

意図のナッシュ・インプリメンテーション・メタファー

犬童健良

関東学園大学経済学部経営学科

〒373群馬県太田市藤阿久200

E-mail: indo@kangaku-u.ac.jp

Abstract

本論文では、意志的主体は脳の代理人であるという仮説の下で、ゲーム理論にもとづくナッシュ遂行理論を用いたエージェントのメンタルモデルを提案した。これは知能の計算メタファーと限界合理性概念という、認知科学と意思決定科学それぞれにおける説明の難点を補うものである。

Nash Implementation Metaphor of Intention

Kenryo INDO

Faculty of Economics, Department of Management, Kanto Gakuen University

200 Fujiaku, Ota si, Gunma ken, 373 Japan

E-mail: indo@kangaku-u.ac.jp

Abstract

In this paper, under the hypothesis that "any intentional state of mind is the agent of brain", a game-theoretic mental model that uses Nash implementation theory has been proposed. as a metaphor of intentional agent, the formalism of mechanism design complements both of counter-intuitiveness, i.e., computational metaphor in cognitive science and bounded rationality in decision science.

1 はじめに

経済学は希少な資源の効率的配分を学ぶ学問であるとされる。ところがサイモンによれば、人間の心、とくにその合理性は思惟の産物 (product) であり、また過程 (process) であり、またそうした心の資源そのものが希少であるという。人々はこれがかつて経済学の方法論の狭い視野を警告するものと考え、じつさいその後多くのエコノミストが限界合理性の問題に少なからず関心を払うようになった。上の警告には、しかし別の示唆的なアイデアが含まれているように思われる。

認知科学の文献で示唆されているように、もし意図というものがそれを形成し、またそれにしたがってふるまうことによって、知的能力に限界のある行為者の合理性に対して何らかの「貢献」を果たすものであるという推測 (Bratman, 1987; Dennett, 1984) が正しいとすれば、経済科学のアナロジーが心の科学に対して、適用できる場面があるのではないだろうか。

とりわけ能力に限界のある行為者の合理性に役立つ、計画と調整のはたらきに注目したブラトマンの意図理論 (Bratman, 1987) では、意思決定論における (帰結主義にかかわる) 最近の論脈と共通する問題が論じられている (Machina, 1989; Hammond, 1989)。また近年の意思決定科学も、期待効用仮説への名だたる反例として知られてきた、アレの背理やエルバーグの背理などで実証されてきた、独立性公理への違反

を示すエージェントの動的な意図（計画された選択と実際の選択）における不整合をどうかわすか、ということにその技術的努力をかたむけてきている。

本論文では、これらの意図理論あるいは限界合理性へのアプローチとは視点を180度変える。わたしたちの出発点は、行為者の認知において主観の限界を生じさせることが、いかに認知システム全体（あるいは脳）にとって合理的でありうるのかを問うことである。

社会科学には計算機科学において通常考えられているものとは異なる、もう一つのインプリメンテーション概念がある。メカニズムデザインあるいはインプリメンテーション理論（遂行理論）と呼ばれる分野では、市場や契約や投票といった社会経済的機構（メカニズム）をゲーム理論および社会選択理論の枠組みでもって形式化する議論がなされてきた。なかんずくナッシュ・インプリメンテーション理論とは、個人の自由な目的追及の結果であるナッシュ均衡と、社会的に望ましい状態とを一致させるようなゲームとしてメカニズムを設計する応用数理である (Osborne and Rubinstein,1994)。

筆者は認知科学の立場から、インプリメンテーション問題を心の社会における意図の調整的機能を形式化する目的で適用しようと試みている (犬童,1995ab)。意図は知能システム内に分散した計算的資源（サブエージェント）によって自発的にそれを遂行させるためのメカニズムであると仮定する。また、それゆえエージェントの自由意志は、脳あるいは知能システム全体から、一定の範囲に限られた権限として、自己意識に委譲されることが前提となる。

