

# Noam Chomsky on Where Artificial Intelligence Went Wrong

The Atlantic, November 2012

**Yarden Katz** - Yarden Katz is a graduate student in the Department of Brain and Cognitive sciences at MIT, where he studies the regulation of gene expression in the developing nervous system and in cancer.

If one were to rank a list of civilization's greatest and most elusive intellectual challenges, the problem of "decoding" ourselves – understanding the inner workings of our minds and our brains, and how the architecture of these elements is encoded in our genome – would surely be at the top. Yet the diverse fields that took on this challenge, from philosophy and psychology to computer science and neuroscience, have been fraught with disagreement about the right approach.

In 1956, the computer scientist John McCarthy coined the term "Artificial Intelligence" (AI) to describe the study of intelligence by implementing its essential features on a computer. Instantiating an intelligent system using man-made hardware, rather than our own "biological hardware" of cells and tissues, would show ultimate understanding, and have obvious practical applications in the creation of intelligent devices or even robots.

Some of McCarthy's colleagues in neighboring departments, however, were more interested in how intelligence is implemented in humans (and other animals) first. Noam Chomsky and others worked on what became cognitive science, a field aimed at uncovering the mental representations and rules that underlie our perceptual and cognitive abilities. Chomsky and his colleagues had to overthrow the then-dominant paradigm of behaviorism, championed by Harvard psychologist B.F. Skinner, where animal behavior was reduced to a simple set of associations between an action and its subsequent reward or punishment. The undoing of Skinner's grip on psychology is commonly marked by Chomsky's 1967 critical review of Skinner's book *Verbal Behavior*, a book in which Skinner attempted to explain linguistic ability using behaviorist principles.

Skinner's approach stressed the historical associations between a stimulus and the animal's response – an approach easily framed as a kind of empirical statistical analysis, predicting the future as a function of the past. Chomsky's

conception of language, on the other hand, stressed the complexity of internal representations, encoded in the genome, and their maturation in light of the right data into a sophisticated computational system, one that cannot be usefully broken down into a set of associations. Behaviorist principles of associations could not explain the richness of linguistic knowledge, our endlessly creative use of it, or how quickly children acquire it with only minimal and imperfect exposure to language presented by their environment. The "language faculty," as Chomsky referred to it, was part of the organism's genetic endowment, much like the visual system, the immune system and the circulatory system, and we ought to approach it just as we approach these other more down-to-earth biological systems.

David Marr, a neuroscientist colleague of Chomsky's at MIT, defined a general framework for studying complex biological systems (like the brain) in his influential book *Vision*, one that Chomsky's analysis of the language capacity more or less fits into. According to Marr, a complex biological system can be understood at three distinct levels. The first level ("computational level") describes the input and output to the system, which define the task the system is performing. In the case of the visual system, the input might be the image projected on our retina and the output might our brain's identification of the objects present in the image we had observed. The second level ("algorithmic level") describes the procedure by which an input is converted to an output, i.e. how the image on our retina can be processed to achieve the task described by the computational level. Finally, the third level ("implementation level") describes how our own biological hardware of cells implements the procedure described by the algorithmic level.

The approach taken by Chomsky and Marr toward understanding how our minds achieve what they do is as different as can be from behaviorism. The emphasis here is on the internal structure of the system that enables it to perform a task, rather than on external association between past behavior of the system and the environment. The goal is to dig into the "black box" that drives the system and describe its inner workings, much like how a computer scientist would explain how a cleverly designed piece of software works and how it can be executed on a desktop computer.

As written today, the history of cognitive science is a story of the unequivocal triumph of an essentially Chomskyian approach over Skinner's behaviorist paradigm – an achievement commonly referred to as the "cognitive revolution," though Chomsky himself rejects this term. While this may be a relatively accurate depiction in cognitive science and psychology, behaviorist

thinking is far from dead in related disciplines. Behaviorist experimental paradigms and associationist explanations for animal behavior are used routinely by neuroscientists who aim to study the neurobiology of behavior in laboratory animals such as rodents, where the systematic three-level framework advocated by Marr is not applied.

In May of last year, during the 150th anniversary of the Massachusetts Institute of Technology, a symposium on "Brains, Minds and Machines" took place, where leading computer scientists, psychologists and neuroscientists gathered to discuss the past and future of artificial intelligence and its connection to the neurosciences.

The gathering was meant to inspire multidisciplinary enthusiasm for the revival of the scientific question from which the field of artificial intelligence originated: how does intelligence work? How does our brain give rise to our cognitive abilities, and could this ever be implemented in a machine?

Noam Chomsky, speaking in the symposium, wasn't so enthused. Chomsky critiqued the field of AI for adopting an approach reminiscent of behaviorism, except in more modern, computationally sophisticated form. Chomsky argued that the field's heavy use of statistical techniques to pick regularities in masses of data is unlikely to yield the explanatory insight that science ought to offer. For Chomsky, the "new AI" – focused on using statistical learning techniques to better mine and predict data – is unlikely to yield general principles about the nature of intelligent beings or about cognition.

This critique sparked an elaborate reply to Chomsky from Google's director of research and noted AI researcher, Peter Norvig, who defended the use of statistical models and argued that AI's new methods and definition of progress is not far off from what happens in the other sciences.

Chomsky acknowledged that the statistical approach might have practical value, just as in the example of a useful search engine, and is enabled by the advent of fast computers capable of processing massive data. But as far as a science goes, Chomsky would argue it is inadequate, or more harshly, kind of shallow. We wouldn't have taught the computer much about what the phrase "physicist Sir Isaac Newton" really means, even if we can build a search engine that returns sensible hits to users who type the phrase in.

