Essays in Philosophy

Volume 6
Issue 1 *The Philosophy of Technology*

Article 6

1-2005

Chess, Games, and Flies

Stefano Franchi
University Of Auckland

Recommended Citation
Franchi, Stefano (2005) "Chess, Games, and Flies," *Essays in Philosophy*: Vol. 6: Iss. 1, Article 6.

Essays in Philosophy is a biannual journal published by Pacific University Library | ISSN 1526-0569 | http://commons.pacificu.edu/eip/

Essays in Philosophy
A Biannual Journal

Vol. 6, No. 1, January 2005

Chess, Games, and Flies

Abstract

Research in Artificial Intelligence has always had a very strong relationship with games and game-playing, and especially with chess. Workers in AI have always denied that this interest was more than purely accidental. Parlor games, they claimed, became a favorite topic of interest because they provided the ideal test case for any simulation of intelligence. Chess is the *Drosophila* of AI, it was said, with reference to the fruit-fly whose fast reproductive cycle made it into a favorite test bed for genetic theories for almost a century. In this paper I will try to show Artificial Intelligence's relationship to games is quite different from what this analogy suggests. In fact, I will argue that AI is, at core, a theory of games and a theory of subjectivity as game-playing.

1. Games, game-playing and game theory.

It is well-known that research in Artificial Intelligence has always had a very strong relationship with games and game-playing. Much work in the field, especially in its founding years, has been devoted to the development of programs capable of playing chess or checkers, programs smart enough to solve puzzles, etc.[1] Chess was a favorite topic, although it was not the only game to receive the researchers' attention. Indeed, one of the brightest successes of early AI was a checkers playing program devised by A. L. Samuel at the IBM Laboratories, while efforts went also into the development of programs capable of solving cryptarithmetic puzzles, the Hanoi tower puzzle, and other "mathematical recreations."[2]

Workers in AI have always denied that this interest was more than purely accidental. Parlor games, they claimed, became a favorite topic of interest because they provided the ideal test case for any simulation of intelligence: a game like chess, for example, can be described by just a handful of rules and yet its complexity defies even the best human minds. However, since we usually take good chess players to be very intelligent human beings *in general*, it follows that a program capable of playing an acceptable game of chess should be considered as instantiating "intelligent behavior" *in general*.[3] Working on chess and other parlor games allowed AI researchers to leave aside unessential complexities to focus on the relevant issues pertaining to human intelligence. This line of defense is best summed up by an analogy that became very popular in the field: chess are the *Drosophila* of AI, it was said, with reference to the fruit-fly whose fast reproductive cycle made it into a favorite test bed for genetic theories for almost a century.[4] As the validity and scope of genetics is not limited to flies, so the scope and validity of Artificial Intelligence's theories is not limited to chess and, more generally, to games.

In this paper I will try to show Artificial Intelligence's relationship to games is quite different from what this analogy suggests. In fact, I will argue that AI is, at core, a theory of games *and* a theory of subjectivity as game-playing. This claim should not be interpreted in a reductive sense: I will not be arguing that the theoretical constructs of AI are only applicable to games. Rather, I will try to show, through a historical and conceptual reconstruction of the genesis of field, that AI came to its basic insights by trying to answer a number of questions that were prompted by its continued interest in chess: what is a game? what does it mean to play a game? Why can we take game-playing as one of the most basic characteristics of human behavior?

Artificial Intelligence is a theory that identifies some of the most basic features of human subjectivity with the basic procedures needed to play a game like chess, I will argue, and it could come to such a conclusion only through a prolonged engagement with chess that produced, at the same time, a theory of games, a theory of game-playing, and a theory of subjectivity as game-playing.[5]

This interpretation will be established through a historical analysis of the genesis of Artificial Intelligence's interest in games and especially in chess. First, I will focus on the biological counterpart of the analogy: the relationship between the fruit-fly and genetics. Then, I will show that chess played a similar role in AI: its peculiar features allowed it to function as the catalyst of a process that took off from the basic insights of von Neumann's game theory and ended with the production of a theory of games and a theory of subjectivity.

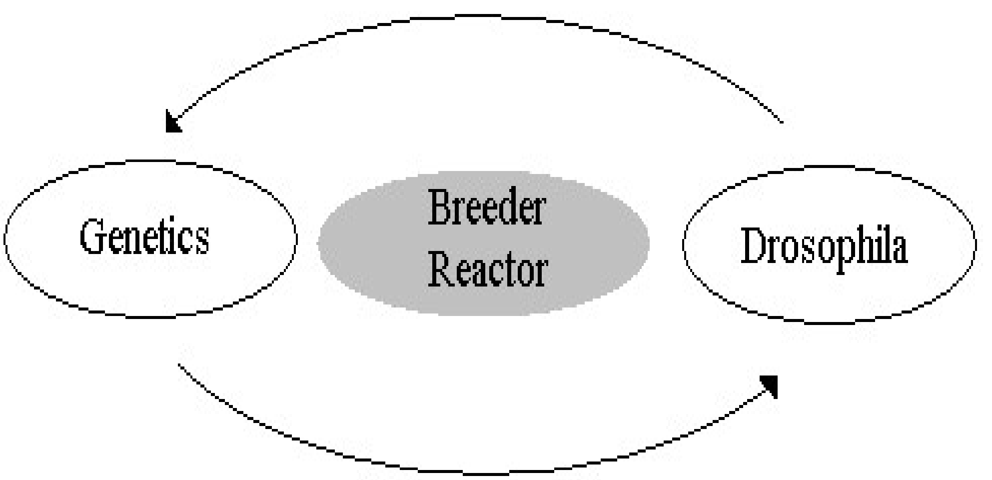
Two orders of reasons make this result quite interesting. The first is purely philosophical: in the last few decades the concepts of game and play have often been used strategically as the pointers toward a new conception of philosophy, an overturning of metaphysics, and other similar claims.[6] Yet, a direct engagement with games and game-playing is almost always missing in all these texts, beyond the occasional passing reference to game theory or to Huizinga's and Callois's classic analysis. The theory of games and game-playing provided by Artificial Intelligence is an important perhaps even indispensable supplement to gain a better understanding of these claims. I will show, in other words, that the analysis of games and game-playing behavior that could make more concrete so many postmodernist and post-structuralist claims is to be found in Artificial Intelligence's treatment of the subject—and not, for example, in classical game theory . Although an explicit discussion and evaluation of what AI says and what Lyotard and Derrida (for example) prescribe is left for another occasion, I think it is important to prepare the ground by sharpening the identity of a candidate. Even more so, in fact, if we consider that by offering a a game-based theory of subjectivity, AI would indeed qualify as an example of a post-modern theory of subjectivity of the kind claimed for by, for example, Lyotard. (and very much in the background of other apparently distant theorizations, like Lévi-Strauss's structuralism, that has nurtured so much French thought.)

The second reason is more methodological in character, although it has some serious philosophical consequences. A detailed reconstruction of the conceptual genesis of Artificial Intelligence will that the complex interaction between a scientific discipline's tools, its object and its theories is never as clear-cut as the traditional distinctions may suggest, in spite of what the scientists themselves may claim. Thus, the immediate goal of this paper is to find out whether chess, and other games, just happened to present favorable features allowing testing of AI programs or whether there has been any kind of backward influence from the characteristics of chess to AI theories. In the latter case, of course, the relevance of game for AI research would take a wholly different meaning, ascending from the role of a classically interpreted instrument to a more substantially theoretical component.

The effort does not entail that my analysis will rely upon a clear cut theoretical distinction between tools and theories that has been extensively, and convincingly, challenged in the last 15 years. It entails, however, that the distinction was (and still is, for the most part) clear-cut in the self-perception of the involved scientists. It is precisely in this sense that AI researchers like Simon wanted to exploit the *Drosophila* analogy: as a rhetorical weapon that, by demoting chess to the role of a theoretically neutral tool chosen for its standard and objective features, saves the seriousness of the theory used to explain chess-competence. Therefore, we are fully allowed to assume a distinction that was internal to the field under investigation as a starting point toward a more comprehensive analysis of the concepts, theories, and tools at work. In other words, the distinction may prove to be as untenable as the works of Latour and Woolgar, Fujimura, and others have shown. The exploration of the consequences of this fact in the specific case of AI is precisely what is as stake in what follows.[7]

2. *Drosophila* in biology.

I will start by taking a closer look at the complex relationship between theory and experimental tools involved in the *Drosophila* case, the bearer of the Artificial Intelligence's analogy. Since AI's case, as I will argue, is even more epistemologically complicated, it will be useful to start from a simpler occurrence of the same relationship. In a recent, comprehensive study, Robert Kohler argues that the relationship between genetic theory and the experimental equipment used to advance it (e.g. the fruit fly) has to be understood as a mutually reinforcing loop that acted in both directions.[8] The interaction between genetics and *Drosophila* can best be understood as a process represented by a diagram (something Kohler does not use) of the following sort:



Three main components make up the picture: genetics, as a discipline with its theories, practitioners, tools, etc.; *Drosophila*, as a living being with its biological, genetic, and social features; and, most importantly, what Kohler calls the "breeder reactor," namely the capability of *Drosophila* plus genetics' manipulations to turn out an ever increasing number of mutants. It is this last element that, although made possible by other two, operates as the engine of the whole process by propelling it back and forth between the two other components. In fact, the most characteristic feature of the *Drosophila*/Genetics interaction is that there is more than one passage in each direction: rather, it is a continuously operating, positive-feedback circle that alters substantially its members each time it goes through a half-revolution. In other words, the relationship is a complex process that produces not only a theory (the genetic theory of the chromosome) as its outcome, but also a new experimental object (the "standard" fly) as well as a new way of doing genetics (the large scale, mass-production, quantitative experiments devised in the "fly room" at Columbia). But let us proceed one step at the time.

The interaction started somewhat casually. At the beginning, as Kohler shows, the introduction of *Drosophila* in Thomas Hunt Morgans's laboratory at Columbia was quite casual, and most likely not determined by theoretical reasons. Rather, it was the nice fit between *Drosophila*'s life-cycle and robustness, on the one hand, and the growing need to provide hands-on laboratory experience to unskilled undergraduate students living on a 9 months schedule, on the other, that was largely responsible for the insect's promotion in the lab.

Drosophila started to assume a more central role when Morgan himself started to work on experimental evolution by trying to induce variations through intensive inbreeding and selection among large populations of flies. For a while, the experiments were inconclusive, until some mutants, among which was the famous *white* eye-color mutant, came up, rather unexpectedly, in the summer of 1910. Two interesting events followed. First, when it became clear that the causes of the mutations were to be found in scaled-up production and not in spontaneous, environment-induced evolution, Morgan's experimental work on *Drosophila* shifted substantially from experimental evolution to hereditary genetics. What we see, in short, is how the unexpected results forced a redirection of the theoretical interest in the scientist's practice.

