D. Baird, A. Nordmann & J. Schummer (eds.), *Discovering the Nanoscale*, Amsterdam: IOS Press, 2004.

ISBN: 1-58603-467-7

Von Neumann, Self-Reproduction and the Constitution of Nanophenomena

Otávio BUENO
Department of Philosophy, University of South Carolina
obueno@sc.edu

Abstract. As part of a larger study of the immediate antecedents of nanoscience and nanotechnology, I examine, in this paper, the role played by John von Neumann's work on self-reproduction in the constitution of these fields (see especially von Neumann 1951, 1956 and 1966). Von Neumann's proposals have always been characterized by an overall unified vision, in which, depending on the domain under consideration, a given logic, specific mathematical theories, probability and the relevant scientific theories were integrated in a clear and well-motivated way. I discuss how this overall vision prompted von Neumann to develop his work on self-reproduction, and how this vision was then transferred to nanoscience and nanotechnology. In particular, I examine the influence of von Neumann's proposals in the development of Eric Drexler's work in molecular manipulation and computation (Drexler 1992). By understanding the influence that von Neumann's work had in nanotechnology and nanoscience, a different – and perhaps slightly more unified – picture of these fields emerges.

Introduction

Despite being relatively new, nanoscale research already involves delicate historical and conceptual issues. Why was 'the' field (to the extant that there is such a well-defined field!) constituted in the way it was? Which criteria have been used to stabilize nanophenomena in their current shape?

In this work, I start to address these issues by discussing some forerunners and immediate antecedents of nanoscale research. In particular, I examine how the interaction between what is physically and mathematically possible, but also impossible, in this domain has shaped the constitution of nanophenomena. Two forerunners, in particular, should be considered: Richard Feynman and John von Neumann. In "There's Plenty of Room at the Bottom", Feynman outlined a vision for the development of nanoscience. He advanced, for the first time, the idea that it should be possible to build objects atom by atom (Feynman 1960). Feynman was concerned with exploring what was physically possible to do at the nanoscale, and he outlined the benefits that should be expected from such a research. Not surprisingly, the nanoscience community took the work as a founding document.

In a series of works on the theory of automata, John von Neumann provided a different picture. He explored what was mathematically and logically possible, but also impossible, to do in the process of building reliable organisms from unreliable components (von Neumann 1951, 1956, and 1966). Although there has been a considerable amount of reflection on Feynman's contribution, especially in the nanoscience community, von Neumann's work has received significantly less attention. By focusing on von Neumann's contribution, a better understanding of the emergence of the theories of automata and self-reproduction is provided. We also obtain a new perspective on the role played by these theories in the constitution of nanotechnology, nanoscience, and the relevant phenomena.

But there's an additional reason to focus on von Neumann's contribution. As will become clear, von Neumann articulated throughout his career a *unified* picture of various domains of science, exploring and establishing connections between apparently unrelated areas. For example, on his view, logic, geometry, and probability are context dependent and should emerge from the formalism of the relevant field to which they are applied (von Neumann 1954). It's not by chance, then, that von Neumann developed quantum logic in the context of quantum mechanics (Birkhoff and von Neumann 1936), and various probability models depending on the particular areas of physics one considers (von Neumann 1937). Furthermore, as we will see, von Neumann showed the need for continuous methods not only in the foundations of quantum mechanics – elaborating the theory of continuous geometry (von Neumann 1960, 1981) – but also in the theory of computation – generalizing the usual discrete approaches found in the area (von Neumann 1951). Von Neumann also searched for a *unified* way of introducing probability in quantum theory, which eventually led him to go beyond his own Hilbert space formalism for quantum mechanics (Rédei 1997). As we will see below, the situation is in no way different when von Neumann developed his theories of automata and self-reproduction. Several moves made by him resulted from the attempt to articulate a *unified* approach to these theories. It's my hope to indicate that, by examining the influence that von Neumann's work had on nanotechnology and nanoscience, it will be possible to see how a more unified picture of *these* fields can emerge.[1]

I first sketch, in Section 1, a conceptual framework in terms of which the study of von Neumann's work will be articulated. The framework combines features of Peter Galison's work in *Image and Logic* (Galison 1997) with some additional aspects of an analysis of scientific practice. I then provide, in Section 2, some of the conceptual background for von Neumann's work on self-reproduction, examining key aspects of his work on large-scale computing machines and the theory of automata. In this way, in Section 3, all the elements to discuss von Neumann's theorem regarding self-reproduction will be on the table. I present the theorem and consider its significance. Finally, in Section 4, I examine the impact of von Neumann's work on self-reproduction to the constitution of nanophenomena, by exploring the role played by this work in Eric Drexler's conceptualization of nanotechnology (see Drexler 1986, 1992). A brief conclusion follows.

1. A Conceptual Framework

To examine von Neumann's contribution, it is useful to have a conceptual framework to guide and give some structure to the questions that will be raised. I'll adopt, in part, a framework Peter Galison developed to describe theoretical practice in microphysics (Galison 1997), but which turns out to be extremely helpful to the historical study of nanoscience. Just as the case of microphysics Galison examines, nanoscience – taken as a discipline – is a genuinely interdisciplinary field, with contributions emerging from a very special combination of chemistry, biology, engineering, and computer science (among other domains). The particular types of interaction among these areas are diverse and complex, just as are diverse and complex the interactions between microphysics and engineering (among other fields) that Galison describes. Although it is a substantive issue how to characterize the particular forms of interactions among fields in nanoscience, there is no question that *there are* such interactions. This becomes particularly clear, for example, in Drexler's work, where different areas of chemistry, biology and computer science are woven together to articulate an account of molecular manipulation and computation (see Drexler 1992). As I'll discuss, von Neumann's approach played a significant role in Drexler's work, providing part of the theoretical context in which the work emerged. To represent the complexity and diversity of this trajectory in the history of nanoscience within a well-structured setting, Galison's approach seems to be perfectly suited.