本論文のモデルは（1）エージェント¹の能力の自覚「できる」を遂行可能性として、また自身の知識や信念の状態を「わかる」を直接顯示として明示的にモデル化する。また（2）シンボル以前の水準の知能からある程度分離独立して動作できるエージェントの意志的はたらきを説明できる、（3）意図が別の意図や欲求と信念と連動しながら、いかにエージェントの合理性に対して貢献できるのかを、心的態度の間の調整のはたらきに着目して論じることが可能である、（4）限界合理的なエージェントの「主観的限界」を、認知システム全体においてそれが果たしている合理的役割から説明できるといった、じゅうらいの「素朴な」心の計算メタファーや限界合理性概念ではほとんど語ることもできなかつた、あるいは解釈が黙示的なままになされていた人間知能の多くの重要な特徴が明らかになる。

本論文の核となる仮説は、「意図ないし意志とは脳ないし知能システム全体によって実現され、志向的システムに委譲された権限である。」あるいはより簡単に、「志向的システムは脳の代理人（エージェント）である。」という一風変わったものである。より一般的には、経営組織や市場メカニズムにおける権限と調整のしくみを、志向的システムのメタファーとして役立てた「語り口」の理論と筆者が呼んでいるものに包括される。おおざっぱに言って、「語り口」とは人が自分自身（あるいは他者）の知能あるいは脳を使うその仕方を総称するもので、意図や意志と呼ばれるものは明らかにその中心となるべきしくみである。「語り口」は素朴な心の計算メタファーにおける「プログラム」に対応する心の機能であるが、計算機がプログラムによって完全に動作を指定されるのに対し、「語り口」はむしろ現実の社会における経営（経営管理）に似た、「どうやったら人に働いてもらうか」にかんする社会的技術を必要とする。

ただし本論文では計算的実現にかんしてほとんどふれない。その代わりに社会科学において研究されたもう一つのインプリメンテーション概念に有益なメタファーを見い出そうとする。しかし、たとえば初期の分散AI研究において提唱された「契約ネット」(Davis and Smith,1983)の実装スタイルや、ヒューイットの「科学世界メタファー」、ミンスキーによって提唱された「知識ライン」(Minsky,1986)などには、本論文に通じる発想がみられるが、本論文の意図の遂行理論とのちがいは、上で述べた通り、その社会的なセンスにおいてである。すなわち本論文では、意図を社会の成員の自由と社会目標とを一致させるためのゲーム理論的調整に見立てる。

¹本論文ではエージェントという用語を、行為者、信念者、熟慮者、主体といった人間行動を説明する理論モデルの意味で使い、とくに混乱を防ぐ必要がないかぎり、これらをいちいち区別しない。

2 「頭を使う」とはどういうことか

意図ないし意志²の特徴は、そのオブジェクト（対象ないし目標）となる認知的機能を実現しようとすることである。いいかえれば意図ないし意志とは、メタレベルの認知的機能、ないしそれに対応する一定の時間において生じている心の状態ないしプロセスのことである。たとえば

$$1 + 2 = ?$$

という問いを対して、私がこれを解こうと試み、そして答えは3であると分かったとする。このとき、この問題の解を求めようとする私の意図は、「1 + 2」をじっさいに計算する手続きからなる「オブジェクトレベルの認知的機能」に、それがどのようなプロセスであるかは私には全く分からないが、何らかの物理的な方法で関わったにちがいない。同じように

$$1 + X = 3$$

という方程式を見て、私がこの式の左辺にある変数Xの値を特定しようという意図をもったとしよう。このとき、私は自分自身で「上の式のXに相当する実数値を求めよ。」という問題を設計し、これを自分自身に対して解かせようとしたことになる。そして時間的に少し後の私自身は、それを自身で解くべき課題として請け負い、「Xは2である。」という答えを導いた。また2番目の問いの答えは上の式の変数Xに2を代入して検算するという別の手続きからなる認知機能を経由して「確かにこれでよい。」という主観的状态がもたらされる。そして私は「この仕事は終わった。」と納得して今一服しているところである。

