovering that unity. They are, after all, a biological system with its specific scope and limits. There are surely no grounds to accept Peirce's belief that human cognitive capacities "were designed to give us a generally veridical view of the world," or any more recent variant of it.

On the contrary, there seems good reason to adopt the conclusions of Locke, Hume, and post-Newtonian science generally that the hopes of the early modern scientific revolution are unattainable: in some fundamental sense the world is unintelligible to us, and we have to reduce our sights to the search for intelligible theories. Even that has been strongly contested by mainstream physicists, for example, in the critique a century ago of atomic theory and even the idea that physics should go beyond establishing quantitative relations between observable phenomena (Heilbron 1964 (Chomsky 1985 for some quotes); Lindley 2001). We cannot easily dismiss the conclusion that motion has effects "which we can in no way conceive motion able to produce" (Locke), or other reactions to the skeptical crises and scientific advances of the seventeenth century. See other replies for some examples. For the basic objections to "physicalism" that Poland outlines, I think contemporary critics can claim little originality.

What conclusions should one draw from the more modest conception of the enterprise of science that took shape in that context? One possibility is Poland's methodological physicalism, based on the "regulative ideals" that he identifies, which perhaps take their place alongside others, such as Galileo's conviction that nature is perfect and that it is the task of the scientist to prove it, an idea with considerable resonance since. An alternative, not inconsistent but with a different emphasis, is that scientific inquiry should be opportunistic, seeking understanding where it can be found, with an eye towards eventual unification but without taking confident stands on matters that go beyond what we comprehend. That was the lesson drawn by eighteenth-century chemist Joseph Black, who recommended that "chemical affinity be received as a first principle, which we cannot explain any more than Newton could explain gravitation, and let us defer accounting for the laws of affinity, till we have established such a body of doctrine as he has established concerning the laws of gravitation" (Schofield

1970: 226) – the course that was actually followed, as chemistry established a rich body of doctrine, its "triumphs . . . built on no reductionist foundation but rather achieved in isolation from the newly emerging science of physics" (Thackray 1970: 279), which was finally unified with chemistry in the 1930s after undergoing revolutionary changes.

We should, I think, take quite seriously the history of the core hard sciences in very recent years, including this dramatic illustration of what Poland calls "downward incorporation." It also seems to me instructive to attend to the striking similarities between debates about the reality of chemistry up until the time of unification, and those concerning language and mind today (Chomsky 2000: 108ff).

Black's "localist" conception seems to me reasonable, justified by the actual history of science, and appropriate for the study of topics at the outer reaches of naturalistic inquiry today. This point of view is consistent with the decision of Poland's methodological physicalist to restrict the term "physical" to "whatever class of entities and principles fundamental physics identifies, without an identification of a particular physical theory." It is also consistent with views of mine to which he alludes (Chomsky 1968, 1980): that we can anticipate a "physical explanation" for what may be discovered about mental aspects of the world for the uninteresting reason that "the concept 'physical explanation' will no doubt be extended to incorporate whatever is discovered in this domain," just as it was extended to accommodate "numerous other entities and processes that would have offended the common sense of earlier generations." One might choose to use the term "physical" only if such discoveries are linked at least weakly with the core physical sciences, which may change in some unanticipated way, as in the past.

I was surprised to learn, from Poland, that these and other remarks of mine have been interpreted to mean that "integration into a physicalist system" is "inevitable" – that is, that unification is necessarily within the scope of human cognitive capacities. He is right to say that that was not the intent; the passages cited explicitly leave open the possibility that such understanding might lie beyond the reach of the "fixed biological systems" of the human mind. That they *might* seems a virtual truism, quite apart from the rich record of antecedents. It carries with it the possibility that human cognitive powers may not be able to discover the "basic science" into which mental aspects of the world are to be incorporated, or the mode of incorporation if indeed the relation holds – not to speak of the question of whether we can even seriously undertake the quest for "the secrets of nature, of things-in-themselves" (Popkin).

"For many," Poland observes, "incorporation of the mental into physics would be the demise of physicalism"; the reference, I presume, is to "downward incorporation," modifying the fundamental theory as when the theory of motion was "incorporated into mechanics," along with many of the great steps in the physical sciences since. If this qualification is dropped, the demise was anticipated

long ago, for convincing reasons. Locke's remarks, just cited, led to his famous suggestion that God might have "superadded" a "faculty of thinking" to matter just as he added motion, with effects inconceivable to us. Similar thoughts were expressed by many other prominent figures (see comments on Lycan for some examples). Formulation aside, the post-Newtonian world allows no alternative. Poland's methodological physicalism would survive the discovery that even for today's physics, "downward incorporation" is required, as in the past in crucial cases.

Poland is right to observe that I have not tried to offer "a logical demonstration" of the conclusions he discusses about *a priori* and *a posteriori* conceptions of the physical. Internal contradictions aside, it has always been possible to hold without inconsistency that current science is essentially "complete." Also without contradiction, one could hold that human scientific inquiry in principle converges to the truth about the world (apart from historical accident). The most that can be said is that such dogmatism is unreasonable. So it seems to me, at least. The sciences are human creations, the best effort of a particular creature to gain understanding and insight into the phenomena of experience, at least some selection of them. How closely it approximates "the secrets of nature, of things-in-themselves," we cannot say – assuming that we can even give some coherent interpretation to the question. Nor can we say how closely these efforts can converge, even in principle, to answers to legitimate questions that we can pose – "legitimate" in the sense that we can at least imagine some possible answers to them, even plainly wrong answers.

More narrowly, as Poland comments, my feeling is that there are no firm grounds "to suppose that contemporary neuroscience is sufficiently mature to underwrite a solution to the unification problem," even in far more elementary cases. A recent review article on the current state of efforts aimed at "breaking down scientific barriers to the study of brain and mind" concludes that "the neuroscience of higher cognitive processes is only beginning" (Kandel and Squire 2000). Basic traditional questions are not even on the research agenda, and even simple ones that might be within reach remain baffling. One prominent cognitive neuroscientist points out that "we clearly do not understand how the nervous system computes," or even "the foundations of its ability to compute," even for "the small set of arithmetic and logical operations that are fundamental to any computation" (Gallistel 1997). Another, who is optimistic about the prospects for bringing the brain sciences to bear even on creativity in the visual arts, nevertheless reminds the reader that "how the brain combines the responses of specialized cells to indicate a continuous vertical line is a mystery that neurology has not yet solved," or even how one line is differentiated from others or from the visual surround (Zeki 1998). One can entertain the idea that "the mental is the neurophysiological at a higher level," but for the present, only as a guide to inquiry, without much confidence about what "the neurophysiological" will prove to be.

Reply to Strawson

If only for my own benefit, I will review the basic framework of Galen Strawson's careful and illuminating paper, or rather the parts that are a preliminary to his final proposal for resurrecting a version of materialism.

Among the many important conclusions Strawson reaches is that Bertrand Russell was basically right in holding that what we know of the non-mental world does not suffice to show whether it is "different in intrinsic character from the world of mind." In Strawson's words, "we have *no* good reason to think that we know anything about the physical that gives us any reason to find any problem in the idea that mental phenomena are physical phenomena"; "physical" understood here as "the ultimate natural-kind term."

It follows that "almost all the present debate about the 'mind–body' problem is beside the point." The basic reasons were understood well enough centuries ago, particularly in the wake of the collapse of the mechanical philosophy, the last serious effort to develop a positive and purportedly complete account of the "material world"; and they have become more compelling as the natural sciences have pursued their quest along strange and often wholly unexpected paths. As Strawson observes, citing Hume, it was never legitimate to claim that matter as then construed is "definitely – knowably – incompatible with some of it also being a matter of conscious goings-on." And with the demonstration of the total failure of the mechanical philosophy, it is no longer possible even to take a weaker stand, postulating the distinct existence of *res cogitans* as an empirical hypothesis, given apparent limits of mechanism.

It is important to stress, as Strawson does, that matter is no more "incompatible with sensation and thought" than with attraction and repulsion (Joseph Priestley). All are elements of reality that may lie beyond what we can intuitively conceive, but that we can hope to investigate and perhaps to comprehend in some different way.

Among the phenomena traditionally "termed mental," in Priestley's phrase, are what Strawson calls "Experiential phenomena." For these phenomena, Strawson provides "some securely anchored, positive descriptive content," and he keeps to these for reasons he mentions, noting also that there is no need to "draw a sharp line" that distinguishes the broader category from the rest of concrete reality. Similarly, no one should try to do so for other aspects of the world that lend themselves to (more or less) integrated study at a particular stage of understanding; as far as I know, it is only for the study of language that such odd proposals have been advanced.

Strawson stresses that the phenomena termed mental should not be counterposed to the physical (at least on materialist assumptions) but to non-mental aspects of the universe of concrete reality, which includes Experiential phenomena. By having experience, he argues, "we are directly acquainted with certain

features of the ultimate nature of reality," namely Experiential features of the world. The argument seems to me persuasive, but we should, nevertheless, be cautious in endorsing "the view that 'experience is really just neurons firing.'" That goes well beyond Strawson's realistic materialism; one may reasonably doubt that current understanding of the brain permits that high a degree of confidence. As Strawson observes, we should not "slide into supposing that the word 'brain' somehow refers only to the brain-as-revealed-by-current physics" ("physics" meaning natural science).

Strawson's realistic materialism assumes further that "each particular mental phenomenon essentially has non-mental being." Specifically, the Experiential phenomena that are "the only real, concrete phenomena that we can know with certainty to exist" have such non-mental being. "At best an assumption," he observes.

Given the lack of a positive determinate account of the non-mental physical, talk of "the hard part of the mind–body problem" (in recent years, conventionally taken to be the problem of consciousness) "can be seriously misleading." The problem is not clearly posed, and we have no good reason to suppose that the specific topic of concern is any harder "than the problem posed for our understanding by the peculiarities of quantum physics, or indeed – Chomsky might say – by the phenomenon of motion." That's a fair assessment of what I've written. It comes down to saying that we still have every reason to take seriously the opinions of Newton, Hume, and others, who in various ways came to the conclusion that motion has effects "which we can in no way conceive motion able to produce" (Locke). Even before Newton, puzzlement about motion was profound. Newton's precursor Sir William Petty described "Springing or Elastic Motions" as the "hard rock in Philosophy." The obscurity was so great, Robert Boyle felt, as to prove the existence of "an intelligent Author or Disposer of Things." Even the skeptical Newtonian Voltaire felt that the mysteries of motion proved that "there is a God who gave movement" to matter (Henry 1986; Kors 1992). One cannot say, I think, that that "hard problem" was resolved. Rather, it led to a significant revision of the enterprise of science (see reply to Poland).

Let's fast-forward to Strawson's final proposal as to how we might retain harmless and illuminating variants of the traditional concepts *materialist* (*physical*) and *monist*. If I understand correctly, Strawson's realistic materialist monism (RMM) adopts three theses: (1) "the working presumption that current physics is genuinely reality-representing," if only through relations of structural correspondence with the nature of matter; (2) that all matter has non-Experiential being, including Experiential phenomena; and (3) that causal interaction holds throughout the phenomena of the universe. Thesis (2) is a bit different from his earlier assumption that "each particular mental phenomenon essentially has non-mental being." Given that Experiential being does not exhaust the (loosely characterized) domain of the mental, thesis (2) appears to allow for the non-

existence of non-mental being; maybe even for the conclusion that "at a very deep bottom," the universe consists of nothing but bits of information (Wheeler 1994). That would be consistent with Strawson's observation that "we really don't know enough to say that there is any non-mental being," but I presume that RMM wants to stake out a more ambitious position.

Strawson suggests that RMM might be understood as my "methodological naturalism ontologized." But that inflates what should be an innocuous proposal: that we should investigate "mental aspects of the world as we do any others, seeking to construct intelligible explanatory theories, with the hope of eventual integration with the 'core' natural sciences." It was worth mentioning only on the assumption (which I argued is correct) that rejection of this stand "dominates much practice." RMM does "ontologize" the methodological stand, in a way that seems to me quite reasonable, though where the chips may fall, who can say?

Reply to Egan

I think that Frances Egan and I agree about basic issues. I'd like to focus, then, on some problems of terminological and conceptual clarification, keeping pretty much to the order of exposition in her careful and illuminating paper.

Let's begin with ethnoscience, which I understand to be the "branch of naturalistic inquiry [that] studies commonsense understanding": how people interpret "the nature and causes of motion, thought and action, and so on ('folk science,' in one of the senses of the term) . . ." It is an open question whether, and if so how, the conceptual resources of folk science relate to those involved in the reflective and self-conscious inquiry found in every known culture ("early science"), and to the particular enterprise we call "natural science," including its subbranch, ethnoscience, which studies "folk science" (Chomsky 1995).

As Egan observes, "the fact that intentional phenomena are involved in our commonsense understanding [does not] entail that they cannot be studied" within naturalistic inquiry; or even that the commonsense concepts will not find a place in their theoretical constructions. Contact mechanics presumably has an essential place in early cognitive development and in "folk science," and could have turned out to be true, as assumed in the early modern scientific revolution. The same could be true of folk psychology, though the history of science suggests skepticism.

I would not suggest that we "relegate our interests in understanding ourselves as rational agents to ethnoscience, and leave the satisfaction of these interests in the hands of novelists and folklorists" (Egan's interpretation). Ethnoscience is an empirical pursuit that seeks to discover how commonsense understanding, in various cultures and settings, seeks to make some sense of how the world works; different naturalistic enterprises seek to discover how the world actually works. Insofar as the latter can shed some light on human thought and action, that's to be welcomed. Egan may have misread some comments of mine about the actual

reach of naturalistic inquiry – anything that might seriously be called "science." Perhaps statements like this:

> Someone committed to naturalistic inquiry can consistently believe that we learn more of human interest about how people think and feel and act by studying history or reading novels than from all of naturalistic inquiry. Outside of narrow domains, naturalistic inquiry has proven shallow or hopeless, and perhaps always will, perhaps for reasons having to do with our cognitive nature. (ibid.)

That, I think, is highly plausible, but it in no sense implies that we should not try to extend naturalistic inquiry to its limits.

Serious ethnoscience may discover that in various cultures (and in some respects universally) people interpret thought and action in terms of something like what we call in English "beliefs" and "plans" and "purposes," or vital forces and hidden agents, applying such notions to humans, animals, and maybe plants, even inanimate objects. Such discoveries leave open the question of what is really happening when people speak and think and act, or flies struggle in spider webs, or plants turn toward the light, or the sun travels on its course through the heavens.

In some cultures, people may consciously invoke some concept of "mental state" in their efforts to make sense of the world. Ethnoscientific inquiry will seek to determine the nature of their concept, just as chemistry seeks to discover the nature of complex molecules. These are hard tasks, not a matter for casual reflection about conventional practices within some professional communities. Other branches of naturalistic inquiry may choose to distinguish mental aspects of the world from chemical or optical aspects, and correspondingly, to refer to some of the states they postulate as "mental states": perhaps the computational states of bees that are involved in performing and interpreting the waggle dance, or the I-languages of humans, or their visual impressions or feelings or choices and plans. If so, this usage is not to be confused with what may be unearthed by ethnoscientific inquiry into commonsense understanding. We take this for granted in investigating motion and its causes, and should adopt the same stance in studying complex organisms, humans in particular.

Accordingly, I do not understand internalism to be a doctrine that denies that "mental states are individuated by reference to features of the subject's environment or social context," or that holds that "subjects in the same internal (neural) state are in the same mental states" (Egan). Internalism studies internal states, including those involved in what tradition (and common sense) often regard as mental aspects of the world. It may develop a concept of "mental state," but if so, that will be a technical construct, finding its place within a particular context of theoretical explanation, denying or asserting nothing about other usages. As for externalism, it faces a choice. If conceived as part of ethnoscience, it is making the factual claim that people (in our culture, or universally) attribute

thoughts, beliefs, etc., which they individuate by reference to environment or social context, and then faces the task of clarifying and defending that empirical thesis. If conceived as part of (biological) psychology, another branch of the empirical sciences, it is making the claim that among the entities of the world, alongside of complex molecules and (maybe) I-languages, are mental states individuated by environment and social context, and it will again have to explain what these entities are, show how they function, and provide empirical confirmation for its conclusions about these matters. I am personally skeptical about the prospects for either of these endeavors (and it is not clear to me that they have even been undertaken). But it should be clear that they are different ones, and that in either case, the conditions of naturalistic inquiry should be satisfied.

Internalism does not, of course, deny that within naturalistic inquiry the world can be investigated in other ways. It does not exclude the study of human or insect societies, or an inquiry that dispenses with people altogether, regarding them at most as (maybe uninteresting) stages of various biochemical processes. It does seem to me that in the study of organisms, their interactions, and their functioning in the world, internalist approaches have a kind of privileged place, and are presupposed, if only tacitly, even when that is strongly denied. That's a distinct thesis, which would have to be defended by an investigation of actual practice.

Turning to "I-language," it may be a bit misleading to say that it plays a role in my "recent work." It's not a new notion; rather, a new term, introduced in the hope of overcoming persistent misunderstanding of the technical notion "grammar" (Chomsky 1985). Since the origins of work in generative grammar in the 1950s, it has been pointed out that the term "grammar" is being used with systematic ambiguity: to refer to the internal states of Jones's faculty of language FL, and to the linguist's theory of that state. But that usage proved confusing. I therefore suggested that we restrict the term "grammar" ("particular" or "universal grammar") to the theories constructed by the linguist, and refer to the internal state that grammars seek to describe as I-languages ("I" to suggest internal, individual, intensional).

I also do not believe that any of this is "radical" or even novel. This seems to me the way empirical linguistics has generally proceeded (whether the fact was recognized or not). In particular, it is not accurate to speak of "rejection of the notion of a 'shared public language'" as peculiarly mine; it is a commonplace of the empirical study of language. In the technical literature one finds such terms as "Chinese" or "Italian," but these are understood to be merely conveniences, not notions with some place in the theoretical explanation of phenomena of language. The burden of proof falls on those who claim otherwise. To my knowledge, the burden has not been met, rarely even recognized (but see Millikan's essay).

Turning to the issues of intentionality that Egan discusses, a number of questions arise when it is asserted that some relation holds between linguistic

expressions (internal objects generated by the I-language) and extra-linguistic entities: the relation *Sounds* holding between "London" and some category N of noises that are produced when the expression is externalized, and the relation *Refers* holding between "London" and some entity E of the extra-mental world. If such assertions are intended to be empirical claims, then before they can be assessed, it is necessary to explain what the linguistic expression is: if it is taken to be an element of a "public language," the problems are severe indeed. Suppose we take the expression to be an element of I-language (hence ignoring most of the discussion of these topics). We then face the next question: what are N and E? With regard to N, the question is never raised. It is taken for granted that there is no sensible notion N. With regard to E, the answer is assumed to be obvious. But it is not obvious in the least, as we discover as soon as we consider how we understand even very simple words: "London," "city," "river," "water," and so on. It is not at all clear that the intuitive underlying picture of word–object relations is any more appropriate than in the case of sound.

There is another way to interpret assertions about the technical notion *Refer*: one can take it to be a relation between an I-language expression and a "semantic value" that is also internal to the mind/brain. That is, I think, the proper way to interpret the valuable and exciting work in formal semantics in recent years, though we should properly regard it as a form of syntax, not referential semantics, in the traditional sense of these terms: it is a study of symbolic objects and their properties – in this case, internal objects, linguistic expressions and semantic values. And postulation of semantic values faces the same challenges as postulation of other theoretical entities: phonemes, atoms, whatever (see Ludlow's essay and reply).

Egan has somewhat misunderstood my exposition at this point. She regards the postulation of a relation *Refer* holding between expressions and "semantic values" as my thesis. But it is not. In the essay to which she refers I express considerable skepticism about it (Chomsky 2000: 38ff and n. 15), simply pointing out that if such relations and entities are postulated, explanation is needed as to what they are. The interpretation that best conforms to practice, I think, takes the account to be internalist, hence a form of syntax, as it would be if a relation *P-Refer* were postulated between phones and "phonetic values," taken to be internal objects that relate somehow to sound. I do bring up the evidence on anaphora that Egan mentions, but not as a "'best case' for [my] thesis," rather, to show that it provides no support for the thesis, and if introduced in this context, raises a welter of questions that are apparently quite pointless. Left to the side are the questions about the use of language to refer. In papers that Egan cites, and other work back to the 1960s, I've tried to indicate some ways in which, I think, these questions can be profitably pursued, drawing on ideas of early modern philosophy (neo-Platonists, British empiricists), and other traditions, including the Aristotelian conceptions developed further in important work by Julius Moravcsik (1975, 1990) and others. This inquiry goes beyond

syntax, but the study of internal objects does not, whether it falls under phonology or internalist semantics. Such work may be of considerable value as a preliminary to the study of language use, and may provide a correct account of internal processes and structures; but that is a different matter.

It is important to distinguish between use of a word to refer to things and "words referring to things," a distinction discussed fifty years ago by P. F. Strawson. The former usage is normal in English (with counterparts in other languages). The latter is a technical innovation that remains obscure and unexplained (as Strawson observed). The distinction should, I think, resolve Egan's uncertainty about how to understand my "general claim that internalist semantics is a form of syntax"; not really a "claim" so much as an observation about the standard technical meaning of "syntax." When the distinction is observed in the case of the quoted sentences (about painting houses brown), I think the unclarity is overcome. The source of Egan's perplexity is the transition from the statement of mine she quotes (about how words are "used to refer") to "talk of a word 'refer[ring]' [or] expressing properties" (Egan). That's the practice I am warning *against*.