The second event is even more interesting. The appearance and complex mechanics of mutant combinations (*white, vermilion, peach* eye-colors, etc.) was at first accounted for in terms of classic neo-Mendelian formulas: that is, by postulating the existence of dominant and recessive traits in particular genes, and by explaining the appearance of visible mutants by their combinations. However, very soon the *Drosophila* colonies started to work all too well for the neo-Mendelian paradigm. Unlike Gregor Mendel's famous green/yellow peas or Cuénot's brown/white/yellow coated mice, Morgan's fruit flies did not stabilize into a few diverse variations of the same morphological element. New mutants continued to come up and each new eye-color (for example) implied that the Mendelian formulas had to be rewritten. Since the formulas explained the visible characters as an exhaustive combination of basic genes (later traits), each unexpected mutation sent the geneticists back to the drawing board.[9] This fact directed Morgan's group away from neo-Mendelian explanations toward a different, and more stable, theory of hereditary variation, e.g. the theory of the gene, and toward the postulation of crossing-over and linkage mechanisms during chromosome splitting. In turn, the shift from one theory to the other forced another change in experimental work: neo-Mendelian searches for possible variations were abandoned in favor of the tecnique of genetic mapping that permitted a more stable, quantitative account of mutations capable of almost indefinite expansion.

This shift, however, forced a change in the relationship between the insect, the theorists' and the theory, that now affected directly the fruit fly. In order to obtain the precise quantitative mapping that the new theory required, the *Drosophila* stocks had to be liberated from all the "genetic noise" generated by other kinds of genes naturally occurring in wild fruit flies. As Kohler puts it, in order to be turned into "a standard experimental instrument [...] every new type of genetic noise—lethals, suppressors, modifiers—had to be identified and eliminated, one by one. *Drosophila*, so to speak,
[10]

had to be debugged." The history of the interaction between genetic theory and *Drosophila* does not stop here. Kohler shows, for example, how the standard form of the "constructed" insect that was essential to the success of the precise, quantitative genetic mapping turned into a major liability when scientists belonging to the fly group tried to expand their research in hereditary genetics toward the genetics of development and evolution, where the relationship between environment and naturally occurring variations became the focus of the analysis.

Four points should be retained from the biological analogy. First, there is more than one passage in each direction, since the interaction itself is a complex process that is constantly retroacting upon itself. Second, and perhaps more importantly, each passage substantially alters at least some relevant aspects of one of the members of the relationship. For example, the "standard fly," an animal painstakingly constructed in the lab throughout a minute debugging of its genetic material, is a different animal from the original fruit fly that entered it. Third, the engine propelling the relationship to move back and forth is constituted by *Drosophila*'s unexpected capability to generate an ever increasing number of mutants when undergoing the intense, large-scale inbreeding and selection procedures devised by the scientists, what Kohler calls the *Drosophila*'s breeder-reactor autocatalytic features. Last, but not least, the output of the whole process should not be forgotten: the theory of the gene. In fact, the theory can best be understood as the object emerging, at the epistemological level, out of the positive-feedback interaction between geneticists and *Drosophila* that the insect's fantastic (and induced) mutational capacities animated.

If *Drosophila* did indeed become the standard testbed of genetic theories for the best part of this century, it is because the insect exhibited the right kind of complexity and the right kind of simplicity: its genetic capabilities were daunting, as the ever new production of mutants clearly showed. Its non-genetic features were extremely simple: its physiology was simple, it was easy to breed the insect on a large scale, and it was possible to purify its genetic stock of all its inessential complexity (the genetic "noise"). A biologist could not have asked for a better experimental setting to test genetic theories.

But this felicitous mix of complexity and simplicity was not simply found in a pre-existing object. In fact, two important features of *Drosophila* stand out. First, the simplicity of the animal was to a large degree constructed by the experimenters themselves via a painstaking debugging that produced a "standard fly". Second, the complexity of the insect's reproduction acted as the catalyst on the whole scientific process and resulted in the production of at least three different kinds of objects: a scientific theory —i.e the theory of the gene—, a new animal—the standard fly—, and a new way of doing genetics—genetic mapping through mass inbreeding.

The similarities with the introduction of the study of chess into Artificial Intelligence, as we will see in detail, are remarkable. What makes AI/chess's case more complex, however, is that, at the beginning of the process, the two terms of the relationship are far less stable than their corresponding biological counterparts. At the dawn of the century, genetics was working along some well stabilized lines: the Darwinian framework plus the great amount of work in heredity that came out of the rediscovery of the Mendelian paradigm. The revolution the fly group started, in other word, took place within a well-established tradition. On the contrary, the difficulties intrinsically inherent in the definition of its topic made Artificial Intelligence far less stable than genetics.[11] Moreover, and as importantly, Artificial Intelligent was born at the same time as, and virtually because of, the interaction with chess.

Even more uncertain is the position of chess. In spite of all the changes it would undergo, *Drosophila*, after all, entered the scientific arena with a well-understood ontological status and at a well-defined epistemological level. No theoretical reasons set *Drosophila* apart from other living beings. It was a living being, but not a "special" living being, like a virus, for example, whose "borderline" zoological status may bias the experiments in a specific direction. Second, *Drosophila* is *inconspicuously* simpler, or more convenient to study, than other beings. Whatever the *Drosophila* does not have in terms of internal functional organization, or does have (reproductive speed, for example) must be theoretically irrelevant, lest biological research wastes its time on a simplistic case instead of focusing its energy on a methodologically simpler object. This second point, in turn, entails that the "object" *Drosophila* must be sufficiently well understood, at least in principle, in order to tear apart its "general biological characteristics" from its "particular drosophiliac" features, so to speak.

Chess does not enjoy the advantages of its flying companion. Even if we restrict the topic of AI to intelligent "behavior," it is not so obvious that playing games is an exemplary case of it. Second, it is not granted that chess is a relevant example of ludic activities in general.[13] Even worse, a case might be mounted that playing chess is a particularly extreme version of playing in general, making therefore chess more like a "virus" than like a "*Drosophila*." Thus, the task of pulling apart the "generally cognitive" characteristics of chess from its "particularly ludic" ones becomes especially thorny. Finally, once the proper understanding of chess is provided, it must be explained what makes it simpler, and why the simplicity is not reductive. The effort focusing on chess as a relevant example of intelligent behavior must start by defining a certain view on the game that selects some relevant features and declares them as relevant, or by assuming which features of chess are relevant and which are not.

Yet, the outcomes of the AI/chess process are at least as clean and polished as the theories and tools produced by the biologist. This entails not only that the process is more elaborated, since the distance traveled is greater. More importantly, it entails that the dynamic interaction between Artificial Intelligence and chess produces theories that redefine substantially their very objects. If the analogy between *Drosophila* in biology and chess in Artificial Intelligence is true, then it will be possible to show that AI built the equivalent of the "standard fly." It constructed a canonical representation of games patterned on chess and made it into the standard case of problem-solving activities. This canonical representation, in turn, left the AI laboratory to become natural once again: it became the standard interpretation of games, problems, and problem solving. In other words, the product of the process is as much a theory of thinking as it is a powerful conceptualization of *Artificial Intelligence* and of *chess*, and (by extension) of games.

3. Games in game theory.

Let us examine how the interaction started. Games were brought onto the scientific limelight by John von Neumann, who provided the basic framework for a mathematical concept of game and proved some important mathematical results about solvability.[14] Von Neumann, however, was interested in a different use of games: he wanted to use them as the basis for a model providing a basic mathematical understanding of the kind of behavior occurring in social interactions like economic transactions and strategic confrontations between social groups. Although he used insights from actual gaming practices—especially from games with less than perfect information, like poker—his work had little to do with psychology and individual cognitive capabilities.[15] It was instead a rather classical example of the applied mathematics he had been pursuing all his life: he used mathematical techniques to turn a potentially useful but imprecise concept into a powerful analytical tool.

Chess, it turns out, presents a rather uninteresting case for classical game theory, since it rather easily proved that there is a strategy that assures White or Black of a minimal gain. like a draw. The result was established by Zermelo in 1913 and states that chess has a value: there is a strategy such that either White can always force a win or a draw or Black can always force a win or a draw. Zermelo's theorem is an existential result: it shows that there must be such a strategy, but it does not provide the means to construct it. In fact, the complexity of the game makes it extremely difficult to find such a strategy, so that chess is theoretically solvable, although, practically, still unsolved.[16] Rather, chess plays an important heuristic function, being one the favorite examples von Neumann and Morgenstern use to illustrate their definitions of games *in general*. Since these definitions will provide the framework for any conceptual understanding of chess in Artificial Intelligence and beyond, it is important to review them, although briefly.

The basic concepts of game theory—*game, play, move, strategy* and *solution*— are defined by von Neumann and Morgenstern in a single page of their work. The concept of game is rather counterintuitive: they distinguish

> between the abstract concept of a *game* and the individual *plays* of that game. The *game* is simply the totality of the rules which describe it. Every particular instance at which the game is played—in a particular way—from beginning to end, is a *play*. (49, their italics.)

There are two important elements to underline. First, in everyday linguistic usage, "game" tends to oscillates between a narrow meaning in which it denotes just the particular instance at which the game is played—we might say, for example, "*a* game of chess", "it was *a* life and death game," etc. —and, on the other hand, a broad meaning in which it denotes the whole complex range of phenomena—the context, so to speak—within which the specific event takes place. Most often, it is not easy to decide exactly how broad is the context picked out by "game" in the latter sense.

For example, the history of chess is not part of the game of chess to many casual chess players. Any player who goes as far into the game as to read an introductory book of chess technique, however, is immediately confronted with "internal" history even in the most technical contexts. Openings, for example, are discussed according to their current viability, or their popularity in the past. Even the basic concepts of the game are very often presented in a historical context. For example, the sometimes subtle difference between attack and defense is often presented by comparing the flamboyant style so popular a century ago with the elaborate, calculated defenses that are now prevailing. Given that many chess players have reached this very elementary level of sophistication, it is probably fair to say that the concept of "game" can be taken to be at least this broad. Or consider the following comment, made by a journalist reporting on one of the final matches of the 1996 world championship between Karpov and Kamsky while trying to explain how they came to a bizarre endgame:

> A typical excessively hypermodern situation arose after 15… Nd8: there was no contact between the warring armies and there was just one open file. It had the look of an extravaganza by Richard Reti from the 1920's.[17]

Thus, at one end of the spectrum we find the narrow concept most economically represented by the score of a game published in a newspaper, or a book. Such a description is *a* game of chess in the sense that it seems perfectly legitimate to call it that way. At the other end, we may have a description as complex as that provided by the Japanese writer Kawabata for a very similar intellectual pastime, the game of *go*.[18] Kawabata—at the time a journalist—provides what anthropologists would call a "thick" description of the Japanese *go* championship that encompasses the whole gamut of experiences evoked by and enshrined into the event, from a technical analysis of the game itself (with *go* scores, diagrams, and all), to the generational clash between the defendant and the young challenger, to the different styles of play throughout the centuries, to the role of *go* in Japan's social and economic life. It is fair to say that one of the goals of the book is to describe, through the magnifying lenses of a single event, *the* game of go.

