人工知能の論理と哲学

ジョン・マッカーシー

人工知能のゴールは、知能を必要とする問題を解き、目標を達成することにおいて、人間よりも有能な機械であります。有用な結果が得られてはいるのですが、最終的なゴールは、まだ大きな概念的進歩を要求しており、その達成は、はるかに遠いものと考えられます。

ゴールに攻め込む道は三つあり、第一の道は人間の神経システムをまねること、第二の道は人間の知能の心理学を研究すること、そして、第三の道は人々がその中で目標を達成しようとしているような常識の世界を理解して、知能的な計算機プログラムを作ることであります。この第三の道が計算機科学としての接近方法であって、これが現在最も成功している方法であり、これについてこの講演の中で議論したいと思います。

人工知能研究者の間でも、機械が世界に関して知っていることを表現するために数理論理の言葉を使うことをどの程度重く見るべきかについて意見が異なっています。手短に言えば、事実については互いに独立に、またそれらが使われることとも独立に学習することができるという利点を持っています。反対の意見としては、人間の推論はそんなに論理的でないと思われる点についてであります。現在のエキスパートシステム技術では、その知識の大部分を表現するために論理的な言語が使われていますが、それ以外に多くのことが計算磯プログラムの中に埋め込まれており、また論理式の並べ方についても多くのことがなされています。私はこの講演の中で論理学的接近の方法を強調したいと考えております。そしてこの機会にふさわしいように私自身の研究も強調したいと思います。

1948年9月に、行動における大脳のメカニズムに関するヒクソン・シンポジウム（Jeffress 1951）のいくつかの分科会に出席したときから、人工知能に興味を持つようになりました。それは、カリフォルニア工科大学で開催されたものであり、また、私が数学科の大学院で勉強を始めたばかりのときでありました。このシンポジウムには、数学者ジョン・フォン・ノイマン（プログラム貯蔵式電子計算機の発明者の一人）、ワレン・マッカロッチ（仮想的神経網が計算の

ためにどのように使われるかについて話していた）と著名な心理学者、神経生理学者が参加していました。フォン・ノイマンの講演の題は、“オートマトンの一般的論理学的理論”でありました。人間の脳に関してわかっていることと、電子計算機の新しい考え方との比較について、たいへん興味をひかれました。この時期には、プログラム貯蔵式電子計算機は建設中でありまして、1号機の完成はその翌年になります。

最近、このシンポジウムの議事録を読みなおしてみましたが、私の記憶は誤っておりまして、知能的計算機を作る試みについては、何も触れられておりませんでした。すべての著名な出席者は、生物学、心理学、および、数学の在来型の研究についてのみ語っていました。フォン・ノイマンは、後にさらに十分発展した二つのアイデア、信頼性の低い部品からいかにして信頼性の高い計算機を作るかという話と、自己増殖機械をどのようにして作るかについて話しています。彼は数理論理学について重要な研究を行った人ですが、このときも、そしてその後も、世界についての事実を論理式で表現することについては何も議論しておりません。彼は計算機の開発には積極的に参加していましたが、計算機が知的に振る舞うためのプログラムの作り方については議論していません。いずれにいたしましても、この会議は私の学問的野心を人工知能へ向かわしめたものでありました。

私は1948年に人工知能について考え始めたわけですが、計算機のプログラムの作り方や、事実を論理式で表現することを考えたわけではありません。私は論理学についてもまた計算機についてもあまり知りませんでした。そのかわり、私は、後に袋小路であったと思うようになった、二つの考え方を追っていました。一つは、オートマトン理論と情報理論とを結びつけようとするもので、この二つの理論は当時新しくて流行的なものでありました。脳は環境に結びつけられた有限オートマトンであり、環境もまた有限オートマトンであると考えていました。環境が何であるかについて脳がはっきりしていないという事実を表現するために、私は環境オートマトンのアンサンブル（確率のついた集合）を考えていました。情報理論はこのアンサンブルに適用され、脳が最初に環境に結びつけられた時である時刻０の時にエントロピーを定義することが許され、その後システムが作動してしばらく経ってから、脳の状態はアンサンブルからどんな環境が選ばれているかに部分的に依存していることになります。これらの時刻のエントロピーの差が、どれだけ脳が環境について学んだかの物差しに

なります。

その後私はこれがまずい考え方であると信ずるようになり、発表することを試みませんでした。困難な点は、世界についての固有の事実を環境のアンサンブルを使ってどのように表現してよいのかわからなかったことであります。後に得られた用語を先回りして使うと、このような表現方法は認識論的に不適切であったということであります。それは実際に得られる情報を表現できていませんでした。