Egan is, however, quite right to say that the remark she cites about relations among expressions and their role in performance "is intended to cover a lot of ground" (and barely touches on the problems) in both of the respects mentioned, rhyme and entailment, and other internalist relations like them. These are serious topics, little understood. Those who believe that some notion of "word–thing reference" enters have to explain what they mean, and provide the evidence. On the sound side of language, no one even tries. On the meaning side, the conclusion is commonly asserted, but without appropriate clarification, to my knowledge.

That brings me to Egan's remarks on content and representation (section 4). Her central thesis is that "representational content serves an important explanatory function in [computational] theories," a position that she contrasts with "Chomsky's claim that questions about the content of the structures postulated are of no theoretical interest." That's not quite my position: rather, my position is as stated in the quote she cites about Marr and Ullman. I do not think that the notion "representational content" has been clarified sufficiently even to have an opinion about the thesis. I entirely agree that we have to go beyond internalist computational theories to find out "what constructing these structures does for the organism." For example, to find out what constructing an internal phonological system does for people, we have to investigate how the phonetic aspects of internally generated expressions enter into articulation and speech perception. The questions that seem to me unanswered are: what is "representational content"? And how does it contribute to this end? In the phonetics case, apparently not at all. What about other cases?

To illustrate the explanatory function of "content ascription" in cognitive theories, Egan gives the example of Marr's "edges," which are "tokened in the

presence of a disjunctive distal property." Evidently, such interactions between an internal computation and a distal property won't be explained within an internalist theory. But what has to be explained goes well beyond the interaction between the postulated internal structure and an external object with an edge. Also to be explained are the interactions typically investigated in this experimental program: with tachistoscopic images, when there is no external object with an edge; or with direct retinal or even optic nerve stimulations if that becomes technically possible. What is the "representational content" in such cases? Or in the case of a Necker cube or Ames room? Or perception of apparent motion when two lights flash? Or perception of depth in a perspective drawing; or a stage more complex, what Rainer Mausfeld calls the "dual character of pictures," namely, "the phenomenon that pictures can generate an in-depth spatial impression of the scene depicted while at the same time appearing as flat two-dimensional surfaces," though in different ways, and only when an appropriate "mental attitude" has been achieved (Mausfield 2001)? The same theory is expected to explain both the "normal case" of co-variation with distal properties in familiar external environments, and the "abnormal cases" that are typically investigated in perceptual psychology. A notion of "representational content" restricted to the former will be of no particular interest, even if it can be devised. It is quite true that a "computational theory that purports to provide an account of human vision would be justly criticized if it failed to recover the spatial properties of the scene before our eyes" – or an account of what the mind/brain constructs (not "recovers") in this and a wide range of other cases.

A better approach, I think, and one that accords more closely with actual empirical practice, is to return to a classical formulation: the senses provide the occasion for the mind to construct ideas (not to be confused with beliefs) by means of its own "cognoscitive powers," largely innate. We definitely want to investigate how postulated computational states enter into performance when we produce or understand an utterance, or perceive a rigid object in motion. But it is not at all helpful, as far as I can see, to introduce the notion "representational content" into such accounts. That would be true even if the notion was properly explained for the "normal case." But I think it is not. Suppose that "the scene before our eyes" is people walking in a crowded street (Descartes's example), or a landscape of forests, lakes, and mountains, or the moon near enough to the horizon to induce the moon illusion, or any other normal case. What exactly is the "representational content"? That problem becomes even more acute when we consider the alleged "representational content" of the expressions of a language, either on the sound or meaning side. No external object can be sensibly identified to satisfy the requirements, at least to my knowledge.

We should also, I think, be cautious about the standard assertion that translation into contemporary philosophical idiom provides an account of how "mental states can have both causal and representational properties." As mentioned, I think the latter are misconceived, but what about the "causal properties"

of mental states? The "causal explanations" and "psychological laws" formulated in this terminology are discovered by reflection, not experiment: we do not discover experimentally, or subject to empirical confirmation, the principle that if Jones believes it will rain and wants to stay dry he'll take an umbrella, *ceteris paribus*. These are our informal modes of giving an interpretation of action as somehow reasonable, formulated in quasi-causal terminology that is fine for ordinary usage, but has little to do with causation in the sense of the sciences (Chomsky 2000: 165ff). Caution is also in order when stating that "of course, we presume that computational mechanisms are *adaptations* . . ." If what is meant is that they evolved, then the statement is innocuous. But "adaptation" may suggest more than that. The bones of the middle ear are a marvelous sound-amplifying system, wonderfully designed for interpreting speech, but they appear to have evolved from the reptilian jaw as a mechanical effect of growth of the neocortex in mammals 160 million years ago. We know far too little about computational systems and their evolution to be able to guess to what extent something similar may be true about them; quite a lot, perhaps, if recent work on optimal design of FL computation is on the right track (the so-called "minimalist program"). More generally, it is a truism that any adaptation takes place within a "channel" of physically available options, and these may have a profound effect on the outcome, as D'Arcy Thompson and Turing argued in classic work. What effect we will not know until we know.

The central problem that troubles me is this. I do not know of any notion of "representational content" that is clear enough to be invoked in accounts of how internal computational systems enter into the life of the organism. And to the extent that I can grasp what is intended, it seems to be very questionable that it points to a profitable path to pursue.

Reply to Rey

Georges Rey covers a good deal of important and often contested ground, and provides a valuable critique of the resort to obscure "intentional talk" in several domains, including an interpretation of linguistic inquiry that has been put forth, though not by me, as he believes. I'll review briefly the picture I've been assuming throughout, focusing on parts that Rey finds problematic, and will then turn to Rey's critique, skipping topics discussed in other comments.

With idealizations that we can ignore, one component of the human organism (mostly the brain, so let's keep to that) is a dedicated faculty of language FL, which passes through a series of states in language acquisition, each an I-language. We seek to discover the nature of these states; e.g., of Jones's current I-language L. Such investigation can be undertaken from various perspectives: among other choices, we can try to construct an account of FL at the level of (I) cells or (II) computational mechanisms or other such conceptions, often called "abstract" in

that they abstract from (often unknown) mechanisms at a "lower level." We also hope ultimately to unify such accounts. Analogues to (II) are familiar in the history of science, chemistry for example (see reply to Poland). The distinctions are not intended to be precise, and are relative to interests: neuroscience is analogous to a level-(II) inquiry relative to particle theory. These remain perfectly legitimate approaches to explaining phenomena of the world even if the unification problem is solved one way or another, often indispensable approaches, for familiar reasons (see Gold and Stoljar 2001 for recent discussion).

In the study of animal intelligence, the choices for (I) and (II) just mentioned are commonly called the "physiological" and "psychological" levels, respectively. Virtually all work keeps to (II); see, e.g., the comprehensive review by Hauser (1998), who calls for research at both levels but presents results on psychological mechanisms, almost exclusively.

Undertaking the study of FL at level (II), we want to find out how the internal structure of Jones's L contributes to accounting for such phenomena as those illustrated in essays and comments here; call them LP (linguistic phenomena). To the extent that this quest is successful, we want to find out (A) how the elements and principles of FL relate to those discovered at level (I) (the unification problem), and (B) how they enter into the many uses of language: specifically (and let's keep to these) (B1) articulation and perception of speech, and (B2) talking about the world. To emphasize that a study of FL is at level (II), it is commonly called a study of the mind/brain.

Assume that L is a computational system that generates expressions E = <PHON(E), SEM(E)>, each an object at the interface with other systems of the organism: PHON(E) interacts with the sensorimotor systems SM, and SEM(E) with the (brain-internal) systems C/I (see Horwich). The study of language use tries to find out how Jones with I-language L proceeds to produce and interpret sounds and tell us about the book he read ((B1) and (B2)) – recognizing that questions of this kind can be studied only in highly restricted ways, even in simple cases (see replies to Lycan and Poland).

Even for vastly simpler systems than I-language, questions (A) remain at the outer limits of research: to repeat, "we clearly do not understand how the nervous system computes," or even "the foundations of its ability to compute," even for "the small set of arithmetic and logical operations that are fundamental to any computation" (Gallistel 1997). There are detailed studies of the computational processes by which an insect carries out path integration, summing values of variables over time and storing them for later use, calculating the sun's position as a function of the time of year and day, etc. (see Gallistel 1999, for review). The computational systems described are properly taken to be an internal component of the insect's nervous system, underlying the capacity for navigation, communication, and other activities. Corresponding events in the nervous system remain unknown. Studies analogous to (B) seek to determine how the computational system is accessed to initiate motion (calls, bird song, etc.); how the

perceptual apparatus is guided by "rules of vision" (Hellman 1998); and so on. But the steps along the way are far from understood, particularly insofar as they involve questions (A). Accordingly, there are well-known gaps in accounts of how such computations lead to action, hard questions for empirical research.

Apart from the complexity, the study of human cognitive capacities faces many further problems, such as bans on direct experimentation and limitations of comparative evidence.

The elementary components of the postulated internal systems of FL are called "features"; the systems are often called "(internal) representations," in approximately the sense in which the computational system involved in insect navigation or bird song is said to be "internally represented," all at the psychological level; this seems to be the sense Rey has in mind when he speaks of "specific representations" that have "contents." The search for unification (namely (A)) is guided by the expectation that corresponding in some manner to such psychological entities there are physiological configurations, also said to be internally represented. The informal usage is standard, and unproblematic if not invested with improper connotations. To take an example virtually at random, recent studies show that if the optic nerve of an animal is "rewired" to connect to the auditory pathway early in life, "the auditory cortex gradually takes on a representation that is normally found in the visual cortex" (Weng et al. 2001); the "representation" is some internal structure R, which is used when the "rewired" animal performs "vision tasks with the auditory cortex." In such performance, R enters into complicated relations with things in the outside world, but it does not "represent" them in anything like the sense in which a photograph of a landscape is said to represent the landscape (not that that notion is trivial).

Uncontroversially, the internal representations PHON(E) are like R in this respect. Features of PHON(E) can be plausibly construed as instructions for articulatory gestures (Halle 1983), but the external manifestation of these varies widely depending on intricate circumstances. I think the same general picture is appropriate for problems of meaning and use, apparently a significant distinction between animal communication and human language (see reply to Ludlow).

We have, then, several distinct uses of the informal notion "represent." In one sense, the psychological and physiological mechanisms are said to be "internally represented," and with an eye towards unification, the psychological mechanisms are said to be "represented in the brain" (though how no one knows, even for elementary cases). The physiological representation R has the properties investigated at level (II), though no one expects point-by-point relations: e.g., that some specific neural structure is dedicated to storing variables, another for calculating the solar ephemeris, etc. And surely no one expects that some isolable part of the organism is dedicated to digestion, or navigation, or language, or any other component that is singled out for investigation in any rational approach to the study of a complex system.

A second usage is of the photo–landscape variety. A technical version of this may be appropriate for animal cognition: a one–one relation between mind/brain processes and "an aspect of the environment to which these processes adapt the animal's behavior" (Gallistel 1990). If so, then for animal cognition the two informal senses of "represent" may be easily confused, since the "internally represented" object R may enter into a "representation relation" (in a different sense) with some aspect of the external environment. One may also, if one likes, say that elements of I-language "represent" sounds and things, having in mind problems (B1) and (B2), but this third usage merely points to problems to be investigated and should not be misinterpreted as a substantive proposal (see reply to Ludlow).

Typically, little care is taken with these informal usages; the basic points are commonly understood. In fairly technical work, the usages are kept distinct. For example, "Logical Structure of Linguistic Theory" (Chomsky 1955), which Rey cites in this connection, postulates internal objects and relations among them; among the internal objects are early counterparts to PHON(E) and SEM(E) (since much revised). Another relation is introduced holding between PHON(E) and outside events, mediated by SM. Both the internal and external relations are informally read "represent." The internal ones are extensively discussed (in effect, they are the generative procedure), but the external one is only mentioned. It is the topic of articulatory and acoustic phonetics, which study how the features of PHON(E) enter into performance (problems (B1)). The broader goal is to reveal the ways in which the I-language (there called "grammar") "provides a basis for a description of how, in fact, language is used and understood" (Chomsky 1955: 75). See Horwich and reply.

For the reasons just reviewed, Rey is exactly right when he observes (or, as he sees it, objects) that I've provided no "general laws of 'representation' or intentionality on which linguistic theory can similarly rely" – similar to Newton's laws. The notions of "representation" and intentionality that he has in mind do not enter into such work, apart from passages that provide informal motivation; accordingly, there is no place for laws governing them. There are specific proposals about the internal structure of the mind/brain, e.g., the computations involved in insect navigation and LP. And there are specific proposals as to how expressions <PHON(E), SEM(E)> are used by a person in making noises and talking about the world, among other actions. But it is understood that a general theory of action lies far beyond reach, even for insects, surely for people. Only selected aspects of these problems are even investigated.

Rey argues that this work is missing a crucial notion, what he calls "intentional content." But his proposal seems to me just to restate open problems in misleading ways. To take his examples, the theory of I-language attributes such features as [+voice] to the mind/brain. The I-language generates expressions E constructed from these elementary features: E is accessed by SM and used to carry out articulatory gestures that eventuate in varied external events. The

general picture is similar to the account of insect navigation or bird song. In these terms, we can offer at least partial accounts for phenomena of the kind Rey mentions. No further illumination is provided if we "clutter up the notation" (to borrow his phrase) in the ways he recommends. What's produced internally must, as he says, "bear some systematic relation to the phenomena in the oral cavity," while also figuring in computations (understood in the standard sense; his qualms about this are misplaced). Current theories give partial answers to these questions, providing computational theories for the mind/brain, and investigating how SM uses the expressions they generate to carry out articulatory gestures. There are many explanatory gaps, among them questions of category (A). We do not help fill the gaps by adding notations and introducing the term "intentional content," which remains unexplained. As elsewhere, we do not expect to find punctate representation of the kind that Rey seems to have in mind when he says that features as [+voice] are "represented in the brain" and asks with what external phenomenon trace "co-varies" (respectively, the two informal senses of "represent"). As noted, similar explanatory gaps arise in the study of insect navigation, bird song, rules of vision, and other phenomena, but here too it would be pointless, or worse, to pursue the terminological course that Rey recommends.

Parenthetically, even if co-variance of trace with external phenomena were a real problem it would not, as Rey asserts, be exacerbated by my "repeated insistence that there actually are no objective phenomena corresponding to representations" (words and other expressions of I-language, or of language in some other sense, maybe "public language"). If by "correspondence" Rey means one–one correspondence to identifiable mind-independent entities, then I do "insist" that no sense has been given to this notion. Dispensing with that obscure doctrine, investigation of problems (B) is concerned throughout with mind-external phenomena and their relations to internal expressions. But calling some of the mind-external phenomena "sounds" and others "things," and introducing some unexplained relations between expressions and these "sounds" and "things," contributes nothing to the study of the problems (B) (see reply to Ludlow). And no problem becomes more severe when the emptiness of these moves is pointed out.

The remainder of Rey's discussion of the alleged need for "intentional content" (section 5.1) seems to me to raise no problems beyond the familiar one: lack of understanding of how to fill the empirical gaps. We need not tarry over the informal usage that concerns him, for example, the common use of the term "phone" that he mentions, applied variously to elements of PHON(E), to what is done by the mouth, and to external events.

The important point is that although there are plenty of problems and gaps, understanding of them is not advanced by introducing terms that are unexplained and notations assigned to problems that are unsolved, such as those of categories (A) and (B).

Rey's discussion of other topics is very detailed, with many references to what

I've written. But his rendition is based on such pervasive misunderstanding that it is difficult to proceed. Much of the misunderstanding can be cleared away by distinguishing informal presentation of a problem from attempts to deal with it, and observing the distinctions among the several senses of "represent" in informal usage. But a good deal is simple misreading. To present a complete review of his citations would take far too much space. I'll give a few representative examples. As far as I can see, no important categories are omitted. For the cases I won't run through, I have to issue a general disclaimer.

Rey's criticisms are mostly variations on a central theme: that I use "intentional notions," but without defining crucial technical terms or facing the problems that arise when we consider "the property by virtue of which representations, thoughts, ideas and propositional attitudes . . . have the *contents* they do." But I do not face these problems because I do not use these notions, apart from informal presentation, at least in anything like the sense he appears to have in mind; and he repeatedly misunderstands references to "technical terms."

For example, Rey complains that I provide "no explicit definitions" for the "technical sense" of the terms "content" and "mental representation" that appear in a passage of mine he quotes. The passage *criticizes* the resort to these terms in work that introduces them "in a technical sense" (meaning "invented") but without explanation. It is Rey and other advocates of "wide and narrow content" and similar "intentional talk" who face the problems that trouble him. Evidently, these are not my problems. We can therefore dispense with his concerns about implicit definition by the "role in [my] theory" of the notions I am criticizing; they have none.

Rey thinks I adopt "intentional talk" by taking language acquisition to be an "inferential process" of "hypothesis confirmation." Here he attributes to me a version of the theory theory, TT, which I clearly do not adopt (see Gopnik). He reaches this conclusion by citing comments in which I use this terminology informally to outline a problem (Chomsky 1965). The discussion that follows, extended in more detail elsewhere before and since, suggests ways to deal with the problem that involve no hypothesis formation and confirmation at all (that's the point of Gopnik's critique). The term "hypothesis" is used here rather in the way Lorenz uses it when he speaks of "the innate working hypotheses present in subhuman organisms" (see reply to Lycan); he does not expect us to conclude that a duckling entertains the proposition that the first thing it sees is its mother. I also made no "more recent retraction," as he asserts, since there was nothing to retract. Once these misunderstandings are cleared up, we can put aside Rey's concerns about the role of confirmation, inference, the nature of LAD, etc.

Rey also finds it "interesting" that I think "there's a sufficiently identifiable 'mind–body problem' to claim that what many of us regard as the one advance we've made on it, the notion of Turing computability, doesn't solve it!"; exclamation point to stress that this contradicts my questioning of the very existence of the problem. He gives citations having nothing to do with any

"mind–body problem"; they refer only to Turing's observation that the question whether machines think is "too meaningless to deserve discussion." Even if we were to agree with Rey and his associates that Turing computability bears on "the mind–body problem" (I don't), the passage he cites would have no relevance to the matter.

Elsewhere Rey argues that "the question whether machines can 'think' is not as simple and terminological as the question of whether planes 'fly'," contrary to what I said in passages he cites, which quote Turing and Wittgenstein and basically agree with their observations that the question is indeed meaningless. Rey's reason is that we do not understand what non-intentional properties two entities must share "to represent the same thing, say, Vienna . . ." That has no bearing on the observations I cited; on the fact – truism, I think – that whether brains and machines think is a matter of terminological choice on a par with the question whether airplanes fly or submarines swim or robots can murder. In my I-language, airplanes fly, submarines don't swim, robots can't murder, and my brain and computer don't think. I might some day replace my word "think" by another one that applies to my brain and my computer, as Turing suggested might happen for computers by the year 2000 (phrasing it a bit differently), and as did happen over time with "fly" in English, but not some other languages. The questions appear to be without significance, along with the debates they have engendered. And while there are surely plenty of problems in accounting for the ways we think and talk about Vienna, I think only confusion is added by introducing Rey's idioms of "representation" and "intentionality," and the entire picture of language–world relations that lies behind such unexplained usages (see replies to other chapters).

Rey cites research that supposedly refutes what he takes to be my objection to empirical research that studies "people decision-theoretically." I have never hinted at such an objection. He is misreading my observation (which he cites) that we should refrain from interpreting informal descriptions that adopt a quasi-causal idiom as empirical hypotheses, when they are not (Chomsky 2000: 166), an observation he misunderstands repeatedly in this essay.

Another discovery is that I "now" (1992) "regard bicycle riding as 'irreducibly cognitive'!" – thus allegedly refuting a discussion of bicycle riding in Chomsky 1980, which Rey misunderstands. He also fails to note that the same point about the "irreducibly cognitive" character of such skills is made there (1980: 3, 53), referring to a more extensive discussion in an earlier article (Chomsky 1975b; also 1969). Throughout, the reasons are the same, and have nothing to do with what Rey is discussing here.

On E-language, Rey completely misunderstands. Denial of the role of unexplained notions and apparently useless (if even meaningful) proposals does not imply that "there is nothing to be right or wrong about."

Rey thinks that if I dispense with the Fregean notion of a "common store of thoughts," then I am compelled to conclude that "there is no common store of

the *principles* of universal grammar *cognized* by normal (neonate) human beings": equivalently (at least if he is using my terminology as I do), we are compelled to deny that the state of FL for a normal infant is an expression of the genes (in good part). Nothing of the sort follows. Perhaps his misattribution to me of something like TT accounts for his fallacious inference. He goes on to conclude that FL can now "be happily characterized without any of [my] controversial mentalistic claims," namely, the reformulations he gives in his terminology of "intentional content" and "representation." I am happy to continue to avoid any claims formulated in the way he proposes, keeping to my own, which are controversial, I hope (that is, not vacuous), but not for the reasons that Rey adduces. The passages he cites are doubtless "puzzling" when translated into his terminology, but that doesn't seem to be my problem.

