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 consciousness 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 computational 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