私のもう一つの良くないアイデアは、より数学的なものでありました。問題が何であるかということ、すなわち知的な機械に解いてもらいたくなるような問題は何かということであります。私は“はっきり定義された問題”は、提起された解をテストするための方法によって与えられることを提案しました。それは良い考え方であって、一つの形、または違った形で存在し続けたのですが、それは明白すぎることであったかもしれません。提起された解をテストする方法は明確に定義されねばなりませんから、チューリング機械で与えることにしました。チューリング機械は1936年に計算機の抽象的な形としてアラン・チューリングによって導入されたものであります。従って、問題解決はチューリング機械によって定義される関数の逆関数を求める問題とみなすことができます。1952年にクロード・シャノンの下で夏休みに働いている間にこのことについて論文（マッカーシー 1956）を書きました。そしてオートマタ・スタディーズ（シャノンとマッカーシー 1956）に入れていただくことにより発表されました。困難点は、やはり認識論的に不適切であることでした。

プログラム貯蔵式計算機が、人工知能の展開に適切な手段であるか否かは、私にとって、まだそれほど明白なことではありませんでしたし、また、オートマタ・スタディーズのどの著者にとってもそうでした。

このことを最初に明確に理解した人は、アラン・チューリングでした。彼の論文“計算機と知能”（1950）は哲学誌に発表され、エドワード・R・ニューマンによって“数学の世界”（1956）に再度載せられるまで知られておりませんでした。私も独自に同じアイデアに到達しましたが、アレン・ニューエル、ハーバート・サイモンもそうでした。おそらくニューエルが計算機が知的な振る舞いをするようにプログラムを作成することを、生涯の仕事とした最初の人でありましょう。チューリングは1954年、計算機が人工知能に貢献しうるようになる前に死去しました。

私の計算機との出会いは、1955年、IBMのナサニエル・ロケスターが、パキプシーにある彼の先駆的な情報研究部に夏の間雇ってくれたときのことでした。私はIBM702のプログラムを作ることを学び、人工知能に対する私の考え方を知能的な計算プログラムを作ることに向けてくれました。私は当時ダートマス大学数学科の助教授でありましたが、1956年に人工知能に興味を持つ人々をダートマスに招待する夏の研究プロジェクトを行うのに、十分なだけ人工知能には興味があるということが、その夏わかったような次第でした。そんなにたくさんの人々を招待できたわけではありませんでしたが、興味を持ってほしいと望んでいた人々を招待することができました。クロード・シャノン、マービン・ミンスキー、オリバー・セルフリッヂ、ナサニエル・ロケスターがロックフェラー財団への提案に協力してくれました。そして、旅費、生活費等にあてる資金7,500ドルが与えられました。

このときに我々のテーマを決めねばなりませんでしたが、私は“人工知能”を選びました。このような大胆な題名を選んだのは、参加者に望ましい影響を与えたいと思ったからです。シャノンと一緒に、オートマタ・スタディーズのために論文を集めたのですが、そのとき私は知能的オートマタを作るためには周辺的な部分にしか関係していないような、オートマタの数学的性質を扱っている論文が多すぎるように思いました。“人工知能”という名前は、帆柱に掲揚された旗でありました。多くの人々はこの名前に批判的でしたが、人々に彼らの現在のプロジェクトと野心的な長期的ゴールとを比べさせるような名前を持つことが、過去においても、また現在でも重要なことであると私は考えております。

ダートマスにおいて発表された論文のうち私にとって最も興味深く思えたのは、ニューエルとサイモンの“数理論理機械”（Logic Theory Machine）でありました。彼らのプログラムは、普通の学生が初等論理学の問題を解いていくプロセスをまねるものでした。このプログラムの作成過程でIPLリスト処理言語を開発し、また再帰的サブルーチンを初めて導入したのです。この他の興味あるプロジェクトとしては、アーサー・サムエルのチェッカーのゲームのためのプログラム、アレックス・ベルンシュタインのチェスのためのプログラムがありました。これらはIBMのプロジェクトでした。マービン・ミンスキーは平面幾何の定理証明機を提案し、計算機に“図形”を記憶させておいて、どの命題が証明を試みるのに十分価するかを決めることに使うようにすべきであると言っ