Rey suggests that we can avoid "all the headaches" he examines by dropping such "intentional idioms" as "knowledge of" in an informal statement that he quotes from Chomsky 1965: 9, which sets out the goals of generative grammar: namely, that it seeks to characterize "the knowledge of language that provides the basis for the actual use of language." We certainly can avoid the "intentional idiom." That's the course followed in the book he cites, and other work that seeks to achieve the goals outlined, eliminating the informal locution "knowledge of language" in favor of the technical notion of states of FL. Thus, we say that Jones has I-language L, meaning that the FL of his mind/brain is in state L; that is the closest counterpart proposed to the English locution "knowledge of language" (one should not be misled by merely terminological changes over the years; see reply to Egan).

Insofar as the broader goal cited from Chomsky 1995 is attained, and the account of Jones's L "provides a basis for a description of how, in fact, language is used and understood" by Jones, we can proceed to give partial, sometimes substantial, explanations for the fact that Jones knows that expressions have the sounds and meanings they do over a wide range (LP); also that he knows how to use them, and so on. For such reasons, it makes good sense to regard FL as a cognitive system, entering crucially into determining Jones's knowledge that so-and-so, how to do such-and-such, etc. As I've discussed elsewhere, I think there is reason to suppose that the picture generalizes in interesting ways.

It is, to be sure, possible to fall into confusion. Thus, one should not be misled by the (rather idiosyncratic) English locution "know a language," concluding that some cognitive relation holds between Jones and an entity ENGLISH that is external to him, as is often done. To the extent that Rey's critique has some force, it is aimed at the wrong target.

To take a final example (unfortunately not the last), Rey complains that I provide no definition of the technical term "cognize." Again, he completely misunderstands. The term is "technical" only in that it was invented for a particular purpose: "to avoid terminological confusion" resulting from idiosyncratic properties of the English word "know," and philosophical theories that

seem to be motivated by them, such as the theory that Jones stands in some cognitive relation to the extra-worldly entity ENGLISH. The term "cognize" is clarified adequately enough for the purpose at hand, namely, informal exposition. It plays no role in any context of theoretical explanation. That much is evident even from the passage that Rey excerpts. In the transition from informal exposition to technical discussion, "Jones cognizes L" becomes "L is the state of Jones's FL." Cognizing, in this sense, does enter into standard cases of knowing that (how, etc.), as just noted, and in the passage Rey cites. He is right, then, to say that "cognize" overlaps with "know" in relatively uncontroversial cases; the overlap was part of the point of introducing it as a technical (invented) term.

In some future psychology, the informal terms "know that (how, etc.)" may also not survive. For now, we have every right to be interested in the phenomena described informally in these terms, and to try to account for them. There seem to be sound reasons to do so by exploring what is called in English "knowledge of language," making use of such concepts as I-language, having an I-language, states of the mind/brain, etc., and showing how development of these ideas can provide an account of what we describe as knowing that . . . (how, why, etc.). There are plenty of headaches along the way, but not of the kind that would arise if the framework is translated into the intentional idioms that Rey mistakenly believes are used.

Rey asks: "if we cognize grammars, do we also cognize laws of gravitation . . .?" That is, if L is the state of Jones's FL, does it also follow that Newton's laws constitute part of the state of some other cognitive system that Jones has, which Jones accesses when he decides to fall from a height? It doesn't seem to be a very serious question. Let's reintroduce the Martian scientist M, who has appeared in other comments (us, in the mode of naturalistic inquiry). M would attribute Newton's laws to some internal system of Jones, or of a rock, if there were reason to believe that they do access this system when they (decide to) fall from a height. But that course is plainly wrong; it suffices to attribute to Jones and the rock the property of having mass. That property, however, will not suffice to account for the differences between benzene and sulfuric acid, or for insect navigation, or for Jones's interpretations of LP. Accordingly, M will attribute to them other properties: for benzene and sulfuric acid, having particular chemical constitutions; for insects, cognizing a system of path integration and calculation of solar ephemeris; for Jones, cognizing L ("cognizing" translated into technical discussion as before). Though differences abound, in the limited present context the similarities are real.

It is worth recalling that attribution of chemical constitution was regarded as highly controversial, if not absurd, not long ago, as absurd as believing in the molecular theory of gases. All of this is dismissed in retrospect as seriously in error, with lessons that are useful, I think, in the present context (see reply to Lycan and references).

Rey opens by formulating a basic question he thinks has remained unanswered

over many years of discussion: do the computational processes postulated have "psychological reality," which means for him that they are "either represented or defined over representations"? He believes that I am using "intentional talk" when I hold that principles and categories of I-language (his term is "grammar") are "psychologically real." I don't use, or even understand, these locutions.

Apart from domain, I do not see what differentiates "psychological reality" from some other kind of reality: say, chemical, optical, or neural reality. Substantive conclusions reached at level (II) satisfy the condition of "psychological reality" insofar as they are true. Categories and rules are postulated as elements of I-languages (states of the mind/brain); whether rightly or not is to be determined by empirical investigation. They are "represented" in the brain only in the standard sense of informal usage, already mentioned. Beyond this, there is no "intentional talk," any more than for insect navigation or bird song, or for that matter, the postulation of entities and properties throughout the modern history of chemistry (see comments on Poland). I do not see any novel questions of "reality."

Rey asks why we should not take realistic talk (what he calls "intentional talk") to be "on a par with the talk of latitude and longitude," or treat it as "metaphor" or an "artifact of our notation." Here he misunderstands the reference he cites to "the computer metaphor," but more importantly, the discussion he cites of the correction by Gilbert Harman. Specifically, he has missed the point of the word "substantive" in the preceding paragraph.

To take a concrete case, suppose we develop an account of LP (say, (1), (2) in reply to Gopnik) that attributes to FL specific principles of anaphora and predication formation, strictly local operations for constructing expressions, etc. Why assume that we are attributing these to the mind/brain? The reasons are straightforward. They are not localized in the foot, and other properties (mass, valence . . .) do not suffice to account for LP. We therefore proceed much as in the case of chemistry, insect navigation, etc. We construct the best explanatory theory we can, and attribute the entities and principles it postulates to FL. We can then test these substantive conclusions by investigating other constructions in the same and different languages, as has been done extensively, revising principles and discovering parametric variation as the inquiry proceeds. We can ask how the conclusions fit into broader studies of language acquisition and deficit, speech and perception, neural mechanisms, etc., topics that have been investigated in interesting and informative ways. In short, we can treat the conclusions as we do other hypotheses about the internal nature of some natural object, not insisting on the "methodological dualism" that holds that the normal procedures of naturalistic inquiry are disqualified in this unique case.

In contrast, if we consider the proposals to be "on a par with talk of . . . longitude" (latitude is a somewhat different matter), then all these avenues of inquiry are cut off at once. Distinct choices of a grid for longitude yield the same substantive empirical consequences; distinct choices about the mechanisms

involved in (1), (2) may yield quite different ones. Longitude is fixed for convenience; the realist interpretation of principles and entities is rich in empirical consequences. Rey's puzzlement is as appropriate as the question why we should assume that benzene and sulfuric acid actually have the structures postulated by nineteenth-century chemistry, instead of taking the structures to be merely arbitrary notation like some longitudinal grid (which no one proposed) or a device for calculating chemical reactions (as indeed was proposed prominently, well into the twentieth century).

In the same connection, Rey asks why we should not content ourselves with saying that the internal system only "satisfies" the principles of language as "digestion and metabolism satisfy principles of chemistry and thermodynamics." The analogy is inappropriate. Scientists studying digestion and metabolism are not content to say that they satisfy chemical and physical laws; that would not even differentiate the systems.

Elsewhere (note 7) Rey asks a different question: why should we not say that the "internal competence" that the linguist's theory describes "*still involves rules merely being satisfied, not represented*, as in the case of digestive and metabolic systems"? Or that insect navigation satisfies "the rules of insect navigation," or that the visual system "satisfies the rules of visual intelligence"? Scientists are concerned with quite different questions. They want to know what it is about the internal structure of components of the body and their interactions that accounts for the fact that "the laws" of digestion, vision, navigation, etc. hold for them, differentially; having mass, electrons, etc., does not suffice. They therefore attribute internal structure to these systems and then explore other evidence that might bear on the accuracy of these different attributions. Unless we again adopt a methodological dualist stance, the same procedures should be followed in the case of humans.

Rey thinks not. In questioning the "psychological reality" of I-language, he asks whether its computations are "really part of the mental life of, say, a child who might be quite incapable of understanding linguistic theory." Are the rules and principles "really *represented*, *followed*, and even *known* innately by all normal neonates"? Or can we be content with a more modest formulation: I-languages are "merely non-representational states of some internal system that only *satisfies* the principles, in the way that processes of digestion and metabolism satisfy principles and laws of chemistry and thermodynamics, presumably without representing anything"? As noted, that makes no sense for digestion and metabolism, but let's modify it to his alternative suggestion about these systems "satisfying the laws" of digestion and metabolism, also inappropriate but not as dramatically so.

Rey thinks these questions remain unanswered for language. I think they have been repeatedly answered, at least insofar as they are clear. "Understanding linguistic theory" is utterly irrelevant. The questions about "mental life," "following," and "knowing" can be considered only insofar as Rey tells us what he means by these terms. If he is using them in their normal sense, there is no

clear or useful question to answer; if in some technical sense, we await an account of what it is, and reasons to take it seriously. I've explained repeatedly what I mean by such phrases, when I use them at all, and in those terms, have given answers that seem to me adequate, along the lines just discussed. Perhaps not, but I see no indication here.

As to whether Jones's I-language is "really *represented*," that depends what one means by "represented." If Rey has in mind the first sense mentioned ("internal representation"), the empirical hypothesis is "Yes, it is" (like the computational mechanisms of insect navigation and bird song and rules of vision). If he has in mind the unification problem (A), then one presumes that there is some answer (as in the other cases), though it need not satisfy the kind of condition of point-by-point representation that his remarks suggest.

Rey believes that intentionality is introduced in saying that PHON(E) "provides information" to SM, and that this informal locution raises the problems he brings up about intentional content and representation; presumably the same should be true, then, when insect scientists say that the postulated computational systems provide information to the motor system. There are plenty of problems, but not the ones that concern Rey. For language, they are the problems addressed by integrating the study of I-language with acoustic and articulatory phonetics, recognizing that problems of type (A) remain open, as in far simpler cases, and that a general theory of action lies beyond anyone's dreams. With regard to SEM(E), the corresponding problems, I think, should be dealt with along the lines indicated in other replies. One who chooses to introduce the terms "intentional content" and "representation" has the responsibility to explain the terms and to show what they contribute to the solution of the problems; how they might, I do not see from the literature (including Rey's comments). Perhaps these further moves will raise the problems Rey perceives, but if so, that's an argument for not following this course.

Another important problem, Rey believes, is that "a typical human mind has thoughts not only about cats and dogs, love and hate, but also about VPs, IPs, traces and c-command." I have no idea what this means. For one thing, people have thoughts, not their minds. And typical people don't have thoughts about the postulated linguistic entities (which don't relate in any simple way to pieces of utterances). In a special mode, I do have thoughts about the linguistic entities: Are they real elements of FL? Is c-command a condition for antecedent–trace relations? Currently, for what it is worth, I suspect not, contrary to what I once thought. I'm sure people have thoughts about cats, love, etc., facts that raise interesting questions, though not the question of whether "all these contents" of their thoughts about cats, traces, etc., "are determined entirely by facts inside their skulls" – at least, until we are told what any of this means. It's easy enough to master the conventional moves with these pieces, but not so simple to say what they are achieving.

Understanding advances not a jot if we rephrase the questions with invented

technical terminology that is not explained, like "intentional content," or succumb to the temptation Rey brings up of thinking that it's a "'primitive' fact" about minds, or biology, that "it just automatically refers to things." People do refer to things, as when I refer to the book I tore to shreds after memorizing and reciting it. But minds and biology don't refer to things, at least in the ordinary sense of "refer"; and if some technical sense is intended, we again have a right to ask what it is, and how introduction of it is justified in some explanatory context – putting aside the matter of what this "'primitive' fact" might be, at least in any naturalistic context.

Rey thinks I have to face the "Quinean challenge" of specifying "just *which* internal relations short of the entire system of them are to serve as the contents of specific representations." There is no challenge until the term "contents" is explained. If the "contents" of the expression "I tore the book to shreds after memorizing and reciting it" involves the book, then there is no coherent question until we are told what the book is, this thing that I can tear to shreds, memorize, and recite – that is, until we deal with the problems of lexical meaning and how it enters into language use, not a meaningful problem for Quine. If Rey has in mind the Quinean challenge to language-specific elements of meaning, I think it is without force for natural language for reasons discussed elsewhere (Chomsky 2000: 61ff), and that the questions should be approached along the lines of Pietroski's essay. If by "representations" Rey means postulated internal objects of I-language (departing radically from Quine), then there is a challenge: discover the relations they bear to others internal to I-language and the interface relations, and then ask how Jones talks about the world (among other actions), equipped with this internal apparatus. The challenge is not Quinean, however.

Rey also thinks that grammatical categories "seem inextricably entwined with the intentional content" of the C/I systems involved in language use. But he hasn't told us how the problems that interest both of us – like thinking about books – are in the least advanced by introducing the unexplained term "intentional content." I-language is "inextricably entwined" with the processes of articulation/perception, and with talking about aspects of the world ((B1) and (B2), above); just how it is entwined is a topic of empirical inquiry, difficult but seriously pursued along lines indicated in other essays and replies here, but not advanced by obscure terminological innovations that have their roots in highly misleading intuitions of the photo-landscape variety, what Horwich calls "referentialism."

Rey is right to say that Newton provided extremely impressive laws concerning gravity, but does not pursue the observation to the relevant conclusion. It is for just this reason that Newton objected to the charge that he was reverting to the occult qualities of "the Aristotelians," who simply gave names to unknown phenomena, and thus "put a stop to the Improvement of natural Philosophy" (*Opticks*). We should, I think, draw the lesson that Newton did: do not place barriers to understanding by giving unexplained names (like "intentional con-

tent") to phenomena that we do not understand. Instead, we should seek to construct "bodies of doctrine" just as Newton did, "deferring" problems that lie beyond reach until we understand enough to raise them seriously (chemist Joseph Black; see reply to Poland). That's the course I and others are trying to follow. I don't see that Rey has raised any problems about it, apart from the huge gaps of understanding that are already familiar.

Rey ends by posing a dilemma: If my approach "requires intentionality and a CRTT," then it's my responsibility to say how I understand such "intentional idioms" as "representations," "cognizing," "information." If it doesn't require intentionality and the CRTT, then I have to say why we should regard grammatical competence "as genuinely part of the *mind*, and not merely as some non-mental system that a mind might happen to exploit." On the second horn, I cannot comment. I have no idea what it takes to be "genuinely part of the *mind*," or what he thinks "the mind" is, or how "a mind" might "exploit" a non-mental system. As for the first horn, I agree – in fact insist – that it's my task to say what I mean by the terms I introduce, though not to introduce and explain intentionality and CRTT (in the sense that Rey may have in mind). As to how well I've done that, others can judge. The questions are not raised here, as far as I can determine.

Reply to Ludlow

Peter Ludlow opens by observing that I am not "hostile towards semantics." That's correct. Most of my work on language, from the outset, has been an effort to explore the form and meaning of expressions and the principles that determine them, core problems of semantics. The numbered examples of the comments are typical illustrations. It's true that I prefer to call this work "syntax," and think the same term might well be used for other investigations that keep to symbolic systems that are internal to the mind-brain, the faculty of language FL or others.

Virtually all contemporary reflections about language and mind take the expressions that have sound and meaning to be in a "public language" that is external to the mind-brain, "a kind of social object the structure of which has been established by convention," say by the pronouncements of the French Academy (Ludlow). It is misleading to say that I regard "the French language, so understood, as devoid of scientific interest." One reason is that the point of view is not specifically mine; it is standard in the empirical study of language. A more important reason is that the French Academy has had little to say about the French language in anyone's sense of the term, informal or technical. It has not even attempted to describe the form and interpretation of expressions of French. It has not been concerned, for example, with the properties of French expressions (including very simple ones) discovered and studied in Richard

Kayne 1975, the work that laid the basis for much of contemporary inquiry into the syntax and semantics of the Romance languages, with far broader impact.

The French Academy issues pronouncements that can be understood by people who already know French; that is, people with a functioning FL that has attained a state that underlies behavior that one might choose (under obscure and variable circumstances) to call "speaking French." The same is true of the *Oxford English Dictionary*, or the most elaborate traditional grammars. This is no criticism. Discovery of the elementary properties of French, or the meanings of words of English, is a challenging task. The French Academy and *OED* have different concerns.

The problem with the "E-language perspective" that Ludlow discusses is not simply the recourse (in some versions) to such factors as social convention and colors on maps. Rather, it does not address the basic properties of language, even the most rudimentary of them: that a language has an infinite variety of expressions, each a form–meaning pair. To play some role in theoretical explanation, an E-language approach has to provide us with at least some indication of the membership of this infinite class. The "public language" approach does not. Nor do other versions of an E-language perspective; Quine's, for example, which takes a language to be an infinite class of well-formed objects. One problem is that natural language appears to lack any relevant notion of well-formedness (Chomsky 1955, and much subsequent work); variations along many dimensions have been a major topic of empirical inquiry, the study of weak and strong islands and their syntactic-semantic properties, for example. A more fundamental problem is that any such approach can be understood only insofar as it provides a finitary account of some infinite class of objects that it has in mind, thus shifting the inquiry to a course that Quine condemned as sheer "folly." Some have used the term "E-language" to refer to actual utterances or texts. People are free to invent terminology as they like, but that was not the meaning I gave to the term when I introduced it; I-language approaches do not differ from others in their concern with such materials.

It is for such reasons that I doubt that there is an E-language perspective, and believe that even those who deny the fact are tacitly assuming the existence of I-languages, internal states of FL. It is only in these terms, as far as I can see, that one can make some sense of work that purports to adopt an E-language perspective, and evaluate its contributions to understanding language and mind.

Suppose we now explicitly adopt the I-language perspective (see Ludlow's note 5). Suppose Jones's I-language generates the expressions (1) ((i) is Ludlow's example):

(1) (i) the flaw in the argument is obvious, but it escaped John's attention
(ii) the fly in the bottle is trapped, but it will escape

(iii) John bought a book for Mary, but she already had it, so after reading it, he shredded it

Each of these expressions has a sound and a meaning; we can think of each E as a pair <PHON(E), SEM(E)> (in the terminology of Horwich and reply). Jones makes use of this information to produce and interpret utterances.

Suppose we want to account for such facts. Here's a simple way. The words of Jones's I-language map onto language-external objects (sounds, things, properties); the sound and meaning of utterances are composed of these in elementary ways. Thus "bottle" in (ii) maps onto a sound; the pronoun "it" maps onto the flaw in (i), the fly in (ii), and the book in (iii), where flaws are the kinds of things-in-the-world that can be in arguments the way flies are in bottles, a book is something that John can buy, read, and shred while Mary still has it. And so on, for every word we choose to look at carefully.

There is a sense in which sounds, flaws, books, . . . "are clearly not logically absurd entities, nor . . . particularly odd entities" (Ludlow). Sounds, for example, are perfectly robust things. We have no problem assigning sounds to the expressions of (1), or none of the above, in a vast range of normal cases. If we are satisfied with this result, we can avoid the hard problems of experimental phonetics.

Proceeding, we can quantify over flaws and average guys; also sounds and books. When we use a nominal phrase X, we can usually say, intelligibly, that the X's are the things we are talking about: flaws, books, the darkening sky, Joe Sixpack, the bank that was burned to the ground after rejecting the loan application, the threat of global warming, a biochemical process that is tragic (life), the purpose of America (the title of a book by one of the founders of realist international relations theory), . . . Again, if satisfied with this result, we can avoid the hard problems of the study of meaning, reference, and language use generally.

If we adopt this course, are we putting forth "substantive metaphysical claims," claims about the constituents of the world? Surely not in terms of the enterprise of natural science; on that there is no dispute.

Let's turn to ethnoscience, which studies "folk science," seeking to determine how Jones and others conceive of the elements of which the world is constituted and their functioning (see reply to Egan). Suppose that ethnoscientific inquiry attributes to Jones internal non-linguistic concepts that are one–one associated with nominal phrases of his I-language, and concludes that these are the constituents of Jones's commonsense understanding of the world. That is a metaphysical claim, on the order of the claim that an insect has a mechanism of path integration, or that Jones's I-language has a phonological system that enables him to distinguish /r/ from /l/ (unlike his Japanese friend). The relation of nominal expressions to these concepts is Ludlow's relation R^0, a syntactic relation that he dismisses.

Suppose ethnoscientific inquiry concludes that Jones also believes that the mind-independent world is constituted of entities that correspond to mind-internal words (or maybe their images under R^0). I presume Ludlow would be skeptical about these conclusions. At least, I would. I doubt that people think that among the constituents of the world are entities that are simultaneously abstract and concrete (like books and banks), or that have the amalgam of properties we discover when we explore the meanings of even the simplest words ("river," "person," "city," etc.). I suspect that serious ethnoscience will find that intuitive commonsense understanding of the world looks quite unlike that.