もちろん、上のようなあえて行った分析者の視点（自分自身に対して私がとった鍵穴実験者の態度）にもかわらず、私自身はなんら自分自身の分裂を感じないし、また私がもう一人の私に対して公式的にある仕事を依頼し、その申出ないし契約が受け入れ、少し後にそれが履行されたなどというシナリオを考えたりはしていなかった。にもかかわらず私の「考え」を客観化してみると、まさに上で述べたような現実の社会における「契約」に似た手順きが起きていたとみなすことができ、またそうみなさざるをえなかった。そしてこれが、本論文でそれを明示的に扱うための精巧なメタファーをわざわざ社会科学から輸入しようという筆者の動機をかりたてている「心の社会的な本性」ないしより単純に「知能の社会性」の特徴である。

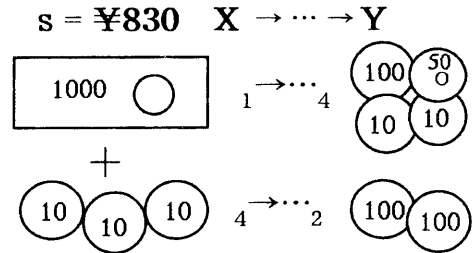
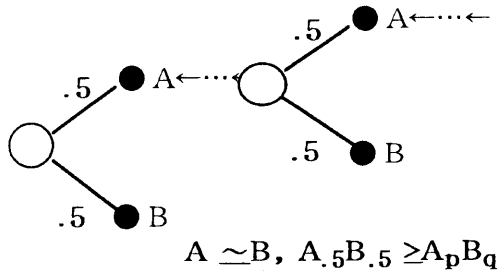
このように本論文では意図というものを、意識において提案され、脳内に散らばっている計算資源を用いるサブエージェント（微視的意志）たちによって請け負われる一種のプロジェクトであり、またこれらのサブエージェントはそれ自身が状況に応じて機能別に分化した専門的知識をうけもつ計算的オブジェクトであると仮定する。この仮定により、わたしたちは知能ロボットに意図という認知的機能を組み込むこと、すなわちAIにおける方法論的慣例をまねてタブラ・ラサからの出発を選ぶ。

2.1 記号水準の暴走：フレーム問題？

さて、どのような手続きが最初の2つの問題それぞれの答えを導いたのか？ 残念ながら私自身はあまり詳しく覚えていないし、それをある程度もってもらいたい脚色をほどこすことによって再現しようという気力もないほど上の2つの問題はあまりに簡単な問題に思われる。またじっさいにそうしようとすれば、私はいらいらして誤りを犯したかもしれない。

たとえば私が最初の問題に対して正しい答えが3であると首尾よく導けたとして、次の問題ではX = 4という答えを導き、この誤った答えを上々の出来だと満足しているとしよう。こうした誤った信念とそれを導いた認知プロセスは私自身によって強く信頼を置かれ、その信頼ないし信用を認知プロセスに「誤って」与えているのは私の意図である。なぜならば意図することなしにこうした誤った手続きと答えに私が

²本論文では意図と意志の厳密な区別はしていないが、たとえば「意志が弱い、強い」とはいうが、意図についての強度を表す日常の用語法はない。



MOM問題(Machina,1989). 1個の品物を, 2人の子AとBのいずれに与えるべきか? しかしあくまで2人を平等に扱わないといけな。そこでMOMはコイン投げで決めることにしたが...

おつり問題(犬童,1995b). 手間を省くために, 支払XとおつりYの枚数を同時に少なくしたい. このときコインをさらに支払うべきか否か? しかし最適化問題を真面目に考えてみると...