ていました。私自身は、事実を証明するために論理を使うことを考え始めていましたが、ダートマスにおける会合で発表するに十分なものは持っていませんでした。

この会合の後、ロチェスターは幾何用定理証明機に関するミンスキーのアイデアのプログラムを作ることを決め、ヘルベルト・ゲレルンターをプログラム作成のため、また私をコンサルタントとして雇いました。私はリスト処理言語に関するニューエル・サイモンのアイデアに深い印象を受けておりましたが、しかし、彼らの使った特定の形式的枠組みが私に訴えるものは何もありませんでした。IBMのプロジェクトであるFORTRANのほうがより魅力的な形を持っていました。そこで私は、ゲレルンターに対して、幾何のためのリスト処理をFORTRANで作ることを提案し、いくつかの基本関数――LISPのcarとcdr――を提案しました。ゲレルンターとカール・ゲルベリッヒがそれを作るとともに、さらにいくつかの関数をつけ加えました。特にconsを関数として作り、従って合成することができるようにしました（このことは、当時必ずしも明白なことではなかったと、あなた方に納得してもらえるような知的情況を説明するスペースを、私は持ち合わせておりません）。

1957年秋にMITに行くまでは、私はフルタイムの計算機学者ではありませんでした。そのころまでには、私は論理が人工知能に対する鍵となる一つの形式であると確信するようになっていました。ニューエルとサイモン（そして間もなく他の数人の人々）は、論理を実行するプログラムをすでに作り上げていましたけれど、誰一人その主題の領域として考える以外は、論理を使うことに興味を持ってはいませんでした。私の論理への転向は、1958年の論文“常識を持つプログラム”（マッカーシー 1959）に述べてあります。

1958年夏を再びロチェスターのIBMグループと過ごしましたが、そこで、ゲレルンターとゲルベリッヒは、平面幾何のプログラムを完成しようとしていました。私は、FORTRANが再帰的リスト処理関数を良い方法で表現していないと確信するようになり、LISPを定式化する研究を始めました。1958年秋にMITに帰ったとき、ミンスキーと私とには、大きな研究室と、キーパンチ機1台と一人の秘書、二人のプログラマーと６人の数学の大学院生とが与えられました。私たちがその前の春に人工知能のプロジェクトを作るためにお願いしてあったことを考えて、文書による要望はまだ出してもいないのに、ジェリー・ウィズナーが決定的かつ速やかな行動をとったのでありました。幸いにして、MITの電子工学研究

所が米国空軍と制限なしの共同研究契約を結んだばかりのときであり、しかも、まだ財源の使用計画がすべて決まっているわけではないというときでありました。このような柔軟性が、米国が他の国々にさきがけて人工知能の研究をスタートさせた理由であると私は考えています。ニューエルとサイモンの研究を可能にしたのも、米国空軍がランド・コーポレーションに与えた柔軟性のある援助によるものでありました。

1958年９月に我々はLISPの研究を始め、そして、1959年初頭には一番最初のインタプリターが稼働していました。最初の考え方はFORTRAN類似の言語でリスト構造をデータとして持つものでありました。しかしながら、大学１年生に対して普通教えられるような代数式の微分のプロセスの論理構造を注目すると、関数を帰納的に使えること、条件式が与えられること、ガーベーヂ・コレクション（ごみ集め）を使うことにより、直接な消去を避けることができること等が可能な場合には、代数式の微分を行う関数は、一つの簡単な式に表させるということが判明しました。この前年FORTRANで書いたチェスのプログラムに関連して私は条件式を発明しました。ラムダ式は、微分の例から示唆を得たもう一つの道具でした。数学者に、普通計算機は、チューリング機械よりもLISPを使うほうがよりコンパクトに記述できることを示すために、私は関数EVALを書きました。我々のプログラマーの一人であったスティーブ・ラッセルは、それを速やかにLISP用のインタプリターに変換しました。「マッカーシー　1977a」にはこの歴史が述べてあります。

1979年初期に、ジェームス・スレーグルは、ミンスキーの指導のもとで、記号積分についての研究を始めました。このころに成果の期待されていたIBMの人工知能の研究が中止になり、その後IBMは1983年までその研究を再開しませんでしたし、まだ小さな役割しか果たしておりません。その中止の理由は、明らかに無知な科学的批判とP.R.との組み合わせでありました。IBMは、計算機が単なるデータ処理機のイメージを持つべきであることを望んでおり、革命的なことは何も望んでおりませんでした。

AIに関する研究は、また時分割方式計算機システムについての最初の提案へ私を導きました。以前の計算機に関する考え方は、プログラムを長時間走らせることに重点をおいていました。そのプログラムは数字で表されたものと考えられていました。プログラムの開発のプロセスは副次的なものであると考えられており、従ってこのプロセスでの計算機との対話も同様にみなされておりま

した。私には、人工知能はまったく異なる接近方法を要求しているように思えました。人工知能の研究を行っている人はその時間の大部分をプログラムの開発と、計算機との対話に費やしていたと思います。研究者には自分の研究室でターミナルの前に座って、自分のペースで自分の都合のよいときに対話することが、午前２時に30分使うことを申し込んだり、カード・デックを納めることよりも、必要なのでした。ここでもこのアイデアが明白なことではないということを信じていただくには、多くの説明を必要とするでありましょう。明白でなかっただけでなく、非常に多くの批判に出会いました。例としては、IBM360の開発者たちは、このアイデアを知ってはいましたが、それを信用しようとしませんでした。幸いにして、デジタル・イクイプメント株式会社のPDP-６という（少し後の）計算機の開発者たちは、MITの卒業生であり、最初から時分割方式を採用しました。