Nonetheless, at some level, we can accept the idea that when we use the words "book" or "river," we are talking about books or rivers, and can happily say that there are books and rivers; similarly for virtually every nominal expression, and, with elaborations, for other expressions. We can accept all this at the level at which we abandon curiosity about language and the mind, about human action and its roots and properties. Though sounds are perfectly robust and simple, much more so than books and rivers and flaws, scientists concerned with the sound aspects of language have not been satisfied with such accounts, any more than Kayne was satisfied with the hints provided by the French Academy or traditional grammars and dictionaries. They have sought to discover the internal entities PHON(E) and to determine how they relate to the kinds of mind-external entities that are studied in the sciences, investigating a relation between internal and external events mediated by sensorimotor systems; a phonological analogue to R^0 has never been considered. They have not been satisfied with the informal notion of thing in the world that is fine for ordinary life but useless if we hope to understand the world – in this case, the robust notion *sound of E*, a fine thing in the world, in informal usage.

Have articulatory and acoustic phonetics followed the right course, or should they just have stopped with the observation that E has a sound, as we all agree? That depends on one's interests. It is possible to disguise lack of interest with technical terminology. In this case, we could invent the term "phonetic value," taking the phonetic value of the expression E to be its sound. Similarly, we can say that the "semantic value of 'water' just is water" (Ludlow).

These moves should, I think, be understood as registering lack of interest in the problems. That may be entirely reasonable; no one seeks to study everything. But we should not mislead ourselves into believing that by invoking sounds, flaws, John Doe, attention, escaping, . . ., or rivers, water, cities, books, trees, . . ., and taking them to be related to pronouns and other words by an invented technical notion called "reference," we have even begun to investigate, let alone to have solved, the problem of how people use language to refer to things in the world, or any other kind of language–world relation. That's taken for granted in the study of sound, and should be in the study of meaning and referring as well, I think.

Consider a Martian scientist M who wants to study humans the way humans

study insects. Suppose M is interested in questions of the kind just raised, and is informed by his human subjects that there is no real problem: we can account for everything by invoking the metaphysical thesis that among the things of the world are sounds, flaws, books, . . ., saying that words like "it" *refer* to these things, and so on. But M wants to know more: What are sounds, flaws, books, . . ., and what is the relation "refer" that has these curious properties? Suppose we assure M again that there's no problem: we all understand very well what sounds, flaws, books, . . . are and how to use "it" to refer to them. True enough, but no help to M, who wants to comprehend what it is that we understand, how we achieved this state of mind, and how our linguistic states relate to the outside world (and from another point of view, the inner world of language-external thought). M will, in short, adopt the stance that humans adopt when studying insects, or the sound side of language; and should adopt if they are interested in meaning and referring as well. M is just us when we are engaged in naturalistic inquiry; that is, seeking to understand something about the world (including the special case of how we engage in self-conscious naturalistic inquiry or unreflective exercises of commonsense understanding, or any amalgam of these enterprises).

Consider Davidson's analysis of the event structure of propositions. I share Ludlow's conclusion that these ideas have been extremely productive, a major contribution to our understanding of language (see Pietroski's essay). But the work remains syntactic in the broad sense, and falls short of significant metaphysical claims, until we are given some further explanation of what events are and of the notion "existence of events," an explanation that responds to the concerns of the Martian M: that is, the kind of explanation we rightly demand when we are told of "the existence of sounds." What is needed is a language-independent account of the postulated "events," either in the domain of ethnoscience or other branches of naturalistic inquiry. We speak freely of events in ordinary life, but that's no help to M. What counts as an event in the relevant sense is not at all obvious, surely no more so than what counts as a book, or a river, or a nameable thing.

In the same connection, consider Ludlow's notion *I-substance* (which he introduces with reservations). I-substances are "the stuff that we are intuitively talking about when we use language (the intuitive referent of 'water')." That's fine, as long as we do not conclude without evidence that our commonsense conception of the way the world is constituted and functions actually employs I-substances, pre-empting the conclusions of ethnoscience and contrary to what even superficial reflection reveals, I believe. The I-substance is what it appears we are "talking about when we use language" only if we misunderstand our use of language, ignoring the warnings of the later Wittgenstein, Gilbert Ryle, and others. The problem of "matching up right" doesn't arise, if we avoid these pitfalls.

There is nothing original about these observations. They are at least as old as Aristotle. He was not satisfied simply to say that a house is a robust entity (just

like a sound or a flaw), but wanted to know what it is. We can "define a house," he concluded, "as stones, bricks and timbers," in terms of material constitution, or as "a receptacle to shelter chattels and living beings," in terms of function and design; and we should combine both parts of the definition, integrating matter and form, since the "essence of a house" involves the "purpose and end" of the material constitution (*Metaphysics*, bk VIII, ch. 3, 1043^{a}; *De anima*, bk I, ch. 1, 403^{b}). That is a step towards understanding, though only very partial. Houses have far more intricate properties, as does every "object," as we discover when we go beyond casual inspection (Chomsky 1975a, 2000, among others). The most elaborate dictionaries, monolingual or pedagogical, never give a hint of these properties, quite rightly; even if they had been noticed, spelling them out would only confuse the user, whose knowledge of these facts comes from "the original hand of nature," in Hume's apt phrase.

We move a step beyond when we reach the seventeenth century and its radical revision of Aristotelian science. One aspect was to shift the burden of explanation of topics such as these from the postulated structure of the world to the structure of the mind. What we can know is determined by the "modes of conception in the understanding," in Cudworth's phrase. That led to illuminating work in the theory of meaning by neoplatonists, empiricists from Hobbes to Hume, and others, influencing Kant, with later resonance until today (see reply to Gopnik for some references). The Martian scientist would be well advised to pursue a similar course, adopting the ethological perspective that seeks to discover the organism's *Umwelt*, its particular mode of interpreting the world.

For non-human animals, so it is alleged, internal computational systems are "representational" in something like the sense of technical referential semantics. Introducing a series of experimental papers, Gallistel (1990) argues that representations play a key role in animal behavior and cognition; here "representation" is understood as isomorphism, a one–one relation between mind/brain processes and "an aspect of the environment to which these processes adapt the animal's behavior," as when an ant "represents" the corpse of a conspecific by odor. For humans, this notion is completely inappropriate, as we see by inspecting even the simplest words, criteria of individuation, the basic mechanisms of intended referential dependence, and other elementary properties of language and its use, even the concept "nameable thing" (as noted in comments that Ludlow quotes).

Pursuing this course, we can learn a lot about humans, and about how the expressions of their I-languages relate to language-external entities – sounds and things in informal usage. But we do not arrive at "substantive metaphysical theses," except about the nature of the human mind and the "modes of conception in the understanding" that enter into constructing the *Umwelt* in which we live and act. Non-human representational systems might yield directly a kind of mind-independent "metaphysical thesis" if Gallistel is correct, but not human language, which does not seem to be a "representational system" in anything like this sense, in either its sound or meaning aspects.

The decision to be satisfied with informal talk about sounds and meanings and things in the case of (1), or "house," or any other aspect of language not only cuts off inquiry into important questions, but leaves us with misleading conclusions. Take "flaw" and "fly." They function in a similar way in (1i, ii), but not in other expressions. Compare (2)–(4):

(2) (i) there is a fly in the bottle
    (ii) there is a flaw in the argument
(3) (i) there is believed to be a fly in the bottle
    (ii) there is believed to be a flaw in the argument
(4) (i) there is a fly believed to be in the bottle
    (ii) there is a flaw believed to be in the argument

4(ii) is deviant, unlike the others. Such constructions as (4) have existential import in some manner beyond (2), (3). The observation generalizes, and falls within an explanatory framework with a variety of consequences, and interesting open problems (Chomsky 2001). Many such aspects of the meaning and structure of expressions, and their use in talking about the world, are likely to be missed if we are satisfied to say that we understand the expressions of our language and need inquire no further, just as the discoveries of articulatory and acoustic phonetics would have been lost if investigators had been satisfied with the observation that expressions have sounds, perfectly robust entities.

For these reasons, I do not entirely agree with Ludlow's critique of model-theoretic semantics. The basic problem seems to me to go beyond the difficulty of identifying the intended model. It has to do with the models themselves. What are the individuals in the model? Are they flaws, books, houses, water, rivers, nameable things, sounds, . . . ? If so, we still have to face the problems of how language is used to refer. If we posit semantic values in the manner suggested, we face the same problems. Either we are positing some mind-internal domain of obscure "individuals" that are related to expressions by something like Ludlow's R^0, and we face the problems of justification that arise when we posit other kinds of syntactic objects (phonology, phrase structure, etc.); or we are choosing to content ourselves with informal talk that would not answer the questions of the Martian scientist or ourselves, as scientists, though as subjects of inquiry we understand this talk very well, just as bees understand the waggle dance; no help to von Frisch. If questioning stops at this point, the enterprise is a form of syntax – which is, of course, in no way to impugn its contribution to the understanding of I-language and its importance as a preliminary to investigation of the use of language to refer and in other ways, just as phonology is an essential prerequisite to the study of how people produce and interpret the sounds of language.

On efforts to translate expressions such as (1) to "logical form," I share Ludlow's skepticism. If the notion "substantive metaphysical thesis" is taken

seriously, I suspect that any such endeavor will soon find itself engaged in something like a Carnapian *Aufbau* project (adding metaphysical claims that Carnap would hardly have welcomed). If the translation project is understood to be in principle part of the empirical sciences, then the operations proposed will at least have to be suitably general and independently justified, like postulated rules of phonology. That may well be possible in such cases as "bad singer"; for (1), it looks hopelessly implausible for reasons Ludlow mentions, and others: e.g., if "flaw" is a modifier like "bad," then we have to be told how the pronoun "it" can be referentially linked to it in (1i), though never to "bad" in the same way. Problems proliferate rapidly, case by case.

Let's turn finally to Ludlow's relations R^i, and the "interpretive puzzle" he poses. R^0 relates a linguistic expression to an entity in an internal symbolic system (call it "S-Mentalese"). A phonetic analogue would relate the expression to an entity in P-Mentalese. S-Mentalese and R^0 have to meet the usual criteria of justification for syntactic theory. Whatever the prospects, I think Ludlow is right to conclude that model-theoretic semantics does not "deliver language/world connections," though not only for the reasons he mentions. And I think the same is true of approaches that invoke "semantic values" without telling us what these are in some terms that would be helpful to the Martian investigator M: that is, to us, if we hope to understand language and thought. If so, then these approaches do not "explore some form of language/world relations" (Ludlow), though they may purport to do so, misleadingly.

Ludlow's R^1 is a "direct" relation (expression, "the world"), R^2 a relation (speaker, expression, context, "things in the world"). One could add others, e.g., the "direct" relation R^3 = (expression, things-in-the-world) and the relation R^4 = (speaker, expression, context, the world). R^4 and (under idealization) R^1 seem to me the appropriate choices for the investigation of sound and meaning.

Ludlow quotes my comment that R^2 is "entirely innocent," because of course "people use expressions to refer to things in particular circumstances." Probably I was not sufficiently clear in the correspondence he cites. The remark quoted is without metaphysical intent; it is intended simply to repeat the truism that ordinary usage is fine as it is, but shouldn't be misrepresented through distorting doctrinal prisms and used to draw illegitimate conclusions about the world or our beliefs about the world.

How should we proceed in studying expression–world relations? On the sound side, it is assumed without comment that the approach should be something like R^4, though actual research introduces radical idealizations, even beyond R^1, to restrict the terrain to something that can be seriously investigated. Though analogies cannot be pursued too far, it seems to me that a somewhat similar approach is appropriate on the meaning-referring side of language. From lexical semantics to more complex expressions, we can investigate the syntactic interface level SEM and the principles that relate it to the rest of FL. We can then ask

how the information in SEM is used in expressing thought, talking about the world, and other kinds of human action, perhaps mediated by a conceptual-intentional system of the mind/brain (see Horwich and reply). It is an open question whether something will emerge that is similar to the referential semantics devised for formal symbolic systems, and that might have some counterpart in animal cognition, making use of a notion similar to the theoretical term "reference." I know of no convincing reason to believe so, and would not be surprised if the quest turns out to be as illusory as the analogue is taken to be (without comment) on the sound side. But far too little is understood to say anything with much confidence.

Reply to Horwich

Paul Horwich's proposals about the faculty of language FL focus on issues of central importance, and also considerable obscurity. I will try to review some of the considerations that seem to me to bear on them.

Horwich notes "a striking convergence of opinion between Chomsky and Wittgenstein regarding the dubious legitimacy of philosophical theories of meaning." That's true, and of course no coincidence. When I began to think about these questions seriously as a student, I was much influenced by recently published work of the later Wittgenstein. My skepticism extended to what Horwich calls "referentialism," though not entirely for the reasons he gives. A more serious problem, I think, is that the best question of use of language – what is the "aspect of the external world" to which a linguistic item is claimed to "refer" – is left unexamined in any serious way, and it is doubtful that this is the right picture to explore at all. See reply to Ludlow. Horwich is right to abandon it, effectively, I think.

Under these and related influences, I assumed from my earliest writings in the mid-1950s a kind of "use theory of meaning," not in Wittgenstein's terms but perhaps not inconsistent with them: the internalized language (I-language) generates syntactic objects, each of which "provides a basis for a description of how, in fact, language is used and understood" (Chomsky 1955: 75; quite a substantial basis, it is argued). The generated objects provide information accessed and put to use by performance systems, including the sensorimotor systems and the conceptual-intentional systems (P/A and C/I in Horwich's notation). In this regard, the theories of sound and meaning are similar. For such reasons, I have an initial sympathy for Horwich's preference for an approach to meaning in terms of "regularity of use." But I do not see how to implement the kind of approach he outlines without returning to the one he rejects.

Horwich's primary concerns are (I) and (II):

(I) the place of meaning in FL and its I-language instantiations;
(II) the nature of meaning.

With regard to (I), he concludes that meaning has no place in FL and I-language, though sound (I-sound) does. He suggests the "very simple picture" VSP in place of the alternative that he reconstructs from my discussion: call that VSP-C. The two differ only in that VSP-C has I-meaning in lexical items (LIs), while VSP is a "simplification," eliminating I-meaning from LIs and hence from the "semantic object SEM(E)" that is constructed from LIs and assigned meaning.

The simplification is real if what is omitted from the lexicon is not reintroduced in some other way. As far as I understand, that happens when we try to implement the account Horwich suggests.

Horwich notes that one might contemplate a further "simplification," eliminating I-sound from LIs as well, but that he excludes as "clearly not feasible." Actually it seems feasible, and has been developed in one of the leading current approaches to these questions: Distributed Morphology, which takes phonological features to be introduced in the course of the computation to PHON(E) (in our shared notation; "feature" is just the technical term for the elementary constituents of LIs and expressions constructed from them). One could go even farther, as discussed in the paper of mine Horwich cites in note 1, placing sound outside FL altogether, in a P(honetic)-Mentalese system that is analogous to the semantic Mentalese that Horwich discusses, a course that has never been pursued (wisely).

I do not entirely understand VSP (hence VSP-C). Thus, I do not see why it has to include the box with "constructions, <CP, LIs>," or why C/I is restricted to belief systems; more seems to be involved. But these questions are marginal to the main issues Horwich raises: (I) and (II). To take one of Horwich's examples, should the lexical item (LI) *persuade* include semantic properties as well as I-sound? And what is the nature of these properties, wherever they are located? To keep the focus on the central issues that Horwich raises, let us assume that LIs include I-sound.

VSP and VSP-C are the same apart from LIs, and for both, LIs include I-sound and also *formal features*, that is, features that enter into the computational procedure. Among the formal features, some – in fact, almost all – have semantic interpretations: for example Tense, or number on nouns. All these elements enter crucially into the procedures that form expressions <PHON(E), SEM(E)>, and they can appear as parts of words in intricate ways. The distinction between VSP and VSP-C is restricted to semantically interpreted features that do not enter into computation: I-meaning, in the technical sense.

The two pictures also agree – continuing with Horwich's exposition – that Peter's I-language "engenders such facts" as that (1) follows from (2):

(1) Mary intended to take her medicine
(2) Mary was persuaded to take her medicine

We want to discover how the I-language engenders such facts.

The account in VSP-C assumes that the LI *persuade* has features that determine the syntactic frames in which it can appear, among them the frame (III) as in (3) (and less transparently, (2)):

(III)	Nominal -	. . . -	Nominal -	[Infinitival Clause]
(3)	Jane -	persuaded -	Mary -	[to join her at the beach]

It also assumes that these features are components of I-meaning (largely or maybe completely), so that the frames are redundant; if I-meaning is omitted, they must be determined in some other way. The LI *persuade* has a causative feature that appears cross-linguistically in varied manifestations (morphological, syntactic, with hidden structure). To first approximation, *persuade* is a causative of *intend*, though general principles of FL induce distinctions having to do with direct causation and other matters.

When embedded in the I-language with its computational procedures, properties of LIs yield the conclusion illustrated in (1)–(2), and many others. Thus it should follow without further specification that *persuade* is an object-control verb: (3) entails that Mary intends to join someone (maybe Jane) at the beach, but entails nothing about Jane's intentions. Interpreting the pronoun as referentially linked to "Jane," we may understand (3) to indicate that Jane intends to join Mary, by virtue of the symmetry of "join," but that is not entailed by the structure of the expression. The pronoun "her" is not used in (3) to refer to Mary, unlike (1)–(2), where that option is open. It should also follow that the agent can be formally omitted but semantically understood, as in (2), in the context: " – in order to win a bet" (understood to mean that the hidden agent of persuading is to win the bet; interesting complications have been studied, which we may put aside). The system must also engender the differences among (3)–(5), and account for the fact that (6) does not relate to (5) as (2) does to its active counterpart, and in fact is highly deviant, unlike (2):

(4) Jane forced Mary to join her at the beach
(5) Jane promised Mary (sincerely) to join her at the beach
(6) Mary was promised to join her at the beach

Furthermore, while (3) entails that Jane persuaded Mary, (7) does not entail that Jane expected Mary, though *expect* allows this interpretation in (8):

(7) Jane expected Mary to join her at the beach
(8) Jane expected Mary

And *persuade* does not allow (9) but *expect* does allow (10):

(9) John persuaded there to be a hurricane tomorrow
(10) John expected there to be a hurricane tomorrow

These are among the earliest questions considered in modern generative grammar.

The expressions exhibit a systematic array of similarities and differences. Some have to do with referential dependence of "her": in (1), (2), (5), the antecedent can be Mary but not Jane (in (5)); in (3), (4), (7) it can be Jane but not Mary. Other distinguishing properties have to do with the intentions of Jane and Mary: in cases (4), (5), (7) Mary's intentions are irrelevant, contrary to (3), from which it follows (as Horwich notes) that Mary intends to join Jane at the beach. With "force," "promise," "expect," no such intention is imputed.

According to VSP-C, such consequences follow from compositional structure and features of LIs that have specific semantic interpretations, but are formal (by definition), insofar as they enter into the computation. Among them are an appropriate version of [causative], [intend], [act], and others (assigned their interpretations and modes of use by C/I and whatever else enters into human action). LIs may have shared but not identical semantic interpretations: e.g., two interpretations of "expect," correlated with the choice of argument (clausal in (7), nominal in (8)). Superficial similarities of structure may conceal crucial distinctions. For example, *expect* and *persuade* both appear in the syntactic frame (III), but in different ways, as illustrated. The features constituting LIs enter into engendering a large variety of consequences for typologically differing languages, for language acquisition and use, and in principle far more.

Note that an adequate account will have to engender the fact that (3) entails that Mary intended to join Jane but not conversely; that (4) entails that Mary joins Jane, not conversely; that (5) entails that Jane intends to join Mary, not conversely. As noted, the symmetry of "join" may yield further conclusions, but these are not entailments of the structures. These conclusions hold even though the relevant circumstances differ only subtly in the entailed and non-entailed cases; and it could easily have been the case that these differences have never arisen in Peter's experience, though his I-language nevertheless determines these properties, by virtue of the innate structure of FL.

Such questions come to the fore as soon as we begin examining Horwich's examples (1)–(2). There is a good deal of empirical work seeking to deal with them. The enterprise faces many problems, but it is not hard to see how it can be carried out, and to a certain extent, it has been, with some success.

Horwich raises two main objections (his (Q1) and (Q2)), but they do not seem to me decisive. With regard to (Q1), we need not expect "the form of an I-meaning [to] look like a classical definition." Some reasons are theory-internal: principles of FL may enrich and refine the semantic properties of LIs and the image SEM determined by computation, as in the case of referential dependence,

the understood subject of the embedded clause, the options for passivization, direct causation, and many other properties. But beyond that, there is no reason to expect that what is informally called "the meaning of X" in English (or partially similar locutions in other natural languages) will be fully determined by the semantically interpreted features of its LI, just as the sound of X is only partially determined by the I-sound. More generally, there is little reason to expect terms of ordinary discourse to find a place in efforts to understand and explain phenomena that fall (loosely) within their range. In the present case, for both sound and meaning, other factors are involved beyond the resources of FL; at the very least, ordinary use of language proceeds against a background of usually tacit belief, as Wittgenstein observed, and when the beliefs are withdrawn, as in such invented cases as Twin-Earth (or talking chairs), judgments commonly fade and apparent consensus is more a matter of convention than fact.