As will become clear, Galison's framework has the advantage of highlighting important features of scientific practice, while still being plastic enough to be applicable to areas other than microphysics. The framework has four main components (see Galison 1997 for details):

(a) *Constraints and contexts*: Despite the plurality of approaches often found in science, scientific practice is *constrained* in various ways. *Theories* impose constraints on the acceptable solutions to problems. But these constraints also indicate how new problems can be solved. *Experiments*, in turn, constrain the way theories are formulated, entertained, tested, rejected or accepted. They also provide new parameters for theory construction. *Instruments* constrain the practice of discovery in laboratories, while they also produce new data for theoretical and experimental research. This is, of course, all done in a *social* context, where *political* considerations play a variety of roles.

In other words, there are several kinds of constraints: *theoretical*, *experimental*, *instrumental*, *political and social*. These constraints play both a negative role of limiting, say, the range of acceptable solutions to various problems, and a positive role of suggesting solutions to new problems.

Not only is scientific practice constrained in the above ways, but it is also something *local* and *contextual*. Different scientific communities have different languages, employ different standards and adopt different norms to conduct their research. It comes as no surprise then that different scientific communities pursue and assess their research according to different criteria. As a result, scientific practice becomes a contextual and local phenomenon. It seems appropriate to examine it in this way.

(b) *Trading zone*: A major challenge to any genuinely interdisciplinary work is to have a common language in which the different assumptions, theoretical commitments and proposals of the various scientific communities in question can be expressed and communicated. In order to develop a genuinely interdisciplinary research – that bridges very different communities (physicists, engineers, mathematicians etc.) – scientists develop a 'trading zone'. In this 'zone', through the development of a simplified language, scientists are able to communicate, despite the (often dramatic) differences in their backgrounds. Of course, the language in question, being extremely simplified, is unable to capture the full content of the theories and methods of the various communities. But the language typically has enough resources to make possible the communication between the members of these communities.

(c) *Image and logic*: In Galison's view, there are two different traditions of instrumentation in physics (Galison 1997). According to the *image tradition*, images of natural processes should be produced with such clarity that these images could serve as evidence for the existence of a new entity. This involves the use of cloud chambers, nuclear emulsions, and bubble chambers. According to the *logic tradition*, evidence is established in a different way. Through the use of electronic counters, coupled in electronic logic circuits, masses of data are aggregated. And through the application of statistical techniques to these data, arguments for the existence of the entities in question are produced. Of course, this whole approach depends on very different instruments than the image tradition, including counters, spark chambers, and wire chambers. These traditions clearly use different tools to achieve their goals, and have succeeded in their own different ways. There is no doubt about the importance of these traditions. In fact, as Galison argues, the history of 20th century microphysics is, in many ways, the history of the vicissitudes of these two traditions of instrumentation. As will become clear, these traditions also found their way into nanoscience and nanotechnology.

(d) *Three levels of analysis*: Throughout the discussion below, three levels of analysis will be explored. The first level examines *theoretical practice*,[2] and it engages the role played by various theories in the formulation of several approaches to nanotechnology. The second level concerns *experimental practice*, the practice of experimentation and its connection to theoretical practice. Finally, the third level addresses *instrumental practice*, exploring the role played by various types of instruments in the constitution of the relevant phenomena. There are, of course, important connections between these three levels of analysis, and the interconnection between them in the context of nanoscience and nanotechnology will be explored in the discussion that follows.

Having briefly indicated the overall framework to be used in this paper, I am now in a position to begin the analysis of von Neumann's work in light of the conceptual setting just presented.

2. Background to von Neumann's Approach

Von Neumann was always concerned with developing new strategies for problem solving, whether such problems involve novel ways of representing the state of a quantum system or the strategic interaction between economic agents. As we will see, it was ultimately this unfathomable interest in heuristics – particularly in the context of mathematics – that led von Neumann to be involved with large-scale, high speed computing. And it was in the context of his work on computing and automatic machines that von Neumann first articulated his approach to self-reproduction. So, I will start by providing some of the background to von Neumann's work on computing machines and the nature of the problems that led him to address the issue of self-reproduction.

In a paper written in 1946 with Herman Goldstine, "On the Principles of Large Scale Computing Machines", von Neumann points out:

> In this article we attempt to discuss [large-scale, high speed, automatic] machines from the viewpoint not only of the mathematician but also of the engineer and the logician, *i.e.* of the more or less (we hope: 'less') hypothetical person or group of persons really fitted to plan scientific tools. We shall, in other words, *inquire into what phases of pure and applied mathematics can be furthered by the use of large-scale, automatic computing instruments* and into *what the characteristics of a computing device must be in order that it can be useful in the pertinent phases of mathematics*. (Goldstine & von Neumann 1946, p. 317; italics added.)

It is important to note that, given the way von Neumann conceptualizes the issue, it concerns people working across very different disciplines: from mathematics through engineering to logic. Clearly, the implementation of a project of this magnitude requires the development of strategies of communication between different fields, and what each participant has to contribute and can get from the project is very different. The concerns of mathematicians are not the same as those of the engineers, which in turn are different from the logicians'. Similarly for the expertise each of them will bring. In the end, the articulation of such an enterprise ultimately demands a 'trading zone'.