図 1: 自己否定的な意図

満足するということはありえないからである。もちろん, それが正しい答えを導いたならば, 私の意図が満足されたことは, まさにその意図の存在を通じて, その手続きを実現した認知的の代行に対して「正しく」貢献を認めていることになる。

それ自体の遂行を意図することがそれ自体を失敗させる, いわば「自己否定的意図」は, 計算量の上限としての客観的限界ではなく, 人間がじっさいに知覚する主観的な「限界」に近いものと思われる。合理性の限界とは, 脳システムが脳の中に散らばった計算的資源を, わたしたちが利用する自由を制限し, 一定の, あるいは状況に応じた権限範囲を定めることによって, 意図の拡散を未然に防ぐメカニズムをもっているだろうという予想は, 道理にかなった一つの仮説である。

道具的合理性は, ひとたび問題や目標が与えられてしまうと, 問題設計の善し悪しや目標設定の妥当性を再考慮できる, メタレベルの問題解決者の視点, ないしメタレベル意思決定者の合理性というものから断ち切られて, 請け負った仕事に専念・没頭している。AIにおけるフレーム問題におけるロボット, これに対してアネット (Dennett,1984) が引き合いに出すある種の手, および意思決定論においてマシーナ (Machina,1989) が例示するMOM問題における非期待効用最大化エージェントの一見愚かな行動計画の反復生成は, みな, そうした意図された目標追及を遂行することがその意図自体に矛盾することに「気付かない」近視眼的行為者の形式化の例題である。

したがって意図の弱さやコミットメントの取消・解消の可能性というものは, 現実の行為者の合理性を論じるためには, むしろ重要なことである。だが中心的な強い意図がいつ弱い周辺の意図によって覆えられるのかという条件を一般化するのは, ゲーム理論におけるフォワードインダクションと同じく難しい。またプリンシパルとしての観点から見れば, 意図はエージェントによる遂行を確実にするような「強さ」をもっていることが望ましいが, 一方エージェントの側の観点でみると受け入れがたい「悪い」目標設定を拒否できることが望ましい。

事前の強い意図が, 後からの弱い意図の侵入に対して免疫でないという問題は, (一見して無関係な先行文脈からの) 独立性の違反に関係しており, 行為者のダイナミックな意図を論じるにあたってとりわけ重要であるので, 後で詳しく論じる。またこのメタレベル問題解決のジレンマはAIにおいてはフレーム問題として, また経済学では限界合理的な意思決定者の限界を定める無限遡及問題として論じられた。しかしこれらは現実の行為者では起こりにくいと考えられる理論上の困難であり, その形式的解決は現実の人間を説明するものとはいえない。一方, 筆者はレジで買い物客がオツリを考えながらお金の出し入れする手間を満足化する問題を考察した (犬童,1995a)。

つまるところ限界合理性とは, シンボル水準知能を, 暗黙知としての適応的知能から分離・離陸させることによって生じる, シンボル水準運用における暴走の危険をヘッジするという認知システム全体からの

必要性により、その運用の権限を限定し、じゅうぶん分権化された状態で意図を形成させることを意味している。またこれはちょうど経営と所有の分離における、経営者のモラルハザードをチェックする株主の監査機能が必要であるといった、いわゆるエージェンシー問題 (principal-agent problem) との間のアナロジーになっており、またそれゆえ非対称情報ゲーム理論によって分析できる問題クラスに属している。

3 ナッシュ遂行にもとづく意図のモデル

ナッシュ・インプリメンテーション理論 (遂行理論)³は、前節で述べたようなエージェンシー問題を特殊な場合として含み、さらに一般的に抽象化された社会選択理論の状況として形式化するものである。