As for "the relation between the I-meanings and the constituents of thought" (Horwich's (Q2)), that seems to be a question of fact – or can become one, insofar as something is said about the nature of I-meaning, and (in language-independent terms) about the constituents of thought. Similar questions of fact arise in connection with I-sound, and have been productively investigated: namely, what is the relation between I-sounds (or more accurately, their image in PHON) and properties of sensorimotor systems that are independent of FL, or even of specifically human biology?

Let's consider now how such issues are addressed by VSP, which excludes I-meaning from LIs, while keeping the I-sound and formal features of VSP-C. VSP takes the formal features to enter into the account of such examples as (1)–(10) in the same way as VSP-C, and also assumes that they receive determinate semantic interpretations (but without intrinsic I-meaning in the LIs; I'm guessing here). VSP and VSP-C differ with regard to semantically interpreted features apart from these: the semantic features distinguishing items with formally identical behavior (a nontrivial theory-internal notion).

A separate question is what meanings and their components are (question (II)). According to VSP, to account for such phenomena as (1)–(10), we assume that "meanings of I-sounds are [not] explicitly represented" in LIs but are "merely constituted by their behavior in the belief system." We "suppose that, in general, I-meanings take the more liberal, flexible form of the specification of a regularity of use, rather than the more constrained form of an orthodox explicit definition"; a more appropriate question, I think, is whether pairing of "an I-sound with the representation of a basic regularity of use" is "more realistic" than pairing of I-sound and semantically interpreted features (some formal), along the lines that have been pursued in work on such problems as (1)–(10) in lexical semantics and the study of computational mechanisms of FL. Since LIs include formal features in VSP (as in VSP-C), a fuller version of the proposal would presumably be that basic regularities of use are paired with LIs constituted of I-sound and formal features.

We therefore have an internal representation of the LI (say, *persuade*) in FL, "the representation of a basic regularity of use" in some other component of mind, and an internal representation of the association between them. That much is required if meanings enter into action and interpretation, and it is assumed in VSP if I understand correctly.

To evaluate these proposals, we need some clarification of the notions in which they are framed: "basic regularity of use," "behavior in the belief system," and so on. We want to know how these properties are characterized in the mind to engender the facts on which VSP and VSP-C agree; and how the child acquiring language selects, specifies, and organizes these properties of the world in the particular manner that we discover on investigation of such phenomena as (1)–(10) in varied languages. The prospects do not seem to me very promising, for a number of reasons.

A narrow problem is that the I-sound to which regularities are paired is invariably distinct from PHON(E), its image under the computational rules. Furthermore, the relation is commonly one–many (or many–many): thus, semantic properties of *persuade* appear in "persuasive" and "persuasion," though I-sound has different images in PHON(E). In this case, the meanings change in complicated and partially idiosyncratic ways, but the full force of the problem is revealed in languages with richer morphology than English, in which various forms of an LI may have determinate meanings though the I-sound is considerably modified in systematic ways, so much so that it may not even be "extricable" from PHON(E). Insofar as we can speak of "regularity of use," it has to be assigned to the underlying element that has these manifestations in PHON(E): its LI.

A larger problem is that the notion "regularity," understood at all literally, does not appear to play much of a role in acquisition of semantic or other aspects of language. One reason is that very limited, scattered, and highly ambiguous experience appears to suffice; word meanings, for example, are acquired very rapidly in highly ambiguous circumstances, with rich delicacy that goes far beyond the hints in the most elaborate dictionary; and the same is true of other properties of expressions, such as those illustrated. Furthermore, whatever "regularities" there may be in experience are not simply *there* to be perceived. Uncontroversially, "regularities" are selected by the organism in ways that depend on its inner nature. The "regularities" of Peter's experience at the beach are radically different from those of his pet cat or the mosquito buzzing around his head. Peter's modes of cognition select certain salient properties and reject others, disregarding overwhelming regularities, and commonly choosing properties that lack regularities (construing the term at all literally).

To cite a typical case, "we might see a single bicycle being ridden on a single occasion and conclude that bicycles are for riding even though the bicycle we have seen is not being ridden most of the time and might be used for other purposes (e.g., for hanging umbrellas on)" (Prasada 2000). We further ignore the

fact that it was raining, it was Tuesday, the rider wore a red hat, a hurricane hit Florida, etc., including innumerable strong statistical regularities. Even selection of a bicycle, with its weird and complex properties, from the welter of undifferentiated experience is a highly complex task. Such matters are little understood. But no one can seriously doubt that for all organisms, what counts as experience is richly determined by internal factors, which construct an organism-specific *Umwelt*. As far as these matters are at all understood, linguistic meaning falls within the same general pattern.

With Peter's experience constructed in such ways, it might turn out that the entire experience that fixed the element "beach" in his mind conforms to the "regularity" that from some perspectives the beach has sand, from others it is near the water and houses, and from all perspectives it is noisy, hot, within walking distance of his house, etc. And his entire experience of the desert could be that it too has sand, though the only nearby water is in a swimming pool, that he gets there by plane to visit his grandparents over Christmas, and it's extremely cold at night. Furthermore, he may believe that all of these "regularities" hold, with equal firmness (and confirmation, in any sense of the term that does not beg all questions). But semantic properties nevertheless emerge, disregarding real regularities and firmly fixed beliefs in favor of features of constructed experience that may exhibit few if any regularities and have no justification in terms of such general procedures of belief-fixation as might exist.

This much is true of even the simplest cases. What already seems highly implausible becomes even worse if we turn to such cases as (3)–(6). Suppose we assume, as is realistic, that nothing in Peter's entire experience yields evidence about the circumstances that differentiate the intentions to join in the various cases. Nonetheless, conclusions of the kind illustrated emerge. This is the normal situation; in fact, artificially simple. The whole picture seems off the mark.

Unquestionably experience is required to fix the I-language within a narrow range determined by principles of FL, including the meanings of its elements. But what seems to be relevant is "salient" experience that has little to do with regularity but a lot to do with intrinsic modes of cognition, including whatever may be provided by FL.

Horwich allows for these (overwhelmingly significant) features of language and cognition. He suggests that for a "mental term 'F' . . . the variety of facts about 'F''s deployment are best explained by the properties of *some* of its tokens – namely that there is a proclivity towards certain beliefs containing them." In the cases mentioned, there might be a proclivity for the belief that the beach is near the water but not that it is near houses, that the desert has sand and not that one gets there by airplane, that intentions are determined by object control with *persuade* and *force* but by subject control with *promise*, that expecting Mary and expecting Mary to leave fall into distinct categories with their peculiar properties, etc. But what is meant here by "proclivity for belief"? It seems to have little to do with belief, or with regularity; and it is hard to see how the

"proclivities" can be spelled out except in the terms of lexical semantics and the computational mechanisms of FL, where the LIs include not only semantically interpreted formal features (as seems unavoidable, at least in the shared framework of VSP and VSP-C) but also apparently something beyond: I-meanings of some sort.

If so, it is not true that "postulation of I-meanings involves an extra level of complexity, and therefore must be justified in terms of explanatory advantages." Rather, it seems to be one method to give some empirical substance to theses that carry the weight assigned to the notions "proclivity for belief," "basic regularity of use," and so on, that are introduced in VSP.

An important question nevertheless remains. It might be, Horwich suggests, that there is "an innate proclivity towards a certain restricted set of conceptual roles," but we need not adopt "the further hypothesis that these are represented in [FL], rather than merely instantiated in the belief system." Conceptual roles might be innate, but some "innate Mentalese term might well possess that role," obviating the need for the further hypothesis that it is represented in FL.

How conceptual roles can be instantiated in belief systems is unclear, for the reasons mentioned: firmly held beliefs involving some "mental term" are commonly irrelevant to its meaning, and relations that are not established by "belief fixation" along any plausible lines appear to be intrinsic to conceptual roles. But if we drop the reference to belief systems, a question still remains, though it seems to me wrongly put. To suppose that conceptual roles are represented in FL is not a "further hypothesis" unless we assume that they are already represented in Mentalese, some system other than FL. Perhaps, but that requires argument, just as it would be necessary to justify the existence of P-Mentalese, in which the phonological role of sound is represented. Furthermore, even if the existence of Mentalese can be established, each LI must contain some kind of "pointer" to select the term of Mentalese in which the right conceptual roles are instantiated. And it would remain a highly obscure empirical problem to determine in what respects the properties of Mentalese differ from those of FL, and how features of "mental terms" distribute between these systems.

As noted, somewhat similar questions arise with regard to sound: in what respects are its properties specific to FL, or alternatively, to language-independent aspects of sensorimotor systems? In the case of sound, the question can be (and has been) seriously investigated. The somewhat similar questions about Mentalese can be posed only insofar as its properties are described in some language-independent way, and compelling reasons are advanced to justify the postulation of this component of mind. The thesis is not implausible, but plausibility falls short of conviction, and even if we decide to accept it, pending a language-independent account of its nature, we cannot inquire into the question of where "innate proclivities" are represented. Questions about the relation of language and thought are venerable, and intriguing no doubt, but it is not easy to see much progress in formulating or answering them in recent discussion.

Similar uneasiness is elicited, at least for me, by Horwich's account of compositionality. We have to account for the fact "that the content of SEM(E) is determined by the contents of its parts" (I am not sure what "content" adds here). This "can be explained trivially," Horwich suggests, if we recognize that understanding an expression "is, by definition, nothing over and above understanding its parts and knowing how they have been combined." It is true that we understand the expressions (1)–(10) when we understand their parts and know how they are combined, but to work this out seems anything but trivial. One could also say that perception of a cube in motion should be explained in terms of firing of cells in the visual cortex and the way the effects are integrated, but there is nothing trivial about the task of explanation, even for perception of a straight line. It might be true that the task can be carried out for language without reference to truth conditions, as Horwich suggests, just as it might be true that it can be carried out *with* reference to truth conditions. But that has to be shown, which doesn't seem easy, in either case.

In brief, I see no realistic alternative to the thesis that LIs contain semantically interpreted formal features, and probably I-meaning – or, what seems empirically indistinguishable in the present state of our understanding, that they contain richly structured pointers of some kind that direct them to specific regions of Mentalese, whch have the required innate properties and structures (as could also be proposed, though pointlessly as all agree, in the case of PHON(E) and P-Mentalese).

Sound and meaning, of course, differ in many ways. Analogies surely break down at some point, though I think it is heuristically useful to pursue them as far as they reach, if only because problems of sound, though deep and difficult, do not seem invested with the mystery that is attributed to problems of meaning, rightly or wrongly. However, care is required in differentiating these systems. In some cases, I am not convinced by Horwich's proposals. To mention one, he points out that "a sound event is a distinctive pattern of air vibrations," though the "thought constituent" correlated with it is individuated not by "its intrinsic properties" but by "its role in the conceptual system." But air vibrations are mind-independent events in the external world, while "thought constituents" are mind-internal: syntactic objects of some system of mind. Their sound counterpart is not external sound events but "sound constituents" that are either elements of PHON(E) or of P-Mentalese (if such exists, which no one assumes). Whether "sound constituents" are individuated by intrinsic properties or by their role in sound systems (rhyme and assonance, positions in structures of oppositions in some versions of structuralism, etc.) is a question of fact, not doctrine. The same is true for "thought constituents" and conceptual systems.

Horwich raises other important issues in passing. I am not always convinced. Take the issue of the language faculty and "states of 'knowledge'." Is the matter just one of *façon de parler* (as Horwich holds) or of theoretical significance? As in every other domain, there is little reason to expect that such English-specific

terms as "know," or its counterparts covering similar semantic fields in other languages, will survive the transition to substantive explanatory theory that attempts to gain some insight into the realm of phenomena that we allude to in using these informal terms. But it seems reasonable to expect that such theory will hope to say something about the fact that (as we informally describe the matter) Peter knows that if (2) then (1) (Horwich's example); and the same in the other cases reviewed. If so, we would like to understand what is the nature of Peter's knowledge that . . ., in these cases; or his knowledge how to use the expressions, or his knowledge of referential dependence, and so on. In these cases, it seems that his knowledge-that (how, of, . . .) is derived from a generative system about which we have a fair amount of understanding, and have ideas about how to learn more. I think there is reason to believe that the conclusions generalize, and may provide a sound conception of the realms of knowledge, belief and others. If that is even possible, then questions of substance arise.

Reply to Pietroski

The project that Paul Pietroski outlines and develops seems to me an eminently reasonable one. Davidson's event analysis has led to major advances in understanding of the nature and use of language. The "minimalist program" is a crucial part of the biolinguistic framework of modern generative grammar (at least as I have always interpreted it), and within that framework, the program is no more controversial than the classic efforts of D'Arcy Thompson and Alan Turing, and a great deal of work since, to show how physical and chemical principles determine fundamental properties of organisms (Belletti et al. 2001, Chomsky forthcoming, and sources cited back to the 1960s). Efforts to integrate these projects are therefore much to be welcomed, and I think Pietroski's contributions, here and elsewhere, take important steps towards that end. His specific proposals raise a number of technical questions, which I will only mention, not pursue, keeping to some general reflections and queries.

Pietroski asks whether event analysis as he and others develop it is a "form of syntax," or whether "we really do need to say that certain verbs are semantically associated with events, in order to explain certain facts about human linguistic competence." The alternatives are not exclusive; both could be correct. We cannot know until an account is given of how the concept "event" employed in event semantics is related to events in some mind-independent sense. The same questions extend to Pietroski's notions of grounding, termination, composition of complex events, criteria of individuation, and so on; and are fundamentally the same as those that arise when we are told that words denote objects, the latter unexamined. Until that further step is taken, the work is "internalist," hence a "form of syntax," though it remains to be determined whether its subject matter is internal to the faculty of language FL or other cognitive systems, an empirical

question about the organization of the mind/brain. And the step does not seem to me a trivial one (see reply to Ludlow).

Pietroski restricts his attention to "first-class analyticities," and I take him to be suggesting that his proposals may be consistent with Quine's theses on analyticity and the theory of meaning generally, which have had such an enormous impact on philosophy of language and mind in the past half-century (Burge 1992). That seems questionable, however: Pietroski's project and proposals go far beyond the bounds that Quine allowed. Pietroski's hypotheses are "analytical," not "genuine," in Quine's sense. They do not keep to the only entity admissible from Quine's "realistic point of view," namely "the right totality of well-formed English sentences." Rather, Pietroski proposes that some particular theoretical account of sentence-generation is correct, that the mechanisms it postulates are real components of the mind-brain, all senseless for Quine. And to reach these conclusions he adduces empirically based "best theory" arguments that are unintelligible on Quine's grounds, at least if given a realistic interpretation. Pietroski's analysis yields "hidden first-class analyticities" only under procedures of lexical decomposition and grammatical analysis which, for Quine, have essentially the status of one or another choice of axioms for some formal system (though textual interpretation is not straightforward because of apparent internal inconsistencies; Chomsky 1975a). It is also not clear to me how Pietroski's project and conclusions are consistent with Davidson's adaptation of Quine's basic theses. In any event, I think there is more to say about these issues, which have considerable import.

Pietroski raises the question whether such expressions as his (12)–(14) fall within his framework of first-order analyticities. If so, "the lexicon may contain far fewer primitive items than one may have thought." One might also consider the kinds of questions that arise in connection with relations of sound: rhyme and assonance, for example. Suppose that for Jones (though not for me), "sad" and "mad" rhyme. That is a matter of phonological structure, syntax in the broad sense. But though phonological structure is a factor in Jones's utterances, it is far from the only one: the sounds he produces when he says these words are highly varied. Bringing in other speakers who share Jones's phonology greatly expands the variation. Phonological primitives are restricted, but they do not come close to determining sound. It is to be expected that however restricted, semantic primitives (whether taken to be lexical "atoms," or components of them) do not determine "meaning" or "perfect paraphrase" as such terms are commonly understood, even if they do determine such semantic relations as entailment. Other factors (background or situational belief, specific concerns, etc.) are typically crucial factors, facts well known to translators, legal analysts and other scholars, and tacitly, to everyone in normal interchange.

The quest for the meaning of "meaning" or of "paraphrase" is illusory, except within a theory of some technical notion or an ethnoscientific inquiry of little apparent general interest. Just what belongs to a theory of rhyme, entailment,

paraphrase, and other such relations is to be discovered by empirical inquiry, not stipulated in advance.

Turning to Pietroski's specific proposals, I should first make it clear that I have no particular claim to the small-v analysis on which he relies; it has multiple origins, including formal semantics (Heim and Kratzer 1998, and sources cited), several of them unrelated to the minimalist program. It is also important to recognize that the latter is a program, not a theory. It has been pursued in many quite different and inconsistent ways, quite properly. Pietroski's choices are different from mine; for my own views on the topics mentioned below and others, see Chomsky (forthcoming) and sources cited.

Throughout, the questions seem to me unsettled. Thus, while in general the "hidden verb" analysis seems to me empirically and conceptually motivated, it certainly has not been definitively established. And within it, many open questions remain. One, with consequences for analysis of overt and hidden causatives, is whether the raising of V to v that both Pietroski and I postulate falls together with incorporation in Baker's sense (as he assumes); and more generally whether non-incorporation head-raising falls within the narrow-syntactic or the phonological component of syntax (Chomsky 2001; for counter-argument, Zwart 2001).

If head movement is narrow-syntactic, it remains to be determined whether it is the initial or raised position that enters into interpretation. Pietroski assumes the latter, for all movement operations. I think both enter, but in different ways: the initial position for argument structure, the raised position for semantic properties of different kinds (old/new information, scope, etc.). This "duality of semantic interpretation" might be the long-sought language-external motivation for transformational operations generally, I suspect. If the approach to the minimalist program I have been pursuing proves correct, the cases Pietroski discusses will be analyzed somewhat differently. There is no trace left by V-to-v movement, no problem of interpretation of trace, and no problem of c-command (it is not quite accurate to say, as Pietroski does, that the trace is c-commanded by the antecedent in head movement; that requires a complication of the definition restricted to this case). There is no interpretation for the v-V complex. The left-hand structure of figure 8.6 is not barred by a constraint on raising involving the adjunct, but because "the water" is not assigned the theme role (or any). Apparent displacement of a verb over an adjunct falls into the much-studied pattern of English Adverb-V-Object vs. French V-Adverb-Object, the Adverb being adjoined to vP in both cases but with different conditions on Tense-V amalgamation (phonological, I suspect). I think Pietroski's basic framework can be translated into these terms, but with modifications.

On ternary branching, the problem of why it is barred remains even if we keep solely to the operation Merge, inescapable in any framework (I take it to be set-formation; if it is complicated to include order, it is concatenation). I think the problem may have a solution in terms of economy of search, within the feature-

matching approach to which Pietroski alludes, though otherwise it remains mysterious. The same approach yields a straightforward account of Burzio's generalization.

More generally, Copy is not a new relation in addition to Merge. It is simply Merge, and is available freely unless there is a stipulation barring it: Copy is simply "internal Merge," forming an unproblematic syntactic object that is on a par with the set {3, {5, 3}}. Its properties, I think, can be reduced to elementary conditions of efficient computation that are in large part language-independent. There is no Delete operation: rather, a specific case of a general economy condition that operations be minimized: in this case, transfer of syntactic objects to the phonological component. Thus "the copy theory of movement" comes essentially free, without stipulation. It follows that an optimal language should make use of the device of "transformation," rather than others that have been proposed, to yield the duality of semantic interpretation. And that seems to me true for natural language. I don't agree that "copies" create problems of interpretation. On the contrary, they give the simplest account of a basic property of language that is dealt with by some device in every approach: the fact that in "what did John see," the scope of "what" is determined by its overt position and the argument structure is determined as in "John saw that." An extra benefit is that we derive a very natural approach to so-called "reconstruction" phenomena, without the mechanisms introduced on other assumptions.

My point here is solely that the questions are unsettled empirical ones. It may turn out that there really is a minimalist answer to them; that is, an answer determined on best-theory grounds with minimal language-specific stipulation. But none of the particular proposals that Pietroski discusses, or the alternatives I have mentioned, can be assigned that status today with much confidence. The research program is at an early stage, and is posing problems that are largely new, apparently quite difficult, and still in the process of being formulated with a degree of clarity. To the extent that it is successful, it will reinforce the long-expressed hope that close study of language, a recent product of the most complex organ known to us, may contribute to the understanding of a basic problem of the biological sciences: to discover how the nature of organisms is determined by fundamental principles that establish the conditions within which evolutionary processes pursue their course (Chomsky 1965: 59).

However the dust settles, I think there is good reason to anticipate that the ideas Pietroski is developing will find an important place, contributing to the realization of the prospect to which he alludes at the end, thus giving some substance to the traditional picture of language as "a mirror of mind."

Reply to Millikan

Ruth Millikan closes her essay by asking whether we disagree, and if so where. I share her lack of confidence about the answers. I'll mention an area of clear agreement, then concentrate on areas of apparent disagreement in an effort at least to sharpen the issues.