Note also the *constraints* on the problem. There is a two-way relation between the computing machines to be devised and their users (particularly if we consider mathematicians): First, mathematics should be developed further by the use of these machines. For example, the solution of some problems that were intractable at the time should be achieved through computing machines. Second, the machines themselves should have an architecture that supports such mathematical developments. In fact, as von Neumann emphasizes, analytical methods at the time were inadequate for the solution of a number of non-linear partial differential equations. Large-scale computing machines were expected to be particularly useful in this context.

In other words, von Neumann's concern with heuristic devices for mathematics motivated him to be involved with high-speed computers. But, in von Neumann's view, the implementation of such computing machines required a theory of automata. And as will become clear in a moment, it was in the context of his theory of automata that von Neumann was led to examine self-reproduction. Now, according to von Neumann, what were the main features of a theory of automata?

In 1951, in an article on "General and Logical Theory of Automata", von Neumann answered this question by putting forward a program to elaborate a whole theory of automata. To develop the theory, von Neumann explicitly invoked two *constraints*: (a) to explore, within certain boundaries, the *analogy with living organisms*, and (b) to use *structures from* (*mathematical*) *logic*. I will elaborate on each of these constraints in turn.

(a) With regard to the *analogy with living organisms*, von Neumann tried to model the functioning of the automaton, in part, in analogy with the functioning of a neuron, and the way in which the latter transmits impulses. Given the remarkable ability that neurons have to transmit impulses and information, it certainly seems to be an appropriate starting point for a theory of automata. This is particularly the case if we first realize the important *differences* (or disanalogies) between automata and neurons. In fact, von Neumann stressed two important dissimilarities.

First, the extremely *small size* of the neuron compared to the vacuum tube (then used in computers). The neuron is not only *smaller*, but *much more efficient* than the vacuum tube. As von Neumann notes: "the basic fact is, in every respect, the small size of the neuron compared to the vacuum tube. […] What is it due to?" (von Neumann 1951, p. 403) This was not a rhetorical question on von Neumann's part. He had a partial explanation for the greater efficiency of neurons in comparison to vacuum tubes, despite the smaller size of the former: it referred to the *materials* that constituted each of them. In the case of vacuum tubes, we have a combination of metals separated by vacuum; in the case of neurons, we have the cytoplasm and nucleus of human cells.

In fact, the different *materials* that characterize neurons and computers amount to a *second disanalogy* between the two. This also helps to explain the difficulties faced at the time in successfully developing computing machines. As von Neumann points out:

> The weakness of this technology lies probably, in part at least, in the *materials employed*. Our present techniques involve the using of metals, with rather close spacings, and at certain critical points separated by vacuum only. This combination of media has a peculiar *mechanical instability* that is entirely alien to living nature. By this I mean the simple fact that, if a living organism is mechanically injured, it has a strong tendency to restore itself. If, on the other hand, we hit a man-made mechanism with a sledge hammer, no such restoring tendency is apparent. (von Neumann 1951, pp. 404-405; italics added.)

That is, the mechanical instability and lack of a self-restoration tendency in computing machines are significant differences between these machines and neurons, and these differences arise, at least in part, from the dissimilar materials used. Note however the strange slipperiness on von Neumann's part from mechanical instability to lack of self-restoration in the case of living organisms. The mechanical instability may lead a given object to suffer some sort of malfunction or to be somehow damaged. But clearly, even in the case of living organisms, this doesn't mean – nor does it entail – that the object in question will go through any self-restoration. The two notions (mechanical instability and lack of self-restoration) are not obviously equivalent.

But the mechanical instability also has a significant consequence for the size of the computing machines. In von Neumann's view, it is in virtue of that instability that the size of computers hasn't been reduced yet. And the instability, in turn, is the outcome of the materials that have been employed:

> It is this mechanical instability of our materials which prevents us from reducing sizes further. […] Thus it is the inferiority of our materials, compared with those used in nature, which prevents us from attaining the high degree of complication and the small dimensions which have been attained by natural organisms. (von Neumann 1951, p. 405)

In other words, von Neumann emphasizes the limitations due to the *materials* used to produce computers – this is a constraint at the *instrumental* level. Moreover, he also highlights the limitations due to the *scale* of the relevant objects (after all, the size of the neuron is an important factor in the successful transmission of the relevant bits of information). Issues about scale also provide a limitation at the instrumental level. By identifying these two instrumental differences between neurons and computers, von Neumann is clear about the areas in which further work still needs to be pursued: to identify new and better materials and, through them, to implement and construct computing machines at a smaller scale.

But, according to von Neumann, there is still an additional constraint to be met. This one arises at the *theoretical* level:

(b) The use of *structures from* (*mathematical*) *logic* is crucial for von Neumann's project. After all, logic provides an overall framework to represent the abstract components of computation and to assess the adequacy of each step. In von Neumann's view, it's only in terms of a mathematical-logical theory of computation that the limitations found in the automata of his time could be overcome. As he points out:

> We have emphasized how the complication [complexity] is limited in artificial automata […]. Two reasons that put a limit on complication [complexity] have already been given. They are the large size and the limited reliability of the componentry that we must use. […] There is, however, a third important limiting factor […]. This factor is of an intellectual, and not physical, character. The limitation which is due to *the lack of a logical theory of automata*. We are very far from possessing a theory of automata which deserves that name, that is, a properly *mathematical-logical* theory. (von Neumann 1951, p. 405; italics added.)

According to von Neumann, it is a *theoretical constraint* on the theory of automata that it be framed in terms of *mathematical logic*. But why should the theory satisfy this constraint?