The point here is not to decide which of the two meanings of "game" is more adequate, since both, of course, are. The point is that von Neumann and Morgenstern's definition does not fall anywhere in the continuum spanned by the common uses of "game."

Instead, they chose the narrow construal and went up one logical level. To say that "game" is the "*totality of the rules which describe it*" means that a game is a set of abstract possibilities describing the possible ways, in fact *all* the possible ways, in which the game can be played, *if the attention is focused on the rules only*. To put it differently, *game*, as they define it, is an abstraction from a concretely given event, that cannot correspond to any possible concrete event because it belongs to a different logical level.

A second related point concerns the concept of *rule*. For von Neumann and Morgenstern,

> the rules of game […] are *absolute commands*. If they are ever infringed then the whole transaction by definition ceases to be the game described by those rules. (*ib.* , my emphasis.)

As a first approximation, this is obviously true: any willful violation of the rules puts the player outside the game and, in most occasions, effectively brings it to an end. If I were to start moving my rook as a bishop, I would quite effectively be putting an end to the game of chess I might have been playing, *unless my opponent were to follow me and do the same*. However, von Neumann and Morgenstern do not directly allow this possibility. Since the game is just the rules that describe it, and since the rules are absolute commands, it follows that any change in the rules will correspond to a different game.

But this is seldom the case. Should we say that the game of chess before the universal adoption of the rule governing the *en passant* capture by pawn is a totally different game, as the theory would require, from the game of chess after the adoption of the rule? Even in the most formal, "intellectual" games, many rules are the results of negotiations between the players, either at the individual or at the social level.[19] Often, moreover, the negotiation involves the most basic rules of the game, as anyone who has ever observed children playing simulation games will witness.[20] In game theory, instead, rules are essentially static and possess no latitude. Even more static, however, are the definitions of *move* and *strategy*.

A move, quite intuitively, is

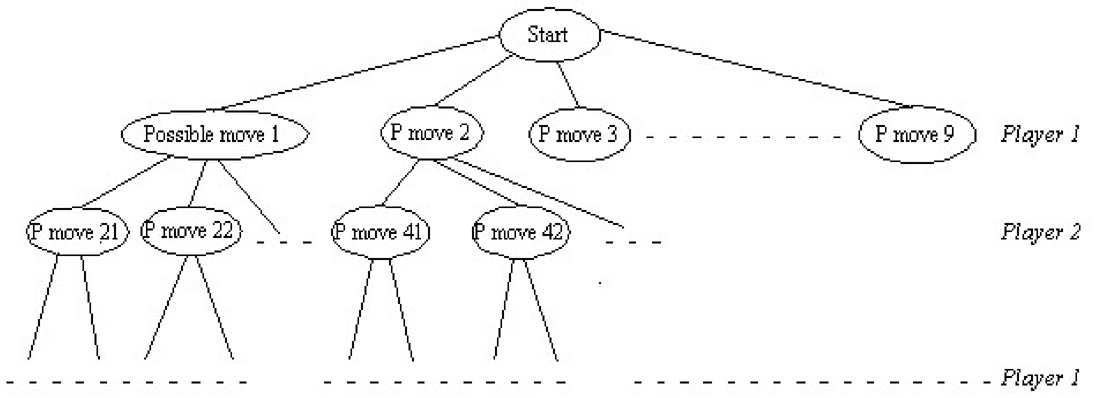
> the occasion of a choice between various alternatives, to be made either by one of the players or by some device subject to chance, under conditions precisely prescribed by the rules of the games. (*ib.*)

Less intuitive is the crucial concept of *strategy* that is built on top of move. Von Neumann and Morgenstern imagine that players may form a plan about their behavior in a game by deciding, beforehand, which alternative they will choose for each set of alternatives presented by each move. They call such a plan a *strategy*. As they observe,

> if we require each player to start the game *with a complete plan of this kind*, i.e. with a strategy, we by no means restrict his freedom of action. […] This is because the strategy is supposed to specify every particular decision only as a function of just that particular amount of actual information which would be available for this purpose in an actual play (79, my emphasis).

What is peculiar about the definition of strategy, is that the concept of move, with its intrinsically related temporal aspect, disappears, at least for any practical purpose. In fact, since a strategy has to be formed before the player begins to play, and since a strategy is the conjunction of *all* the choices that the player will make, it follows, trivially, that the whole course of action has to be planned beforehand. This conception of *strategy* differs somewhat from the usual meaning of the term: it is not a general rule of behavior that the player chooses at the beginning as a general maxim to guide him through the actual mechanisms of the actual game, when it will be tailored to the specific circumstances. Thus, "take control of the center" is considered to be a good strategic rule in chess, but its concrete realization is not given in advance. Instead, the formulation of a game-theoretic strategy requires that all possible games must have been plotted in advance in order to select the advantageous from the disadvantageous behaviors. There is no game-theoretic strategy that prescribes to "take control of the center;" rather, there is a set of strategies that describe the sequence of moves that actually occupy the center. In other words, the concept of strategy is intrinsically linked to the idea of game as a set including all possible games. It follows that, in order to choose a strategy, all possible games must have been devised even if the player is to play a single game. To put it differently, the games as described by von Neumann's and Morgenstern's theory are purely static, since the dynamic element of the interaction provided by the actual exchange of moves has to be planned before the interaction starts.

The typical graphical representation of a game present in *Game Theory and Economic Behavior* will make this point clear. Consider the children's game of TicTacToe.[21] It would be natural to try to analyze this game by proceeding as follows: "let's imagine the possible first moves by the first player, then let's imagine the possible countermoves by the second player, then let's imagine the possible counter-countermove by the first player, etc. A graphical representation of this procedure would bring us to a tree-like representation, in which each level stands for a move, in order to capture the dynamic aspect:



The ovals, or nodes, represent possible configurations of the game, and the lines connecting them stand for the action performed when a cross or naught is drawn. We can consider the initial position of the game as the root of the tree, the first row as the possible first moves by the first player, the second row as the possible countermoves by the second player, the third row as the possible counter-counter-moves by the first player, etc.

Instead, the standard game-theoretic description is a matrix where each row represents one player's strategy (not move) and each column stands for the other's, with each single cell representing the final result. If we assume that the complete strategies for the first TicTacToe players are represented by rows, the second player's by columns, wins by W, losses by L, and draws by D, we may obtain a representation like the following: (which represents only a fragment of the complete game):

Player 1	Player 2	Strategy 2.1	Strategy 2.2	Strategy 2.3	....
Strategy 1.1		W	L	D	...
Strategy 1.2		D	D	W	...
Strategy 1.3		W	L	W	...
....		...	...	...	...

The matrix representation is usually called the "normal" or "normalized" form of a game, as opposed to the "extensive" form (e.g. the tree-like representation). Although von Neumann and Morgenstern introduce the extensive representation of games at first (in TGEB), they quickly shift to the normal form in an effort to simplify the formalism. In fact, the introduction of the concept of strategy as a "complete plan of actions" by a rational player is what allows the reduction of a game from the extensive to the normal form.

Notice that I am not accusing von Neumann and Morgenstern of having used a somehow "wrong" representation. That would not only be silly, but also plainly wrong, from the point of view of game theory. In fact, many of the sophisticated distinctions and concepts provided by the theory (dominant strategy, zero-sum, equilibrium, etc.), are immediately evident when the game is represented in normal form and are rather obscure in the extensive form. The point here is to show the relationship between game-theory and chess (as an instance of a large class of games) and how the former basic concepts make the latter quite irrelevant. In this context, moreover, the issue is not that the matrix does not permit to express the complete game of chess; the tree-like one does not either, since in both case we would have to do with immensely large objects. The point is that the normalized form does not allow us to think *about* the complexity, since there are no moves in it, only outcomes. In the case of games whose complexity lies along the temporal axis (e.g. the moves), the game itself escapes the formalism.

We have now come to the last basic concept of the theory. A *solution* for a game is, intuitively, a strategy that can maximize, for a player, his minimum guaranteed payoff: the payoff that a player may won no matter what the opponent(s) strategy may be. The problem of game theory, then, is to find if, for a given game, a solution exists. On the basis of the previous definitions, von Neumann and Morgenstern are able to provide a taxonomy of games along different, independent, dimensions: number of players, perfect or imperfect information (like most card-games), cooperative vs. non-cooperative, zero-sum (one player's loss is the other player's win) and non-zero-sum. For each class, they set to solve the problem of game theory. The simplest case turns out to be the class of 2-person, zero-sum, perfect information game, whose best example is, yet again, chess. For this class of games, the theory provides a theorem, sometimes called the Fundamental theorem or the maximin theorem, that assures of the existence of a solution in any case.[23]

After this excursus on the basic concepts of game theory as presented by von Neumann and Morgenstern in their 1944 work, let us go back to chess and examine how it moved out of game theory's concerns to become a central topic in Artificial Intelligence's research.

First, it should be noted that from the strictly game-theoretical point of view chess is a quite uninteresting if not altogether trivial example.[24] In fact, chess is, essentially, no more complex than TicTacToe. The only difference is that the greater complexity of the game translates into a bigger matrix: more strategies are available to the players because there are more moves and because every move offers more alternative choices. However, since any game-theoretic treatment of a game requires the preliminary computation of the strategies, the complexity of chess escapes the theory. To win (or at least not to lose) at chess is enough to calculate the outcome of every possible strategy available to White and Black and then compute which one to use by means of the "minimax" technique they describe.[25] One of the first results of game theory (Zermelo's theorem, in fact) assures us that there is a strategy such as either Black can always force a win or a draw, or White can always force a win or a draw. The theorem states that chess has a value (W, L, or D) but the technique that would discover it is not applicable to chess, since it would require the preliminary computation of the outcome of any possible chess game. As a consequence, chess becomes only slightly more interesting than TicTacToe and in the context of game-theory serves, at most, the same original purpose of *Drosophila* at Columbia: it is a convenient didactic tool that may come handy since most students are already familiar with it, but is quite unfit for serious theoretical work. It is interesting to see what Von Neumann and Morgenstern are perfectly aware of the almost paradoxical status of chess within game theory as a solvable game with an unknowable solution, and it is interesting to see how they comment this interesting fact:

> This shows that if the theory of chess were really fully known there would nothing left to play.[...] But our proof, which guarantees the validity of one (and only one) of these three alternatives, gives no practically usable methods to determine the true one. *This relative, human difficulty necessitates the use of those incomplete, heuristic methods of playing which constitute "good" Chess*.[26]