3.1 モデル

社会の成員を番号 $i \in I = \{1, 2, \dots, n\}$ で表わす。この社会における選択対象 (社会の結果) を $a \in A = \{a_1, \dots, a_m\}$ とする。各成員 i はこの集合 A の要素の間で、自分にとっての好みを比較した選好順序 R_i をもつものと仮定する。つまり $a R_i b$ は、 i が a を b よりも好ましいと思っていることを表わす。 $R = (R_1, \dots, R_n)$ を選好プロファイルと呼ぶ。各 R_i (および R) のすべての可能性を Q_i (および Q) とする。

このとき、社会選択ルール (SCR) を f とすると、 f は選好プロファイル R から選択対象の集合 A の部分集合への写像として表わされる。 $f: Q \rightarrow A; f(R) \subseteq A, R \in Q$ 。

一方、ゲーム形式 g は、各 i の戦略集合 S_i 、 $S = (S_1, \dots, S_n)$ として、 S から A への写像として表わされる。 $g: S \rightarrow A; g(s) \in A$ 。ゲームの結果としてのナッシュ均衡 $NE(g, R)$ は、ゲーム形式 g と選好プロファイル R によって定まる最適応答である。

以上の設定の下で、社会選択ルール f の遂行可能性を考える。

定義 1 社会選択ルール f がナッシュ遂行可能であるというのは、任意の $R \in Q$ に対して、 $f(R) = g(NE(g, R))$ となるような、あるゲーム形式 $g: S \rightarrow A$ が存在する場合である。またこのとき、 g は f をナッシュ遂行するという。

定義から分かるとおり、社会選択ルールを遂行しようとする社会プランナー (当局者) にとって、社会の成員のもつ好みというのは直接観察できないプライバシーに属するものである。この分権化された社会において、自由な個人の目的追及と、社会選択ルールとが一致する保証はないわけであるから、制度的工夫としてゲーム形式 g をメカニズムとして設計する権限だけを当局者がもって、あとは自発的に人々が社会選択ルールにしたがうような制度的工夫をしようというのが、ナッシュインプリメンテーションの発想である。

また、社会選択ルール f を遂行するゲームをじっさいに開発する場合、各 i の選好 R_i を報告させるゲーム (直接顯示メカニズム) とそこで全員が正直に自分の好みを報告する正直均衡による遂行性がしばしば論じられる。これは顯示原理によって、正直遂行できない f はどのように g を作ってもナッシュ遂行できないことが知られているためである。

ナッシュ遂行と直接顯示メカニズムは、それぞれ心の社会における意図 (ないし意志) のはたらきと自己意識についての2つのメタファーの組み合わせに対応する (図)。

³ナッシュ・インプリメンテーションについての最初の一般的結果はマスキン (Maskin, 1977) による。必要十分条件についての厳密な結果は後の研究者 (Dutta and Sen, 1991; Moore and Reullo, 1990) が行った。またさまざまなナッシュ均衡の修正概念にもとづくメカニズムが考案され、遂行可能な社会選択ルールの範囲が飛躍的に広がった。グローブズメカニズムは、公共プロジェクト選択に特定化して研究されたナッシュ均衡の下での遂行メカニズムである。それ以前はナッシュ均衡を使わず、支配均衡を用いた遂行理論が研究されていた。クラークのポポタル・メカニズムはとくに有名である。また社会選択理論における操作不可能性—自分の好みについて嘘をつくことを通じて、自分にとって都合の良い社会的選択結果を得られるか否か—にかんする一連の不可能性定理 (Gibbard, 1973; Satterthwaite, 1975) が知られる。