This is a point where von Neumann's search for a unified account plays a significant role. A theory of automata should provide an account of reasoning processes, accommodating the way in which knowledge can be represented and inferences obtained. Mathematical logic is, of course, particularly useful for that, even though, given the way in which it has traditionally been formulated, it has a major limitation:

> Everybody who has worked in formal logic will confirm that it is one of the technically most refractory parts of mathematics. The reason for this is that it deals with *rigid, all-or-none concepts*, and has *very little contact with the continuous concept of the real or of the complex number*, that is, with mathematical analysis. Yet analysis is the technically most successful and best-elaborated part of mathematics. (von Neumann 1951, p. 406; italics added.)

What von Neumann proposes is to re-conceptualize the logical tradition in terms of analysis, and elaborate a theory of automata in this new setting. Properly characterized, mathematical logic could overcome its traditional all too rigid outlook. By incorporating resources from real and complex analysis, mathematical logic could become still more useful to model the complexities inherent in reasoning and in the representation and transferring of information.

The incorporation of analysis into logic is also achieved by the development of set theory, in which results from both real and complex analysis can be formulated and established. In the 1920's, von Neumann provided an extremely elegant axiomatization of set theory (von Neumann 1925), a system that is now called von Neumann-Bernays-Gödel. It was an important feature of von Neumann's work that his system was *finitely axiomatizable*, given that the main rival system of set theory at the time, the one provided by Zermelo, *couldn't* be finitely axiomatized (Zermelo 1908). With infinitely many axioms, it's not possible to express a system of set theory as the conjunction of its axioms – unless one invokes some admittedly artificial devices, such as introducing infinitary languages, that arguably no human could ever actually use.[3] Given the motivation to use the resources of mathematical logic to develop a theory of computation, devices of this nature wouldn't be of much use for von Neumann.

The emphasis on continuous methods rather than discrete ones is an important component of von Neumann's overall approach, and it is a unifying theme throughout much of his work. This emerges from von Neumann's emphasis on the resources for modeling provided by analytical methods. For example, in the 1930's, von Neumann developed the theory of *continuous geometry*, a generalization of projective geometry involving a *continuous* number of dimensions (for an overview, see von Neumann 1960 and 1981). The development of this kind of geometry emerged from von Neumann's work in the foundations of quantum mechanics. It was the result of his attempt to develop a mathematically unified account of quantum theory, going beyond his previous work on the Hilbert spaces approach (von Neumann 1932). In terms of continuous geometry, and using what we now call von Neumann algebras, von Neumann showed how quantum probability could emerge from the formalism of quantum theory in a natural way – even when one considered quantum systems with infinite degrees of freedom. This result couldn't be obtained using the Hilbert space formalism (see Rédei 1997 and 1998).

It is in this context of trying to extend the logical paradigm of his time to incorporate analysis, and searching for a better, more sophisticated theory of automata that von Neumann faced an additional alleged limitation to that theory. Given the analogy with living organisms that motivated so many aspects of the theory of automata, it's not surprising that von Neumann considered an additional putative dissimilarity between living organisms and computing machines. Living organisms have the ability to reproduce, and some to self-reproduce. Given that automata are artificial entities, does that mean that they are in principle unable to self-reproduce? In von Neumann's view, the answer is *negative*. This provides, of course, additional evidence for the analogy between living organisms and computers. To show why this is the case, von Neumann was led to study the properties of self-reproduction in the context of his theory of automata.

3. Von Neumann and Self-Reproduction

Von Neumann starts his analysis of the notion of self-reproduction by identifying a difficulty that the notion seems to face. It addresses the very possibility of devising self-reproducing automata, given a "degenerating tendency" regarding the complexity of the automata involved in the task:

> If an automaton has the ability to construct another one, there must be a decrease in complication [complexity] as we go from the parent to the construct. That is, if A can produce B, then A in some way must have contained a complete description of B. […] In this sense, it would therefore seem that a certain degenerating tendency must

> be expected, some decrease in complexity as one automaton makes another automaton. (von Neumann 1951, p. 415)

This is, of course, an objection against the possibility, in principle, of self-reproducing automata. If the degree of complexity has to decrease as we move from the parent automaton to the offspring, we won't have a case of self-reproduction, given that the offspring is not of the same kind as the parent, but is a less complex type of object.

In response to this objection, von Neumann relied, once again, on the analogy with biological organisms:

> Although this has some indefinite plausibility to it, it is in clear contradiction with the most obvious things that go on in nature. Organisms reproduce themselves, that is, they produce new organisms with no decrease in complexity. (von Neumann 1951, p. 415)

But this response doesn't completely settle the issue, as von Neumann was certainly aware. After all, even if organisms reproduce themselves without decreasing the complexity of the offspring, why would that establish that *artifacts*, such as *automata*, could also self-reproduce?

To answer this question, von Neumann has to tackle head-on the problem of the possibility of self-reproducing automata. In fact, he proves that it is *mathematically* possible for an automaton to self-reproduce. To establish this result, von Neumann generalizes a theorem first proved by Turing regarding the existence of "universal automata" (a particularly strong kind of Turing machine). An automaton is said to be *universal* if it can produce any sequence that can be produced by any automaton. In other words, a universal automaton is at least as effective as any conceivable automaton – including one that is twice its size and complexity! How is this possible? By using an idea of Turing's:

> Turing observed that a completely general description of any conceivable automaton can be […] given in a finite number of words. This description will contain certain empty passages – those referring to the functions […] which specify the actual functioning of the automaton. When these empty passages are filled in, we deal with a specific automaton. As long as they are left empty, this schema represents the general definition of the general automaton.
>
> Now it becomes possible to describe an automaton which has the ability to interpret such a definition. In other words, which, when fed the functions that […] define a specific automaton, will thereupon function like the object described. […] This automaton, which is constructed to read a description and to imitate the object described, is then the universal automaton in the sense of Turing. (von Neumann 1951, p. 417)

But there is a significant limitation in Turing's conception. As von Neumann notes, Turing's proposal is too narrow in one important respect. To function as a self-reproducing automaton, a computing machine has to yield as output *another automaton*, rather than, say, simply a sequence of numbers (typically, zeros and ones). Talking about Turing's machines, von Neumann insists:

> His automata are purely computing machines. Their output is a piece of tape with zeros and ones on it. What is needed […] is an automaton whose output is other automata. (von Neumann 1951, p. 418)

To establish the possibility of automata that generate other automata, and thus to dispel any worries regarding the latter, while substantially extending Turing's view, von Neumann provides his theorem regarding self-reproduction.