This brief quote substantially ends any discussion of chess in von Neumann and Morgenstern's work. Game theory, at least as envisioned by its founders, has no room for this kind of human limitation that may make game-playing difficult. Instead, it focuses on other, qualitatively different kinds of complexities. Once they established the existence of a solution for the simplest class of two-person zero-sum games (to which chess belongs), von Neumann and Morgenstern tried to extend their analysis, first to multi-person games and then to non-zero-sum games.[27] This program brings game-theory farther and farther away from real games people play and into more and more complicated examples of "games" that have little, if anything, in common with games as we know them. In fact, the only link between "games" and the "theory of games" is provided by the general definition illustrated above. An immediate consequence of the program is that any interest in "empirical" games is lacking in game theory. A game like chess, therefore, lacks any interest for game-theorists and, as a consequence, the game-theoretic community who set to work to expand von Neumann and Morgenstern's basic theory, in the 1950s, did not initiate any backward loop toward a direct investigation of the game itself with the tools they had provided.[28]

The gap between the complexity of a concrete game of chess and the sophisticated machinery provided by von Neumann and Morgenstern was so paradoxically big that it started to be used as an objection against game-theory. Martin Shubik, for example, remembers that when he went to Princeton to study economics and game theory with Morgenstern, immediately after the war, he encountered considerable skepticism in the economics department. The standard objection raised by the economists was: if "game theory could not even solve the game of chess, how could it be of use in the study of economic life, which is considerably more complex than chess?"[29] The game theorists responded with what has now become the standard reply to be found in any textbook on the subject; for example, Roger McCain writes:

> Game theory may be about poker and baseball, but it is not about chess, and it is about such *serious interactions* as market competition, arms races and environmental pollution. But game theory addresses the serious interactions using the *metaphor* of a game: in these serious interactions, as in games, the individual's choice is essentially a choice of a strategy, and the outcome of the interaction depends on the strategies chosen by each of the participants. [30]

It should be noted that both the economists' and the game-theorists' critiques converge on decreeing the theoretical irrelevance of chess. For the former, the inability of game theory to solve the game is a sure sign that the theory is totally off the mark: it has better focus on serious, concrete economic facts instead of getting lost in analysis that cannot even come to terms with recreational activities. Games should be abandoned altogether, chess first and foremost. For the latter, chess offers the wrong kind of complexity, a complexity peculiar to the non-serious sphere, but the basic metaphor can be exported to the serious realm of strategic interaction, where it will prove its worthiness. On both counts, however, chess is not to be taken seriously, and anyone working on them should be prepared to defend his work against forthcoming criticism. Chess is off the loop of serious theoretical work, to put it differently, and can serve, at most, as a didactic introduction to the basic "metaphor" of the field to be offered in the first chapter of game-theory textbooks.[31] Chess's case is certainly not unique in the field, and many other "real" games have endured the same (mathematical) treatment: they are picked up as a source of intuition and then relegated to an introductory, paradigmatic function. Poker, for example, has become the "standard" non-perfect information game; "heads and tails" the standard example of the simplest example to require the use of mixed strategies; more recently, the game of "Nim" has become the paradigmatic introduction to combinatorial game theory. However, what is characteristic of, and perhaps unique to, chess, is that it experienced a second scientific life. That life began immediately after the publication of *Game Theory and Economic Behavior* and began, once again, with John von Neumann.

4. The *Drosophila* of AI.

Chess was brought back into the scientific limelight by von Neumann's intellectual evolution and his personal involvement with the RAND corporation in Santa Monica, a research institute heavily funded by the military (specifically, the Air Force). RAND had been created in 1946 in order to preserve in postwar years the intense collaboration among scientists that had proved so fruitful to the national military interests during War World II.[32] Santa Monica became one of the centers— with Princeton and, on a lesser scale, Michigan— for the development of game-theory, and von Neumann was immediately hired as a consultant. In the immediate postwar years, however, his own interests started to shift toward the design and construction of electronic computers and, consequently, toward the automatic resolutions of problems posed by, among other things, game-theory. The possibility was far from being purely theoretical, since the combination of von Neumann theory and military interests had brought about the construction of one of the first computers at RAND, appropriately called (apparently against his protestations), the JOHNNIAC.

As a consequence of von Neumann's shifting interests, chess (and games) were brought into the orbit of computer science, or rather into the orbit of computers, while being understood within the conceptual framework of game theory. It is in this context that he started lecturing on chess again, every time emphasizing the immense complexity of the task ahead. Von Neumann, it should be remarked, lectured on the difficulty of practically finding the "solution" of chess whose existence was guaranteed by the fundamental minimax theorem. Therefore, he was not interested in a chess-*playing* computer since he was trying to find chess-*solving* algorithms that a machine could use.

The "optimizing" perspective of the theory, however, when combined with the complexity of the game, made the task seem impossibly prohibitive, even on a fast computer. Moreover, the algorithm designer could find no help in the study of the expertise of a good chess player, because there is no direct link between the algorithm a computer will use and the way the game is actually played. Humans can play chess well, sometimes exceedingly so—but they will never play it perfectly. A game-theoretic computer, instead, is stuck in a paradox: it will either play a perfect game or an exceedingly bad one.[33] The two conceptual steps toward the solution of the paradox were taken, in rapid succession, by Turing and Shannon and then by Simon and Newell in the early 1950s. It is their work that closes the circle and allows chess to become as productive for Artificial Intelligence as *Drosophila* had been for genetics.

In a few years span, Alan Turing and Claude Shannon published two papers describing the basic design principles that would allow a computer to *play* chess, although they never came to write the actual programs. However, their use of the game is still outside Artificial Intelligence, broadly construed: they want to prove, by demonstrating the practical possibility of a chess-playing machine, that computers are able to manipulate symbols and not just number-crunchers perennially devoted to solving differential equations as most people, at the time, took them to be. The quality of the game played by a machine, its intelligence, so to speak, was thus less important than the demonstration that such a machine could exist.[34] To be more precise, a chess-playing machine would perhaps demonstrate some intelligence, but the degree of intelligence would not be tied to the quality of the game played.

Chess is still used in the "exemplary" mode as an instance of a more general paradigm, namely symbolic thinking, which can be tested on the simpler domain provided by games. Chess may enter the computer scene because the problem of "solving" the game, crucial to game theory, has been put aside. "Playing" the game becomes the central concern, because being able to play an intellectual game like chess is taken to be a significant test of cognitive abilities. Paradoxically, the quality of the game being played is less important that the fact that a machine can "play" the way humans are supposed to. Turing, for example, tested his design for a chess playing computer against a person who had never played the game before and was taught the rules only a few hours earlier. The experiment is significative and could not be farther away from game theory's concerns. What is tested is the program's ability to follow the rules of chess, and perhaps pursue a goal, i.e. its ability to *think*. It is not the performance level that counts, but the proof that a machine may think. Game theory wanted to find solutions for the games of war and stock bargaining, and did not care if an untrained human secretary (Turing's test case) was or was not capable of understanding their rules.[35] In the passage from von Neumann to Shannon and Turing, instead, the emphasis shifts from trying to *solve* the game to trying to understand the *process* that allows the player to play it (and sometimes to win it).

It is remarkable how little of the conceptual apparatus of game theory needs to be abandoned in order to pursue this new goal. The concept of solution has to be modified in order to find a new conceptualization of "game" that would allow the theorist to find a winning strategy for the game of chess (as von Neumann wanted), while allowing a computer to play the game (as Shannon and Turing now desire). In fact, the definition of a different concept of solution is precisely one of the first backward effects from chess to Artificial Intelligence, as we will see. Even more remarkable, however, are the consequences of the approach that retains most of the theoretical framework while redefining the concept of strategy. From a mathematical tool that strives to provide a computable models of economic interactions, the redefined theory of games will transform itself into a sweeping theory of that most cherished human feature—thinking. The first steps toward that goal were taken by Claude Shannon, who was to be followed by and then by the three-some composed of Herbert Simon, Allen Newell and Cliff Shaw, all of whom were working at RAND, either full-time or as consultants, in 1950s.

In two short articles written in 1950, Claude Shannon entertains the possibility of a thinking machine, and focuses his attention explicitly on a computer playing chess. If such a machine existed, he suggested, its behavior would probably be considered "by some psychologists" as a thinking process. Shannon begins his discussion from a game theoretic description of chess in extensive form, although his attention shifts quickly from the formal theoretic structure, the game, to the agent of the playing process. The machine, he says, would first have to try out in the abstract various possible solutions to a given problem (e.g. a move in the chess game), in order to evaluate the results of these trials and then to carry out the chosen solution.[36] The background of the article is in place: the construction of a chess-playing machine is being examined in order to illuminate the possibility of a thinking machine in general, i.e. a machine able to exhibit internal processes and structures essentially similar to human cognitive processes. I stress this point because in order to appreciate what Shannon is about to say on the internal structure of the machine we should not forget that the implicit referent are the human cognitive processes, and the structures and algorithm he presents are supposed to help either in a better understanding of the human mind or in an actual reduplication of (parts of) it.

The basic problem, according to Shannon, can be split into three parts: (a) a translation system from chess positions to sequence of numbers must be chosen, (b) a strategy must be found for choosing the moves to be made, and (c) this strategy must be translated into a sequence of elementary computer orders, i.e. a program (2126). The first and third problems are irrelevant for what interests us here, concerned as they are with the problem of translating from human level to machine level.[37]

Shannon describes the basic problem of chess playing as follows:

> A straightforward process must be found for calculating a *reasonably good move* for *any given* chess *position*. (2127, my emphasis).

Although Shannon is thinking "the game of chess" in game-theoretic terms, he abandons in a single gesture two deeply interrelated tenets of the theory. First, he shoves aside the concept of a "solution" for the game of chess and proposes instead to find "*reasonably* good moves." Second, he abandons the game-theoretic and deeply static concept of strategy by focusing the attention, instead, on *moves*. In other words, Shannon is proposing to shift the attention from the static matrix of a game to the dynamic process that makes the game. This move allows him to see chess's taunted complexity in a totally new light. Or rather, it allows him to provide a framework to think such a complexity. This last point becomes clear when we look at how he frames the solution for his point (b) above:

> The program designer can employ here the principles of correct play that have been evolved by expert chess players. These empirical principles are a means of bringing some order to the maze of possible variations of a chess game. Even the high speeds available in electronic computers are helplessly inadequate to play perfect chess by calculating all possible variations of a chess game. In a typical chess position there will be about 32 possible moves with 32 possible replies— already this creates 1024 possibilities. Most chess games last 40 moves or more for each side. So the total number of possible variations in an average game is about 10^{120}. A machine calculating one variation each millionth of a second would require over 10^{95} years to decide on its first move! Other methods of attempting to play perfect chess seem equally impracticable; we resign

> ourselves, therefore, to having the machine play a reasonably skillful game, admitting occasional moves that may not be the best. *This, of course, is precisely what human players do: no one plays a perfect game*" (2127, my emphasis)[38]

Thus, the tree representing the complete extensive form of the game would have, as Shannon reports, approximately 10^{120} nodes (80 half-moves with 32 possible choices = ^{3280} » 10^{120} different chessboard configurations). Set issues of size aside for a moment, however, and focus on the structure.