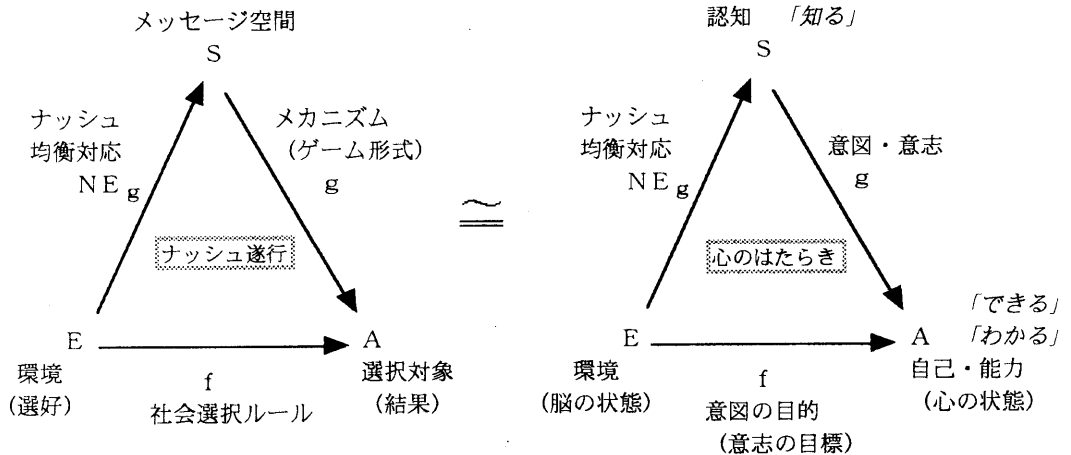


図 2: ナッシュ遂行と顕示機構にもとづく心のモデル

3.2 直接顕示メカニズムとしての心

人は自分の脳の物理的状態とそれを支えている神経細胞の物理的活動について、直接モニターしたり制御することはできない。いいかえれば、私たちの脳システムと自己意識の間には、情報の非対称性が存在している。にもかかわらず、私たちは自分の身体感覚や意図したことにもとづいて、自分の体を思うとおりに動かしたり、考えを述べたり、話したりすることができるのに十分な「情報」を心のはたらきによって得ている。

脳のなかに分散して私たちの意識を実現している何らかの機能的実体——これを脳細胞とか、モジュールという脳のメタファーで言い表すこともできる——は、それが適応的にふるまえる状況ごとに専門化されたエージェントとみなせるが、私たちの社会とことなり、言語によるコミュニケーションができない、このため心の社会は、一種の非協力ゲームの形式に適合する。

自身の考えについて知っている、あるいは少なくとも部分的に正しく認知できるということとしての自己意識の特性は、けして神秘などではなく、心の状態の計測（顕示）メカニズムとして論じることができる。エージェントの自己意識とは、意志的システムが利用したい脳内に分散した（サブ）エージェントに対し、その私的情報（エージェントのタイプは心の状態に対応）を正しく報告させる直接顕示メカニズム（direct revelation mechanism）である。

また認知の全体システムが自己意識を実現するとき、もちろん神経活動のすべてをわたしたち認知主体の意識に知らせるようなことはしていない。その意味で、非協力ゲーム理論に一種の身振り言語によるコミュニケーションを導入した非対称情報ゲーム（ベイジアン・ナッシュ遂行）としても形式化できる。心の社会のサブエージェントの状態は「語りえないこと」あるいは直接観測できない真の自然の状態 θ であり、それによって実現されるエージェントの認知は暗黙知であるが、そのふるまい x （戦略）の選択を通じて私たちに「自分の心の状態」を知らせている——つまり内観 x は心の状態 θ のサポートである。

3.3 間接メカニズムとしての心

また意識は、潜在的な計算複雑性が表面化しにくいようにする何らかの仕組によって支えられているはずである。これは環境に依存した形で情報を圧縮することによって可能になる。じっさい抽象的な環境で設計されたマスクンのメカニズムでは選好プロフィールを含む表明をさせるため、メカニズムとして messy

であるが、公共財の最適配分問題に特化したナッシュ遂行であるウォーカーのメカニズムでは、最低限の情報伝達でリンドール配分を実現できる。こうした環境ないし身体性をうまく利用することで効率的情報伝達を行えるという発想は、生態学的知能とよばれることがあるが、状況意味論における「制約」の概念や認知工学における「アフォーダンス」概念について、ナッシュ遂行は一つの別解釈を提供している。