The key ideas of the theorem are very clear, as von Neumann clearly indicates (see von Neumann 1951, p. 420). Let *A* be an automaton with the property that, when supplied with the description of any other automaton, it constructs that object. Let *B* be an automaton that can copy any instruction *I* that is furnished to it. Combine the automata *A* and *B* with each other, and with a control mechanism *C*. *C* does the following. Let *A* be supplied with an instruction *I*. Then *C* will first make *A* construct the automaton described by the instruction *I*. Then *C* will make *B* copy the instruction *I*, and insert the copy into the automaton that has just been constructed by *A*. Finally, *C* will separate this construction from the system $A+B+C$, and take it as an independent object. Call *D* the total aggregate $A+B+C$.

In order to function, the aggregate $D = A+B+C$ must be supplied with an instruction *I*. Of course, this instruction has to be inserted into *A*. Now form an instruction I_D, which describes this automaton *D*, and insert I_D into *A* within *D*. Denote the aggregate which now results by *E*.

> *E* is clearly self-reproductive. Note that no vicious circle is involved. The decisive step occurs in *E*, when the instruction I_D, describing *D*, is constructed and attached to *D*. When the construction (the copying) of I_D is called for, *D* exists already, and it is in no [way] modified by the construction I_D. I_D is simply added to form *E*. Thus there is a definite chronological and logical order in which *D* and I_D have to be formed, and the process is legitimate and proper according to the rules of logic. (von Neumann 1951, p. 420)

Note the role played by mathematical logic throughout this construction. The process of construction of self-reproducing automata is modeled by the process of construction of the cumulative hierarchy in set theory (whose development von Neumann was, in part, responsible for in the 1920's). The set-theoretic cumulative hierarchy is constructed by stages, and at each stage, only sets that have already been constructed in previous stages can be used. (In this way, set-theoretical paradoxes can be avoided.) Similarly, to avoid a vicious circle in the construction of self-reproducing automata, von Neumann is very clear about what is constructed in each stage. As he makes it clear, the construction of the new automaton *E* is only possible *after* the construction of the automaton *D* and the instruction I_D, and *D* and I_D basically encompass *all* that is needed to construct *E*. So, the possibility of constructing self-reproducing automata is definitely open.

This result raises a number of questions, and someone may be tempted to use them to undermine the significance of the theorem. For example, what is special about the fact that it is mathematically possible to construct an automaton that self-reproduces? To be completely convinced of the possibility of self-reproduction isn't it enough just to look at nature, with the astonishing spectacle of organisms that reproduce themselves? Why do we need a mathematical theorem to prove such an obvious fact?

It's important to note, in response, that these questions miss the point of von Neumann's result. There is no doubt that nature provides a remarkable variety of self-reproducing systems. But, as noted above, in nature, we are not talking about *artifacts*; we are considering *living creatures*. There is no doubt that living beings of the appropriate sort (*e.g.* which are members of the same species) can reproduce, and in some cases even self-reproduce. What is definitely *not* obvious is that *artifacts*, such as an automaton, could do the same – even in principle. And this is the point of establishing von Neumann's theorem.

I'm here assuming, with von Neumann, a distinction between *natural* and *artificial* systems. The distinction is, of course, *vague*. It's vague in the technical sense that there are clear-cut cases of natural systems (such as untouched parts of the *Amazon jungle*); clear-cut cases of artificial systems (such as the *software* I am using to edit this paper); and cases in which it is not clearly determined whether they constitute natural or artificial systems (such as a *glass of beer*). Despite the vagueness of the distinction between the natural and the artificial, it is important to recognize that *there is still a distinction*. Typically, those who deny the natural/artificial distinction are only denying that there is a *sharp* distinction here.[4] But to appreciate the significance of von Neumann's theorem, all that is needed is the existence of an *unsharp* distinction. After all, as long as some automata are on the artificial side of the divide – and this is precisely the case considered by von Neumann – it is indeed not obvious why they should self-reproduce.

Note also that granting the existence of the distinction between the natural and the artificial in *no way* undermines von Neumann's use of natural processes to model various aspects of artificial systems (such as the automata he studies). Any analogy has its limitations – there are always *negative* analogies – and, as noted above, von Neumann is perfectly aware of them. But these limitations don't undermine the existence of the *positive* analogies: the common features that natural and artificial systems share, despite their differences. And these common features ground, in part, the way in which von Neumann models his automata.

Finally, note that von Neumann is particularly concerned with establishing the logical – but also mathematical – possibility of self-reproducing automata. This is the reason why he emphasizes the fact that *no vicious circle* is involved in the process of construction of a new automaton from another. So, in principle, it's not *logically* impossible to develop self-reproducing automata. And it's not *mathematically* impossible either. The construction carried out in von Neumann's proof is articulated in a simple mathematical setting. In fact, as already noted, the construction is modeled in set theory. As a result, nothing in the proof is incompatible with classical mathematics.