It turns out that related disagreements have been pressing biologists who try to understand more traditional biological systems of the sort Chomsky likened to the language faculty. Just as the computing revolution enabled the massive data analysis that fuels the "new AI", so has the sequencing revolution in modern biology given rise to the blooming fields of genomics and

systems biology. High-throughput sequencing, a technique by which millions of DNA molecules can be read quickly and cheaply, turned the sequencing of a genome from a decade-long expensive venture to an affordable, commonplace laboratory procedure. Rather than painstakingly studying genes in isolation, we can now observe the behavior of a system of genes acting in cells as a whole, in hundreds or thousands of different conditions.

The sequencing revolution has just begun and a staggering amount of data has already been obtained, bringing with it much promise and hype for new therapeutics and diagnoses for human disease. For example, when a conventional cancer drug fails to work for a group of patients, the answer might lie in the genome of the patients, which might have a special property that prevents the drug from acting. With enough data comparing the relevant features of genomes from these cancer patients and the right control groups, custom-made drugs might be discovered, leading to a kind of "personalized medicine." Implicit in this endeavor is the assumption that with enough sophisticated statistical tools and a large enough collection of data, signals of interest can be weeded it out from the noise in large and poorly understood biological systems.

The success of fields like personalized medicine and other offshoots of the sequencing revolution and the systems-biology approach hinge upon our ability to deal with what Chomsky called "masses of unanalyzed data" – placing biology in the center of a debate similar to the one taking place in psychology and artificial intelligence since the 1960s.

Systems biology did not rise without skepticism. The great geneticist and Nobel-prize winning biologist Sydney Brenner once defined the field as "low input, high throughput, no output science." Brenner, a contemporary of Chomsky who also participated in the same symposium on AI, was equally skeptical about new systems approaches to understanding the brain. When describing an up-and-coming systems approach to mapping brain circuits called Connectomics, which seeks to map the wiring of all neurons in the brain (i.e. diagramming which nerve cells are connected to others), Brenner called it a "form of insanity."

Brenner's catch-phrase bite at systems biology and related techniques in neuroscience is not far off from Chomsky's criticism of AI. An unlikely pair, systems biology and artificial intelligence both face the same fundamental task of reverse-engineering a highly complex system whose inner workings are largely a mystery. Yet, ever-improving technologies yield massive data related to the system, only a fraction of which might be relevant. Do we

rely on powerful computing and statistical approaches to tease apart signal from noise, or do we look for the more basic principles that underlie the system and explain its essence? The urge to gather more data is irresistible, though it's not always clear what theoretical framework these data might fit into. These debates raise an old and general question in the philosophy of science: What makes a satisfying scientific theory or explanation, and how ought success be defined for science?

I sat with Noam Chomsky on an April afternoon in a somewhat disheveled conference room, tucked in a hidden corner of Frank Gehry's dazzling Stata Center at MIT. I wanted to better understand Chomsky's critique of artificial intelligence and why it may be headed in the wrong direction. I also wanted to explore the implications of this critique for other branches of science, such as neuroscience and systems biology, which all face the challenge of reverse-engineering complex systems – and where researchers often find themselves in an ever-expanding sea of massive data. The motivation for the interview was in part that Chomsky is rarely asked about scientific topics nowadays. Journalists are too occupied with getting his views on U.S. foreign policy, the Middle East, the Obama administration and other standard topics. Another reason was that Chomsky belongs to a rare and special breed of intellectuals, one that is quickly becoming extinct. Ever since Isaiah Berlin's famous essay, it has become a favorite pastime of academics to place various thinkers and scientists on the "Hedgehog-Fox" continuum: the Hedgehog, a meticulous and specialized worker, driven by incremental progress in a clearly defined field versus the Fox, a flashier, ideas-driven thinker who jumps from question to question, ignoring field boundaries and applying his or her skills where they seem applicable. Chomsky is special because he makes this distinction seem like a tired old cliché. Chomsky's depth doesn't come at the expense of versatility or breadth, yet for the most part, he devoted his entire scientific career to the study of defined topics in linguistics and cognitive science. Chomsky's work has had tremendous influence on a variety of fields outside his own, including computer science and philosophy, and he has not shied away from discussing and critiquing the influence of these ideas, making him a particularly interesting person to interview. Videos of the interview can be found [here](#).

*I want to start with a very basic question. At the beginning of AI, people were extremely optimistic about the field's progress, but it hasn't turned out that way. Why has it been so difficult? If you ask neuroscientists why understanding the brain is so difficult, they give you very intellectually unsatisfying*

*answers, like that the brain has billions of cells, and we can't record from all of them, and so on.*

Chomsky: There's something to that. If you take a look at the progress of science, the sciences are kind of a continuum, but they're broken up into fields. The greatest progress is in the sciences that study the simplest systems. So take, say physics – greatest progress there. But one of the reasons is that the physicists have an advantage that no other branch of sciences has. If something gets too complicated, they hand it to someone else.

*Like the chemists?*

Chomsky: If a molecule is too big, you give it to the chemists. The chemists, for them, if the molecule is too big or the system gets too big, you give it to the biologists. And if it gets too big for them, they give it to the psychologists, and finally it ends up in the hands of the literary critic, and so on. So what the neuroscientists are saying is not completely false.

However, it could be – and it has been argued in my view rather plausibly, though neuroscientists don't like it – that neuroscience for the last couple hundred years has been on the wrong track. There's a fairly recent book by a very good cognitive neuroscientist, Randy Gallistel and King, arguing – in my view, plausibly – that neuroscience developed kind of enthralled to associationism and related views of the way humans and animals work. And as a result they've been looking for things that have the properties of associationist psychology.