The tree can be created one step (one node, that is) at the time by the recursive application of the rules of the game. From any given position any legal move of one of the pieces on the chessboard will produce a new position, and so on. The complete tree is of course impossible to create, but the search procedure does not have to rely on a complete tree if it settles for less than optimal results. It may create just a few positions per turn, deciding to explore only one or two moves beyond the current one. Even this is not a simple task to accomplish, at any rate: if we consider an average of thirty-two different possibilities moves from a standard chess position, the machine has to consider 32^2=1024 positions just for one move, $32^{4\ x}$ or more than million for just two moves, etc. On the other hand, a less-than-complete game theoretic matrix is useless, since it does allow the specification of the strategies available to the players.

The second thing to notice is that the tree does not represent the "perfect" game, or indeed any game, but the collection of all possible chess games, from the most trivial one to the grandmaster's masterpiece. In fact, every complete vertical path of the tree begins with the initial position and terminates, after a variable number of moves, either with a victory for white or black or with a draw, and as such stands for a complete game. The tree as a whole represents the space of "chess" as such. Individual games can be recomprehended as proper parts of the complete structure (i.e. as complete vertical paths.) The perfect game would become possible *after* the complete tree is in place, since to play it is necessary to know the possible outcome of all possible moves, i.e. it is necessary to read the tree from the bottom up. This is evident from a moment of reflection upon the tree-like structure: White, for example, will know for sure if it can always force a win, only if it examines every possible countermove by Black at any possible point in the game (i.e. at any level in the tree). In other words, White has to examine every single node in the tree, before moving a piece. This is why Shannon points out that the first move would take the machine 10^{95} years to complete.

Since game-theory reasons in terms of "strategies," it is forced to build the complete tree *before* it can apply its theoretical tools since a strategy, according to von Neumann and Morgenstern's definition, has to take care of all possible occurrences. The complexity of chess, therefore, lies outside the theory itself and has to be tamed before the latter can display its prowess. For Shannon, instead, the complexity is expressed in terms of the number of *states*—that is, chess positions—not strategy, and this makes all the difference. States can be generated on the fly by the application of the rules of chess, which form a small manageable set. A proficient (although not perfect) chess-playing computer will be able to base its game on only a very small fragment of the complete chess-game. The complexity of chess is now an integral part of the theory itself: it is the "combinatorial explosion" in the number of states that a small number of rules can generate. Its problem, therefore, becomes how to tame such a complexity by finding ways to keep the size of the tree under control.

By accepting most of the basic game-theoretic apparatus and, at the same time, relinquishing the crucial concept of strategy, Shannon has provided a framework that makes chess's complexity thinkable. Moreover, his suggestions about how to tame such complexity will make the study of chess relevant and will be crucial in establishing a strict connection between Artificial Intelligence and chess. "This [e.g. finding moves that may not be best] is precisely what human players do: no one plays a perfect game," he says. Therefore, the "solution" of the game of chess might be found in the human way of playing chess and, on the reverse, the performance of a chess-playing computer may be measured against human performance.

This last point, however, is still a hint in Shannon's paper and will not receive its full development until the works of Simon and Newell. What, on the other hand, seems undoubtedly his, is the "discovery" of the combinatorial explosion of states as the result of the dynamic application of the rules. Although this may seem quite a discouraging result, in fact it plays a role analogous to the breeder reactor's properties we had seen at work in the *Drosophila*/genetics interaction. The "complexity" of the insect lay in its ability to produce an ever increasing number of mutations when interacting with the experimental methods of the geneticists. These mutations, by challenging the scientists' tools, were forcing them to refine their theories to keep up with the insect's capacities. Similarly, chess's "ability" to turn out ever new forms of combinatorial explosions under the scientists' pressure to keep it under control forces them to devise more and more refined methods (read, "theories") to keep up with the game's "complexity." In both cases, the productive relationship can be established because there is a conceptual framework that makes the object's complexity thinkable and that allows the search for a solution to start. In genetics' case, this was the theory of the gene, in Artificial Intelligence is the conception of a game as a *search-space*, e.g. as a space of single, static positions generated recursively by the application of the rules of the game.

The framework provided by Shannon shifts all the emphasis in the study of chess from "solving the game," to "searching the space for an acceptable solution." The problem of Artificial Intelligence becomes how to search that space for a solution as good as a human being could find, providing thus, in a single stroke, a theory of thinking and a theory of chess. It is the solution of this problem that attracts the attention of Herbert Simon and Allen Newell.

Around 1950, Herbert Simon was a young economist working in the "backwaters" area of industrial organizations and had already published a substantial work, *Administrative Behavior*. In the book, a study of the way employees work in very large and generally public structures, he had argued that real problem-solving decision cannot and do not happen by finding the optimal solution to a given problem. In real-life situations, problem solvers (like managers) have to give up the hope to find the best solution because they can count only upon limited information and do not have the best possible strategies available to them. They can only rely on rules-of-thumbs, on heuristic rules derived from past experiences to take decisions. The solutions they are looking for will be "satisfactory" solutions that are "good enough" for the given, specific situation in which the problem has to be solved and not optimal solutions valid in general. Simon called "satisficing," as opposed to "optimizing" this typical organizational behavior.[39] This approach is very consonant with most of game theory's approach to economic behavior, in its insistence on seeing rationality as always bounded by very effective constraints. Simon was in fact an early reader of von Neumann and Morgenstern's work, and published a very positive review of the book, the very first one to come out, in fact, in 1945.[40] However, he was very critical of the concept of "solution" of a game, a concept that he associated with the impossible search for an "objective" optimal rational behavior as it could be judged from a third-person perspective. Instead, he insisted on *satisficing* as the best expression of first-person, subjective rationality that would prove much more useful in the comprehension of intelligent, decision-making behavior.[41]

Allen Newell's work provided a different angle on the same issue: how to find an effective way to tame the intrinsic complexity of the search space that has now become expressible. Newell, 10 years Simon's junior, had been a student of the mathematician George Polya while a physics undergraduate at Stanford in the late 1940s, and had become well-acquainted with his work on heuristics. Polya had diverted his attention from the then common study of the formal structures of proofs as they are presented to the mathematical community (and as formal logic studies them), and focused instead on the "solving methods" that bring the mathematician to the discovery of those proofs. He had provided an articulate summary of heuristic methods that, although not guaranteeing a solution, can provide the crucial insights that will lead to it.[42] It is easy (it always is, *post factum*) to see the convergence between Simon's concerns and the basic premises of Polya's studies that the young Newell had absorbed.

After a year of graduate school at Princeton—where he met most of the leading game-theorists, but that left him very dissatisfied with the purely mathematical approach to game theory that was dominant in those years—Newell joined RAND to work on "applied" mathematics. There he met Herbert Simon, who, although teaching in Pittsburgh, had been hired as a consultant and started to spend a few summers there.[43]

In his autobiography, Simon notes that he got involved with RAND through the Cowles commission and started to attend summer seminars in Santa Monica in 1952. The first product of his interaction with the game-theorists was the article "A Behavioral Model Of Rational Choice," originally a RAND technical report. The paper contained an appendix on chess that was stimulated by a talk on the topic given by von Neumann that summer. Simon writes: "I thought that von Neumann was overestimating the difficulties substantially, and moreover I believed I had some solutions for them which I proposed in the appendix [later excised from the public version]." He later recalls that immediately after he met Newell, whom he recognized as being far more advanced on the topic, he put his plans for a running chess-playing program on hold until they began collaborate intensely on the project toward the end of 1955. Soon joined by Cliff Shaw, a system programmer at RAND, they began working toward the implementation of a program incorporating the profoundly different interpretation of game-theory we have just outlined. The report of their effort was published in 1958 as "Chess-Playing Programs and the Problem of Complexity," and marks a landmark in the AI literature on chess. In fact, it marks a landmark in AI in general since, by bringing to a conclusion the early phase of the interaction with games, it provides the guidelines of a general theory about problems solving and its measurement through the application of computer programs to chess that will remain stable for a long time to come.

At the beginning of their ground-breaking article, Simon, Newell and Shaw state clearly what is at stake in the research: the program's ability to play chess provides a measure of "recent progress in understanding and mechanizing man's intellectual attainments."[44] The argument supporting the claim is quite simple: "Chess is the intellectual game *par excellence*," they say, and

> Without a chance device to obscure the contest, it pits two intellects against each other in a situation so complex that neither can hope to understand it completely, but sufficiently amenable to analysis that each can hope to outthink his opponent. [...] If one could devise a successful chess machine, one would seem to have penetrated *the core of human intellectual endeavor*. (39, my emphasis)[45]

First of all, the scope of the project has substantially changed, veering off from applied mathematics to, should we say, philosophy: the researchers' aim is no longer, or not only, to discover useful techniques to analyze complex situations, but rather to penetrate the "human intellectual core." This is possible because the emphasis has shifted from the general inquiry into rationality that was specific to the original game-theoretic project to a more specific, and yet more ambitious, search for the subjective roots of that rationality. In other words, Simon and Newell will strive to obtain a description of the rational agent *as it works*, that is, a description of the process that brings rationality about.

In this opening act, therefore, we see that the "core" of thinking, or the core of intellectual behavior, is constituted by problem solving, and problem solving is well exemplified by taming the complexity of a chess game. A theory of thinking must therefore provide not only a *explanation* of how the goal, the solution, can be reached, but also a replication of the performance. Since the solving process is what is at stake, the process itself will have to be replicated, not just the underlying principle. The computer assumes then a central role, because it is only by simulating on a computer the "blind" search of a solution abstractly possible in the game-tree that we can expect a true confirmation of a theory.[46]

The second point to emphasize concerns the basic problem that such a theory must solve: combinatorial explosion, or complexity, whose measure is given by the sheer, directly unmanageable size of the game-tree. Newell, Simon, and Shaw's answer to the problem proceeds along two lines: first of all, we have better consider only a subset of all possible chess positions, and more precisely only those which have a meaning for a normal chess game. This is where "heuristics" and "satisficing" come into play. No exact, optimal rule for identifying this subset exists, of course, but one can rely on the knowledge gathered by past chess experts and accumulated in a small number of rules of thumb, or general strategic guidelines. In particular, the authors compile a number of overall "goals" that a chess player must achieve, loosely speaking, in order to win. For example, "keep the king safe," "develop the pieces," "do not block your own pawns," etc., are goals of this kind. Only moves that contribute toward the satisfaction of the specified goals are considered, thus greatly reducing the size of the search. In terms of the diagram, these rules amount to a pruning of the tree along the horizontal axis. The second pruning strategy proceeds along the other axis: instead of considering the whole game—which is, on the average, 80 levels deep—we consider only the first four or five levels below the level at which the move is being played.

