

# 新たなモラル・ハザード概念を分析するための エージェント・モデルの提案

松村 良平, 木嶋 恭一, 中野 文平, 猪原 健弘

## 1. はじめに

組織において最も重要な問題の一つに、いかに成員を動機づけるかというものがあり、これを分析するためのツールとして、エージェント・モデルと呼ばれるものが存在する[1, 2]。これは、プリンシパルと呼ばれる経済主体が、自らの目的にそった成果をエージェントと呼ばれる経済主体にあげてもらうために、どのようにインセンティブを与えるか意思決定しようとしている状況をモデル化したものである。このモデルの特徴としては、プリンシパルがエージェントの行動を完全には観察し得ない、つまり情報の非対称性が存在する、ということがあげられる。この特徴ゆえ、エージェントはプリンシパルの望む行動をとらず、別の自らの望む行動をとる可能性が存在する。これがエージェントのモラル・ハザードと呼ばれるものである。従来までのモデルでは、このモラル・ハザードとして、エージェントの「怠ける」という行為のみが想定されていた。しかし、成員が組織目的に沿わない行動をすることで、組織に利益をもたらさなばかりか、不利益をもたらすことさえあることを我々は現実に目にすることがある。たとえば、情報処理技術者が上司の目を盗んでクラッキングなどの行為をして問題を起せば、組織の不利益になることは明らかである。また、研究開発組織などでは、個人目的と組織目的の葛藤が大きいと一般にいわれているが、組織が応用研究や製品開発を望んでいるのに、個人研究者が、組織の望まない基礎研究を進めることで他の研究者へのモラルに悪影響を及ぼすことがあれば、これは問題となろう[3]。さらに、官僚制が機能しにくいような組織で、

個人的な信条により官僚的なリーダーシップを発揮するマネージャーも、組織に対して不利益な行為をとっていることになる[4]。このように我々の周りには、怠けるだけでなく、組織に不利益をもたらす可能性をもったエージェントが存在している。本論文は、このような行動をも含めた、より広い意味でのモラル・ハザードの概念を表現し、分析できるモデルを提案しようというものである。

新しいモデルでは、この状況を表現するために、エージェントのもつ個人目的というものを考える。これはつまり、組織の目的に直接かなうものとは別に、成員自らの非金銭的欲求を大きく満足させるような目的のことである。たとえば、研究開発組織において、組織は応用研究や新製品の開発を望んでいるのに、個人が基礎研究を望んでいる場合、この基礎研究は個人目的に当たる。また、情報処理技術者が組織の目を盗んでクラッキング行為を行うときのクラッキング行為も個人目的に当たる。さらに、官僚制が機能しにくいにもかかわらず、個人的な信条に従い、官僚的リーダーシップを発揮するマネージャーにとって、このようなリーダーシップを発揮することは個人目的に当たる。

個人目的に努力を投ずることによるモラル・ハザードには、二つの面が考えられる。一つは、個人目的に希少資源である時間とエネルギーを投入することにより、組織目的に投入する時間やエネルギーが減少してしまうというものであり、もう一つは、上に述べたように、個人目的そのものが組織に不利益をもたらす場合である。前者については、著者らは[5]で分析を行った。本論文では、後者の影響のみを分析の対象とする。

エージェントが二つ以上の目的に努力量を投入する状況を扱うモデルとしては、まず文献[6]が存在しているが、文献[5]では、エージェントの個人目的、非金銭的動機づけという要素を取り入れ、これを発展させた形になっている。しかし、エージェントの個人目

まつむら りょうへい, きじま きょういち, なかの ぶんべい, いのはら たけひろ  
東京工業大学  
〒152-8552 東京都目黒区大岡山 2-12-1  
受付 01.9.17 採択 02.5.7

的がプリンシパルにとって直接不利益をもたらす可能性に焦点を当て分析しているものは現在までのところ存在しない(文献[5]では、エージェントの個人目的は、組織への成果をもたらさないと仮定していた)。

本論文の構成は以下の通りである。節2では、エージェンシー・モデルの解説、一般モデルの紹介、そして、分析モデルの提案を行う。節3で、比較静学分析を行い、節4で結論を述べる。

## 2. モデル

### 2.1 エージェンシー・モデルについて

エージェンシー・モデルとは、前節で述べたように、ある経済主体(プリンシパル)が、自らの目的にそった行動を別の経済主体(エージェント)に依頼するという状況で、どのようにインセンティブを与えれば効率的にエージェントを動機づけられるのかを分析するためのモデルである。通常、プリンシパルはエージェントのあげた成果から、エージェントの行動(努力水準)を特定化できないことが仮定される。このように成果と行動が一对一に対応しないような状況では、エージェントは必ずしもプリンシパルの望むような行動をとる保証はない。怠けてしまうかもしれないし、組織にとって不利益な行動をするかもしれない。このモラル・ハザードを防ぐために効果的なのは、与える給与を強く成果に依存させることである。しかし、このやり方はエージェントに大きなリスクを負担させることにもなり、契約を成り立たせるためにエージェントに一定以上の効用(留保効用)を与えなくてはならないと考えれば、その分だけ大きなリスク・プレミアムを払うということになる。このような状況で、どのようにインセンティブを与えればプリンシパルの効用が最大となるのかを分析するのがエージェンシー・モデルを用いた理論の主目的の一つである。