This raises the issue of what is mathematically and logically *impossible* to achieve according to von Neumann's approach to self-reproduction. The answer emerges from the mathematical framework von Neumann uses to articulate his proof. The limitative results from mathematical logic regarding what *cannot* be computed clearly apply to the program of self-reproduction he devised. Von Neumann is, of course, perfectly aware of this fact as well. And he tries to overcome some of these results by insisting that a new framework for computation is required, one that is based on analysis rather than on combinatorial systems. In this way, by emphasizing the continuous nature of the computational processes, a more refined, and more powerful, approach to computation could be provided.

To sum up the discussion so far: von Neumann employed, in a particularly fruitful way, structures from mathematical logic, such as Turing machines suitably adapted and set-theoretical constructions. These structures provided an important *constraint* at the *theoretical level* for his work. Clearly, von Neumann's contribution sides with the *logic* tradition of computer making. But von Neumann is also changing and restructuring this tradition, by broadening the logical tools used, and bringing logic closer to analysis than to combinatorics. Moreover, von Neumann also emphasized the *instrumental constraints* imposed by the *materials* used in the construction of computing machines and the *scale* of the components that were employed at the time. He clearly highlighted the need for the development of better materials. Now, with the *mathematical* possibility of self-reproducing automata, the case is open for their *physical* construction. Although there is still a long way to go, at least the first step was taken.

4. Von Neumann, Drexler and Nanophenomena: Roots to Nanoscience

What is the impact that von Neumann's work had on nanoscience and nanotechnology? To answer this question, I will discuss the influence of this work in the development of a very interesting approach to nanoscale phenomena: Eric Drexler's vision for the field (see Drexler 1992 and 1986).[5]

Drexler is very clear about the nature of his investigation. It is what he calls *theoretical applied science* (Drexler 1992, pp. 489-491). This is a "mode of research which aims to describe technological *possibilities* as constrained not by present-day laboratory and factory techniques, but by physical law" (Drexler 1992, p. 489; emphasis added). The goal, then, is to examine what is feasible, given *physical* constraints on the phenomena under investigation, rather than technological limitations that might be present at the time of the research. The typical product of theoretical applied science, similarly to theoretical physics, is not a family of experimental results, but a "theoretical analysis demonstrating the *possibility* of a class of as-yet *unrealizable* devices, including estimated lower bounds on their performance" (Drexler 1992, p. 489; the first emphasis added). In other words, theoretical applied science is concerned with the study of technological *possibilities*, which immediately links it with research in science and engineering. Talking about theoretical applied science, Drexler points out that

> Its technical content (drawing extensively from physical theory and experimental results) and the nature of its product (knowledge, rather than hardware) link it closely to scientific research. Yet it is also closely akin to engineering: studying technological possibilities poses problems of design and analysis. The products of theoretical applied science can be termed *exploratory designs*, although some take the form of a rather abstract analysis. (Drexler 1992, p. 490)

Now among these exploratory designs, Drexler studies nanomechanical computational systems (Drexler 1992, pp. 342-371), molecular assemblers (*ibid.*, pp. 372-410), and molecular manufacturing systems (*ibid.*, pp. 411-441). And throughout the elaboration of these designs and analyses, a crucial component plays a significant role. Just as von Neumann had done in the context of his theory of automata, Drexler also explores *analogies between biological phenomena and events at the nanoscale* as a guiding principle in theory construction. Similarly to von Neumann's approach, this also includes examining significant *dissimilarities* between biological phenomena and some constructions at the nanoscale. For my current purposes, it is enough simply to illustrate this move with a typical example.

In his discussion of the research that forms the foundation for his own approach, Drexler notes:

> Most experimental research in molecular electronics has focused on the development of molecules that exhibit useful electronic properties in thin films or in microscale aggregates; some proposals, however, have focused on the construction of computational devices in which individual molecules or moieties would serve as signal carrying and switching elements. (Drexler 1992, p. 509)

A seminal work by Robinson and Seeman is then referred to. In this work, the design of a biochip is described through the formulation of a self-assembling molecular-scale memory device (see Robinson and Seeman 1987). Clearly, the strategy consists in exploring biological grounds for molecular electronics. As Drexler notes, works such as this

> have suggested various combinations of chemical synthesis, protein engineering, and DNA engineering to make *self-assembling systems on a broadly biological model*. This objective is a form of molecular systems engineering (though not of machines or manufacturing systems) and the capabilities required would resemble those discussed in Chapter 15 [the chapter on macromolecular engineering in Drexler's 1992 book]. (Drexler 1992, p. 509; emphasis added.)

In other words, it is through a biological model that molecular systems engineering could be implemented, even though the particular goal that Drexler has – namely, to develop molecular manufacturing systems – had not been pursued before.

As passages such as the above indicate, in Drexler's approach to nanotechnology, there is a significant integration between several areas of research (Drexler 1992, pp. 507-511). In fact, the foundation for Drexler's approach emerges from related research on a number of areas, in particular, chemistry, molecular biology, protein and mechanical engineering, as well as computer science and proximal probe technologies (especially, the use of scanning tunneling and atomic force microscopes). Interestingly enough, the integration between these areas is achieved through an overarching vision, dominated by the goals of theoretical applied science, in which the autonomy of each area is preserved, and the relevant results from each area are invoked to establish the new outcomes. This means that a genuine 'trading zone' has to be created, which requires meeting constraints at the theoretical and instrumental levels.