実際にはオン・ライン計算機を与えるには、二つの方式がありました。一つは時分割方式であって、大型の計算機をその計算時間を分割して多くの使用者にスイッチしながら与えるものであり、もう一つは、計算機TX-０とTX-２を開発したウェス・クラークによってMITのリンカーン研究所で初めて提案されたパーソナル・コンピューターでありました。TX-０、TX-２はともに単一のユーザーにオンラインで使わせることを目的とした、非常に高価な機械でした。パーソナル・コンピューターは、1970年代に入る以前では経済的でありませんでしたし、多数の人々にオンライン・サービスを提供することと、少数の人々に高価なパーソナル・コンピューターを与えることとの間の競合は現在でも存在しています。

人工知能における論理

1959年の“常識を持つプログラム”の論文では、次のことを述べています。

「アドバイス・テーカーはある形式言語で書かれた文を操作することにより問題を解くために提案されたプログラムである。他のプログラムや形式言語を操作するために提案されたプログラム（ニューエル、サイモン、ショーの数理論理機械、およびゲレルンターの幾何のプログラム）と、アドバイス・テーカーとの違いは、前者が形式的体系が問題の主題であり、発見的手法はプログラム自身の中に埋め込まれているのに対して、後者では手続きはできるかぎりその言語自身の中で記述され、特に発見的手法はすべて言語自身の中で記述され

ていることである。

アドバイス・テーカーに期待されている主要な利点は、その振る舞いが、その記号的環境と、それに何が望まれているかということを物語る文を作ることによって、改良されうることにある。このような文を作るために、そのプログラムに関する知識、あるいはアドバイス・テーカーが以前に持っていた知識は、ほとんど要求されない。アドバイス・テーカーは、それに対して語りかけられたことが何であるにしても、それらとそれ以前の知識の論理的帰結の相当広い集合を使うことが可能である。この性質は常識を持っているものとし、ある人間を記述することにかなり共通のものを持っていることが期待されている。

したがって、我々は、『もし一つのプログラムがそれに語りかけられた事柄、およびすでにそのプログラムが知っている事柄からの帰結を十分幅広くそれ自身で自動的に演繹するならば、そのプログラムは常識を持っている』と主張するのである」

人工知能において論理式を広く使う主要な理由は、1958年のころよりも現在のほうが、研究者によって、より良く理解されていることです。情報を宣言的な文で表現することは、情報を計算機のプログラムの切片、または表で表現することよりもはるかに柔軟性を持っています。論理式は特定のプログラムが有用であることよりもはるかに広い前後関係において真になりうるものであります。事実の供給者は、受け取る側の関数を、また受け取る側がどのように使用するか、あるいはそれを使うかどうかについて、それほど理解している必要はありません。同一の事実は種々な目的に使われてよいのです。何故ならば事実の集まりからの論理的帰結が入手可能だからであります。

アドバイス・テーカーという趣意書は1958年当時大きな野望でありましたし、現在でもそのように考えられていますが、その実現からはまだ遠く離れています。その言語自身の中で、ゴールへ到達する道程を案内するような発見的手法を表現するという決勝点はまだ実現から程遠いのは確かであります。

（マッカーシー 1959）における定式化は、一連の行動によってゴールに到達するという理論の素描にすぎません。それは1964年のスタンフォード・レポートの情況計算法において、より明確となり、かつ（マッカーシーとヘイス　1969）に発表されました。その基本的なアイデアは、情況sにおいて出来事eが起こったときに情況s’を表現するのに、式

s ’ =result (e,s)

を用いることであります。出来事のうちでよく研究されているのは、行為の場合であります。そして、我々が情況s’を推論する前に情況sを満たすべき条件が存在するのが普通であります。情況計算法は行為、およびその他の出来事を推論することの特別の場合を具体化しています。第一に、出来事は不連続に起こることであると考えられます。出来事は一つの情況の下で起こり、次の情況へ結果をもたらし、出来事が起こっている間のことを推論する必要はありません。第二に、ある時刻に唯一の出来事が起こると考えます。同一時刻に共同に作用するような出来事は分析されません。