エージェンシー・モデルの基本構造は次の通りである。まずプリンシパルが、成果に応じたインセンティブの与え方を規定したインセンティブ・システム(以下IS)をエージェントに提示する。次にエージェントは、このISのもとで自らの効用関数の値を最大にするような行動を決定し、この行動をとったときの効用関数の値が、留保効用以上になれば契約に応じ、下回れば契約に応じない。このことをかんがみて、プリンシパルは自らの効用関数の値を最大にするようにISを決定する。

### 2.2 一般モデル

この節では、節1で述べた新たなモラル・ハザード概念を表現し、より包括的にインセンティブ問題を分析するための一般モデルを提案する。従来までのモデルとの決定的な違いは、エージェントが怠けるという形でのモラル・ハザード以外にも、彼らが組織の不利益になる行動をとる可能性もあることを許容した点である。ここでは、エージェントは組織目的以外にも個人目的をもつ。それは、組織に正の利益をもたらすかもしれないし、不利益をもたらすかもしれない。本モデルでは、エージェントの複数の目的を明示的に設定し、それらへの努力水準、そしてそれらの努力がもたらすプリンシパルへの正の利得のみならず、損失をも考えることで、このような新しいモラル・ハザード概念を表現し、新たな分析視点を提供しようというものである。

以下、本論文で用いる記号を説明していく。記号は、確率変数 $\theta$ および、分析モデルで用いる $p_i$ を除いてすべて、正の実数値をとるものと仮定する。

- $e_i$ : 個人目的のための努力水準
- $e_o$ : 組織目的のための努力水準
- $U_i(e_i)$ : 個人目的のための努力が組織に対してもたらす成果

これは、 $e_i$ の増加関数になることもあれば減少関数になることもあると仮定する。

- $U_o(e_o)$ : 組織目的のための努力が組織に対してもたらす成果

これは当然、 $e_o$ の増加関数と仮定する。

- $f$ : 固定給

成果の大きさにかかわらずきまって支払われる給与。

- $U_p(e_i, e_o, \theta)$ : プリンシパルの観察する成果

エージェントのあげた成果が、プリンシパルによって正確に測定されるということは、現実には難しい。通常、エージェントの業績を測定する際に、何らかの誤差が生ずるものである。 $\theta$ はこの誤差を表すパラメーターである。

- $s$ : 業績給の配分係数( $0 < s < 1$ )

エージェントのあげた成果のうち $s$ をエージェントが、 $1-s$ をプリンシパルが得る。

- $M_p((1-s)U_p - f)$ : プリンシパルの金銭的効用関数
- $M_a(sU_p + f)$ : エージェントの金銭的効用関数
- $N(e_i, e_o)$ : エージェントの非金銭的効用関数

エージェントは、各目的に努力を投ずることにより、金銭的効用のみならず、非金銭的効用をも得る。これ

は、仕事の面白さ、達成感といった正の効用と、疲労、余暇の減少といった負の効用の双方を含むものである。

・  $B$ : 留保効用

エージェントは、総効用がこの留保効用以上にならなければ契約に応じない。そのような最低水準の効用を  $B$  で表す。

これらの記法を用いて、プリンシパルのインセンティブ・システム設計に関する最適化問題は、次のように表現される。

$$\begin{aligned} \max_{f,s} \quad & M_p \\ \text{s.t.} \quad & M_a + N \geq B \\ & e_i, e_o \in \arg \max_{e_i, e_o} (M_a + N) \end{aligned}$$

### 2.3 分析モデル (関数の特定化)

前節で、一般モデルの提案を行ったが、このままの形では分析を行うことは困難である。関数にある程度制約を置き、モデルを単純化することによって、具体的な分析を試みたい。ただし、特に非金銭的効用関数については、ケースバイケースで様々なモデル化が可能であるので、今回のモデル化はあくまで、ある特別なケースのみを表していることを断っておく。状況に応じたモデル化が必要になるであろう。

今回の研究の主目的は、新たなモラル・ハザード概念をも表現し得るモデルを提案し、それによって、従来のモデル分析とは異なる指針が得られることを示すことにあり、現段階では、あらゆるケースのモラル・ハザードを分析しようというものではない。