Despite the fact that work in so many fields forms the foundation for Drexler's proposal, the direction he favors differs from the traditional approach. The divergence from the latter view emerges from an important methodological difference between the traditional approach, which is *top-down* oriented, and Drexler's *bottom-up* methodology (Drexler 1992, p. 508). According to the top-down approach, favored for example by microtechnology, we start with "large, complex, and irregular structures", and we try to reduce their sizes; the challenge, then, is to "make imprecise structures smaller" (*ibid.*, p. 508). This differs significantly from Drexler's bottom-up approach. According to this approach, which is ultimately grounded on chemistry, we start with "small, simple, and exact structures", and we try to increase their size; the challenge, then, is to "make precise structures larger" (*ibid.*, p. 508).

As we saw, in dealing with his theory of automata, von Neumann clearly recognized the importance of *size* as a limitative constraint on the efficiency of artificial computing machines. Even though von Neumann's overall approach seemed to be closer to the top-down strategy (recall his discussion of how different materials may allow a decrease in the automata's size), his theorem regarding self-reproducing automata was based on a clearly *bottom-up* construction. In fact, the whole point of using the set-theoretic cumulative hierarchy as a model for von Neumann's mathematical construction of self-reproducing automata was exactly to ensure a bottom-up approach. As we saw, in this way, von Neumann avoided the objection that self-reproducing automata involved a vicious circularity.

Given considerations such as these, it comes as no surprise that Drexler clearly acknowledged the importance of von Neumann's work for the development of his own approach to nanotechnology. First, von Neumann's work on the theory of automata is quoted in both *Nanosystems* (Drexler 1992) and in *Engines of Creation* (Drexler 1986). Moreover, in a private communication (September 12, 2003), Drexler pointed out: "I'd been familiar with the outlines of [von Neumann's] work on self-replicating systems before my own work turned toward nanotechnology, hence it was part of the intellectual foundation for that work." As for the significance to nanotechnology of von Neumann's work on self-reproduction, Drexler is also very clear: "[von Neumann's] work originated the idea of non-biological self-replicating systems, which was central to early concepts for the implementation and use of large-scale systems of productive nanomachinery." However, Drexler currently thinks that, contrary to "widespread impressions that [he, Drexler] had a role in forming", self-replication is "not, in fact, necessary for the implementation and use of large-scale systems of productive nanomachinery". In fact, in his present view, self-replication is feasible, potentially safe, but ultimately unnecessary (Phoenix and Drexler 2004). Despite this, Drexler notes that von Neumann's work "strongly influenced [nanotechnology]", even though, as far as he knows, it "did not anticipate its essential features".

I think this establishes, without doubt, the importance that von Neumann's work had in the constitution of a significant approach to nanotechnology – namely, Drexler's – and hence, indirectly, to the overall construction of the field. Far more could be said here, of course, beyond the methodological and conceptual strategies linking von Neumann and Drexler.[6] But the existence of these shared methodological strategies, although not conclusive on its own, is already significant. In fact, it would be misleading to claim that all that Drexler's and von Neumann's approaches had in common was the fact that they followed the old idea that 'technology imitates nature.' The way in which von Neumann and then Drexler use aspects of natural systems to model technical devices – being sensitive to *scale* and to the *materials* of the relevant objects – is remarkably similar. And together with Drexler's explicit acknowledgement of the role of von Neumann's work in his own, the shared methodological strategies undoubtedly establish the historical link between von Neumann and Drexler.

5. Conclusion

Von Neumann clearly had a unified approach to the various foundational issues he addressed, from quantum mechanics to the theory of automata. In his view, logic, mathematics, probability and the relevant scientific theories need to be articulated in a unified and well-motivated way. As noted above, in von Neumann's view, the notion of probability in quantum mechanics should emerge naturally from the formalism of the theory, even when we consider quantum systems with infinitely many degrees of freedom. To fully articulate this view, von Neumann then developed a completely new branch of mathematics: continuous geometry. Similarly, in the case of his theory of automata, by re-conceptualizing the paradigm of mathematical logic of his time – through the exploration of the resources of analysis and set theory – von Neumann was able to show the mathematical possibility of self-reproducing automata.

Von Neumann's work, and his unified vision for theoretical research, later informed important parts of Drexler's approach to nanotechnology. Just as von Neumann, Drexler also insisted on the importance of exploring analogies with biological systems in the modeling of nanophenomena. And just as von Neumann, Drexler also noted the important limitations of such analogies, and what can be learned from them as guidance for future research. Moreover, just as von Neumann had a unified picture of theoretical research, Drexler also has a unified picture of nanotechnology, one in which the various areas involved – from chemistry through molecular biology to computer science – have to be integrated, even though their autonomy should be preserved along the way. Trading zones have to be constructed to implement the details of such a vision, just as trading zones had to be elaborated in von Neumann's own implementation of his vision for the theory of automata.

Although there is much more to be said, I hope I said enough to *motivate* the idea that a slightly more unified picture of nanotechnology and nanoscience *can* emerge when these fields are examined from the historical perspective suggested here. Of course, identifying the particular historical trend highlighted here is only the first, but a necessary, step in this process – and I plan to explore these issues further in future work. The roots to nanoscience, from von Neumann to Drexler, are rich, sophisticated, unified, and definitely worth exploring. There is a lot there.

Acknowledgment

Earlier versions of this paper were presented at the conferences 'Discovering the Nanoscale', at the University of South Carolina in March 2003, and 'Discovering the Nanoscale, II', at the Technical University of Darmstadt, Germany, in October 2003. I wish to thank both audiences for extremely helpful comments and probing questions that led to substantial revisions. Special thanks go to Arne Hessenbruch, Ann Johnson, and Richard Ray for their skepticism, and to Eric Drexler for particularly valuable information. Joachim Schummer commented on my paper in Darmstadt, raising thoughtful and challenging points. Support for this work was given by the National Science Foundation through a NIRT grant.