Three notions figure prominently in Millikan's proposals: convention, public language, and (primary) function. We agree that a good part of recent discussion of these topics should be discarded, and that her revision (here and elsewhere) is a substantial improvement. In particular, her work has provided a more principled account of a useful notion of "function" than anything else I know. I am not persuaded, however, that the improvement is leading down a productive path for the particular case at hand.

I should clarify that I have no problem with describing behavior as conforming to conventions. Much of what I've written about the topics Millikan addresses is quite conventional; for example, what she takes to be my "inveigh[ing] against both commonsense and technical notions of public language or 'externalized language'." I'd prefer to put it differently: I endorse the conventional views adopted in the empirical study of language, and we agree that we should disregard what she calls "common views of public language."

Furthermore, I do not reject Millikan's core belief that "language is so obviously conventional"; in fact, I endorse it in the sole passage of mine she quotes dealing with language and convention, and in exactly her terms: one of the "conventional aspects of language" is that one "calls a chair 'a chair'," or to take her example, that we "call a spigot 'a spigot'." Millikan seeks to explain "what has brought Chomsky to deny these rather mundane truths." In this case, at least, there is nothing to explain, though considerably more to explore.

I hope it is also clear that I think that commonsense notions merit attention. That is, again, apparent in the passage of mine on "public language" that she cites. We should attend to the fact that such notions are not uniform; even the word "language" does not translate easily into closely related languages, and conceptions vary widely in other cultural settings, facts of some interest (they enter into Saussurean linguistics in crucial ways, as well known). As she quotes, I think we should also take seriously the teleological aspect of commonsense notions: the fact that if Gianni is a normal four-year-old child in Venice, we have no way of referring to the language that he has acquired, but can only say that he is on his way to acquiring some version of what we here loosely call "Italian" (perhaps a second language for Gianni). All of this could fall within an interesting chapter of what I've been calling "ethnoscience" in these replies: the study of commonsense understanding and its cultural varieties.

We do not, however, confuse such investigation of "commonsense notions of public language" with the effort to gain some theoretical understanding of what

language is and how it is acquired and used, just as the physical sciences no longer rely on the commonsense understanding that objects fall to their natural place, or that a cannon ball shot straight into the air from a moving platform (a vehicle, or the earth if it is moving) will fall behind the gun, or that air is a pure simple substance.

As for "technical notions" of public or externalized language, I do not inveigh against them but rather ask for clarification. To take examples from the philosophical literature that I've discussed (Chomsky 1985, among others), suppose Jones proposes that "linguistics," as distinct from "psychology" and science in general, must restrict itself to certain specified categories of evidence. We have a right to ask for justification for what seem to be odd and pointless terminological proposals, which depart sharply from the practice and concerns of much linguistic research, traditional or contemporary. Or suppose Jones takes a natural language to be a set of "grammatical" or "well-formed expressions." Unless Jones offers some indication of a generative procedure or a criterion of well-formedness, the technical notion remains at the level of examples and hints. Jones might also be asked to explain why he believes that there is such a category in natural language, contrary to what has been assumed in work on generative grammar since its modern origins (Chomsky 1955). Introductory expository passages have misled many readers, but the point was emphasized from the start, and, to my knowledge, has not been challenged in any relevant way. A generative procedure does determine a significant category of entities, namely, the set of "expressions" that it *strongly* generates, what I've called "the structure of the language." That's a potentially well-defined notion, but I have never inveighed against it; rather, I have always regarded it as a primary object of inquiry in the study of I-language, though I think we should construe expressions as higher-order properties of the I-language, rather on a par with possible trajectories of Halley's comet in the solar system, or possible visual images for cats (vs. bees) (Chomsky 2001).

Similar questions arise in interpreting Millikan's notions of convention and public language. "What is conventional," she writes, "is an activity or a pattern of activities"; "public language" is to be defined in terms of "conventional," in this sense. But we face the same problem as with other technical notions of "externalized language." What constitutes "a pattern of activities," and how is the infinite variety of such patterns determined? If the notion is to contribute to theoretical understanding, we need some indication of how such questions are to be answered. One answer is some version of I-language; nothing else is suggested here or in the sources cited, or elsewhere to my knowledge.

To make the problem concrete, consider Millikan's example (1), adding similar ones modeled on (2) in the reply to Gopnik (ADJ = "unlikely" or "impossible"):

(1) the missionary (M) is ready to eat
(2) (i) his enemies are ADJ to eat M
(ii) he is ADJ to eat M

(iii) M is ADJ to eat
(iv) M is ADJ to expect anyone to eat
(v) M is ADJ to meet anyone who will eat
(vi) M is ADJ to be eaten

We agree that (1) falls under distinct patterns. What are they? How do these patterns yield the interpretations of (1) or (2)? What is the significance of the fact that the patterns conform to those found in a variety of other constructions (interrogatives, relatives, comparatives, parasitic gaps, . . .), which sometimes have overt indication that an element ("who," "how many books," "pictures of which," . . .) is displaced from the position in which it receives its semantic role in corresponding indicative constructions? These are the kinds of questions that have been raised and studied since the origins of modern generative grammar almost fifty years ago.

We have two choices at this point. We can content ourselves with the observation that the phenomena fall into conventional patterns of activity; or we can try to find the principles that yield these consequences, thus giving some substance to the notions "convention" and "pattern of activity." But now we are back to the investigation of I-language (however conceived), and to the conclusion that it is an internal state of individuals who have come to know that instances of these patterns have specific forms and meanings over an unbounded range. We quickly find that the principles of I-language that enter into the use and interpretation of these patterns are not conventional in even the loosest sense of the term (I think Millikan agrees). And though the point is contested, I see no serious alternative to some version of what critics call "the innateness hypothesis" (see Gopnik and reply).

Suppose we add Millikan's proposal that some of the patterns of conventional activities are "reproduced . . . in certain respects," guided by "the young child's language faculty as a filter through which language conventions are to be transmitted," "a faculty engaged in the accumulation of a larger and larger repertoire of conventional patterns it can recognize and reproduce on demand," conventions that are "complex, and not particularly systematic" and constitute the "public languages" German, French, and others. Each such public language is "real sort of stuff in the real world," consisting "of actual utterances and scripts."

From the standpoint of individual psychology, Millikan continues, "public language is merely a stimulus to transition from the initial state of the language faculty" to later states. I understand this to mean that some selection from "public language" is the stimulus. This selection is what is often called "primary linguistic data (PLD)": whatever parts of the blooming, buzzing confusion are selected by the child's mind and interpreted in specific ways to induce state changes of FL. FL maps PLD to I-language. But in Millikan's account, it also maps PLD to a finite and growing "repertoire of conventional patterns"

accumulated in the child's mind. Some array of such finite repertoires constitutes such "public languages" as German, each a finite collection of actual spoken and written utterances.

The same problems arise, but now in a more severe form. From the earliest stages, Gianni's "repertoire" is unbounded, facts familiar to Galileo and Descartes, if not before. Insofar as Gianni can "recognize and produce . . . conventional patterns," it is not because they are in some "repertoire," stored or not. Rather, the new utterances that Gianni constantly recognizes and produces are derived from some internal generative system. Gianni cannot recognize and produce utterances of the public language Italian as spoken elsewhere, just as I can't understand what I hear in the streets of Glasgow, or parts of the United States, where people are talking what we all call "English." If English "consists of actual utterances and scripts," does it include what appears in this book, or what I heard someone say at the bus stop yesterday, or what people will say tomorrow for the first time in human history (using new "patterns," in any sense of the term that has been given any substance), or any other instances of the unbounded array of utterances and scripts of which ordinary language use is a sample? Furthermore, if the public language Italian is "merely the stimulus" for state transition for Gianni, then it is his accidental and finite linguistic experience, of which he knows essentially nothing; apart from the margins, we cannot remember whether we have heard the utterances (1) and (2), or any others. And the selection of public language that stimulated Gianni's state transitions only weakly resembles the one that did so for the kid down the street, let alone in Naples.

The problems that arise at once seem overwhelming, and I don't see how they are addressed, even approached, by invoking "patterns," "coordination," "convention," and the like.

We should also take some care with the notion of "convention," much too casually invoked in the comment of mine that Millikan quotes. We can say that it is conventional to call a chair "a chair," as both Millikan and I do, but only if we deprive the term "conventional" of connotations of agreement, compact, intention to conform, etc. The linkage of sound and meaning in the lexicon of Gianni's I-language (or in this case, his cousin Charlie's) is conventional only in the sense of Saussurean arbitrariness: FL does not determine that sound and meaning must be linked to yield English "chair," though it does impose intricate constraints on the choice of sound and meaning complexes that can enter into such linkages. It is an additional empirical claim to hold that intentions enter into establishing this linkage when Charlie acquires the lexical item "chair," one that seems to be dubious at best, at least if the term "intention" has something like its ordinary sense; and if some technical notion is intended, we want to know its meaning, the context of theoretical explanation in which it is introduced, and the justification for these moves, all lacking as far as I am aware.

Putting intricacies aside, there are many steps in cognitive development that

are "arbitrary" in something like the sense of lexical linkages: a songbird nestling picking up one or another "dialect," which it uses only after maturation; Gianni's coming to have the /r/–/l/ distinction, but not his cousin Gigi who grew up in Tokyo; acquisition of the lexical item "book" by Charlie but "libro" by Gianni; setting of the head parameter (the same way for Gianni and Charlie, but differently for Gigi) and the null-subject parameter (the same at least superficially for Gianni and Gigi, but different for Charlie); and so on. In all cases, experience plays a role in determining how these steps are taken. There are all sorts of differences among the cases. Calling some of them "conventional" seems to add nothing in the way of understanding; and any further empirical claims (e.g., that some involve intention) have to be justified.

As we have all learned from John Austin and others, departures from common usage often carry a cost: in this case failure to make important distinctions (of the kind Millikan mentions in other connections) between normal usage of the term "conventional" and the departure from it when we term lexical arbitrariness "conventional." Take, say, the word "terror." The term has a literal meaning which we (more or less) share in our internal lexicons. But it also has a conventional usage that crucially departs from the literal meaning: "We must recognise," political scientist Michael Stohl observes, "that by convention – and it must be emphasised *only* by convention – great power use and the threat of the use of force is normally described as coercive diplomacy and not as a form of terrorism," though it commonly involves "the threat and often the use of violence for what would be described as terroristic purposes were it not great powers who were pursuing the very same tactic," in accord with the literal meaning of the term in our internal lexicons (Stohl 1988; more generally, George 1991). There are also important distinctions between lexical decisions that really do involve intention and others that appear to be much like the "tuning" of the system in the case of the songbird, establishing a phonological system, and parameter setting. There was a classic debate, for example, over whether the phrase "United States" should be taken to be singular or plural in the opening words of the preamble to the US Constitution, a debate serious enough to have been a leading factor in a murderous and destructive civil war.

Even in cases of much lesser moment, such distinctions are worth observing. There is a substantial difference between the ways most of us learned the words "chair" and "spigot," or "livid" and "disinterested," and the ways in which some of us may have replaced the latter two by new lexical items with a different meaning, with an intention to conform to instruction and authority. But the more important point is that invoking "convention" and "patterns" terminates inquiry where it should begin, and leaves us with a notion of "public language" that seems to me to lie somewhere between highly problematic and unintelligible.

Before we turn to the third notion, "function," another clarification may be in order. I have not "denounced the notion that the purpose of language is communication," but again, have only asked for clarification. I don't know what

is "the purpose" of the skeleton, or the ear, or insect wings; they seem to serve many purposes, and their complex evolutionary origins are not closely related to various functions now or in earlier stages. Furthermore, what are in some intuitive sense their current "primary functions" seem often only weakly related, if at all, to "reproducing" or "sustaining" them in anything like Millikan's sense, though they may well have played some role, maybe an important one, in their evolutionary history, an entirely different matter. In the case of language, from some reasonable points of view, its "primary function" is internal dialogue: virtually all use of language falls into this category, as can easily be verified by self-observation (though I do not suggest that this is *the* or even *a* "purpose of language"). On the role of social interaction in sustaining language, Millikan has some suggestions to which I will return, again with a skeptical note.

Millikan points out that "learning language is not merely acquiring an I-language." That is unquestionably true, if by "learning language" we mean what is described as such in common discourse. Narrowly technical work insists that no one has an I-language, which is a high-order idealization (introduced with the understanding that "idealization" is a necessary prerequisite to gaining some grasp on reality).

Millikan asserts that "learning language is essentially coming to know various *public* conventions and, with trivial exceptions, these conventions are around to learn only because they have functions"; and she seeks to explain why I "deny these rather mundane truths." Insofar as I understand these observations, I have little to say about them. As noted, I do not deny but rather assert the "mundane truths" in some cases, though with qualifications. I do not see how we have progressed if we introduce the technical term "public conventions" to cover the various ways (all? only some? which?) in which experience enters into fixing properties of FL or other cognitive systems, failing to make what seem to be important distinctions.

In some respects, Millikan's terminology comes close to being complementary to what seems to me common usage. If I understand her correctly, the principles of anaphora that enter into interpretation of (2i, ii) are "conventions of grammar"; in my informal usage, these are no more "conventions" than properties of the mind/brain that determine the framework within which the visual system attains a certain state (but not other imaginable ones). The observation generalizes.

Furthermore, it does not seem to me a "mundane truth" that in the (fairly typical) cases mentioned, the phonological distinctions, setting of the head parameter, fixing of Saussurean arbitrariness, principles of anaphora, etc. are "there because they have functions" – that is, because they "enable various kinds of communication." Even granting the notion "function" (obscure in this context, I think), I do not understand the force of the word "because." Does it add empirical content? Has some evidence been discovered to verify the causal relation asserted? Do we have some idea how to obtain such evidence? Is it "because of these functions" that Charlie came to know how to interpret the

pronoun in (2i, ii), or the adjectival constructions throughout (1)–(2), or the complex phonetic and semantic properties of "book"? Again, the questions proliferate. I do not see how the "third alternative" that Millikan proposes avoids the problems that (we agree) undermine efforts to show that "language functions would have to reduce to or be derived from speaker intentions."

We can agree that the indicative mode can be used to provide information, and that people process expressions in the indicative mode in accord with their internal resources and sometimes use the information to modify their beliefs. But what more is being said – and specifically, what is the force of "because" – in the claim that "H responds to the indicative sentence by translating it into belief in accordance with certain semantic rules because, in H's experience, responding selectively to indicative sentences in this way has often enough resulted in the appropriation of useful information"? Has this paraphrase of a near truism added some substantive content? If so, do we have evidence to confirm or refute what is added, or any idea of how to obtain such evidence?

There may be some sense in which "public conventionality is of the very essence of human languages." Thus, although there are attested cases of development of apparently normal language by a few children in the absence of any relevant experience from outside (Goldin-Meadow and Feldman 1977; Goldin-Meadow and Mylander 1990), there are none of children in complete isolation. Presumably some interaction is necessary (though no one really knows, since isolation imposes extreme psychic trauma). If that is what is meant, then "public conventionality" is being used in a new technical sense, and it is no more "the very essence of human languages" than having a sensory system with at least limited functioning and not too much brain damage, also essential for language acquisition. If that is not what is meant, then what was "the very essence" of the languages constructed by the children in Goldin-Meadow's cases? In what sense is "public conventionality" the "very essence" of what Gianni, Gigi, and Charlie acquired? I have no opinion on the matter, and do not see how others can.

It is entirely reasonable to suppose that without meaningful social interaction (Millikan's "communicative coordination"), a child's FL would not develop, and it is possible that it "would atrophy," like other systems. If Gianni joins his cousin in Tokyo at age four, his Venetian dialect would probably be lost undetectably (by present means at least). If he does so at age twenty, the situation is much more complex. Suppose that after moving to Tokyo, Gianni's only contact with his native language is reading Italian journals. He might be more fluent in Italian after forty years in Tokyo than when he arrived, though no social interaction has taken place. We could, of course, stretch the meaning of the term "communicative coordination" to cover this and many other cases, even internal dialogue, but it's not clear that anything is achieved beyond some degree of conformity to dubious philosophical convention.

Furthermore, even if Gianni stays in Venice and lives a normal life there, with plenty of social interaction, is it true that this interaction is what sustains the

specific "choices" made as his I-languages matured? Do I rely on "communicative coordination" to sustain my understanding of the word "table" or "circumspection," or my reliance on the head-first parameter? It's one thing to claim that fluency will erode with lack of use, but quite another to hold that "communicative coordination" is required to sustain fluency, point by point or even in some vague way. Millikan's reference to expletives seems to me to obscure the generality of the problem.

I do not see what follows about the "very essence" of language beyond the observation that experience plays a role in fixing I-language and that social interaction may also be required (as is also true in some form for vision, motor coordination, etc.); and that if some cognitive capacity is not used, it may not be readily accessible, though using it in isolation apparently suffices for FL.

Millikan points out that "pen" is pronounced differently in Tennessee and Connecticut, but tells us that "it's exactly the same word." There is also a word "milkshake" that means one thing in Philadelphia and another elsewhere, as I learned to my dismay on arriving in Boston long ago. The word "bonnet" is pronounced differently in London and Boston, and is used with a different meaning, but we can, if we like, say that "it's exactly the same word." Such informal modes of individuation are fine, but they are routinely abandoned, and rightly, as soon as one undertakes serious inquiry into the nature, use, and acquisition of language (just as notions like "Italian" or "Chinese" are abandoned). There are no general answers to the questions Millikan raises about individuation of linguistic elements, though they can be studied within the framework of I-language and various more complex constructions built on it in ways that depend on our particular interests and concerns.

Millikan asserts that we "type words" in terms of their "lineage," bringing out their "conventionality, hence public nature." Since all these terms are being used in some technical sense that I only partially grasp, I have no opinion about the conclusions. Some such course is followed in etymological studies, quite constructively, but Millikan is suggesting something different. To the extent that I understand it, I do not see how the enterprise will avoid the kinds of difficulties that have been repeatedly raised (Mercier 1993, 1994, 1998; Segal 2000). Even if it can, it seems clear that it departs sharply from the empirical study of the nature, use, and acquisition of language, at least within the "biolinguistic" framework of the past half-century (Jenkins 2000), and far more broadly. Plainly, Gianni and his cousins know nothing about lineage.

At the end, like Millikan, I am not at all confident about whether we disagree, and if so, where. I am unable to formulate any clear disagreement about matters of theory or fact. I don't see any disagreement about how one would proceed to inquire into the various aspects of language. As best I understand, her discussion provides an informal way to characterize empirical inquiry into language. Much seems to me lost; I do not see what is gained.

Reply to Gopnik

Alison Gopnik counterposes what she takes to be two alternatives in the study of cognitive development: the innateness hypothesis (IH-G) and the theory theory (TT). By IH-G I mean the version of an innateness hypothesis IH that Gopnik formulates and attributes to me and others who work in the same general framework, what Jenkins (2000) calls "the biolinguistic framework." The topics she addresses are of considerable significance, but I have some problems understanding her alternatives.

There has been a great deal of controversy about IH, but with an oddly one-sided character. IH has been sharply criticized, but it has never been formulated or defended. True, specific proposals have been made about innateness: with regard to the growth of limbs, the circulatory and visual systems, motor organization, the waggle dance of bees, the human faculty of language FL, and much else. But no one has ever told us what IH is supposed to be.

To her credit, Gopnik does offer a definition of a particular version of IH, namely IH-G, which is defined by the principles (I):

(I) (i) "human minds are highly constrained innately";
(ii) "we can only formulate a very small set of possible representations and rules";
(iii) "the constraints remain unchanged throughout life";
(iv) "representations and rules are not inferred or derived from the input."

It is difficult to evaluate (iv). Everyone agrees that attained states are "derived from the input" in some way, but not by inference. Nonetheless, (iv) hints at a distinction that is significant and that has been extensively discussed, between approaches that have been described as "rationalist" vs. "empiricist" in essence; for good reasons, I believe (Chomsky 1965, 1966, 1975a, among others). But as TT is formulated (see below), it may well fall on the same ("rationalist") side of the divide as the parameter-setting models, or their predecessors, to which Gopnik proposes TT as an alternative. I don't know of other relevant ways of elaborating condition (iv), and will therefore put it aside.

As for (iii), I know of no version of IH for FL that insists on it as stated here, though something similar has indeed been proposed (and denied) for language acquisition within the approaches that Gopnik is criticizing. This "uniformity hypothesis," however, does not reach as far as (iii), and furthermore, has been proposed as an empirical hypothesis, not part of a definition of IH for FL. But it is not clear how TT differs in these respects (see (IIIii), below).

Accordingly, I will restrict IH-G to (i) and (ii).

Principle (i) is the special case, for this specific domain, of what I am calling IH:

(II) Organisms and their components are highly constrained innately.

Without further elaboration, IH is a truism (hence (Ii) as well) and (ii) tells us nothing more. With these emendations, IH-G is the same as IH, so I will not distinguish them further.