The effectiveness of these strategies, and especially their adequacy as a model of chess competence in particular and thinking processes in general, can of course be seriously questioned. Indeed, they have been questioned more than once.[47] But my point is different. Heuristics adds the final touch to the interaction between AI and chess that allows the closure of the feedback circle and triggers even more theoretical attention on the game. Heuristics is about chess (king's safety, etc.), and it is about reducing chess's intrinsic complexity. Chess, as a consequence, is studied more and more as a space to search on the basis of more and more sophisticated heuristic functions. This is what chess is for AI, and will ever be. But not only: a whole, concrete view of games emerges, games as the closed space that isolates it from anything outside and makes potentially meaningful whatever lies inside. On the rebound, AI is precisely what comes out of this article, as I have already suggested above: heuristic search in a search-space game-theoretically defined. "Thinking" must be explained in terms of satisficing a set of rigid constraints by searching heuristically the space that those constraints (i.e. rules) define. And the best way to provide and test a detailed theory of thinking is by writing a computer program that will effectively search such a space.

5. Conclusion.

Seen from the inside, i.e. from the standpoint of its practitioners, AI is the discipline that can finally provide a rigorous, scientific explanation of the most elusive phenomenon ever to confront scientific progress: mind itself. A popular account of the field, for example, calls it "the mind's new science."[48] Under this common interpretation, AI is the theoretical effort striving to conquer the last topic still unexplained by Western science and philosophy. Marvin Minsky, for example, called the "AI problem one of the hardest science has ever undertaken."[49] (some "ultimate frontier" rhetoric is far from absent in all popular and not so popular discussions of the topic). In this paper I have tried to show that Artificial Intelligence's scope cannot be contained within broad definition. Since the very beginning, researchers like Herbert Simon, Allen Newell, and their followers were trying to develop a general theory of human interaction centered on the activity of problem-solving as the most conspicuous manifestation of human nature. This effort, in turn brought them to develop a theory of games that, although built upon the central intuitions of classical game theory, pushed them in a new direction and brought them to see most of mental life as a sophisticated game played between the subject and the environment.

In concluding, I will say a few words on one important consequence of the interpretation suggested here. AI's ambitious research programme was bound to bring it in conflict with the discipline that had tried for centuries to provide a satisfactory account of human nature: philosophy. Researchers within the field of AI claim that their discipline is the "negation" of philosophy: they have transformed the idle, "armchair" speculations of two thousands years of philosophy into a serious, empirically-grounded scientific effort that will finally unveil the mysteries of the mind. The opposition thus built rests on two claims: (a) that AI is a science in the traditional sense, and most importantly, (b) that philosophy's inquiry into the working of the mind are (or were) a hopelessly misguided (because a-priori) scientific effort. In a nutshell, this interpretation identifies the object of traditional philosophy (or most of what goes under this name) as a narrowly construed phenomenal field, e.g. the "mind." The field could perhaps have been empirically investigated, given the appropriate intellectual and social conditions, in a proper scientific (e.g. "empirical") way. Unfortunately, either because of its basic theoretical leanings or because the times were not yet mature for real science, philosophy decided to use the wrong methodological approach and resorted to "armchair speculations" that doomed its efforts and kept the "mind" wrapped in a web of mysteries for two thousand years. According to this interpretation, AI and philosophy come to have the same content but a different form: they are (or were) both interested in providing a scientifically reliable account of the mind, but where philosophy proceeds a priori AI proceeds scientifically on solid empirical grounds.

The interpretation sketched above, and the role played by games and game playing in the construction of Artificial Intelligence's theories suggests a different answer. Artificial Intelligence, we said, can still be seen as a discipline that shares the content with philosophy but adds a different form. The meaning of this claim varies with the scope of "content." In the classical interpretation, the basic assumption is that content of philosophy, say of Descartes, can be brought over to science by a suitable change of methods because it is epistemologically equivalent to a scientific hypothesis on a par with those used in the empirical sciences. However, if the subject matter of philosophy, its content, is more properly understood, as truth in the broadest sense, then the meaning of the claim that philosophy has the same content as AI changes radically because it forces us to reverse the sense of the relationship between AI and philosophy.

Far from being a scientific reduction of raw ideas provided by philosophy, AI is an attempt, and cannot be anything but the attempt to provide a metaphysics with non-philosophical means. Artificial Intelligence is then a non-philosophy in the technical sense of determinate negation suggested above: it stands in a determined opposition to philosophy because it receives from the former the full scope of its inquiry, but it rejects the philosophical "tactics," in Dennett's words, i.e. the a-priori investigations that have always characterized the philosophical demarche. Engineering techniques, i.e. the implementation of computer programs, fill up the gap that is thus left open: they are the right tools for the job, so to speak, since they allow AI to answer its questions synthetically, instead of proceeding a-priori. Thus, the crucial relevance of engineering techniques within a generally non-engineering context that I emphasized above, is explained. The production of artifacts is only secondarily relevant to Artificial Intelligence, while the methodological reliance on synthetic methods in service of "philosophical" goals is all that matters.[50]

Stefano Franchi
University Of Auckland

Notes

1. Artificial Intelligence's birth date is usually taken to be 1956, when a summer school at Dartmouth organized by John McCarthy and Claude Shannon gathered for a few weeks the researchers who were working on the mechanization of intelligence: Marvin Minsky, Allen Newell, Herbert Simon, Claude Shannon, John McCarthy, A. Samuel, etc. See John McCarthy, Marvin Minsky, Nathaniel Rochester and Claude Shannon "A Proposal for the Dartmouth Research Project on Artificial Intelligence," available at *URL:http://www-formal.stanford.edu/jmc/history/dartmouth.html*, and Pamela McCorduck, Machines Who Think (San Francisco: W. H. Freeman and Company, 1979) 93-114. For reasons that will be apparent in what follows, my genetic reconstruction will be focused on what happened in the decade that led to the Dartmouth seminar.

2. On checkers see A. L. Samuel, "Some Studies on Machine Learning using the Game of Checkers, *The IBM Journal of Research and Development*, 3, July 1959, 211-229, later in Edward Feigenbaum and Julian Feldman, *Computers and Thought* (San Francisco: McGraw Hill, 1963) 71-105. Interest in checkers has recently been revived by the exploits of a team based at the University of Alberta, Canada, whose program, Chinook, is the first world Human-Machine champion (i.e. it

won the title in a tournament open to both "human and non-human forms of intelligence"). The literature on computer chess is immense. For a brief synopsis on the history of checkers playing programs, see L. Stephen Coles, "Computer Chess: The *Drosophila* of AI," *AI Expert*, 9, 4 (April 1994) 25-32. The first studies on chess-playing machines were by Turing and Shannon, who provided the basic framework, later substantially enriched by Herbert Simon, who wrote the first actually running program in 1955 (with Allen Newell, and Cliff Shaw), first at the Rand Corporation and later at Carnegie Mellon; see Allen Newell, J.C Shaw, and Herbert Simon, "Chess-Playing Programs and the Problem of Complexity," *The IBM Journal of Research and Development*, 2, October 1958, 320-335, also in E. Feigenbaum and Feldman *Computers and Thought...* , 39-70. Computer chess has now become almost a sub-discipline of computer science, with its specialized journal, conferences, etc., although it does not enjoy any longer the "exemplary status" that was its own. For an early history of the development of machine chess see Allen Newell and Herbert Simon, *Human Problem Solving* (Englewood Cliffs, NJ: Prentice Hall, 1970) 663-707; more recently, David E. Welsh, *Computer Chess* (Dubuque, Iowa: W.C. Brown Publishers, 1984); for a recent assessment of the possible evolution of the field see Mikhail Donskoy and Jonathan Schaeffer, "Perspectives on falling from grace," T. Anthony Marsland and Jonathan Schaeffer, eds., *Chess, Computers, and Cognition*, (Berlin: Springer-Verlag, 1990), and Herbert Simon and Jonathan Schaeffer, "The Game of Chess," *Handbook of Game-Theory with Economic Applications*, Robert J. Aumann and Sergiu Hart, eds., (Amsterdam: Elsevier Science Publishers, 1992) vol. 1, 1-17. The Hanoi tower gathered attention only later: in 1958 Simon learned of its use in experimental psychology and started to work on it, with Allen Newell, to test heuristic search capabilities of their early programs. See Herbert Simon, *Models of My Life* (New York: Basic Books, 1991).

3. Alan Turing's insights on computers playing chess and his hand simulation are presented in Bertran Bowden, *Faster than Thought, a Symposium on Digital Computing Machines* (London: Pitman, 1953) 288-295.

4. Herbert Simon, one of those most responsible for chess's popularity, used the analogy widely since the early 1960s, when AI's interest in chess was still at its peak. He reports that he used it routinely in the question and answer sessions after his talks, to defend the study of chess and other games as a worthwhile research-project. See, for example, Herbert Simon, *Models of My Life…*, 327. In a private communication, Simon reported that he is not sure whether he or Allen Newell actually coined the analogy. According to John McCarthy (personal communication), the analogy itself might have been first proposed by Alexander Kronrod, then the head of the group that wrote the Soviet chess program at the Institute of Theoretical and Experimental Physics in Moscow in 1965-68. Kronrod, however, might have gotten it from Herbert Simon, as McCarthy acknowledges. A rather elaborate defense of chess as a research tool is contained in Allen Newell, J.C Shaw, and Herbert Simon, "Chess-Playing Programs…" The metaphor is still actively used, in spite of chess' smaller role in contemporary AI. Some researchers in AI have suggested that although the theoretical impulse of chess within AI is all but exhausted, the game of go will soon become AI's new *Drosophila*.

5. Thus, AI provides the theory of games that game theory is not because its necessary focus on actual game-playing behavior forced it to abandon from the start the assumption of optimal rational behavior that is (or has been until very recently) a constitutive part of traditional game theory. I will come back to a fuller discussion of this point in section 3 below.

6. It will be enough to remember here J-F. Lyotard's claim (in *La condition postmoderne* (Paris: Seuil, 1979) 16) that a theory of games and agonistic behavior is what is needed in order to reach an understanding of the postmodern condition; or the many remarks by the early Derrida (not only in "L'écriture, le signe et le jeu dans le discours des sciences humaines," but also in *De la Grammatologie* and *La pharmacie de Platon*: "Jeu est la cyphre de l'ouverture anti-dialectique par excellence"); Michel Foucalt's essential reference to the "agonal character" of the interplay between power and fredom (in *Le sujet et le pouvoir*), Heidegger's remarks in *Der Satz vom Grund*, Lévi-Strauss's insistence on games throughout its work, etc.

7. See for example Bruno Latour and Steve Woolgar, *Laboratory Life: the Construction of Scientific Facts* (Princeton: Princeton UP, 1986);Joan Fujimura, *Crafting Science* (Cambridge: Harvard UP, 1996); Joan Fujimura and Adele Clarke, eds., *The Right Tools for the Job: At Work in Twentieth Century Life Sciences* (Princeton: Princeton UP, 1992).