*Like Hebbian plasticity? [Editor's note: A theory, attributed to Donald Hebb, that associations between an environmental stimulus and a response to the stimulus can be encoded by strengthening of synaptic connections between neurons.]*

Chomsky: Well, like strengthening synaptic connections. Gallistel has been arguing for years that if you want to study the brain properly you should begin, kind of like Marr, by asking what tasks is it performing. So he's mostly interested in insects. So if you want to study, say, the neurology of an ant, you ask what does the ant do? It turns out the ants do pretty complicated things, like path integration, for example. If you look at bees, bee navigation involves quite complicated computations, involving position of the sun, and so on and so forth. But in general what he argues is that if you take a look at animal cognition, human too, it's computational systems. Therefore, you want to look the units of computation. Think about a Turing machine, say, which is the simplest form of computation, you have to find units that have properties like "read", "write" and "address." That's the minimal computational unit, so

you got to look in the brain for those. You're never going to find them if you look for strengthening of synaptic connections or field properties, and so on. You've got to start by looking for what's there and what's working and you see that from Marr's highest level.

*Right, but most neuroscientists do not sit down and describe the inputs and outputs to the problem that they're studying. They're more driven by say, putting a mouse in a learning task and recording as many neurons possible, or asking if Gene X is required for the learning task, and so on. These are the kinds of statements that their experiments generate.*

Chomsky: That's right..

*Is that conceptually flawed?*

Chomsky: Well, you know, you may get useful information from it. But if what's actually going on is some kind of computation involving computational units, you're not going to find them that way. It's kind of, looking at the wrong lamp post, sort of. It's a debate... I don't think Gallistel's position is very widely accepted among neuroscientists, but it's not an implausible position, and it's basically in the spirit of Marr's analysis. So when you're studying vision, he argues, you first ask what kind of computational tasks is the visual system carrying out. And then you look for an algorithm that might carry out those computations and finally you search for mechanisms of the kind that would make the algorithm work. Otherwise, you may never find anything. There are many examples of this, even in the hard sciences, but certainly in the soft sciences. People tend to study what you know how to study, I mean that makes sense. You have certain experimental techniques, you have certain level of understanding, you try to push the envelope – which is okay, I mean, it's not a criticism, but people do what you can do. On the other hand, it's worth thinking whether you're aiming in the right direction. And it could be that if you take roughly the Marr-Gallistel point of view, which personally I'm sympathetic to, you would work differently, look for different kind of experiments.

*Right, so I think a key idea in Marr is, like you said, finding the right units to describing the problem, sort of the right "level of abstraction" if you will. So if we take a concrete example of a new field in neuroscience, called Connectomics, where the goal is to find the wiring diagram of very complex organisms, find the connectivity of all the neurons in say human cerebral cortex, or mouse cortex. This approach was criticized by Sidney Brenner, who in many ways is [historically] one of the originators of the approach. Advocates of this field don't stop to ask if the wiring diagram is the right level*

*of abstraction – maybe it's not, so what is your view on that?*

Chomsky: Well, there are much simpler questions. Like here at MIT, there's been an interdisciplinary program on the nematode *C. elegans* for decades, and as far as I understand, even with this miniscule animal, where you know the wiring diagram, I think there's 800 neurons or something ...

*I think 300..*

Chomsky: ...Still, you can't predict what the thing [*C. elegans* nematode] is going to do. Maybe because you're looking in the wrong place.

*I'd like to shift the topic to different methodologies that were used in AI. So "Good Old Fashioned AI," as it's labeled now, made strong use of formalisms in the tradition of Gottlob Frege and Bertrand Russell, mathematical logic for example, or derivatives of it, like nonmonotonic reasoning and so on. It's interesting from a history of science perspective that even very recently, these approaches have been almost wiped out from the mainstream and have been largely replaced – in the field that calls itself AI now – by probabilistic and statistical models. My question is, what do you think explains that shift and is it a step in the right direction?*

Chomsky: I heard Pat Winston give a talk about this years ago. One of the points he made was that AI and robotics got to the point where you could actually do things that were useful, so it turned to the practical applications and somewhat, maybe not abandoned, but put to the side, the more fundamental scientific questions, just caught up in the success of the technology and achieving specific goals.

*So it shifted to engineering...*

Chomsky: It became... well, which is understandable, but would of course direct people away from the original questions. I have to say, myself, that I was very skeptical about the original work. I thought it was first of all way too optimistic, it was assuming you could achieve things that required real understanding of systems that were barely understood, and you just can't get to that understanding by throwing a complicated machine at it. If you try to do that you are led to a conception of success, which is self-reinforcing, because you do get success in terms of this conception, but it's very different from what's done in the sciences. So for example, take an extreme case, suppose that somebody says he wants to eliminate the physics department and do it the right way. The "right" way is to take endless numbers of videotapes of what's happening outside the video, and feed them into the biggest and fastest computer, gigabytes of data, and do complex statistical analysis – you know, Bayesian this and that [Editor's note: A



modern approach to analysis of data which makes heavy use of probability theory.] – and you’ll get some kind of prediction about what’s gonna happen outside the window next. In fact, you get a much better prediction than the physics department will ever give. Well, if success is defined as getting a fair approximation to a mass of chaotic unanalyzed data, then it’s way better to do it this way than to do it the way the physicists do, you know, no thought experiments about frictionless planes and so on and so forth. But you won’t get the kind of understanding that the sciences have always been aimed at – what you’ll get at is an approximation to what’s happening.

And that’s done all over the place. Suppose you want to predict tomorrow’s weather. One way to do it is okay I’ll get my statistical priors, if you like, there’s a high probability that tomorrow’s weather here will be the same as it was yesterday in Cleveland, so I’ll stick that in, and where the sun is will have some effect, so I’ll stick that in, and you get a bunch of assumptions like that, you run the experiment, you look at it over and over again, you correct it by Bayesian methods, you get better priors. You get a pretty good approximation of what tomorrow’s weather is going to be. That’s not what meteorologists do – they want to understand how it’s working. And these are just two different concepts of what success means, of what achievement is. In my own field, language fields, it’s all over the place. Like computational cognitive science applied to language, the concept of success that’s used is virtually always this. So if you get more and more data, and better and better statistics, you can get a better and better approximation to some immense corpus of text, like everything in The Wall Street Journal archives – but you learn nothing about the language.