Gopnik believes that IH was proposed to account for the nature, variety, and acquisition of I-languages (understood to be internal computational systems that generate expressions with form and meaning), and other questions concerning them. That is not accurate. IH is unquestioned in the study of organisms. Specific versions of IH have been devised for the study of FL: for example, to explain the fact that in a sense of "know that" that seems relatively uncontroversial (Gopnik may disagree), I and others have come to know that (1). And furthermore, with regard to (2), with V = "forget" (or "understand," "forgive," etc.), we know that in (i) but not (ii) the pronoun can be understood to refer to Clinton; that (iii) is about Clinton's forgetting (etc.) someone or something but not (iv)–(vi); and that (iv) is about Clinton's expecting something. And we know further that if "unlikely" is replaced by "impossible" throughout, the meanings change radically: (i)–(ii) and (v)–(vi) are gibberish, while (iii) and (iv), respectively, have to do with someone else forgetting (etc.) Clinton or expecting someone to do so. Furthermore, some of the expressions for which these judgments are made are unnatural and require some thought, and therefore are particularly interesting, because they stretch the limits of knowledge of language – of the ability to use the resources of I-language; thus (iv), with "impossible," is unnatural, but with attention it becomes clear that the meaning is determinate, and the distinction between (iv) and (v) is quite clear:

(1) "chase" rhymes with "lace" and entails "follow" (with a certain intent, etc.)
(2) (i) his enemies are unlikely to V Clinton
(ii) he is unlikely to V Clinton
(iii) Clinton is unlikely to V
(iv) Clinton is unlikely to expect anyone to V
(v) Clinton is unlikely to meet anyone who will V
(vi) Clinton is unlikely to be V-ed

Furthermore, the crucial underlying principles appear to be universal for FL and the variations among languages appear to be quite restricted. Such properties of FL are what we would expect if human mental faculties are part of the biological world, satisfying IH as does everything else.

These examples are typical of the concerns of what Gopnik calls "syntax" – as I do too, recognizing, however, that it is common to refer to the topics as semantics, since they crucially have to do with the meanings of expressions and

the structural conditions that play a substantial role in determining how they are used: to refer, and in other ways.

As Gopnik observes, IH is intended to have much broader scope, and does. It is taken for granted, and sharpened in specific ways, in the work that has led to "the greatest successes in cognitive science . . . particularly vision" (Gopnik): for example, study of the effects of very early experience on distribution of receptors in the visual cortex, recognition of faces and body schemata, the interpretation of 2D images as rigid 3D bodies in motion, and so on. More generally, it is refined in specific ways to provide the "problem-specific structure" of learning processes "that makes learning possible," constraining attainable solutions (Gallistel 1999), analogues of properties found throughout the biological world.

The versions of IH developed for FL adopted from the start the view that Gallistel calls "the norm these days in neuroscience": the "modular view of learning" that takes the brain to be a complex of "specialized organs," "learning mechanisms" that we can regard as "organs within the brain [that] are neural circuits whose structure enables them to perform one particular kind of computation," as they do more or less reflexively, apart from "extremely hostile environments," by virtue of internally directed processes and environmental effects that trigger and shape development. In humans, one of these "learning mechanisms" is FL (see, e.g., Chomsky 1975a), which yields such outcomes as those just illustrated on the basis of experience and an innate initial state that constrains the course of development and possible outcomes.

One might go on to adopt Gallistel's stronger thesis that the operation of these cognitive organs exhausts cognitive development. As he puts it: "To imagine that there exists a general purpose learning mechanism in addition to all these problem-specific learning mechanisms . . . is like trying to imagine the structure of a general purpose organ, the organ that takes care of problems not taken care of by adaptively specialized organs like the liver, the kidney, the heart and the lungs," or a "general purpose sensory organ, which solves the problem of sensing" for the cases not handled by the eye, the ear, and other specialized sensory organs. Nothing like that is known in biology. "Adaptive specialization of mechanism is so ubiquitous and so obvious in biology, at every level of analysis, and for every kind of function, that no one thinks it necessary to call attention to it as a general principle about biological mechanisms." Accordingly, "it is odd but true that most past and contemporary theorizing about learning" departs so radically from what is taken for granted in the study of organisms – a mistake, he argues (Gallistel 1997, 1999).

All of this falls within the ethological framework developed by N. Tinbergen and others, which was an important formative influence on generative grammar at its origins in the 1950s. Of particular interest is what C. H. Waddington called "canalization": processes whereby the endstate of development is relatively independent of environmental perturbations; that is, heavily constrained by genetic constitution, hence particularly informative about them (cf. Ariew 1999).

Let's return to a closer look at (I) and TT. In (Ii), the word "highly" is unobjectionable, but as an impressionistic (though not worthless) qualification; there is no metric. That aside, (Ii) (= IH) is a truism. If not highly constrained innately, human minds could attain no regular form or states, nor play any role in human life, for familiar reasons. TT thus cannot be an alternative to IH, so construed, but only a special version of it; and as described, it is.

Gopnik's TT adopts the principles (III):

(III) (i) "the representational system will change in relatively orderly, predictable, and constrained ways";
(ii) "there may still be some overall constraints of representations that are generated; not every logically possible theory will be formulated and tested by humans."

(IIIi), shared with the other versions of IH that Gopnik is considering, essentially restates (Ii). (IIIii) also falls within IH, if we interpret "may still be" as "still are," as we must if TT is to have any substance. It is therefore not true that innateness in the sense of TT is not "a claim about the constraints on the final state of the system" (Gopnik). TT differs from the versions of IH that Gopnik has in mind in ways to which I'll return, but not in this respect.

Like other versions of IH, TT faces the task of determining the innate properties – the structures of the "initial state" – that account for the fact that changes are "relatively orderly, predictable, and constrained," and for the "overall constraints" on attainable outcomes to which (IIIii) refers.

TT differs from IH, according to Gopnik, in that IH requires (Iiii). As noted, that is incorrect. Like TT, every other version of IH (including those that adopt the "uniformity hypothesis") assumes that constraints can change, though only within innately determined limits. In the case of language, for example, it is universally recognized that the shaping of the phonological system by early experience strongly influences the way speech is perceived and interpreted in later life, by changing the constraints of the initial state, and the same appears to be true far more generally. It is also possible that new constraints arise through maturation, and that others are modified in many ways: for example, by parameter setting that opens the option of new parameter choices (see Baker 2001 for review).

TT also differs from IH, Gopnik argues, in that "the range of theories that can be generated" within the constraints of TT "is very much wider than the limited set of representations that are possible on innateness views." This cannot be correct; as Gopnik formulates TT, it is an "innateness view." Fortunately; if it were not it would have no substance. But putting that aside, the term "very much wider" is unclear (which is not to dispute the intuitive thrust of Gopnik's observation). Specific proposals of the theories of vision, language, etc., seek to discover the restrictions, as does TT. There is no meaningful numerical com-

parison of, say, the range of outcomes allowed by specific theories of vision or language with the range of outcomes permitted by TT. In the case of language, it has always been assumed that the range is infinite: there is no limit on the number of monomorphemic nouns in the lexicon of languages of the English type. With regard to the operations that form expressions (e.g., (2)), until about 1980 the approaches Gopnik has in mind assumed that there are infinitely many possible I-languages; more recent approaches that suggest that the number of options may be finite are working hypotheses about a framework for theory (the "principles and parameters" framework), an important conclusion about IH for FL if true. As for the range of outcomes permitted by the TT variant of IH, nothing can be said until more information is provided about the constraints.

The model for TT is natural science, which is also constrained by some version of IH. But when we proceed beyond the surface, the import of this observation is not at all clear. Consider the example that Gopnik rightly singles out for special mention, the visual system. Normal growth of the visual system is based on highly restricted inputs and takes place, reflexively, over a brief period; and even if "higher-level vision" introduces familiar objects as "templates" of some kind, that too is highly constrained in time and range of data, and surely in possible outcomes (Ullman 1996). That is why vision was taken long ago as a suggestive model for study of FL within the biolinguistic approach (as was scientific inquiry; see below). But although adding more time may change little, radical tampering with inputs may. It might, for example, convert parts of the dedicated auditory system to something like a visual system (see reply to Rey); or, conceivably, allow mammals to develop compound eyes, given the commonality of fundamental genetic structure that has recently been discovered. Scientific inquiry proceeds self-consciously over many centuries, building on earlier achievements and with greatly varied data selected theory-internally. The fact that it appears to yield varied outputs tells us very little about the differences – which surely exist – between the innate basis for scientific inquiry and the kinds of "learning mechanisms" to which Gallistel was referring (all within IH).

I do not want to suggest that Gopnik is mistaken to stress the asumption of heavy innate constraints on outcomes in the theories of language she is considering, as in all more or less explicit versions of IH: for vision, language, motor organization, recruitment communication of some species of bees (Nieh and Roubik 1998), etc. On the contrary, I think these qualitative observations are quite important. Elsewhere I have argued that ethologists were right to maintain that modern biology provides empirical substance to a version of traditional conceptions of *a priori* knowledge, most explicitly Konrad Lorenz, who refers to Kant, though I think a better choice would have been the seventeenth-century sources from which Kant drew (Chomsky 1966, 1968, 1975a; McGilvray 1999). It is, I think, correct and important to stress the strong rationalist strains underlying contemporary biology (and including what is misleadingly called

"enlightened empiricism," when it passes beyond vague allusion). But these are different topics.

Without proceeding further, I can see only one substantial difference between TT and other versions of IH. TT holds that the initial state for human cognitive development consists of "initial innate theories," and that successive states are attained (under whatever constraints (III) imposes) by the same "powerful cognitive processes" and "powerful learning mechanism" responsible for theory formation in the sciences. That is a departure – in several important respects – from other versions of IH, and from what Gallistel describes as "the norm these days in neuroscience" and the rest of biology.

To evaluate the proposal, we have to know what TT means by "theory," and what it assumes about theory formation in the sciences. I'll return to the second question. As for "theories," Gopnik defines them as "abstract, coherent, systems of entities and rules, particularly causal entities and rules," that can be used to predict, interpret, and explain evidence. A theory, then, is (at least) a set of assertions in some symbolic system S from which conclusions can be deduced (or at least assigned some rational warrant) that are then subject to confirmation or revision on the basis of evidence formulated within the resources of S.

If so, then TT appears to hold that the initial state of the human visual system incorporates some such "initial innate theory," along with a "powerful learning mechanism" that enables it to construct other theories (in the same sense) that fix the distribution of horizontal and vertical receptors in the first few weeks of life; that interpret binocular presentations, faces and bodies of persons (with certain orientations and actions) a few months later; that interpret presentations as 3D rigid objects in motion; and so on. The child verifies, rejects, and modifies these theories by the methods of the sciences. Kittens appear to do something rather similar, so the null hypothesis would be that they too have innate scientific theories and use methods of the sciences to construct new ones and verify them, as when they decide to stop at a visual cliff, and so on.

To select examples with a more "cognitive feel," if "the infant's perception is guided by a conception of physical objects" and "they conceive of the world as consisting of units that are internally cohesive and separately movable," disregarding other salient features of the environment (Spelke 1985), TT holds they have used the methods of science to modify their "initial innate theories" and confirm the new theories they have constructed that yield these consequences. And if kittens do something similar, we presumably draw the same conclusions about their cognitive development. Similarly, according to TT, we cannot accept the standard interpretation of the conclusion that "humans have a specialized, modular mechanism for the socially crucial task of detecting gaze direction, which works on the assumption that the dark area [in a photograph of a face] is the iris," a "cognitively impenetrable" process that accounts for some new illusions (News and Comment 2001). We must rather say that the child modifies its initial innate theories to construct a theory that has these consequences,

making use of the methods of Galileo, Newton, Lavoisier, Darwin, Einstein, and other scientists, famous or not, whatever these methods may be.

In another domain, Jacques Mehler and his associates showed that infants a few days old make interesting distinctions among languages and between language and the same signals reversed (see Mehler and Dupoux 1994, for review of this and many other aspects of infant cognition). According to TT, they do this by applying the methods of science to their innate theories, developing a theory of language with these outcomes in the first few days of life, or maybe prenatally. It has recently been shown that Tamarin monkeys perform in a similar way without training (Ramus 2000), so they too presumably have developed theories of the same kind, in the same way.

None of this seems very plausible to me, and I presume none of it is intended. But I do not understand why.

Specific proposals about cognitive development have to be evaluated case by case, and TT variants of IH can be compared with others to the extent that the constraints of the initial states are made explicit. Instead of running through examples, let's look at the core notion: formation of explicit scientific theories. Gopnik observes that cognitive naturalism should apply to the "powerful learning mechanisms" that allow human minds to develop scientific theories "from limited output, namely evidence." I entirely agree, and do not understand why Gopnik seems to think otherwise. For many years, I have been suggesting that a potential topic of cognitive science is the innate human "science-forming faculty" (SFF) that enables people, in certain complex "problem situations," to devise theoretical constructions and to select certain data that can be interpreted as evidence bearing on them (Chomsky 1968, 1975a, 1994, and elsewhere). TT holds that SFF is also responsible for cognitive development generally; on that I have no opinion, lacking relevant information about SFF (hence about TT).

We should, however, be clear about the surprising claims of TT. Unlike other variants of IH, TT holds that the human infant (and apparently other animals) literally entertains explicit theories of the kind constructed by scientists, and devises and verifies them in the same ways, but unconsciously and presumably beyond accessibility to consciousness. The products of SFF, in contrast, are entirely conscious. That is quite a radical difference between the two categories that TT seeks to assimilate: science and cognitive development. And it raises many other questions, to my mind at least. Other variants of IH, including the study of language, do not make such claims about entertaining, constructing, and verifying theories. They do not take the initial mechanisms to be theories, for humans or other animals; and they assume that these mechanisms ("learning mechanisms" in Gallistel's sense) are modified by the triggering and shaping effects of the environment in ways that have been studied extensively in many domains, and that appear to have little resemblance to theory formation in the sciences.

Gopnik feels that I should not find this radical innovation of TT problematic,

because I have not adopted the doctrine that the contents of the mind must be "accessible to consciousness" (and do not even feel that it has been given a coherent formulation). But these conclusions move too quickly among different questions. Gopnik is surely right to say that scientists lack "conscious access to their theory-formation mechanisms," but lack of conscious access to the theories constructed is quite a different matter. Also quite different is the matter of lack of conscious access to the internal computational systems that enter into ordinary cases of knowing that (how, why, etc.); what is involved in (1) and (2), for example. That seems to me unproblematic, probably the standard case. But the innovation of TT does seem to me problematic, requiring clarification and justification.

It should be noted, again, that this innovation of TT holds only for cognitive development, not for theory construction in the sciences. That fact alone raises questions about the effort to assimilate the two. But the questions become more serious when we consider what little is known about SFF.

A classical model for SFF is C. S. Peirce's account of "abduction" (Peirce 1957). Peirce holds that theories are constructed by a "guessing instinct" (abduction) that provides hypotheses to test. Successful theory construction can be explained only by assuming that "Man's mind has a natural adaptation to imagining correct theories of some kinds." This innate property of mind "puts a limit upon admissible hypotheses." It accounts for the fact that "men of surpassing genius" had to make only a few guesses "before they rightly guessed the laws of nature," despite highly inadequate data, including often disconfirming data that are shelved. The very rapid success results from the fact that the innate "natural beliefs" are true, Peirce held, a "logical necessity" because the mind is a product of nature. The general picture seems to me on the right track, and has been adduced as a suggestive analogue to language acquisition (Chomsky 1968). But many qualifications are necessary, including several just mentioned. Furthermore, the history of science shows that most theories (perhaps all) are false, not true; and the fact that mind is a product of nature tells us nothing about the validity of "natural beliefs."

Peirce recognized that abduction is a mystery. It remains so, as has been stressed particularly by Jerry Fodor (recently Fodor 2000; I should register my disagreement with Gopnik's dismissive comment on Fodor). As far as I can see, TT adopts the Peircean framework with no substantive change, except to hold that these capacities suffice to account for cognitive development generally, a conclusion that cannot be evaluated until the mysteries of abduction are unraveled. Gopnik regards the existence of these capacities, a matter on which all agree, as "a kind of demonstration proof" for the general applicability of TT; it is not. She also paraphrases Peirce's invalid argument for the truth of "natural beliefs," stating that the capacities of SFF "were designed to give us a generally veridical view of the world." The history of science shows nothing of the sort, and even where theories are (more or less) veridical, it is hard to see how adaptive

"design" enters into explaining their origin. To mention a classic case, what process of adaptation to life 100,000 or a million years ago led Newton to postulate a principle of universal gravitation that he, along with other eminent physicists of his day and later, considered to be "so great an Absurdity that I believe no Man who has in philosophical matters a competent Faculty of thinking, can ever fall into it"?

If SFF conforms generally to Peirce's expectations, as apparently assumed by TT (see (III)), then it follows that "some things are incomprehensible." Gopnik regards this as unwarranted "*a priori* pessimism." On the contrary, it is warranted on the most elementary shared assumptions: that humans are biological organisms, not angels. For this reason, it is rather misleading to say that "arguments about our innate inability to understand certain things are bound to be unconvincing." The general conclusion should be uncontroversial; specific empirical arguments, which are rare, have to be evaluated like all others. And the general conclusion is optimistic, not pessimistic; if it were not for its innate constraints, SFF could achieve nothing, just as an unprogrammed computer can do nothing and a embryo without functioning genes could not become a fly or an elephant. What falls within the scope of SFF, of course, is a matter of fact, not *a priori* pronouncements. As a matter of fact, there is some reason to take quite seriously Hume's conclusion that by showing "the imperfections of the mechanical philosophy," Newton "restored [Nature's] ultimate secrets to that obscurity in which they ever did and ever will remain," a conclusion that entered deeply into the modern scientific outlook and modified the goals of science in important ways. But the general point is hardly in doubt.

Gopnik believes that TT "clearly can be, and has been, applied to understanding semantic and lexical capacities." To illustrate, she cites accounts that "suggest that many nouns encode 'natural kinds'," and that acquisition of early words is "closely related to the child's developing theoretical understanding" in various domains. TT, however, leads to the conclusion that nouns should typically not encode natural kinds, if these are kinds of nature, postulated in the natural sciences – hardly a transparent notion (Bromberger 1996). The reason, which Gopnik outlines, is that the data available to the child are, obviously, radically different from what is considered in the mature sciences, so if TT is correct in supposing that the child and the physicist are using the same procedures, the outcomes should be radically different as well. That's a fortunate conclusion for TT, because beyond an early descriptive level, the sciences have little use for the concepts of commonsense understanding. Perhaps earth, air, fire, and water are elementary constituents of the world in some versions of commonsense understanding, but even though the picture persisted into the eighteenth century in Western science, it has of course been thoroughly abandoned.

Nevertheless, a close relation between lexical acquisition and the child's understanding (whether "theoretical" in the sense of TT or not) is surely to be expected, not only for early words. A child lacking any understanding of agency,

intention, and causality could hardly learn the word "persuade," and the same holds from the simplest examples to the most complex. That observation, however, leaves unanswered the questions of lexical semantics: for example, the nature and origin of the ways that agency enters into nameability of even simple inanimate objects, and the intricate semantic structure of the simplest words, far beyond anything discussed in the most elaborate dictionaries, topics addressed in interesting ways in classical and early modern philosophy (Chomsky 1966, 1975a, 2000; McGilvray 1999; Moravcsik 1975, 1990).

Gopnik is right to conclude that the fact that scientists lack conscious access to the mechanisms of SFF does not undermine the claim that such mechanisms exist; they surely do. The same holds generally of mechanisms postulated in the various applications of IH. The problem in the case of SFF is different. It is not clear that much if anything has been learned about the mechanisms of abduction since Peirce formulated the problem. Historians of science have done very insightful work on the development of theories: for example, I. B. Cohen's classic studies on how the revolutionary "Newtonian style" allowed Newton to study "a mathematical attractive force" even though he found the concept "to be abhorrent and not admissible in the realm of good physics" (Cohen 1985); or George Smith's careful analysis of how J. J. Thompson devised a series of experiments that laid the basis for the modern concept of the electron (Smith 2001); and a great many other cases. But these leave us in the dark about the "theory-formation mechanisms" that Peirce called "abduction." The SFF is a potential subject of inquiry in the cognitive sciences, but there is little to say about it now. Accordingly, I do not see how one can have a general opinion about TT, though particular instances can be evaluated on their merits.

In the light of TT, questions about poverty of stimulus, rationalist and empiricist approaches, the nature of knowledge, and other general topics remain as they were, as far as I can see. TT, as it now stands, seems very far from challenging 'the innateness hypothesis as a broad claim about the nature of language and mind.'

References

American Academy of Arts and Sciences 1998: *The Brain*, *Daedalus*, 127 (2), Spring.

Ariew, A. 1999: Innateness is Canalization: In Defense of a Developmental Account of Innateness. In V. G. Hardcastle (ed.), *Where Biology Meets Psychology*, Cambridge, Mass.: MIT Press.

Baker, M. C. 2001: *The Atoms of Language*. New York: Basic Books.

Belletti, A., Chomsky, N., and Rizzi, L. 2001: *Su natura e linguaggio*. Siena: University of Siena Press (Italian and English). Extended version, Cambridge: Cambridge University Press.

Bromberger, S. 1996: Natural Kinds and Questions. In M. Sintonen (ed.), *Essays on Jaakko Hintikka's Epistemology and Philosophy of Science*, Amsterdam and Atlanta: Rodopi.

Burge, T. 1992: Philosophy of Language and Mind. *Philosophical Review*, Centennial Issue, January.