脳ないし知能システムそのものは重並列処理されているが、意識的な問題解決ないし意志と意図が、少なくとも近似的に逐次処理されるように実現されていなければならない。また生体にとって、環境への適応に支障をきたすような意図、あるいはフレーム問題のように意図そのものにとつて、意図を遂行しようとするものの逐次性が引き起す処理負荷が意図の遂行を不可能にするブレイクネックを避けるために、その裁量範囲を予め設定しておくことは、むしろ合理的であろう。すなわち意図は、権限範囲の設定によって限界のある行為者の合理性に対して貢献する。そしてこれはまさに一種の社会的調整のための技術である。

4 「語り口」の理論へ向けて

あえてじゅうらいの計算機（プログラム）メタファーに対応させると、「語り口」の理論において、行為者の意図や意志が果たす役割は、計算機メタファーにおけるプログラムの役割にあたる。おおざっぱに言えば、これはそれぞれの知的動作の機構に「特定の仕事をさせる」ための方法という意味においてである。しかし著者が着目する点は、人間はコンピュータとちがって、指示されたとおりのことをしない傾向がある、いかえれば行為者あるいは熟慮者にとって、同じシンボルが文脈によって全く異なる意味をもつ場合があるということである。たとえば同じ内容を喋っても、権威ある a さんと素人の b さんの助言は、処方すべき問題を抱えた c さんにおいて、異なる影響を及ぼすかもしれない。

たとえば c さんはデスクのパソコンにインストールされた電子メールソフトの使い方が分からなくて困っているとき、コンピュータに強いが自分よりも年若い a さんよりも、たとえ適切な情報が得られないとしてもマージャン友達の b さんに聞こうと思うかもしれない。

また別の例として私が毎週どの曜日のどの時間に自分のオフィスに出かけるか、またそれは電車やバスではなく自分の車を運転して行くのだということは、みなすでに決まってしまうているが、私は今の仕事はどうにもつまらないと感じていると仮定してみよう。それでもなお、私は出勤日の朝目覚めたとき、どのように自分を会社に出かけさせるかにかんしては選択の余地があるだろう。

私は自分を、行かなければ「クビになる」とか、「仕事は戦争だ」、「負け犬になるな」などといった半ば強迫観念であるような語り口によって、自分をくたくたの出勤遠征にかりたてることもできるだろうが、おそらくはもう少しましな、より積極的になれる別の語り口をさがすだろう。あるいはそれも見つからないなら、そういったことをいっさい考えない「機械人形の比喩」を自分自身に対して適用し、かつそのことを意識に昇らせないおこうと決め込むこともできるだろう。

こうしたことは計算機メタファーではまったく扱われない。

意図や意志はしばしば行為者自身が、自身に対する指令や計画や助言あるいは警告といった、言語行為 (Searle, 1969) に準ずる形式でもって思惟することを通じて実現される。こうした意図や意志を含む志向システムを「語り口」と呼ぶことができるだろう。

また意図はしばしばそれを続けることによって、意図をもつ行為者自身の能力を安定させる。いかえれば意図は、行為者が自分の欲求や信念の状態の下で、成し遂げようとする目標に向かうというだけでなく、むしろそれ自体によって行為者の効用や信念が行為者自身にとって明確になるように、行為者の置かれた問題状況そのものを定義する。

意図を遂行しようとしている間に、その課題が思っていたよりも困難であることが分かったときに、その意図を保持（または変更）することの可能性に対する態度、つまり関与 (commitment) の程度はケースによるとともに、客観的な選択の可能性とは一般に異なっている。しかしより重要なことは、意図を変更

したりあるいは保持するために、因果を含める「語り口」の選択によって、いかに知能のパフォーマンスが左右されるのか否かということである。これはメカニズムが自発的なエージェントの協調を助けることも、あるいは逆にさまたげになることもあるということに対応する。