情況の一つの特徴は、概念的には強調されるけれども、実際の公理化にあたってはそれほど役割を持っていないことにあります。情況は、無限に詳しいものと考えられるものであって、例えばテーブルの上の一つの積木の例であれば、積木は特定の場所にあり、詳しい形の記述や原子段階に至るまでの材料の区分等々であります。したがって、定式化は情況の詳細を的確に知ることではなくて、情況について、それを部分的に特徴づける事実を知ることについてだけ行われます。この点に関しては、情況計算法は物理学で使われたり、また、例えば重力天文学などの科学哲学で議論されている数学的模型とは本質的に異なるものであります。物理学のモデルではどの天体を考慮に入れるか、あるいは、どれを質点として取り扱えばよいのか、質量分布のモーメントを考えればよいのかといった種々のことを決めるのが普通であります。これに対して、情況計算法は、どんなときにも新しい詳細をつけ加えることができるような模型を与えることを目指しています。私は、この開放性が、推論を人間が行うにしても、また機械が行うにしても、常識・情報・状況の本質的な特性であると主張するものであります。私は、情況のように無限な詳細をもつ存在を豊かな存在と呼び、これを完全に記述可能な不連続な存在と対照して考えたいと思います。推論は豊かな存在を近似するための不連続な存在を使うだけではなくて、推論の過程の間に近似のレベルを変えることができるものであります。

情況計算法は、レゾリューションによる定理証明法と、彼が考案した『解答抽出法』を取り入れた方法でもって、コーデル・グリーン（1969）の学位論文で初めて使われました。しかし、彼とSRIにおける協力者たちは、彼らの定理証明システムが実用的であるとは言い難いほどの量の検索を行っていることに気づきました。私の意見としては、探索を制御するような発見的事実をまったく

使わなかったということが理由であると考えていますが、しかし、それをどのように行うべきかの提案は持っていません。現在では私はもう少し多くのアイデアを持ってはおりますが、事実を使って推論を制御する確定的な提案をするにはまだ十分ではありません。

ついに、ファイクスとニルソン（1971）はSTRIPSと呼ばれる強化された定式化に到達しました。そこでは、レソリューションによる定理証明は単一の情況の性質に対する推論に使われ、一つの情況から次の情況へといくには行為の記述を直接解釈するプログラムを使っています。その帰結として、より速いプログラムが得られましたが、二つ以上の情況に関する論理式は許容しませんでした。制御の問題が現在でも未解決なので、行為に関して推理を行うような実用的な人工知能システムはすべて、制限された言語を使っています。

1970年の後半に至って、形式化された非単調推論の導入は人工知能における論理の使用を革命的なものにしました（マッカーシー　1977b、1980、1986;ロイター1980;マックダーモットとロイド　1980参照）。在来の論理は次のような意味で単調であります。

式pが式の集合Aから推論されたときには、Aを含むような式の集合B（A⊂Bと記号で表す）からもpが推論される。

もし推論が論理的演繹であれば、AからpをBからの証明はそのままBからの証明として役立つものであります。もし推論がモデル論的であるならば、すなわち、Aのすべてのモデルについてpが真であるならば、Bのすべてのモデルにおいてpは真になるはずであります。なぜならば、Bのモデルの集合は、Aのモデルの集合の部分集合になっているからであります。したがって、在来の論理の単調性は基本的なものであって、論理的体系の詳細には依存していません。

人間の推論の多くは在来の論理に対応して単調であるけれども、いくつかの重要な人間の常識的推論は非単調であります。前提から結論に到達することができ、しかし、他の式を前提に付加すると同じ結論を導くことができないことがあります。例えば、私が車を持っていることを貴方が知っていて、時には私の車に乗せてもらうことができると結論しても、その車が実は故障を直してもらうために修繕屋に預けてあることを貴方が知れば、もはや同じ結論を得ることはできないのです。貴方の心にあったのは、車の所有者に車に乗せてもらうことを頼むための一般的な規則ではなくて、確率的規則であったのだと主張することによって、単調性を救うことができると考えている人々も存在するので

す。この方法の詳細な認識論を研究すること、すなわち、どんな確率的文を使うべきかを明確にすることを試みて成功した人はいません。実際に確率的理由もまた非単調な手法に終わってしまうように思えるのです。いずれにしても人工知能の研究は、まっすぐに非単調な論理的推論の形式化へ向かつて進んでいます。

非単調推論の形式化は、速やかな展開の下にあり、多くの種類のシステムが提案されてきています。私は、限界付け（circumscription）と呼ばれる方法に集中して話を進めようと思います。それは広く認められており、今も、最も活発に研究されているものであります。このアイデアは、仮定されている文のモデルの中から“望ましい”、あるいは“標準的な”モデルを抜き出すためのものであります。“望ましい”モデルは、ある種の極小条件を満たすものであります。何を極小化すべきかについては、完全に一般的にはまだ決められておりませんし、問題によって異なるかもしれません。しかし、これまで研究されてきた多くの領域で、異常さの極小化、ないし、ある種類の存在の集合の極小化を使った極めて一般的な理論が得られています。この考え方は、まったく新しいものではなく、例えばオックハムの剃刀、“必要な範囲を超えて実存を繁殖させてはならない”は、これを論理の中で形式化することを試みれば、このような極小条件に到達します。