・  $U_i(e_i) = p_i e_i$

$p_i$ : 個人目的のための努力の生産性

個人目的のための努力は、組織に対して正の効用をもたらすこともあれば、負の効用をもたらすこともあるので、 $p_i$  の符号は特定化しない。正の値も負の値もとれる。これにより、怠けるだけでなく組織に不利益をもたらすこともあるエージェントを、モデル化できるのである。

・  $U_o(e_o) = p_o e_o$

$p_o$ : 組織目的のための努力の生産性

組織目的のための努力は、組織に対して正の効用のみをもたらすと考えるのが自然なので、 $p_o > 0$  を仮定する。

・  $U_p(e_i, e_o, \theta) = (U_i + U_o)\theta$

$\theta$  は平均分散  $z^2$  の正規分布に従う確率変数とする。これより、 $U_p$  は平均  $U_i + U_o$ 、分散  $(U_i + U_o)^2 z^2$  の確率変数となる。 $z$  は観測のノイズの大きさを表すパ

ラメーターである。実際には、成果は投じた努力量だけでなく、環境の影響も受ける。つまりエージェントにとっては、測定の誤差だけでなく、一生懸命努力しても良い成果が得られないという環境の不確実性によるリスクも存在する。このように考えると、理想的には関数  $U_i, U_o$  も努力水準だけでなく、環境の影響を表す変数を含む確率変数と考えるのが望ましい。しかし、分析が極端に複雑になるのを避けるため、ここでは測定の際のノイズ  $\theta$  のみを考える。

・  $M_p = (1-s)(U_i + U_o) - f$

通常のエージェンシーモデルと同様、プリンシパルはリスク中立的で、得られる金銭の期待値のみに関心があるものと仮定する。

・  $M_a = s(U_i + U_o) + f - rs(U_i + U_o)z$

エージェントの金銭的効用関数は、得られる金銭の期待値とその標準偏差の定数倍の差で定義する。これは、線形トレード・オフ・モデルと呼ばれるものである [7]。確率変数  $s(U_i + U_o)\theta$  の平均は  $s(U_i + U_o)$  で、標準偏差は  $s(U_i + U_o)z$  である。ただし、 $rz \geq 1$  のとき、エージェントは組織目的のために努力をすればするほど、得られる金銭的効用が減る、あるいはいくら努力をしても金銭的効用は変わらないということになってしまうので、 $rz < 1$  を仮定する。

・  $N = u_i e_i + u_o e_o - c_i e_i^2 - c_o e_o^2$

$e_i$  や  $e_o$  が小さいとき、 $N$  はこれらの増加関数と考えるのが自然である。しかし、これら努力水準が大きくなると、疲労、余暇の減少などによる不効用が急速に増加すると考えられる。そこで、次の仮定を満たすのが望ましい。即ち、

(1)  $e_i$  や  $e_o$  が小さいとき  $\frac{\partial N}{\partial e_i} > 0, \frac{\partial N}{\partial e_o} > 0$ .

(2)  $e_i$  や  $e_o$  が大きいとき  $\frac{\partial N}{\partial e_i} < 0, \frac{\partial N}{\partial e_o} < 0$ .

$u_i, u_o$  をそれぞれ、個人目的および組織目的のための努力がもたらす非金銭的動機づけの大きさと呼ぶことにする。 $c_i, c_o$  は両目的のもたらす不効用の大きさを表すパラメーターである。 $u_i > u_o, c_i > c_o$  と仮定するのは自然であろう。

この関数では、 $e_i$  と  $e_o$  の相互作用はまったくない。つまり、個人目的のために努力することによって、組織目的のために努力する時間が減少するというようなことは考えていない。たとえば、個人目的のために努力量を投入しても、そのための疲労により組織目的のための努力量が減るようなことがない状況を考えてい

るわけである。総労働時間に関する制約が少ないケース、個人目的のための努力が気分転換になるというようなケースが本モデルでの想定状況に近い。逆に、総労働時間が一定水準をこえてはならないというような制約があったり、個人目的のための努力による疲労が組織目的遂行の障害になるというようなケースは、本モデルでの想定状況から遠くなる。

個人目的がモラル・ハザードの原因となるのは、そのような行動に時間を費やすことで、組織目的への努力に支障が出るというものと、実際にそのような行為が組織に不利益をもたらすというものの二通りが考えられる。ここでは、後者のみについて考えているというわけである。

### 3. 分析結果

前節で導入された最適化問題を解くと、

$$s = \frac{p_i c_o \{p_i(1-rz) - rz u_i\} + p_o c_i \{p_o(1-rz) - rz u_o\}}{(p_i^2 c_o + p_o^2 c_i)(1-r^2 z^2)}$$