A very different approach, which I think is the right approach, is to try to see if you can understand what the fundamental principles are that deal with the core properties, and recognize that in the actual usage, there’s going to be a thousand other variables intervening – kind of like what’s happening outside the window, and you’ll sort of tack those on later on if you want better approximations, that’s a different approach. These are just two different concepts of science. The second one is what science has been since Galileo, that’s modern science. The approximating unanalyzed data kind is sort of a new approach, not totally, there’s things like it in the past. It’s basically a new approach that has been accelerated by the existence of massive memories, very rapid processing, which enables you to do things like this that you couldn’t have done by hand. But I think, myself, that it is leading subjects like computational cognitive science into a direction of

maybe some practical applicability...

*..in engineering?*

Chomsky: ...But away from understanding. Yeah, maybe some effective engineering. And it's kind of interesting to see what happened to engineering. So like when I got to MIT, it was 1950s, this was an engineering school. There was a very good math department, physics department, but they were service departments. They were teaching the engineers tricks they could use. The electrical engineering department, you learned how to build a circuit. Well if you went to MIT in the 1960s, or now, it's completely different. No matter what engineering field you're in, you learn the same basic science and mathematics. And then maybe you learn a little bit about how to apply it. But that's a very different approach. And it resulted maybe from the fact that really for the first time in history, the basic sciences, like physics, had something really to tell engineers. And besides, technologies began to change very fast, so not very much point in learning the technologies of today if it's going to be different 10 years from now. So you have to learn the fundamental science that's going to be applicable to whatever comes along next. And the same thing pretty much happened in medicine. So in the past century, again for the first time, biology had something serious to tell to the practice of medicine, so you had to understand biology if you want to be a doctor, and technologies again will change. Well, I think that's the kind of transition from something like an art, that you learn how to practice – an analog would be trying to match some data that you don't understand, in some fashion, maybe building something that will work – to science, what happened in the modern period, roughly Galilean science.

*I see. Returning to the point about Bayesian statistics in models of language and cognition. You've argued famously that speaking of the probability of a sentence is unintelligible on its own...*

Chomsky: ..Well you can get a number if you want, but it doesn't mean anything.

*It doesn't mean anything. But it seems like there's almost a trivial way to unify the probabilistic method with acknowledging that there are very rich internal mental representations, comprised of rules and other symbolic structures, and the goal of probability theory is just to link noisy sparse data in the world with these internal symbolic structures. And that doesn't commit you to saying anything about how these structures were acquired – they could have been there all along, or there partially with some parameters being tuned, whatever your conception is. But probability theory just serves as a kind of*

*glue between noisy data and very rich mental representations.*

Chomsky: Well... there's nothing wrong with probability theory, there's nothing wrong with statistics.

*But does it have a role?*

Chomsky: If you can use it, fine. But the question is what are you using it for? First of all, first question is, is there any point in understanding noisy data? Is there some point to understanding what's going on outside the window?

*Well, we are bombarded with it [noisy data], it's one of Marr's examples, we are faced with noisy data all the time, from our retina to...*

Chomsky: That's true. But what he says is: Let's ask ourselves how the biological system is picking out of that noise things that are significant. The retina is not trying to duplicate the noise that comes in. It's saying I'm going to look for this, that and the other thing. And it's the same with say, language acquisition. The newborn infant is confronted with massive noise, what William James called "a blooming, buzzing confusion," just a mess. If say, an ape or a kitten or a bird or whatever is presented with that noise, that's where it ends. However, the human infants, somehow, instantaneously and reflexively, picks out of the noise some scattered subpart which is language-related. That's the first step. Well, how is it doing that? It's not doing it by statistical analysis, because the ape can do roughly the same probabilistic analysis. It's looking for particular things. So psycholinguists, neurolinguists, and others are trying to discover the particular parts of the computational system and of the neurophysiology that are somehow tuned to particular aspects of the environment. Well, it turns out that there actually are neural circuits which are reacting to particular kinds of rhythm, which happen to show up in language, like syllable length and so on. And there's some evidence that that's one of the first things that the infant brain is seeking – rhythmic structures. And going back to Gallistel and Marr, its got some computational system inside which is saying "okay, here's what I do with these things" and say, by nine months, the typical infant has rejected – eliminated from its repertoire – the phonetic distinctions that aren't used in its own language. So initially of course, any infant is tuned to any language. But say, a Japanese kid at nine months won't react to the R-L distinction anymore, that's kind of weeded out. So the system seems to sort out lots of possibilities and restrict it to just ones that are part of the language, and there's a narrow set of those. You can make up a non-language in which the infant could never do it, and then you're looking for other things. For

example, to get into a more abstract kind of language, there's substantial evidence by now that such a simple thing as linear order, what precedes what, doesn't enter into the syntactic and semantic computational systems, they're just not designed to look for linear order. So you find overwhelmingly that more abstract notions of distance are computed and not linear distance, and you can find some neurophysiological evidence for this, too. Like if artificial languages are invented and taught to people, which use linear order, like you negate a sentence by doing something to the third word. People can solve the puzzle, but apparently the standard language areas of the brain are not activated – other areas are activated, so they're treating it as a puzzle not as a language problem. You need more work, but...

*You take that as convincing evidence that activation or lack of activation for the brain area ...*

Chomsky: ...It's evidence, you'd want more of course. But this is the kind of evidence, both on the linguistics side you look at how languages work – they don't use things like third word in sentence. Take a simple sentence like "Instinctively, Eagles that fly swim", well, "instinctively" goes with swim, it doesn't go with fly, even though it doesn't make sense. And that's reflexive. "Instinctively", the adverb, isn't looking for the nearest verb, it's looking for the structurally most prominent one. That's a much harder computation. But that's the only computation which is ever used. Linear order is a very easy computation, but it's never used. There's a ton of evidence like this, and a little neurolinguistic evidence, but they point in the same direction. And as you go to more complex structures, that's where you find more and more of that.

