人　工　知　能　95－12
（１９９４．　７．　２１）

協働のための「語り口」としての自己意識

犬童 健良
関東学園大学経済学部経営学科
e-mail:　indo@tansei.cc.u-tokyo.ac.jp

はじめに

著者の研究スタンスは、自己意識の認知的モデルを探究するにあたって、市場理論や非協力ゲーム理論（あるいはメカニズムデザイン論）とのアナロジー[注1]から出発し、心の社会における調整(coordination)として捉えることであった[4,5]．しかしこれら合理性に基づく協調のための一連の語り口は、じつは分権的調整を支える論理的な正当化の信念(justifying belief)にはなりえない[注2]ことが最近のゲーム理論の文献[2]で論じられている．ここで協働(cooperation)は、むしろ調整のしくみを通じて各個人の能力の限界が克服されるときに興味ある現象であったことを思い出さずにはいられない．じっさい協働はC.I.Barnardから、後にH.A.Simonがその定式化を受け継いだ、限定された合理性(bounded rationality)の概念や満足化原理が論じられたオリジナルの文脈であった．しかし、そもそも合理的でない人がどうやって自分に許された合理性のうち、協働のために放棄すべきかを定められるというのだろうか．

分権化された知能の協働

そこで論じられた現実の人々の協働は、ただ複雑な作業を分解して単純な作業を各個人に分担させることとは異なっている．協働に参加する個人には、(a)誘因(incentive)と意志(intention)、ならびに(b)社会的調整のための知能-----少なくともそれが依存する調整機構[注3] (coordination mechanism)ないし制度(institution)にみあうだけの----を自らが備えていることが要請される．分散AIの研究者はこうした社会科学の諸分野で基礎となっている思想に必ずしも十分な注意を払っていないが、関心はもっている．たとえばGasser[3]は、BeckerやGersonが批判的に検討したShelling流のコミットメント(commitment)の概念、つまり複数背景(settings)をもつ交渉者の態度の一貫性(consistency)についての説明および広域のコミットメント概念としての主権(sovereignty)の概念----権威の一種である---を積極的に引用している．しかしGersonが援用している自我(I; ego)と社会的自己(Me; self)の理論のG.H.Meadには若干触れているのみであり、協働系(cooperative system)あるいは公式組織の理論のC.I.Bernardは忘れられており、またエージェントの目標や信念にもとづく行動モデルとしてのゲーム理論----Schellingの議論が出てきた文脈である----については全般に無関心である．こうした社会科学における伝統的な語り口としての合理性概念に基づく志向的説明についての無視は、それがもっとも技術的に洗練された理論体系であることからして、理解しずらいものがある．だがこれまでの純粋に知識工学手法とは別の理由から、数理モデルによる諸アプローチは協働できる個人およびその自己意識の特徴を捉え切れていない．

注1　市場（価格）あるいは組織（権威）によって代表される分権化された社会調整の方法は、その抽象化されたバージョンについて比較的数理モデル化しやすかったために今日まで生き延びている．また分権化(decentralization)によって自己完結化できる複雑なタスクを自律的知能の単位・エージェントに割り当てることにより、意識は計算複雑性を回避できるのではないかという思いつき自体は、著者のオリジナルではない．それはむしろ経済理論における個人主義の古き良き信仰の習慣であったが、市場による分権的調整は残念ながら情報的に効率的にはなりえないというのがメカニズムデザイン論の分野での一般的な結果である．価格決定力を持たないほど微視的多数のエージェントから取り引きに関連するすべての情報を受取って均衡解を計算することは、実質的にはloose couplingにはなりえない．

注2　非協力ゲーム理論ではかならずしも微視的でない影響力をお互いに持つ知的エージェントが戦略的相互作用をする状況が----シグナリングと呼ばれる意図理解コミュニケーションも含めて分析される．しかし、それが提案するゲームの解概念(solution concept)そのものが自己意識の問題に直接かかわる難問----人々の合理性(rationality)の仮定とその共通知識(common knowledge)の仮定の非両立を抱えていることが最近の研究から分かってきている．ゲームプレイヤーは相手の思考のモデルを明示的に用いた推論をするために、相互推論の階層が無限に深くなる．しかしそれでも論理的な意味での共通知識には達しない．またこれはゲームプレイヤーの知的推論をアルゴリズム化することによって計算論でよく知られる決定不可能性の問題に帰着される[2]．他者行動の推測に依存して自身の行動を決めるエージェントの信念は自己言及の構造を持つために、一種の志向システムとしての合理的エージェントがその目標を１つに決めることができない．

語り口----自己意識の顕現化、他者との共変[5]