Chomsky, N. 1955: Logical Structure of Linguistic Theory. Manuscript, Harvard. Most of a 1956 revision was published by Plenum and University of Chicago Press, 1975 (the latter with index). Microfilms of 1956 version at Harvard and MIT libraries.

Chomsky, N. 1965: *Aspects of the Theory of Syntax*. Cambridge, Mass.: MIT Press.

Chomsky, N. 1966: *Cartesian Linguistics*. New York: Harper & Row.

Chomsky, N. 1968: *Language and Mind*. New York: Harcourt Brace Jovanovich. Enlarged edition 1972.

Chomsky, N. 1969: Knowledge of Language. *Times Literary Supplement*, May 15.

Chomsky, N. 1975a: *Reflections on Language*. New York: Pantheon.

Chomsky, N. 1975b: Knowledge of Language. In K. Gunderson (ed.), *Language, Mind, and Knowledge*, Minneapolis: University of Minnesota Press.

Chomsky, N. 1980: *Rules and Representations*. New York: Columbia University Press.

Chomsky, N. 1985: *Knowledge of Language*. New York: Praeger.

Chomsky, N. 1994: Naturalism and Dualism in the Study of Language and Mind. *International Journal of Philosophical Studies*, 2 (2), September, 181–209. Reprinted in Chomsky 2000.

Chomsky, N. 1995: Language and Nature. *Mind*, January. Reprinted in Chomsky 2000.

Chomsky, N. 1997: Language and Cognition. Address to Cognitive Science Society (July 1990). In D. M. Johnson and C. E. Erneling (eds.), *The Future of the Cognitive Revolution*. New York: Oxford University Press.

Chomsky, N. 2000: *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press.

Chomsky, N. 2001: Derivation by Phase. In M. Kenstowicz (ed.), *Ken Hale: A Life in Language*, Cambridge, Mass.: MIT Press.

Chomsky, N. (forthcoming): Beyond Explanatory Adequacy. In A. Belletti (ed.), *Structures and Beyond: Current Issues in the Theory of Language*, Oxford: Oxford University Press.

Cohen, I. B. 1985: *Revolution in Science*. Cambridge, Mass.: Harvard University Press.

Fodor, J. 2000: *The Mind Doesn't Work that Way*. Cambridge, Mass.: MIT Press.

Gallistel, C. R. 1990: Introduction. In Gallistel (ed.), *Animal Cognition*, *Cognition*, 37 (1–2), November (special issue).

Gallistel, C. R. 1997: Neurons and Memory. In M. S. Gazzaniga (ed.), *Conversations in the Cognitive Neurosciences*, Cambridge, Mass.: MIT Press.

Gallistel, C. R. 1999: The Replacement of General-Purpose Learning Models with Adaptively Specialized Learning Modules. In M. S. Gazzaniga (ed.), *The Cognitive Neurosciences*, 2nd edn. Cambridge, Mass.: MIT Press.

George, A. 1991: *Western State Terrorism*. London: Polity.

Gold, I. and Stoljar, D. 2001: A Neuron Doctrine in the Philosophy of Neuroscience. *Behavioral and Brain Sciences*, 22 (5), October, 809–69.

Goldin-Meadow, S. and Feldman, H. 1977: The Development of Language-Like Communication Without a Language Model. *Science*, 197, 401–3.

Goldin-Meadow, S. and Mylander, C. 1990: The Role of Parental Input in the Development of a Morphological System. *Journal of Child Language*, 17, 527–63.

Halle, M. 1983: On Distinctive Features and their Articulatory Implementation. *Natural Language and Linguistic Theory*, 1, 91–105.

Hauser, M. 1998: *The Evolution of Communication.* Cambridge, Mass.: MIT Press.

Heilbron, J. L. 1964: A History of the Problem of Atomic Structure from the Discovery of the Electron to the Beginning of Quantum Mechanics. Ph.D. dissertation, University of California, Berkeley.

Heim, I. and Kratzer, A. 1998: *Semantics in Generative Grammar.* Oxford: Blackwell.

Hellman, D. 1998: *Visual Intelligence.* New York: Norton.

Henry, J. 1986: Occult Qualities and the Experimental Philosophy. *History of Science*, 24, 335–81.

Jenkins, L. 2000: *Biolinguistics.* Cambridge: Cambridge University Press.

Kandel, E. and Squire, L. 2000: Neuroscience. *Science*, 290, 1113–20.

Kanwisher, N. and Downing, D. 1998: Separating the Wheat from the Chaff. *Science*, 282, 57–8.

Kayne, R. 1975: *French Syntax*. Cambridge, Mass.: MIT Press.

Kors, A. 1992: The Atheism of D'Holbach and Naigeon. In M. Hunger and D. Wootton (eds.), *Atheism from the Reformation to the Enlightenment*. Oxford and New York: Oxford University Press.

Koyr, A. 1957: *From the Closed World to the Infinite Universe.* Baltimore: Johns Hopkins University Press.

Lange, F. 1925: *The History of Materialism.* London: Kegan Paul, Trench, Trubner.

Lindley, D. 2001: *Boltzmann's Atom.* New York: Free Press.

Lycan, W. G. 1987: *Consciousness.* Cambridge, Mass.: Bradford Books/MIT Press.

Mausfeld, R. 2001: Conjoint Representations and the Mental Capacity for Multiple Simultaneous Perspectives. Manuscript, Institute for Psychology, Kiel.

McGilvray, J. 1999: *Chomsky.* Malden, Mass.: Blackwell.

Mehler, J. and Dupoux, E. 1994: *What Infants Know.* Cambridge, Mass.: Blackwell.

Mercier, A. 1993: Normativism and the Mental. *Philosophical Studies*, 72, 71–88.

Mercier, A. 1994: Consumerism and Language Acquisition. *Linguistics and Philosophy*, 17, 499–519.

Mercier, A. 1998: On Communication-Based *De Re* Thought, Commitments *De Dicto*, and Word Individuation. In R. Stainton (ed.), *Philosophy and Linguistics*, Boulder: Westview.

Moravcsik, J. 1975: Aitia as Generative Factor in Aristotle's Philosophy. *Dialogue*, 14, March, 622–8.

Moravcsik, J. 1990: *Thought and Language*. London: Routledge.

Mountcastle, V. 1998: Brain Sciences at the Century's Ebb. In *The Brain*, *Daedalus*, 127 (2), Spring.

News and Comment 2001: *Trends in Cognitive Science*, 5 (4), April.

Nieh, J. and Roubik, D. 1998: Potential Mechanisms for the Communication of Height and Distance by a Stingless Bee, *Melipona Panamica*. *Behavioral Ecology and Sociobiology*, 43, 387–99.

Peirce, C. S. 1957: The Logic of Abduction. In V. Tomas (ed.), *Peirce's Essays in the Philosophy of Science*, New York: Liberal Arts Press.

Popkin, R. 1979: *The History of Scepticism from Erasmus to Spinoza.* Berkeley: University of California Press.

Prasada, S. 2000: Acquiring Generic Information. *Trends in Cognitive Sciences*, 4, 2, Feb.

Ramus, F., Hauser, M., Miller, C., Morris, D., and Mehler, J. 2000: Language Discrimination by Human Newborns and by Cotton-Top Tamarin Monkeys. *Science*, 288, 349–51.

Russell, B. 1927: *The Analysis of Matter*. London: Kegan Paul, Trench, Trubner.

Schofield, R. 1970: *Mechanism and Materialism*. Princeton: Princeton University Press.

Segal, G. 2000: *A Slim Book about Narrow Content*. Cambridge, Mass.: MIT Press.

Smith, G. 2001: Thomson and the Electron: 1897–1899. In J. Buchwald and A. Warwick (eds.), *Histories of the Electron: The Birth of Microphysics*, Cambridge, Mass.: MIT Press.

Spelke, E. 1985: Perception of Unity, Persistence and Identity. In J. Mehler and R. Fox (eds.), *Neonate Cognition: Beyond the Blooming Buzzing Confusion*, Hillsdale, NJ: Lawrence Erlbaum.

Stohl, M. 1988: States, Terrorism and State Terrorism. In R. O. Slater and M. Stohl, *Current Perspectives on International Terrorism*. New York: Macmillan.

Thackray, A. 1970: *Atoms and Powers*. Cambridge, Mass.: Harvard University Press.

Ullman, S. 1996: *Higher Level Vision*. Cambridge, Mass.: MIT Press.

Weng, J., et al. 2001: Autonomous Mental Development by Robots and Animals. *Science*, 291, 599–600.

Wheeler, J. 1994: *At Home in the Universe*. Vol. 9 of *Masters of Modern Physics*. New York: American Institute of Physics.

Zeki, S. 1998: Art and the Brain. *Daedalus*, 127 (2), Spring.

Zwart, C. J. W. 2001: Syntactic and Phonological Verb Movement. *Syntax*, 4 (1), 34–62.

Major Works by Noam Chomsky

1957: *Syntactic Structures*. The Hague: Mouton.
1964: *Current Issues in Linguistic Theory*. The Hague: Mouton.
1965: *Aspects of the Theory of Syntax*. Cambridge, Mass.: MIT Press.
1966: *Cartesian Linguistics: A Chapter in the History of Rationalist Thought*. New York: Harper & Row.
1966: *Topics in the Theory of Generative Grammar*. The Hague: Mouton.
1966: *Perspectives on Vietnam* [Microform] Speech by Noam Chomsky as part of a program presented by the Faculty Peace Committee, November 10, 1966, at the University of California, Berkeley. Berkeley: Academic Publishing.
1968: *Language and Mind*. New York: Harcourt, Brace & World.
1969: *American Power and the New Mandarins*. Harmondsworth: Penguin.
1969: *I nuovi mandarini; gli intellettuali e il potere in America*. Torino: G. Einaudi.
1969: *L'Amérique et ses Nouveaux Mandarins*. Paris: Editions du Seuil.
1970: Notes on Anarchism. *New York Review of Books*, 14 (10), May 21, pp. 31–5.
1970: *At War with Asia*. New York: Pantheon Books.
1970: *Two Essays on Cambodia*. Nottingham: Bertrand Russell Peace Foundation for *The Spokesman*.
1971: *Chomsky: Selected Readings*. Ed. J. P. B. Allen and P. Van Buren. London and New York: Oxford University Press.
1971: *Problems of Knowledge and Freedom*. New York: Pantheon Books.
1972: (with R. Jakobson and M. Halle). *Hypothèses, trois entretiens et trois études sur la linguistique et la poétique. [Traduction,] présentations et contributions de Jean-Pierre Faye, Jean Paris, Jacques Roubaud, Mitsou Ronat*. Paris: Seghers, Laffont.
1972: *Language and Mind*. New York: Harcourt Brace Jovanovich.
1972: *Studies on Semantics in Generative Grammar*. The Hague: Mouton.
1972: *Syntactic Structures*. The Hague: Mouton.
1973: *Conoscenza e libertà*. Torino: Einaudi.
1973: *For Reasons of State*. New York: Pantheon Books.
1974: *Peace in the Middle East? Reflections on Justice and Nationhood*. New York: Vintage Books.
1974: *Proceso contra Skinner*. Barcelona: Editorial Anagrama.

1975: *The Logical Structure of Linguistic Theory*. New York: Plenum Press.
1975: *Reflections on Language*. New York: Pantheon Books.
1977: *Essays on Form and Interpretation*. New York: North-Holland.
1978: *Human Rights and American Foreign Policy*. Nottingham: Spokesman Books.
1978: *Intellectuals and the State*. Baarn: Wereldvesten.
1978: (with J.-P. Vigier). *Verso la terza guerra mondiale?* Milano: Gabriele Mazzotta. Translation from French by Sergio Mancini. Reprint. Originally published: Paris: François Maspero, 1976.
1979: *Language and Responsibility: Based on Conversations with Mitsou Ronat*. Translated from the French by John Viertel. New York: Pantheon Books.
1979: *Morphophonemics of Modern Hebrew*. New York: Garland Publishing Co.
1979: (with E. S. Herman). *After the Cataclysm: Postwar Indochina and the Reconstruction of Imperial Ideology*. Boston, Mass.: South End Press.
1979: (with E. S. Herman). *The Political Economy of Human Rights*. Montreal: Black Rose Books.
1979: *Reflexiones sobre el lenguaje*. Translation by J. A. Argente and J. M. Nadel. Barcelona: Editorial Ariel.
1979: *The Washington Connection and Third World Fascism*. Boston, Mass.: South End Press.
1981: *Lectures on Government and Binding*. Dordrecht: Foris.
1981: *Radical Priorities*. Montreal: Black Rose Books.
1982: *Some Concepts and Consequences of the Theory of Government and Binding*. Cambridge, Mass.: MIT Press.
1982: *Superpowers in Collision: The Cold War Now*. Harmondsworth and New York: Penguin.
1982: *Towards a New Cold War: Essays on the Current Crisis and How We Got There*. New York: Pantheon Books.
1983: *The Fateful Triangle: The United States, Israel, and the Palestinians*. Boston, Mass.: South End Press.
1984: *Modular Approaches to the Study of the Mind*. San Diego: San Diego State University Press.
1985: *Turning the Tide: U.S. Intervention in Central America and the Struggle for Peace*. Boston, Mass.: South End Press.
1986: *Barriers*. Cambridge, Mass.: MIT Press.
1986: *Knowledge of Language: Its Nature, Origins, and Use*. New York: Praeger.
1986: *Pirates and Emperors: International Terrorism in the Real World*. New York: Claremont Research & Publications.
1986: *The Race to Destruction: Its Rational Basis*. Nottingham: published by *The Spokesman* for the Bertrand Russell Peace Foundation.
1987: *The Chomsky Reader*. Ed. James Peck. New York: Pantheon Books.
1987: *La quinta libertad: la política internacional y de seguridad de Estados Unidos*. San Salvador, El Salvador: UCA Editores.
1987: *On Power and Ideology*. Boston, Mass.: South End Press.
1987: *Turning the Tide: The U.S. and Latin America*. Montreal: Black Rose Books; Cheektowaga, NY: University of Toronto Press.
1988: *The Culture of Terrorism*. Boston, Mass.: South End Press.
1988: *Language and Politics*. Montreal and New York: Black Rose Books.
1988: *Language and Problems of Knowledge*. Cambridge, Mass.: MIT Press.
1988: (with E. S. Herman). *Manufacturing Consent: The Political Economy of the Mass Media*. New York: Pantheon Books.
1988: *Nuestra pequena region de por aqui: politica de seguridad de los Estados Unidos*. Managua: Editorial Nueva Nicaragua.
1989: *Necessary Illusions*. Boston, Mass.: South End Press.

1991: *Pirates and Emperors: International Terrorism in the Real World*. Revised edition. Montreal and New York: Black Rose Books.
1991: *The Sound Pattern of English*. Cambridge, Mass.: MIT Press.
1991: *Terrorizing the Neighborhood: American Foreign Policy in the Post-Cold War Era*. Stirling: AK Press; San Francisco: Pressure Drop Press.
1992: *What Uncle Sam Really Wants*. Berkeley: Odonian Press.
1992: *Chronicles of Dissent: Interviews with David Barsamian*. Monroe: Common Courage Press; Stirling: AK Press.
1992: *Deterring Democracy*. New York: Hill and Wang.
1993: *Language and Thought*. Wakefield, RI: Moyer Bell.
1993: *Lectures on Government and Binding: The Pisa Lectures*. 7th edition. Berlin and New York: Mouton de Gruyter.
1993: *Letters from Lexington: Reflections on Propaganda*. Monroe: Common Courage Press.
1993: *The Prosperous Few and the Restless Many*. Berkeley: Odonian Press.
1993: *Rethinking Camelot: JFK, the Vietnam War, and U.S. Political Culture*. Boston, Mass.: South End Press.
1993: *World Order and Its Rules: Variations on Some Themes*. Belfast: West Belfast Economic Forum and Centre for Research and Documentation.
1993: *Year 501: The Conquest Continues*. Boston, Mass.: South End Press.
1994: *Critical Assessments*. London and New York: Routledge.
1994: *Keeping the Rabble in Line: Interviews with David Barsamian*. Monroe: Common Courage Press.
1994: *Manufacturing Consent: Noam Chomksy and the Media*. The companion book to the award-winning film by Peter Wintonick and Mark Achbar. Montreal and New York: Black Rose Books.
1994: *Secrets, Lies and Democracy*. Tucson: Odonian Press.
1994: *World Orders, Old and New*. New York: Columbia University Press.
1995: *The Minimalist Program*. Cambridge, Mass.: MIT Press.
1995: *Secrets, Lies and Democracy*. Interviewed: David Barsamian. Tucson: Odonian Press.
1995: *Chiapas Insurgente*. Navarra: Txalaparta.
1996: *Powers and Prospects: Reflections on Human Nature and the Social Order*. Boston, Mass.: South End Press.
1996: *Class Warfare: Interviews with David Barsamian*. Monroe: Common Courage Press.
1997: *Media Control: The Spectacular Achievements of Propaganda*. New York: Seven Stories Press.
1997: *The Cold War and the University*. New York: New Press.
1999: *Profit over People: Neoliberalism and Global Order*. New York: Seven Stories Press.
1999: *The New Military Humanism: Lessons from Kosovo*. Monroe: Common Courage Press.
1999: *The Umbrella of U.S. Power: The Universal Declaration of Human Rights and the Contradictions of U.S. Policy*. New York: Seven Stories Press.
2000: *Chomsky on Miseducation*. Ed. and introduced by Donaldo Macedo. Lanham, Md.: Rowman & Littlefield.
2000: *New Horizons in the Study of Language and Mind*. Cambridge and New York: Cambridge University Press.
2000: *Rogue States: The Rule of Force in World Affairs*. Cambridge, Mass.: South End Press.
2000: *The Architecture of Language*. Ed. Nirmalangshu Mukherji, Bibudhendra Narayan Patnaik, and Rama Kant Agnihotri. New Delhi and New York: Oxford University Press.
2001: *9-11*. New York: Seven Stories Press; London: Turnaround.
2001: (with David Barsamian). *Propaganda and the Public Mind: Conversations with Noam Chomsky*. Cambridge, Mass.: South End Press.

Chomsky Essays, Collections, Interviews, Introductions, etc.

Assmann, H. (ed.), with N. Chomsky, et al. 1978: *Carter y la logica del imperialismo*. Ciudad Universitaria Rodrigo Facio, Costa Rica: Editorial Universitaria Centroamericana.

Bates, G. (ed.), 1991: *Mobilizing Democracy: Changing the U.S. Role in the Middle East: Interviews and Essays Featuring the Thoughts of Noam Chomsky and Others*. Monroe: Common Courage Press.

Bertrand Russell Centenary Symposium on Spheres of Influence in the Age of Imperialism, Linz, Austria 1972: *Spheres of Influence in the Age of Imperialism; Papers Submitted to the Bertrand Russell Centenary Symposium, Linz, Austria, September 11th to 15th, 1972 by Noam Chomsky and Others*. Nottingham: Bertrand Russell Peace Foundation for *The Spokesman*.

Brooks, D. and Verdecchia, G. 1991: *The Noam Chomsky Lectures*. Toronto: Coach House Press.

Chomsky, N. et al. 1970: *Trials of the Resistance. Essays by Noam Chomsky [and Others]*. With an introduction by M. Kempton. New York: New York Review, distributed by Random House.

Dieterich, H. 1985: *Centroamerica en la prensa estadunidense: entrevistas y textos de Noam Chomsky*. Coyoacan, Mexico: Mex-Sur Editorial.

Dieterich, H. 1986: *Nicaragua, la construcción de la sociedad sin clases*. Interview with Noam Chomsky: Prologue by G. Selser. Mexico City: Unomásuno.

Grewendorf, G. 1995: *Sprache als Organ, Sprache als Lebensform, Interview mit Noam Chomsky, Über Linguistik und Politik*. Frankfurt am Main: Suhrkamp.

Kim-Renaud, Y.-K. 1986: *Studies in Korean Linguistics*. Followed by an interview with Noam Chomsky. Seoul: Hanshin Publishing Company.

Morley, M. and Petras, J. 1987: *The Reagan Administration and Nicaragua: How Washington Constructs its Case for Counterrevolution in Central America*. With an introduction by N. Chomsky and an afterword by M. Parenti. New York: Institute for Media Analysis.

Piattelli-Palmarini, M. (ed.), 1979: *Théories du langage, théories de l'apprentissage: le débat entre Jean Piaget et Noam Chomsky*. Translation from the English by Y. Noizet. Paris: Editions du Seuil.

Piattelli-Palmarini, M. (ed.), 1980: *Language and Learning: The Debate Between Jean Piaget and Noam Chomsky*. Cambridge, Mass.: Harvard University Press.

Ramos-Horta, J. 1987: *Funu: The Unfinished Saga of East Timor*. Preface by Noam Chomsky. Trenton: Red Sea Press.

Rieber, R. W. (ed.), in collaboration with G. Voyat. 1983: *Dialogues on the Psychology of Language and Thought: Conversations with Noam Chomsky, Charles Osgood, Jean Piaget, Ulric Neisser, and Marcel Kinsbourne*. New York: Plenum Press.

Index

Abbreviations used in the index: CRTT = Computational Representational Theory of Thought; LWI = language/world isomorphism; VSP model = very simple picture model.