That's, in my view at least, the way to try to discover how the system is actually working, just like in vision, in Marr's lab, people like Shimon Ullman discovered some pretty remarkable things like the rigidity principle. You're not going to find that by statistical analysis of data. But he did find it by carefully designed experiments. Then you look for the neurophysiology, and see if you can find something there that carries out these computations. I think it's the same in language, the same in studying our arithmetical capacity, planning, almost anything you look at. Just trying to deal with the unanalyzed chaotic data is unlikely to get you anywhere, just like as it wouldn't have gotten Galileo anywhere. In fact, if you go back to this, in the 17th century, it wasn't easy for people like Galileo and other major scientists to convince the NSF [National Science Foundation] of the day – namely, the aristocrats – that any of this made any sense. I mean, why study balls

rolling down frictionless planes, which don't exist. Why not study the growth of flowers? Well, if you tried to study the growth of flowers at that time, you would get maybe a statistical analysis of what things looked like.

It's worth remembering that with regard to cognitive science, we're kind of pre-Galilean, just beginning to open up the subject. And I think you can learn something from the way science worked [back then]. In fact, one of the founding experiments in history of chemistry, was about 1640 or so, when somebody proved to the satisfaction of the scientific world, all the way up to Newton, that water can be turned into living matter. The way they did it was – of course, nobody knew anything about photosynthesis – so what you do is you take a pile of earth, you heat it so all the water escapes. You weigh it, and put it in a branch of a willow tree, and pour water on it, and measure you the amount of water you put in. When you're done, you the willow tree is grown, you again take the earth and heat it so all the water is gone – same as before. Therefore, you've shown that water can turn into an oak tree or something. It is an experiment, it's sort of right, but it's just that you don't know what things you ought to be looking for. And they weren't known until Priestly found that air is a component of the world, it's got nitrogen, and so on, and you learn about photosynthesis and so on. Then you can redo the experiment and find out what's going on. But you can easily be misled by experiments that seem to work because you don't know enough about what to look for. And you can be misled even more if you try to study the growth of trees by just taking a lot of data about how trees growing, feeding it into a massive computer, doing some statistics and getting an approximation of what happened.

*In the domain of biology, would you consider the work of Mendel, as a successful case, where you take this noisy data – essentially counts – and you leap to postulate this theoretical object...*

Chomsky: ...Well, throwing out a lot of the data that didn't work.

*...But seeing the ratio that made sense, given the theory.*

Chomsky: Yeah, he did the right thing. He let the theory guide the data. There was counter data which was more or less dismissed, you know you don't put it in your papers. And he was of course talking about things that nobody could find, like you couldn't find the units that he was postulating. But that's, sure, that's the way science works. Same with chemistry. Chemistry, until my childhood, not that long ago, was regarded as a calculating device. Because you couldn't reduce to physics. So it's just some way of calculating the result of experiments. The Bohr atom was treated that way. It's the way

of calculating the results of experiments but it can't be real science, because you can't reduce it to physics, which incidentally turned out to be true, you couldn't reduce it to physics because physics was wrong. When quantum physics came along, you could unify it with virtually unchanged chemistry. So the project of reduction was just the wrong project. The right project was to see how these two ways of looking at the world could be unified. And it turned out to be a surprise – they were unified by radically changing the underlying science. That could very well be the case with say, psychology and neuroscience. I mean, neuroscience is nowhere near as advanced as physics was a century ago.

*That would go against the reductionist approach of looking for molecules that are correlates of...*

Chomsky: Yeah. In fact, the reductionist approach has often been shown to be wrong. The unification approach makes sense. But unification might not turn out to be reduction, because the core science might be misconceived as in the physics-chemistry case and I suspect very likely in the neuroscience-psychology case. If Gallistel is right, that would be a case in point that yeah, they can be unified, but with a different approach to the neurosciences.

*So is that a worthy goal of unification or the fields should proceed in parallel?*

Chomsky: Well, unification is kind of an intuitive ideal, part of the scientific mystique, if you like. It's that you're trying to find a unified theory of the world. Now maybe there isn't one, maybe different parts work in different ways, but your assumption is until I'm proven wrong definitively, I'll assume that there's a unified account of the world, and it's my task to try to find it. And the unification may not come out by reduction – it often doesn't. And that's kind of the guiding logic of David Marr's approach: what you discover at the computational level ought to be unified with what you'll some day find out at the mechanism level, but maybe not in terms of the way we now understand the mechanisms.

*And implicit in Marr it seems that you can't work on all three in parallel [computational, algorithmic, implementation levels], it has to proceed top-down, which is a very stringent requirement, given that science usually doesn't work that way.*

Chomsky: Well, he wouldn't have said it has to be rigid. Like for example, discovering more about the mechanisms might lead you to change your concept of computation. But there's kind of a logical precedence, which isn't necessarily the research precedence, since in research everything goes on at

the same time. But I think that the rough picture is okay. Though I should mention that Marr's conception was designed for input systems...

*information-processing systems...*

Chomsky: Yeah, like vision. There's some data out there – it's a processing system – and something goes on inside. It isn't very well designed for cognitive systems. Like take your capacity to take out arithmetical operations..

*It's very poor, but yeah...*

Chomsky: Okay [laughs]. But it's an internal capacity, you know your brain is a controlling unit of some kind of Turing machine, and it has access to some external data, like memory, time and so on. And in principle, you could multiply anything, but of course not in practice. If you try to find out what that internal system is of yours, the Marr hierarchy doesn't really work very well. You can talk about the computational level – maybe the rules I have are Peano's axioms [Editor's note: a mathematical theory (named after Italian mathematician Giuseppe Peano) that describes a core set of basic rules of arithmetic and natural numbers, from which many useful facts about arithmetic can be deduced], or something, whatever they are – that's the computational level. In theory, though we don't know how, you can talk about the neurophysiological level, nobody knows how, but there's no real algorithmic level. Because there's no calculation of knowledge, it's just a system of knowledge. To find out the nature of the system of knowledge, there is no algorithm, because there is no process. Using the system of knowledge, that'll have a process, but that's something different.