8. Robert Kohler, *Lords of the Fly* (Chicago: Chicago UP, 1994).

9. Kohler, for example, reports Morgan's desperation when, after having reworked the Mendelian formula for the third time in two years of intensive labor, yet another eye-color mutant popped up unexpectedly. See Morgan and Bridges' report quoted in Robert Kohler, *Lords of the Fly*..., 60-61.

10. Robert Kohler, *Lords of the Fly* 66.

11. Artificial Intelligence is often defined as the discipline aiming at the construction of artifacts whose observable behavior, in some relevant domain, would be considered intelligent if performed by a human being. If this were the case, then AI would just be a branch of engineering that does not require any deep understanding of intelligence, nor of its (possibly psychological, possibly physiological) processes underlying it. The additional assumption shared by several (but by no means all) researchers in field, and by all the founding fathers, is that AI will succeed because its artifacts will implement a more or less complete theory of human rationality. In other words, AI artifacts will not just emulate intelligence, they will explain it. AI thus requires a definition of "intelligence" and 'intelligent behavior" and it is by no means obvious that playing chess is such a perspicuous example of such a behavior that a theory of chess-playing would qualify as a theory of intelligence. Recent developments in AI, in facts, have vigorously denied this last assumption: see, for instance, the nouvelle AI perspective as presented in Rodney Brooks, *Cambrian Intelligence* (Cambridge, Mass.: MIT Press, 1999).

12. Suffice here to remember that Herbert Simon and Allen Newell began working on chess around 1952 at the RAND Corporation and traveled to the founding event of Artificial Intelligence, the 1956 Dartmouth summer seminar, well armed with ideas and first results. Claude Shannon, himself one of the four organizers of the seminar, is responsible for the first introduction of chess into computer science, although from a different perspective than Simon's and Newell's.

13. Roger Callois, *Les jeux et les hommes* (Paris: Gallimard, 1958), and Elliott Avedon and Brian Sutton-Smith, *The Study of Games* (New York: J. Wiley, 1971) provide a broad overview of games and play activities in general.

14. The history of game theory is actually richer, and it includes, at least, the contributions by the

French mathematicians Emile Borel in the 1920s and 1930s. Borel provided the first clear definition of pure and mixed strategies and proved the minimax theorem for some limiting cases. However, it is John (then Johann) von Neumann who proved the general form of the theorem in 1928, while still at Göttingen. Unfortunately, the historical literature on game theory is not that abundant. A detailed reconstruction of a immediate context preceding and following the publication of von Neumann and Morgenstern's major work in 1944 is provided by Robert J Leonard in "Creating a Context…," 29-76. See also Johann von Neumann "Zur Theorie der Gesellschaftsspiele," *Mathematische Annalen*, 100 (1928), 295-320. See also Émile Borel, "La théorie du jeu et les équations intégrales à noyau symétrique gauche," *Comptes Rendus de l'Académie des Sciences*, 173 (1921) 1304-1308; "Sur les jeux où interviennent l'hasard et l'habilité des jouers," *Elements de la théorie des probabilités*, (Paris: Librairie Scientifique, 1924). For a discussion of Borel's and von Neumann's different perspectives on games see Luca Dell'Aglio, "Divergences in the History of Mathematics: Borel, von Neumann and the History of Game Theory," *Rivista di Storia della Scienza* 3.2 (1995) 1-45.

15. In fact, it has been argued that von Neumann and Morgenstern emphasis on strategy vs. Moves in *Theory of Games*, and their shift away from minimax interpretation was precisely an attempt to move away from the psychological dimension generating the "circularity of the 'I think he thinks that I think' logic of strategically interdependent situations"; Andrew Schotter, "Oskar Morgenstern's Contribution to the Development of the Theory of Games," Roy Weintraub (ed.), *Toward a History…*, 106.

16. Ernest Zermelo "Über eine Anwendung der Mengenlehre auf die Theorie des Schachspiels," *Proceedings of the Fifth International Congress of Mathematicians*, 1913, vol. 2, 510-504. Zermelo was another mathematician of the Hilbert school, like von Neumann.

17. Robert Byrne, "Karpov Adjourns Play in a Drawish Endgame," *The New York Times*, Tuesday, July 9, 1996, B2.

18. Yasunari Kawabata, *The Master of Go* (New York: Knopf, 1972).

19. Alan Aycock, "Derrida/Fort-da: deconstructing play," *Postmodern Culture*, 3, 2, January 1993, contains one of the few anthropological analysis of chess that examines the concrete event of chess playing and offers a very different perspective on rules perfect closure, etc. See alsoAlan Aycock, "Finite Reason: A Construction of Desperate Play," Play and Culture 5, 2, (1992)182-208 for additional data on the multiplicity of chess scenarios (tournament play, *Blitzkrieg* games, etc.) and the different conventions enforced by the players in the different occasions. More general critiques of the conception of rule outlined above can be found in the present debate on Artificial Intelligence, its ambitions and shortcomings. See for example H.M. Collins, *Artificial Experts, Social Knowledge and Intelligent Machines* (Cambridge: MIT Press, 1990).

20. Note that this extremely static interpretation of the rules by von Neumann and Morgenstern is often attributed to an excessive reliance of their theory on real games, whereas they should have paid more attention, it is argued, to the economic behavior that the theory allegedly models. See, for example, Philip Mirowsky "What Were von Neumann And Morgenstern Trying To Accomplish?" *Toward A History Of Game…*, 141. In other words, "mere games" like chess have fixed rules, whereas "real games" like war are subject to continuous negotiations. Instead, I think it

will be progressively clearer that, if a criticism can be raised, is that they did not pay enough attention to "mere games as they are actually played" as John McCarthy recently confirmed when commenting upon Deep Blue's defeat of Kasparov: "computer chess has developed much as genetics might have if the geneticists had concentrated their efforts starting in 191o on breeding racing *Drosophila*. We would have some science, but mainly we would have very fast fruit flies." See John McCarthy, "AI as Sport," revue of Monty Newborn, *Kasparov versus Deep Blue. Computer Chess Comes of Age* (New York: Springer Verlag, 1996), *Science*, 6/6/1997, emphasis in the original.

21. The first game for which von Neumann and Morgenstern presents a a "normal" (i.e. matrix-like) representation is actually "Matching pennies." In this game, the two players agree that one will be "even" and the other will be "odd." Each one then shows a penny. They are shown simultaneously, and each player may show either heads or tails. If both show the same side, then "even" wins the penny from "odd;" if they show different sides, "odd" wins the penny from "even. Since this game has only one move, the difference betwenn extensive (i.e. tree-like) and "normal (matrix-like) representations is purely esthetical. In fact, it is a game in which the usual concept of strategy does not apply, since each player has no planning to do. See John von Neumann and Oskar Morgenstern, *The Theory of Games and Economic Behavior* (Princeton: Princeton UP, 1944) 94, 111. Hereafter referred to as TGEB.

22. The label "normal" is a clear enough indication of which representation von Neumann and Morgenstern thought the correct one. Notice that current standard terminology in game theory has abandoned this approach and prefers to talk of strategic games vs. extensive games, i.e. of games in which "each player chooses his plan of action once and for all, and all players' decisions are made simultaneously" versus games that specify "the possible orders of events" (Martin Osborne and Ariel Rubinstein, A Course in Game Theory (Cambridge: MIT Press, 1994) 3.) Recent work in game theory has indeed questioned the assumption of perfect rationality that underlies the very concept of strategy in a game represented in normal (or strategic) form. Kenneth Binmore goes as far as to claim that serious progress in game theory has been hampered by the lack of serious modeling of the players,. He then proceeds to recover Herbert Simon's distinction between substantial and procedural rationality (i.e. the distinction between "optimizing" and "satisficing" that I will discuss below) in the context of contemporary game theory. Much game-theoretical work on "bounded rationality" (another, and perhaps "the" Simonian concept) proceeds along similar lines. See Kenneth Binmore, "Modeling Rational Players" I and II, *Essays on the Foundations of Game Theory* (Cambridge, Mass.: Blackwell, 199) 155-231; and the special 1994 issue on bounded rationality of the *Journal of Institutional and Theoretical Economics*.
However, even unorthodox game theorists like Binmore accept the idea that although optimal rationality cannot be assumed, it will eventually be reached at the end of a learning process. See Kenneth Binmore and Larry Samuelson, "An Economist's Perspective on the Evolution of Norms," *Journal of Institutional and Theoretical Economics*, 150, 1 (1994), 45-63, where the authors claim that "although we deny that real people are natural gamesmen, it does not follow that we think that the optimizing paradigm that has served economists so well should be abandoned. Experimental work [by Binmore, ndr] shows that people can and do find their way to one of the equilibria of the game by trial-and-error learning provided that:
- The problem they have to solve is not too complicated
- Adequate incentives are provided

- They are allowed to play the same game many times (against a new opponent each time" (46). Thus, the path toward a solution to the problem of bounded rationality is: learning will teach common people to reach their game-theoretic equilibria. (Cf. the debate in early AI between search and learning).

23. The more general form of the minimax theorem for two-person zero-sum games is essentially what von Neumann proved in his 1928 paper. The result had already been shown to hold for chess, (which is an instance of a class of games providing an additional constraint, i.e. perfect information) by Ernst Zermelo, in 1913, in "Über eine Anwendung…" as reported above.

24. In fact, chess enjoys an almost paradoxical position in Game theory. Zermelo proved that it is solvable in principle, i.e. he proved that there must be a strategy that guarantees at least a draw either to black or to white. However, no one knows neither the value of this solution nor the strategy in question. "Furthermore—as a game theorist stated— were the solution known, watching a chess game would be about as interesting as watching paint dry: and nobody would play chess at all if it were known to how to play it optimally." But of course people do play chess, and they do find it interesting. Conclusion? Game theory has basically nothing to say about chess, except exploiting it as a typical example of a problem that it cannot solve. See Kenneth Binmore, *Fun and games: a text on game theory* (Lexington, Mass.: D.C. Heath, 1992).50-51.

25. Or, to say it in game-theoretical jargon, "assign a value to each one of the terminal nodes and then minimax your way back to the beginning." See TGEB 143-165.

26. TGEB 125, my emphasis.

27. The two steps are in fact closely linked in TGEB, since the programmatic strategy is to treat a n-person non-zero sum game as a n+1-person zero-sum game. This strategy has been questioned in the following development of game theory. See for example TGEB, chp 11 and Andrew Schotter, "Oskar Morgenstern's contribution to the development of the theory of games," in Roy Weintraub, ed., *Toward a History ....*".

28. The most well-known debate in the years immediately following the publication of TGEB is the discovery, by Arnold Tucker, of the famous Prisoners' dilemma, an example of a two-person, incomplete information game with no equilibrium point that was bound to become one of the paradigmatic "thought experiments" in game-theory. In spite of being formally called a "game," however, it is doubtful whether the Prisoners' dilemma should be called a "game" in the common sense of the word. See H. Kuhn and A. W. Tucker, *Contributions to the Theory of Games*, vol. 2, (Princeton: Princeton University Press, 1953) and the popular work by Richard Powers, *Prisoner's Dilemma* (New York: Beach Tree Books, 1988). A brief historical recollection on the invention of the dilemma can be found in Howard Raiffa," Game Theory at the University of Michigan, 1948-1952," *Toward a History*, Roy Weintraub, ed., 171-173.