論理における極小化は、数学の正当的な内部発達からくるよりも、応用と関連して発見されつつある数学の一分野のもう一つの例であります。もちろんこの反対のこともさらに大きいスケールで起こりつつあります。純粋に数学的な理由のために開発された多くの論理的概念が人工知能に対して重要であることが判明しているのです。

非単調推論のもう少し具体的な例として、川を渡ることのできるボートの条件について考えてみましょう。我々は一つのボートについて何か悪い条件であるような事柄を知っています。例えば、水漏り、ボートの種類によってオールが無いか、エンジンが無いか、または帆が無いといったことであります。これらの事柄のいくつかを公理の集合として、リストアップすることはかなり便利でありましょう。しかし、あらかじめリストに取り入れることを期待できる事柄のほかに人間の推論は、さらに他の事柄も、もしそれが起きればそれを許容できるけれども、それらをあらかじめ考えることは期待できません。例えば、川の真ん中に柵が張られているなどです。こういったことは、“ボートが川を

横切ることを妨害する事柄”の集合、すなわち、克服すべき障害の集合を極小化することによる限界付けによって克服されます。もし推論者が特定の場合について何も知らなければ、そのボートを使うことができると推測を持つでありましょうし、他方、一つの障害を知っているならば、極小化したとき別な結論が得られるでしょう。

このことは、非単調推論が厳密であることよりも予測的である事実をよく説明しています。実際に、ある種の数理論理学的システムは厳密な意味では拡大できないことが示されています。すなわち、ある種の完全性をもつ場合であります。

式を用いないで物理学の基礎を議論することのように、上記のような種類の議論を式を用いないで行うことは誤りであります。不幸にして多くの人がこのことに気づいていません。したがって“イェールの射撃問題”と呼ばれる、簡単な例のブラディミール・リフシッツ（1987） による形式化を述べましょう。ドゥリュー・マクダーモット（1987） は、人工知能における論理の使用に対して落胆し、特に非単調の形式化について落胆した人であるが、彼は挑戦してこの問題を発明しました。（しかし、形式化した部分は一頁ですむ）変化に関して事実を公理化するというやや古い方式はこの問題に対しては役に立ちません。リフシッツの方法は、ここでうまく働くけれども、私はさらに一部変更することが必要であると考えています。

初期情況では、弾の込めていない鉄砲と、フレッドいう一人の人がいます。鉄砲が弾込めされると、しばらく待ってから、鉄砲はフレッドに照準を合わせ、そして、発射されます。そして、望ましい結論はフレッドが死ぬことであります。形式ばらないで言えば、規則は次の通りです。（１）一人の生きている人は彼に何かが起こらないかぎりそのまま生きている。（２）弾込めは、鉄砲に弾が込められていることの原因になる。（３）弾が込められている鉄砲は、何かが弾をぬくまで、弾込めされたままである。（４）射撃は鉄砲を弾込めされていない状態にする。（５）弾込めされている鉄砲を他人に向けて発射すれば、その人は死ぬ。我々は次のように推論するつもりである。フレッドは鉄砲が発射されるまでは生きている。何故ならば彼に何かが起こるようなことは何も推論できないからである。鉄砲はそれが発射されるまで弾込めされている。なぜなら鉄砲に何かが起こるとは推論できないからである。フレッドは鉄砲が発射されたときには死にます。この推論の非単調な部分は“起こること”を極小化

する、あるいは、“理由が無ければ何も起こらない”と仮定することにあります。

論理式は上の5個の前提を表現するように意図されています。しかし、それらは、他の事象が起こらないということを陽に言ってはいないのです。例えば、フレッドは防弾チョッキを着ていないとか、あるいは防弾チョッキの性質についても何も言っていません。それにもかかわらず、人間は情況に述べられていないことが存在するのでなければ、フレッドは死ぬと結論するのです。困難な点は、これらの文が数理論理学的用語を使用すると、意図していない極小モデルを許容してしまうという点にあります。すなわち、しばらく待っている間にある明記されていない理由により鉄砲は弾込めされていない状態になり、フレッドは生きたままになってしまうことが起こりうるのです。非単調は形式化である限界付けや、またはライターの欠席裁判論理では、問題を上記のように定式化するけれども、“異常な”結果を極小化するのは二つの可能性についてであって、一つではないのです。意図されていない可能性というのは、しばらく待っている間に不可思議にも鉄砲が弾込めされていなくなることであります。