*But since we make mistakes, isn't that evidence of a process gone wrong?*

Chomsky: That's the process of using the internal system. But the internal system itself is not a process, because it doesn't have an algorithm. Take, say, ordinary mathematics. If you take Peano's axioms and rules of inference, they determine all arithmetical computations, but there's no algorithm. If you ask how does a number theoretician applies these, well all kinds of ways. Maybe you don't start with the axioms and start with the rules of inference. You take the theorem, and see if I can establish a lemma, and if it works, then see if I can try to ground this lemma in something, and finally you get a proof which is a geometrical object.

*But that's a fundamentally different activity from me adding up small numbers in my head, which surely does have some kind of algorithm.*

Chomsky: Not necessarily. There's an algorithm for the process in both cases. But there's no algorithm for the system itself, it's kind of a category

mistake. You don't ask the question what's the process defined by Peano's axioms and the rules of inference, there's no process. There can be a process of using them. And it could be a complicated process, and the same is true of your calculating. The internal system that you have – for that, the question of process doesn't arise. But for your using that internal system, it arises, and you may carry out multiplications all kinds of ways. Like maybe when you add 7 and 6, let's say, one algorithm is to say "I'll see how much it takes to get to 10" – it takes 3, and now I've got 4 left, so I gotta go from 10 and add 4, I get 14. That's an algorithm for adding – it's actually one I was taught in kindergarten. That's one way to add.

But there are other ways to add – there's no kind of right algorithm. These are algorithms for carrying out the process the cognitive system that's in your head. And for that system, you don't ask about algorithms. You can ask about the computational level, you can ask about the mechanism level. But the algorithm level doesn't exist for that system. It's the same with language. Language is kind of like the arithmetical capacity. There's some system in there that determines the sound and meaning of an infinite array of possible sentences. But there's no question about what the algorithm is. Like there's no question about what a formal system of arithmetic tells you about proving theorems. The use of the system is a process and you can study it in terms of Marr's level. But it's important to be conceptually clear about these distinctions.

*It just seems like an astounding task to go from a computational level theory, like Peano axioms, to Marr level 3 of the...*

Chomsky: mechanisms...

*...mechanisms and implementations...*

Chomsky: Oh yeah. Well..

*..without an algorithm at least.*

Chomsky: Well, I don't think that's true. Maybe information about how it's used, that'll tell you something about the mechanisms. But some higher intelligence – maybe higher than ours – would see that there's an internal system, its got a physiological basis, and I can study the physiological basis of that internal system. Not even looking at the process by which it's used. Maybe looking at the process by which it's used maybe gives you helpful information about how to proceed. But it's conceptually a different problem. That's the question of what's the best way to study something. So maybe the best way to study the relation between Peano's axioms and neurons is by watching mathematicians prove theorems. But that's just because it'll



give you information that may be helpful. The actual end result of that will be an account of the system in the brain, the physiological basis for it, with no reference to any algorithm. The algorithms are about a process of using it, which may help you get answers. Maybe like incline planes tell you something about the rate of fall, but if you take a look at Newton's laws, they don't say anything about incline planes.

*Right. So the logic for studying cognitive and language systems using this kind of Marr approach makes sense, but since you've argued that language capacity is part of the genetic endowment, you could apply it to other biological systems, like the immune system, the circulatory system....*

Chomsky: Certainly, I think it's very similar. You can say the same thing about study of the immune system.

*It might even be simpler, in fact, to do it for those systems than for cognition.*

Chomsky: Though you'd expect different answers. You can do it for the digestive system. Suppose somebody's studying the digestive system. Well, they're not going to study what happens when you have a stomach flu, or when you've just eaten a big Mac, or something. Let's go back to taking pictures outside the window. One way of studying the digestive system is just to take all data you can find about what digestive systems do under any circumstances, toss the data into a computer, do statistical analysis – you get something. But it's not gonna be what any biologist would do. They want to abstract away, at the very beginning, from what are presumed – maybe wrongly, you can always be wrong – irrelevant variables, like do you have stomach flu.

*But that's precisely what the biologists are doing, they are taking the sick people with the sick digestive system, comparing them to the normals, and measuring these molecular properties.*

Chomsky: They're doing it in an advanced stage. They already understand a lot about the study of the digestive system before we compare them, otherwise you wouldn't know what to compare, and why is one sick and one isn't.

*Well, they're relying on statistical analysis to pick out the features that discriminate. It's a highly fundable approach, because you're claiming to study sick people.*

Chomsky: It may be the way to fund things. Like maybe the way to fund study of language is to say, maybe help cure autism. That's a different question [laughs]. But the logic of the search is to begin by studying the

system abstracted from what you, plausibly, take to be irrelevant intrusions, see if you can find its basic nature – then ask, well, what happens when I bring in some of this other stuff, like stomach flu.

*It still seems like there's a difficulty in applying Marr's levels to these kinds of systems. If you ask, what is the computational problem that the brain is solving, we have kind of an answer, it's sort of like a computer. But if you ask, what is the computational problem that's being solved by the lung, that's very difficult to even think – it's not obviously an information-processing kind of problem.*

Chomsky: No, but there's no reason to assume that all of biology is computational. There may be reasons to assume that cognition is. And in fact Gallistel is not saying that everything is in the body ought to be studied by finding read/write/address units.

*It just seems contrary to any evolutionary intuition. These systems evolved together, reusing many of the same parts, same molecules, pathways. Cells are computing things.*

Chomsky: You don't study the lung by asking what cells compute. You study the immune system and the visual system, but you're not going to expect to find the same answers. An organism is a highly modular system, has a lot of complex subsystems, which are more or less internally integrated. They operate by different principles. The biology is highly modular. You don't assume it's all just one big mess, all acting the same way.