Notes

1 I am not trying to provide here yet another 'founding myth' for nanoscience and nanotechnology. I'm only suggesting that, by exploring a particular trend in the recent history of these areas, it's possible to conceive of a different, more unified, picture for nano-scale research. In this paper, I only identify this particular historical trend, and sketch its major features. To articulate the details of the resulting (more unified) picture will have to wait for another occasion.

2 I owe this term to R.I.G. Hughes, who is currently developing a fascinating account of theoretical practice in physics (see Hughes 2004).

3 Other devices could be invoked here as well. For example, if the language in question has a truth predicate, it is possible to assert an axiom scheme that encompasses infinitely many sentences at once, such as 'Every sentence of the form '*P* or not-*P*' is true'. But with a truth predicate, problems such as the liar paradox emerge, and thus the truth predicate itself becomes suspicious. Alternatively, one could use a device such as the substitutional quantifier. Despite the name, this is not exactly a quantifier, but a technique to generate infinite conjunctions. As a result, roughly speaking, it ends up facing similar problems as the use of infinitary languages.

4 It is hard to believe that anyone would claim that untouched parts of the Amazon jungle form an *artificial* system, or that the software used to edit this paper is a *natural* system! This would be the outcome of the denial that there is *any* distinction between natural and artificial systems.

5 This is, of course, only one possible trend to explore. As Joachim Schummer pointed out to me, it's worth examining the rediscovery of von Neumann's theory of automata by people working on 'Artificial Life' in theoretical biology in the 1980s, and then studying the impact of their work in theoretical nanobiotechnology. I plan to explore this point in future work.

6 For example, one could analyze the details of the arguments used by Drexler for the possibility of assemblers and compare them with von Neumann's argument for the existence of self-reproducing automata. Due to limitations of space, I'll be unable to do that here, but I hope to explore this issue elsewhere.

References

Birkhoff, G. & von Neumann, J.: 1936, 'The Logic of Quantum Mechanics', *Annals of Mathematics*, **37**, 823-843.

Drexler, E.: 1986, *Engines of Creation: The Coming Era of Nanotechnology*, New York: Anchor Books. (Expanded edition, with a new afterword, 1990.)

Drexler, E.: 1992, *Nanosystems: Molecular Machinery, Manufacturing, and Computation*, New York: John Wiley & Sons.

Feynman, R.: 1960, 'There's Plenty of Room at the Bottom', *Engineering and Science*, **23**, 22-36.

Galison, P.: 1997, *Image and Logic: A Material Culture of Microphysics*, Chicago: University of Chicago Press.

Goldstine, H. & von Neumann, J.: 1946, 'On the Principles of Large Scale Computing Machines', reprinted in: J. Glimm, J. Impagliazzo & I. Singer (eds.), *The Legacy of John von Neumann*, Providence, RI: American Mathematical Society, 1990.

Hughes, R.I.G.: 2004, *Theory and World*, unpublished manuscript, University of South Carolina.

Phoenix, C. & Drexler, E.: 2004, 'Safe Exponential Manufacturing', *Nanotechnology*, **15**, 869-872.

Rédei, M.: 1997, 'Why John von Neumann did not Like the Hilbert Space Formalism of Quantum Mechanics (and What He Liked Instead)', *Studies in History and Philosophy of Modern Physics*, **28**, 493-510.

Rédei, M.: 1998, *Quantum Logic in Algebraic Approach*, Dordrecht: Kluwer Academic Publishers.

Robinson, B.H. & Seeman, N.C.: 1987, 'The Design of a Biochip: A Self-Assembling Molecular-Scale Memory Device', *Protein Engineering*, **1**, 295-300.

von Neumann, J.: 1925, 'An Axiomatization of Set Theory', reprinted (in English translation) in: J. van Heijenoort (ed.), *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, pp. 394-413.

von Neumann, J.: 1932, *Mathematical Foundations of Quantum Mechanics*, Princeton: Princeton University Press (English trans. by Robert T. Beyer, of the original German edition was first published in 1955).

von Neumann, J.: 1937, 'Quantum Logics. (Strict- and Probability- Logics)', unpublished manuscript, von Neumann Archives, Library of Congress, Washington, DC (brief summary by A.H. Taub in: J. von Neumann, *Collected Works, vol. IV. Continuous Geometry and Other Topics*, Oxford: Pergamon Press, 1962, pp. 195-197).

von Neumann, J.: 1951, 'The General and Logical Theory of Automata', in: L.A. Jeffress (ed.), *Cerebral Mechanisms in Behavior: The Hixon Symposium*, New York: John Wiley, pp. 1-31.

von Neumann, J.: 1954, 'Unsolved Problems in Mathematics', typescript, von Neumann Archives, Library of Congress, Washington, DC.

von Neumann, J.: 1956, 'Probabilistic Logics and the Synthesis of Reliable Organisms from Unreliable Components', in: C.E. Shannon & J. McCarthy (eds.), *Automata Studies*, Princeton: Princeton University Press, pp. 43-98.

von Neumann, J.: 1960, *Continuous Geometry*, Princeton: Princeton University Press.

von Neumann, J.: 1966, *Theory of Self-Reproducing Automata*, Urbana: University of Illinois Press (ed. A.W. Burks).

von Neumann, J.: 1981, 'Continuous Geometries with a Transition Probability', *Memoirs of the American Mathematical Society*, **34**, 1-210.

Zermelo, E.: 1908, 'Investigations in the Foundations of Set Theory I', reprinted (in English translation) in: J. van Heijenoort (ed.), *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, pp. 200-215.