イェールの射撃問題のためのリフシッツの因果公理

リフシッツの公理は情況計算法を使うけれども、その他に定義を持たない記法として述語*causes*を導入しています。

以下は（リフシッツ　1987）からの引用である。

『射撃問題に対する我々の公理は三つのグループに分類される。第一のグループは初期情況に関する次のものである。

holds（*alive,* So），（Y　1，　1）

¬*holds*（*loaded,* So）．（Y　1，　2）

第二のグループは我々に変数が行為によってどのように影響するかを教えてくれる。

causes（*load,　loaded,　true*），　（Y　2，　1）

causes（*shoot,　loaded,　false*），　（Y　2，　2）

causes（*shoot,　alive,　false*）．（Y　2，　3）

これらの公理は行為が成功裡に実行されたときの効果を記述する公理である。しかし、それらはどんなときに行為が成功できるかについては何も主張していない。この情報は次のように別に与えられる。

precond（*loaded, shoot*）．（Y ２，４）

最後のグループはより一般的性質をもつ二つの公理から成る。我々は次のような略記法を用いることにする；

success（a，s）≡∀ f（*precond*（f，a）⊃ *holds*（f，s）），

affects（a，f，s）≡ *success*（a，s） ∧∃v，*causes*（a，f，v）．

第一の公理は変数の値が、この変数に影響を与える行為が実行された後、どのように変わるかを記述している。

success（a，s）∧*causes*（a，s，v）

⊃（*holds*（f，*result*（a，s））≡V≡*true*）．　　（Y ３，１）

（vは二つの値*true,false*を変域としている。Y ３，１における≡は、vが*true*のとき*holds*（f，*result*（a，s））に帰着され、vが*false*のときこの式の否定に帰着される）もし変数が影響を受けないときにはその値は同じ値にとどまる。

affects（a，f，s）⊃（*holds*（f，*result*（a，s））≡*holds*（f，s）），（Y ３，２）』

*causes*と*precond*とに関して極小化を行えば、適切なことが起こる。これらの公理と限界付けの政策はこの問題を解いてしまうけれども、この形式化の下で常識的知識の大きな本体を他の不愉快な驚きを得ることなく我々が書き下すことができるか否かは疑問点として残ります。望ましいモデルを決定するために、外の世界だけに関する公理でよいのか、それとも、推論を行う目的に関する情報を公理が含まねばならないのかといったことが現在研究されつつある問題であります。さらに常識推論のために要求されている形式的な極小原理を探究する多くの可能性か存在しています。

結びと注意

１．世界に関する事実を表現することに論理を使うことの進歩はいつも遅々としたものでした。アリストテレスは形式的方法を発明しませんでした。ライプニッツは人間に関する事柄についての議論を計算に置き換えることを望んだ人ですが、ニュートンとともに発明した無限小計算より技術的にはるかに容易であるにもかかわらず、命題計算を発明しませんでした。ブールは命題計算を発明し、彼の本に“思考の法則”と名づけたが、述語計算を発明しませんでした。フレッゲとその後継者たちは、非単調推論を形式化することの必要性も可能性も見い出さなかったのです。私には、これら

のアイデアのいずれもが前段階の発見者にとって許容されるべきものであったと思えるのです。例えば、ライプニッツは、もしすでに提案されていたなら、命題計算を許容していたでありましょう。したがって、我々は、示唆されたときに明白である人間の思考過程に関する多くの事実を定式化することは、本当に難しいことであると結論しなければなりません。

形式化された非単調推論についても、認識論的に適切であり、改良が容易であるように、常識的世界についての一般的な事実を表現することには、いくつかの障害が存在します。このことを、私は主要な革新がまだ起こっていないことの印であるととっており、これに対する私の示唆を私の専門的な講演で述べるつもりであります。

私はいつも、科学の進歩が実際よりもなぜもっと速やかに行われないのだろうかと疑問に思っています。いくつかの科学の領域では、器械の設置や実験の技法によって進歩が制限されています。例えば、分子生物学ではこのことは正しいでしょう。他の科学では、進歩につながるようなアイデアから離れた方向に注意を向けるような、誤ったアイデアによって進歩が制限されることもあります。このようなことが起こりますと、一つの発見がなぜそれが為されるよりも20年前、ないしそれ以上前に為されなかったかということは難しくなります。例えば非単調推論がなぜ1920年代に形式化されなかったかという技術的理由は無いのです。