29. Martin Shubik "Game Theory at Princeton, 1949-1955: A Personal Reminiscence," Roy Weintraub, ed., *Toward a History of Game Theory*, 152.

30. Roger A. McCain, "Lecture notes of a class in Economics and Game Theory," URL: *http://william-king.www.drexel.edu/top/class/game-toc.html*.

31. Note that, in game-theory, this tends to be case still now. Robert Aumann and Sergiu Hart, for example, are editing a monumental three-volume work, *Handbook of Game Theory with Economic Applications* (New York: North Holland 1991-forthcoming) which opens with a chapter on chess co-authored by Herbert Simon and Jonathan Schaeffer. The contribution is presented as follows: "Historically, the first contribution to Game Theory was Zermelo's 1913 paper on chess, so it is fitting that the "overture" to the Handbook deals with this granddaddy of all games. The chapter covers chess-playing computers. *Though this is not mainstream game theory*, the ability of modern computers to beat some of the best human chess players in the world constitutes a remarkable intellectual and technological achievement, *which deserves to be recorded in this handbook*." (iv, my emphasis).

32. See Philip Mirowski, "When Games Grow Deadly Serious…," for a detailed analysis.

33. Most of the early attempt are discussed by Newell and Simon in *Human Problem Solving*, 671-678. One of the first real programs to play chess was programmed at the Los Alamos labs in 1956. See J. Kister, Peter Stein, Stan Ulam, W Walden, and M Wells, "Experiments in Chess," *Journal of the ACM*, 4 (1957) 174-177.

34. This may explain why neither Shannon nor Turing thought so important to actually write the programs; also, it may explain why the crucial concept of heuristics rules, the technique that would allow a dramatic improvement in the programs' performances, was missing, in fact was the only missing element, from Shannon's design principle. Claude Shannon, "A Chess-Playing Machine," *Scientific American*, February, 182 (1950) 48 ff., later in James Newman ed., *The World of Mathematics* (New York: Simon and Schuster, 1956) vol. 4; Claude Shannon, "Game-Playing Machines," *Journal of the Franklin Institute*, 260, 6 (1955) 447-453; Both essays are now available in Claude Shannon, Collected Papers, N. J. A. Sloane and Aaron D. Wyner, eds. (New York, NY: IEEE Press 1993).

35. On this issue see Alison Adam, "Constructions of Gender in the History of Artificial Intelligence," *IEEE Annals of the History of Computing*, 18, 3 (1996) 47-53 and id. *Artificial knowing : gender and the thinking machine* (New York : Routledge, 1997).

36. Claude Shannon, "A Chess-Playing Machine...," 2132.

37. However, the problems (a) and (c) will prove very important in later development of AI. The first one stands for the whole issue of knowledge representation, e.g. the problem of how to translate the relevant information into an efficient set of data structures that a computer can manipulate. The second problem, in turns, can be seen as pointing toward finding an adequate computer language to express the program itself, an issue that has kept AI busy for many years to come, from the development of IPL, the ancestor of list-processing languages that Simon, Newell, and Shaw invented, to the development of LISP by John McCarthy. Thus, Shannon's choice of the search problem (i. e. (b)) as the most important one might seem to reduce Artificial Intelligence to the problem solving activities that became the specific characteristics of the "Carnegie" school of AI, i.e. the school led by Simon and Newell. But in fact, its choice makes clear the different epistemological relevance of the problems at hand. Problem (b), searching in a state-space, is conceptually prior to both (a), representation of the states, and (c), efficient processing, as Shannon

immediately sees. It is precisely because searching turns out to be the problem, and a much harder problem that early AI had thought, that expressive representation of knowledge and efficient processing become important. Marvin Minsky, who, although close to Shannon in the years under consideration, became known for his work on “frames,” e. g. a knowledge representation technique conceptually belonging to problem (a), was very much aware of this fact. In an influential article published in 1960, “Steps toward Artificial Intelligence,” Marvin Minsky argues that search is the basic approach, sound but inefficient. The other fields of research are basically attempts to reduce the massive inefficiency of thorough searches through either an appropriate reduction of the search space (planning, learning) or an adequate improvement of the navigation through the search space (pattern-recognition). See Marvin Minsky, “Steps toward Artificial Intelligence,” Edward Feigenbaum and Julian Feldman (eds.), *Computers and Thought* (New York: MacGraw-Hill, 1963) 406-450. This essay had been circulating in draft form, as a technical report, since late 1956, e.g. immediately after the Dartmouth seminar and was instrumental in providing a first organization of the field. Minsky has repeated, and in fact broadened, his claim in Society of Mind where he calls the possibility of space-searching the “puzzle-principle that is philosophically basic to AI since it establishes the possibility of a creative machine insofar as it guarantees the existence of a solution that the machine might find but the programmer does not know about. See *Society of Minds*, 73-74. For a history of the development of LISP see Herbert Stoyan, “Early LISP History,” available at URL: *http://www8.informatik,uni-erlangen.de/html/lisp/histlit1.html*, and Herbert Stoyan, LISP-Anwendungsgebiete, Grundbegriffe, Geschichte, (Berlin: Akademie Verlag:,1980).

38. Shannon was actually pessimistic about the size of chess’s search space. Its size is now generally taken to be of the order of O(1044). See Jonathan Schaeffer, *Experiments in Search and Knowledge*, Ph.D. Dissertation, University of Waterloo, 1986; Jonathan Schaeffer, *One Jump Ahead* (New York: Springer Verlag, 1997); and Jonathan Schaeffer, Joseph Culberson, Noeman Treolar, Brent Knight, Paul Lu and Duane Szafron, “Reviving the game of checkers” D.N.L. Levy and D.F. Beal, eds., *Heuristic Programming in Artificial Intelligence*; *The Second Computer Olympiad* (London: Horwood, 1991) 119-136. Checkers’ search space is approximately 5X1020, with approximately 10^{18} legal positions, i.e.. states).

39. Herbert Simon, *Administrative Bbehavior* (New York: MacMillan, 1947).

40. Herbert Simon, “review of *Theory of Games and Economic Behavior*,“ *The American Journal of Sociology*, 1945.

41. See for example Simon’s contribution to the Santa Monica conference in 1952, now in Robert Thrall, ed., *Decision Processes* (New York: Wiley, 1954). See also his comment on those years in an interview to Vernon Smith: “I was profoundly dissatisfied with the concept of “solution” in von Neumann and Morgenstern—it seemed to me to confirm the complexity of the problem rather than solve it.” in Vernon L. Smith, “Game Theory and Experimental Economics,” Roy Weintraub, ed., *Towards a History*, 253.

42. George Polya, *How To Solve It* (Princeton: Princeton UP, 1944). Polya writes in the preface to his celebrated book: “This study [of solving methods], that some authors call heuristics, now is out of fashion. However, it has a glorious past and, perhaps, a future.” (iii).

43. Herbert Simon, *Models Of My Life*,166 and 202.

44. Allen Newell, J.C Shaw and Herbert Simon, “Chess-Playing Programs…,” 39.

45. Note, also, that this approach marks quite a distance from the current standard interpretation of intelligence as whatever behavior can fool a human being into thinking that s/he/it is human, e.g. the so-called Turing test. In its early phase, AI was explicitly much more ambitious, since it was not just the outcome of the intelligent process counted, but also, and especially, the process itself.

46. See, for example, Newell and Simon's reservations, along these lines, against Turing hand-simulation of his chess-playing programming in *Human Problem Solving*, 671 ff.

47. See, e.g. Hubert Dreyfus’s books and especially Mind over machines (New York: The Free Press, 1986) where an alternative model of human expertise is presented (and later attributed to the co-author of the book, Stuart Dreyfus). It should be noted that most researchers in AI (including its designers) agree that Deep Blue uses a fundamentally simpler approach closer to Shannon’s original proposal than to Simon and Newell’s more sophisticated goal-based heuristics, and attribute most of the merit for its victory over Kasparov to the sheer power of the hardware and the patient work of the engineers who tuned the software in order to take full advantage of that raw speed. Herbert Simonb, however, disagreed and claimed that Deep Blue was really using heuristic reasoning, in spite of IBM’s understandable and quite successful marketing rhetoric. Fora detailed technical discussion see Murray Campbell, A. Joseph Hoane, Jr., and Feng-hsiung Hsu “Deep blue,” *Artificial Intelligence*, 134(1-2):57–83, 2002. For a recent debate on what Deep Blue’s victory really meant for AI’s prospects see Hubert Dreyfus and Daniel Dennett “Did *Deep Blue’s* win over Kasparov prove that Artificial Intelligence has succeeded? A debate,” in Stefano Franchi and Güven Güzeldere, eds., *Mechanical Bodies, Computational Minds* (Cambridge, Mass.: MIT Press, 2004).

48. For example by Howard Gardner in his well-known book: *The Mind’s New Science: A History of the Cognitive Revolution* (New York: Basic Books, 1985).

49. From an interview contained in Gina Kolata, “How Can Computers Get Common Sense?” *Science*, 217 (1982).

50. This conclusion about the philosophical status of Artificial Intelligence agrees, although it has proceeded along very different lines, with the famous diagnosis about Cybernetics pronounced by Heidegger in the “The End of Philosophy and the Task of Thinking” and elsewhere. In the *Der Spiegel* interview, for example, Heidegger claims that, in the modern era, metaphysics will achieve its long standing goal by dissolving into the sciences, for example into psychology, logic, political science. The role of philosophy, i.e. the role of an overarching investigation that, similarly to old-fashioned philosophy, harmonizes the results of the sciences into a global vision, will be taken up, according to Heidegger, by Cybernetics, the “technological” (i. e. engineeristic) science of control, or the science that “corresponds to the determination of man as an acting social being.” The difference here is that Heidegger stresses the continuity between philosophy and cybernetics by interpreting the latter as the final stage of the history of metaphysics, while my interpretation emphasizes, instead, the dis-continuity between the two. Thus, although I agree with Heidegger on the philosophical assessment of Cybernetics (or of Artificial Intelligence, since the difference does not really matter in this context) I would need a Hegelian *Aufhebung* to bridge the dis-continuity between philosophy and non-philosophy and bring my interpretation in tune with Heidegger. And I

am not sure that such a move would be granted by the context of this analysis. See Martin Heidegger, “Only A God Can Save Us,” transl. Maria Alter and John Caputo, *The Heidegger Controversy*, Richard Wolin, ed., (New York: Columbia University Press, 1991) 107ff. and "The End of Philosophy and the Task of Thinking," *Basic Writings* (New York: Harper and Row, 1977) 376.