*No, sure, but I'm saying you would apply the same approach to study each of the modules.*

Chomsky: Not necessarily, not if the modules are different. Some of the modules may be computational, others may not be.

*So what would you think would be an adequate theory that is explanatory, rather than just predicting data, the statistical way, what would be an adequate theory of these systems that are not computing systems – can we even understand them?*

Chomsky: Sure. You can understand a lot about say, what makes an embryo turn into a chicken rather than a mouse, let's say. It's a very intricate system, involves all kinds of chemical interactions, all sorts of other things. Even the nematode, it's by no means obviously – in fact there are reports from the study here – that it's all just a matter of a neural net. You have to look into complex chemical interactions that take place in the brain, in the nervous system. You have to look into each system on its own. These chemical interactions might not be related to how your arithmetical capacity

works – probably aren't. But they might very well be related to whether you decide to raise your arm or lower it.

*Though if you study the chemical interactions it might lead you into what you've called just a redescription of the phenomena.*

Chomsky: Or an explanation. Because maybe that's directly, crucially, involved.

*But if you explain it in terms of chemical X has to be turned on, or gene X has to be turned on, you've not really explained how organism-determination is done. You've simply found a switch, and hit that switch.*

Chomsky: But then you look further, and find out what makes this gene do such and such under these circumstances, and do something else under different circumstances.

*But if genes are the wrong level of abstraction, you'd be screwed.*

Chomsky: Then you won't get the right answer. And maybe they're not. For example, it's notoriously difficult to account for how an organism arises from a genome. There's all kinds of production going on in the cell. If you just look at gene action, you may not be in the right level of abstraction. You never know, that's what you try to study. I don't think there's any algorithm for answering those questions, you try.

*So I want to shift gears more toward evolution. You've criticized a very interesting position you've called "phylogenetic empiricism." You've criticized this position for not having explanatory power. It simply states that: well, the mind is the way it because of adaptations to the environment that were selected for. And these were selected for by natural selection. You've argued that this doesn't explain anything because you can always appeal to these two principles of mutation and selection.*

Chomsky: Well you can wave your hands at them, but they might be right. It could be that, say, the development of your arithmetical capacity, arose from random mutation and selection. If it turned out to be true, fine.

*It seems like a truism.*

Chomsky: Well, I mean, doesn't mean it's false. Truisms are true. [laughs].

*But they don't explain much.*

Chomsky: Maybe that's the highest level of explanation you can get. You can invent a world – I don't think it's our world – but you can invent a world in which nothing happens except random changes in objects and selection on the basis of external forces. I don't think that's the way our world works, I don't think it's the way any biologist thinks it is. There are all kind of ways

in which natural law imposes channels within which selection can take place, and some things can happen and other things don't happen. Plenty of things that go on in the biology in organisms aren't like this. So take the first step, meiosis. Why do cells split into spheres and not cubes? It's not random mutation and natural selection; it's a law of physics. There's no reason to think that laws of physics stop there, they work all the way through.

*Well, they constrain the biology, sure.*

Chomsky: Okay, well then it's not just random mutation and selection. It's random mutation, selection, and everything that matters, like laws of physics.

*So is there room for these approaches which are now labeled "comparative genomics", like the Broad Institute here [at MIT/Harvard] is generating massive amounts of data, of different genomes, different animals, different cells under different conditions and sequencing any molecule that is sequenceable. Is there anything that can be gleaned about these high-level cognitive tasks from these comparative evolutionary studies or is it premature?*

Chomsky: I am not saying it's the wrong approach, but I don't know anything that can be drawn from it. Nor would you expect to.

*You don't have any examples where this evolutionary analysis has informed something? Like Foxp2 mutations? [Editor's note: A gene that is thought to be implicated in speech or language capacities. A family with a stereotyped speech disorder was found to have genetic mutations that disrupt this gene. This gene evolved to have several mutations unique to the human evolutionary lineage.]*

Chomsky: Foxp2 is kind of interesting, but it doesn't have anything to do with language. It has to do with fine motor coordinations and things like that. Which takes place in the use of language, like when you speak you control your lips and so on, but all that's very peripheral to language, and we know that. So for example, whether you use the articulatory organs or sign, you know hand motions, it's the same language. In fact, it's even being analyzed and produced in the same parts of the brain, even though one of them is moving your hands and the other is moving your lips. So whatever the externalization is, it seems quite peripheral. I think they're too complicated to talk about, but I think if you look closely at the design features of language, you get evidence for that. There are interesting cases in the study of language where you find conflicts between computational efficiency and communicative efficiency.

Take this case I even mentioned of linear order. If you want to know which

verb the adverb attaches to, the infant reflexively using minimal structural distance, not minimal linear distance. Well, it's using minimal linear distances, computationally easy, but it requires having linear order available. And if linear order is only a reflex of the sensory-motor system, which makes sense, it won't be available. That's evidence that the mapping of the internal system to the sensory-motor system is peripheral to the workings of the computational system.

*But it might constrain it like physics constrains meiosis?*

Chomsky: It might, but there's very little evidence of that. So for example the left end – left in the sense of early – of a sentence has different properties from the right end. If you want to ask a question, let's say "Who did you see?" You put the "Who" in front, not in the end. In fact, in every language in which a wh-phrase – like who, or which book, or something – moves to somewhere else, it moves to the left, not to the right. That's very likely a processing constraint. The sentence opens by telling you, the hearer, here's what kind of a sentence it is. If it's at the end, you have to have the whole declarative sentence, and at the end you get the information I'm asking about. If you spell it out, it could be a processing constraint. So that's a case, if true, in which the processing constraint, externalization, do affect the computational character of the syntax and semantics.