２．人工知能の研究の1950年代にとった形は、そのころの古参の科学者たちの考えとはまったく異なっていました。私、および他の人工知能研究者達もそうだと私は思うのですが、先輩の科学者たちに従ったり反対したりするのでなく、むしろ避けたのでした。このことは可能でした。なぜなら我々は研究資金や大学における地位に関して彼らに従属していなかったからです。このような助言は一般的には良くないかもしれません。しかし、人工知能に関しては、諸先輩――フォン・ノイマン、ウィーナー、マッカロッチ――によって提案されたアイデアのいずれもが、これまでのところ実を結ぶものではなかったということがはっきりしてきたと思えるのです。私の意見としては、チューリングだけが、人工知能の研究がこれまで到達した主要な結果へ、我々を導いてくれるような研究の方向づけをした人でありました。もちろん、古い提案を否定するものでもありませんし、また新しい人々によって、実り多いものと再発見されるかもしれません。コネク

ショニズムはこうした意味での試みとも考えられます。

３．私の人工知能の研究成果は、ある種の哲学的な物の見方に従っています。しかしどの程度まで哲学が私の研究に影響したかを言うことは難しいと思います。私自身の主観的な印象では、哲学は重要でした。

ａ．我々人間の世界に関する知識は感覚データやその要約から成り立っているものではありません。物質的対象、あるいは、人間的目的といったものは、それらに対する我々の感覚的印象よりもはるかに安定したものであります。このことは人間の情況に対してだけのことではありません。このことは、我々が作ることができるであろうロボットにとっても当てはまるのであります。赤ん坊が物質が存在することを経験から学ぶことも確かなことでありますが、しかし、世界の構成は我々の遺伝子によって存在することも確かなことであります。物質的対象やその他のことを想定する動物というのは、進化論的有利性を持っていると考えられます。人工知能の観点から言えば、このことは、実際の世界に存在する構造について想定されることを、プログラムの中に組み入れることを示唆しています。もし我々が知識を表現するために論理式を使おうとするならば、物質的対象を変域にもつ変数や、行動や信念といった他の存在物などを変域とする変数が存在しなければなりません。

このことは明白であると思われるかもしれませんが、しかし、それは他のいくつかの見方と対照的であります。例えば、人間あるいはロボットの経験は感覚的印象の列として表現することができるのでありますから、知能は、現時点までの値から、ある列の未来の価を予測する能力から成り立っていると提案する人達もいました。このことは、世界について何も仮定していないので、極めて注意深い哲学者がよくとる立場かもしれません。系列の未来を予測するようなプログラムを作る努力は有益ではありませんでした。この実験の結果から明白でないものが得られた試しはなく、実験者達はその後どのように実験を続けるべきかわからなくなったのです。

感覚データを実証論的に強調することは、19世紀の観念論的哲学に対する20世紀初頭の反動の剰余であると私には考えられます。前世紀の哲学は多くの漠然とした公準的存在を含んでおり、それに対する反応は、直接観察できる存在だけを基礎として許容する形をとったのでした。

ｂ．精神的資質、例えば信念を、人間がお互いに帰することは理解、予測、他

の人の振る舞いに影響を与えるものに帰することの有利性によって保証されています。このようにすることへの傾向が遺伝的であるか否かは知りませんけれども、自閉症の子供に関する一つの説明として、このような機構の正常な発育ということが言えるかもしれません。それ自身の振る舞いを説明するシステムに、精神的資質を帰することは適正なことであります。このアイテアは哲学では(デネット　1971，1981)、人工知能では(ニューウェル　1980)、(マッカーシー1979a) で展開されています。しかし、哲学ではまだ一般的に許容されているとはいえません。

c．我々が種々なシステムに帰する、とりわけ精神的資質を含む多くの資質は、現象の情況を理解させてくれるが、世界の状態に関しては正確な定義を与えることができないような近似的理論によってのみ意味を持っているのであります。(マッカーシー　1979a)を参照してください。

d．今日人工知能は、一般的推論を行う人間の能力を持つようなシステムを作り上げることからは、まだ程遠いと言わざるをえません。このことは、哲学者に推薦して良いような謙遜さを生むものである。信念のような概念に関係する人工知能システムは、本当の意味で完全に一般な信念を意味することはできません。なぜなら本当の意味での信念については何らの合意もないからであります。実際に計算機にインプレメントするために必要なだけの精密さのレベルにおいて、候補者さえないのであります。したがって、人工知能は制限された近似的な概念しか持ちえません。もっとも、すべての人間も近似的にしかこうしたものを持っていないとも考えられます。哲学者が一般的に信念を理解しようとしている試みは、単に彼らをして概念を発見するのではなく、概念を作ることに努力させることになっているのかもしれません。人間が実際に行動すること、あるいは、計算機がそのようにするようプログラムされていることに対応する完全に一般な信念という概念は、存在しないのかもしれません。