5 おわりに

わたしたちは自分の意図や考えについて知ることができ、また他者のそれについても、その人のふるまいや言葉で述べたことにもとづいて、意図を理解したり、推測することができる。しかし、意外なことに心の科学としての認知科学や人工知能では、「分かる」(知識)とか「できる」(能力)といった内観とそれを実現する意志や意図が成立するしくみそのものは説明されてこなかった。

意図や意志の理解は、また認知科学と人工知能が慎重に避けてきた心の中心的機能への接近としての意義がある。診たところ、認知科学におけるパラダイムの対立は、(1) 自分(あるいは他者)の考えていることが分かる、また(2) 自身の意志または意図もってふるまえるという2つのメタレベル認知的機能としての、心のはたらきの最も基本的な性質に言及するための理論装置をもっていなかったことが混乱の一因と考えられる。

デネットのいう志向システム(Dennett,1987)の理論的説明は認知科学の大きなテーマであり、おそらく多くの認知科学者がそう認識し、あるいは難問であるから手を着けずに置こうと決め込んでいるためか手着かずである。フォーダーの有名な予想では、認知科学の成功の見込みは認知プロセスのモジュラリティの減少関数である(Fodorの第一法則)。モジュラリティは入力系のように領域特異的で、情報遮蔽的で、かつ生得的に確定している認知プロセスほど高い。意図は明らかに中央系の中心部分であるから、フォーダーが正しいとすれば認知科学の成功の見込みはきわめて低いということになる。

この現状は、心というモノを対象とせず、心のはたらき(つまり心の機能)について、計算機の実現をモデルに客観化する学問としての道を歩んできた人工知能と認知科学の伝統に根差しているのかもしれない。科学では語り得ぬものは沈黙されるべきかもしれない。しかし、おそらくだれもがほとんど直観的に疑う余地のないと思われる直観に忠実なところから、ささやかな試みを始めることは可能ではなかったろうか。

参考文献

- [1] Bratman, M.(1987): *Intention, Plans, and Practical Reason*. Harvard University Press. [門脇峻介・高橋久一郎訳(1994).『意図と行為：合理性、計画、実践的推論』。産業図書.]
- [2] Dennett, D. C. (1984): *Cognitive Wheels: the Frame problem of AI*. In Dennett(ed.) *Minds, Machines and Evolution*. Cambridge University Press. [信原幸宏訳(1987).「コグニティブ・ホイール：人工知能におけるフレーム問題」。現代思想 15(5),128-150.]
- [3] Dennett, D. C. (1987): *The Intentional Stance*. MIT Press. [若島正・河田学訳(1996)『志向姿勢の哲学』。白楊社.]
- [4] Gibbons, R.(1992): *Game Theory for Applied Economists*. Princeton Univ. Press. [福岡正夫・須田伸一(訳)『経済学のためのゲーム理論入門』創文社.]
- [5] Machina, M. J.(1989): Dynamic Consistency and Non-expected Utility. *Journal of Economic Literature*.(also in Bacharach and Hurley(eds.). *Foundations of Decision theory*. Blackwell.
- [6] Osborne, M. J. and Rubinstein, A.(1995): *A Course in Game Theory*. MIT Press.
- [7] Simon, H. A.(1978): Rationality as process and as product of thought. *American Economic Review*,68(2), 1-16.
- [8] 犬童 健良(1994b): 協働のための「語り口」としての自己意識。SWOPP 琉球'94.
- [9] 犬童 健良(1994a): 貢献認知ゲームとしての会話(2): 「AハB。」文における認知的補償のゲーム理論モデル。関東学園大学紀要第23集.
- [10] 犬童 健良(1995a): 協働系内エージェントの認知モデル。経営情報学会 1994 春季全国大会発表.
- [11] 犬童 健良(1995b): POSゲーム(1): お釣りの問題と「意図」のゲーム論モデル。miemo.