There are cases where you find clear conflicts between computational efficiency and communicative efficiency. Take a simple case, structural ambiguity. If I say, "Visiting relatives can be a nuisance" – that's ambiguous. Relatives that visit, or going to visit relatives. It turns out in every such case that's known, the ambiguity is derived by simply allowing the rules to function freely, with no constraints, and that sometimes yields ambiguities. So it's computationally efficient, but it's inefficient for communication, because it leads to unresolvable ambiguity.

Or take what are called garden-path sentences, sentences like "The horse raced past the barn fell". People presented with that don't understand it, because the way it's put, they're led down a garden path. "The horse raced past the barn" sounds like a sentence, and then you ask what's "fell" doing there at the end. On the other hand, if you think about it, it's a perfectly well formed sentence. It means the horse that was raced past the barn, by someone, fell. But the rules of the language when they just function happen to give you a sentence which is unintelligible because of the garden-path phenomena. And there are lots of cases like that. There are things you just can't say, for some reason. So if I say, "The mechanics fixed the cars". And you

say, "They wondered if the mechanics fixed the cars." You can ask questions about the cars, "How many cars did they wonder if the mechanics fixed?" More or less okay. Suppose you want to ask a question about the mechanics. "How many mechanics did they wonder if fixed the cars?" Somehow it doesn't work, can't say that. It's a fine thought, but you can't say it. Well, if you look into it in detail, the most efficient computational rules prevent you from saying it. But for expressing thought, for communication, it'd be better if you could say it – so that's a conflict.

And in fact, every case of a conflict that's known, computational efficiency wins. The externalization is yielding all kinds of ambiguities but for simple computational reasons, it seems that the system internally is just computing efficiently, it doesn't care about the externalization. Well, I haven't made that a very plausible, but if you spell it out it can be made quite a convincing argument I think.

That tells something about evolution. What it strongly suggests is that in the evolution of language, a computational system developed, and later on it was externalized. And if you think about how a language might have evolved, you're almost driven to that position. At some point in human evolution, and it's apparently pretty recent given the archeological record – maybe last hundred thousand years, which is nothing – at some point a computational system emerged with had new properties, that other organisms don't have, that has kind of arithmetical type properties...

*It enabled better thought before externalization?*

Chomsky: It gives you thought. Some rewiring of the brain, that happens in a single person, not in a group. So that person had the capacity for thought – the group didn't. So there isn't any point in externalization. Later on, if this genetic change proliferates, maybe a lot of people have it, okay then there's a point in figuring out a way to map it to the sensory-motor system and that's externalization but it's a secondary process.

*Unless the externalization and the internal thought system are coupled in ways we just don't predict.*

Chomsky: We don't predict, and they don't make a lot of sense. Why should it be connected to the external system? In fact, say your arithmetical capacity isn't. And there are other animals, like songbirds, which have internal computational systems, bird song. It's not the same system but it's some kind of internal computational system. And it is externalized, but sometimes it's not. A chick in some species acquires the song of that species but doesn't produce it until maturity. During that early period it has the song, but it

doesn't have the externalization system. Actually that's true of humans too, like a human infant understands a lot more than it can produce – plenty of experimental evidence for this, meaning it's got the internal system somehow, but it can't externalize it. Maybe it doesn't have enough memory, or whatever it may be.

*I'd like to close with one question about the philosophy of science. In a recent interview, you said that part of the problem is that scientists don't think enough about what they're up to. You mentioned that you taught a philosophy of science course at MIT and people would read, say, Willard van Orman Quine, and it would go in one ear out the other, and people would go back doing the same kind of science that they were doing. What are the insights that have been obtained in philosophy of science that are most relevant to scientists who are trying to let's say, explain biology, and give an explanatory theory rather than redescription of the phenomena? What do you expect from such a theory, and what are the insights that help guide science in that way? Rather than guiding it towards behaviorism which seems to be an intuition that many, say, neuroscientists have?*

Chomsky: Philosophy of science is a very interesting field, but I don't think it really contribute to science, it learns from science. It tries to understand what the sciences do, why do they achieve things, what are the wrong paths, see if we can codify that and come to understand. What I think is valuable is the history of science. I think we learn a lot of things from the history of science that can be very valuable to the emerging sciences. Particularly when we realize that in say, the emerging cognitive sciences, we really are in a kind of pre-Galilean stage. We don't know what we're looking for anymore than Galileo did, and there's a lot to learn from that. So for example one striking fact about early science, not just Galileo, but the Galilean breakthrough, was the recognition that simple things are puzzling.

Take say, if I'm holding this here [cup of water], and say the water is boiling [putting hand over water], the steam will rise, but if I take my hand away the cup will fall. Well why does the cup fall and the steam rise? Well for millennia there was a satisfactory answer to that: they're seeking their natural place.

*Like in Aristotelian physics?*

Chomsky: That's the Aristotelian physics. The best and greatest scientists thought that was answer. Galileo allowed himself to be puzzled by it. As soon as you allow yourself to be puzzled by it, you immediately find that all your intuitions are wrong. Like the fall of a big mass and a small mass,

and so on. All your intuitions are wrong – there are puzzles everywhere you look. That's something to learn from the history of science. Take the one example that I gave to you, "Instinctively eagles that fly swim." Nobody ever thought that was puzzling – yeah, why not. But if you think about it, it's very puzzling, you're using a complex computation instead of a simple one. Well, if you allow yourself to be puzzled by that, like the fall of a cup, you ask "Why?" and then you're led down a path to some pretty interesting answers. Like maybe linear order just isn't part of the computational system, which is a strong claim about the architecture of the mind – it says it's just part of the externalization system, secondary, you know. And that opens up all sorts of other paths, same with everything else.

Take another case: the difference between reduction and unification. History of science gives some very interesting illustrations of that, like chemistry and physics, and I think they're quite relevant to the state of the cognitive and neurosciences today.