われわれは協働において生じている共変(covariance)に注目している．それは次のように言い直せるかもしれない．エージェントを社会状況に埋め込むこと[184]と、その自己意識の顕現化[185](revelation)とは両立しうるだけでなく、相互補完する．そこで以下の仮説を提唱する．１つの知能体は、その社会で普及している語り口[1,5](narratives)を模倣できたときに、それが特定するエージェント(agent)になる．それによって、同じ語り口のなかで特定される他のエージェントと同様に、自分の行為についての意図（または志向）に基づく説明を利用できるようになる．そこで自己意識はこのような他者と自己を互換的にする志向のソフトウエアと考えることができ、それが作動する環境としての語り口とともに、模倣によって普及する．またこれら語り口と志向的ソフトには、ちょうど現実の計算機のＯＳとアプリケーションの相互補完関係と同じような、共変するバラエティがある．

おわりに

ところで、かつて機械の知能の出現に人々が驚いたのは、それがこなす仕事の複雑さではなく、むしろ仕事のソフトウエア化によって獲得された汎用性・柔軟性ではなかったか．計算機はヒトを除く進化した知能を一気に追い越して、この知能のテストをクリアした．また今までのところ、これが唯一本質的とみなせる人と計算機の類似だろう．　もう１つの類似はこうした万能性あるいは原理的計算可能性に課されている制約である．計算機と人間はともに、それぞれ別の理由から、何にでもなれるわけではない．仮に記憶容量その他の計算的資源を逐次拡張できると仮定しても、実際に計算機が実行するプログラムを設計しコーディングするのは、結局は人である．また構文的に正しいプログラムとしてひとたび命令が与えられれと計算機は仕事を忠実に遂行する．人間の場合、上司が指示する仕事を請け負うかどうかは、権限ないし権威(authority)[186]の受容、あるいはより広い意味で契約(contract)の問題であろう．

参考文献

[1]Abell,P., The syntax of social life: the theory and method of comparative narratives, Oxford, Clearendon press(1987).
[2]Binmore, K., Modeling rational players, parts I and II, *Economics and Philosophy* 3(1987):179-214; 4(1988):9-56.
[3]Gasser, L., Social concepts of knowledge and action: DAI foundations and open systems semantics, *Artificial Intelligence* 47(1971):107-138.
[4]Indo, K., Communication devices for agent society, in proceedings of *CEMIT'92*(1992):445-448.
[5]犬童健良、協働系内エージェントの認知モデル：内生的制約と合理性のコミュニケーション的基礎、１９９４年春季全国研究発表大会発表要旨(1994):183-186、経営情報学会.

[184] 専門の職業に相当期間コミットして、ふつうの人は専門的知識と実用的技能をバランスさせる．またこうしたコミットメントが分権化された社会において情報の非対称(asymmetric information)を発生させている．専門領域に密着した技能がクオリファイされたインプット／アウトプットを生成する手続きは暗箱化する傾向があるが、しかしそれは情報の経済分析でいうところの、価値があるために私的情報(private information)として秘匿されるのではなく、むしろ本人にとってさえ形式化して伝達することが困難である暗黙知なのである．

[185] ここでいう自己意識の顕現化とは、より平たく言えば、人は他者と比較可能なインプット／アウトプットを請け負うことができるとき、はじめて個人(individual)になるということである．欧米の企業風土では自分の仕事に適度のプライドをもたない人を仕事相手として尊敬(respect)されないようである．もっともこのような個人の協働適性観には文化差や個人差が見られる．日本の文化では、むしろ思いやりや甘え、あるいはわいがやといった概念が適応的な個人の典型イメージであり、自己主張の強い合理主義的個人は会社組織などにおいては敬遠される傾向があるといわれる．

[186] 権威とはある個人がなすべき問題解決を自身ではなく他者が定義し、それがしかるべき正当化の信念をもって受け入れられることである．自由の放棄としての調整者の権威(authority)の受容はとりわけ以下の２点で興味ぶかい．１つは自分の解くべき問題を権威あるいは調整者(coordinator)によって（部分的に）定義してもらうことによって、仕事をする個人の情報処理負担を軽くすること．つまり決定の代行---作業の代行ではなく---である．またそれによって主体としての特徴---合理性(rationality)を一部放棄している．もう１つは権威の共通知識性を通じて、自分の仕事と他者の仕事との間に比較可能性と予見可能性を確立させることである．これによって個人は自分の専門的作業に専念することが無意味な労役になる心配を感じなくてすむことになる．無関心圏(zone of indifference)に入る指令については、その定義上、権限関係は問題なく機能する．一方、与えられた指令が個人の自律を脅かすならば、人々は自分が合理的にふるまえる作動範囲を維持する理由でそれを拒否できる．しかし権限関係ないし階層組織において重要なのは、これら無関心圏と拒否圏の中間の水準である．またこの論理的には決定困難である中間領域こそ、個人の自己意識における決定(decision)あるいは意志が効いている領域である．