となる（導出過程は文献[5]と同様なので、こちらを参照されたい）。ここでは、業績給の導入効果について分析するため、シェア  $s$  に注目する。ここで、 $s$  と  $u_o, u_i$  の関係に着目してみる。

$$\frac{\partial s}{\partial u_o} = \frac{-p_o c_i r z}{(p_i^2 c_o + p_o^2 c_i)(1-r^2 z^2)} < 0.$$

これより、組織目的のための努力が、大きな非金銭的動機づけをもつときは、業績給の導入が効果的でないことがわかる。この結果は、著者らが別のモデル研究により、得ているものであり、また直感的にも自然であろう[5]。

$$\frac{\partial s}{\partial u_i} = \frac{-p_i c_o r z}{(p_i^2 c_o + p_o^2 c_i)(1-r^2 z^2)}.$$

$p_i$  の符号は正の場合と負の場合が存在するが、 $s$  と  $u_i$  の関係はこの  $p_i$  の符号によって変わってくる。 $p_i$  が正の値をとり、さらに組織目的のための努力が大きな非金銭的動機づけをもつときは業績給の導入は効果的ではない。また、 $p_i$  の値が負で、さらに組織目的のための努力が大きな非金銭的動機づけをもつとき、業績給の導入は効果的となる。

ここで、理解を深めるために、数値例を用いて以上の結果を検証してみたい。

$p_i$  が正のとき

ケース 1:  $p_i=1, p_o=2, c_i=1, c_o=2, u_i=0.07, u_o=0.05, rz=0.9$  のとき

$$s=0.34$$

ケース 2:  $p_i=1, p_o=2, c_i=1, c_o=2, u_i=0.07, u_o=0.06, rz=0.9$  のとき

$$s=0.32$$

ケース 3:  $p_i=1, p_o=2, c_i=1, c_o=2, u_i=0.1, u_o=0.05, rz=0.9$  のとき

$$s=0.29$$

$p_i$  が負のとき

ケース 4:  $p_i=-1, p_o=2, c_i=1, c_o=2, u_i=0.07, u_o=0.05, rz=0.9$  のとき

$$s=0.56$$

ケース 5:  $p_i=-1, p_o=2, c_i=1, c_o=2, u_i=0.07, u_o=0.06, rz=0.9$  のとき

$$s=0.54$$

ケース 6:  $p_i=-1, p_o=2, c_i=1, c_o=2, u_i=0.1, u_o=0.05, rz=0.9$  のとき

$$s=0.61$$

ケース 1 とケース 2 の比較およびケース 4 とケース 5 の比較より、 $p_i$  の符号にかかわらず、 $s$  は  $u_o$  の減少関数となっていることがわかる。一方、ケース 1 とケース 3 の比較およびケース 4 とケース 6 の比較より、 $p_i$  の符号によって、 $s$  と  $u_i$  の関係がまったく異なってくるのがわかる。

ところで、これらの比較静学分析結果と同様のものは、組織目的に関わる要素 ( $e_o, p_o, u_o, c_o$ ) を除いたとしても得ることができる。その意味で、これらの要素は分析結果を得るために必要不可欠なものではない。しかし、 $p_i$  が負のとき、組織目的への努力水準がないとしたら、プリンシパルは負の成果しか得られなくなってしまふ。その意味で現実的に意味のあるモデル設定をするという目的のために組織目的に関わる要素は必要となってくる。本論文のオリジナリティは、ただ単に複数の目的を考慮に入れたモデルを提案したことではなく（これは文献[5, 6]によってすでに提案されている）、エージェントの個人目的が負の生産性をもちうるケースも考慮に入れるという意味で、モラル・ハザード概念を拡張した点にある。

### 4. 結論

前節の比較静学分析の結果をまとめると、次の通りである。

1. 組織目的のための努力が、小さな非金銭的動機づけしかもたないとき、業績給導入が効果的となる。
2. 組織に不利益をもたらすような個人目的のための努力が、大きな非金銭的動機づけをもつとき、業績

給導入が効果的となる。

3. 組織に利益をもたらすような個人目的のための努力が、小さな非金銭的動機づけしかもたないとき、業績給導入が効果的となる。

1は、今回のモデル分析によって新たに得られた知見ではないが、2, 3は新たなモラル・ハザード概念を考えたモデルによる分析によってはじめて得られる知見である。まず2であるが、組織に不利益をもたらすような個人目的のための努力が、大きな非金銭的動機づけをもつとき、業績給の割合を高くして、組織目的への金銭的動機づけを高めると同時に、私的行為による成果の落ち込みを、金銭的報酬を減少させることによってカバーするのが効果的になるということをいっている。次に3であるが、組織に利益をもたらすような個人目的のための努力が、小さな非金銭的動機づけしかもたないとき、業績給の導入により金銭的動機づけを高める必要が高いということをいっている。

このように、個人目的のための努力が、組織に対して利益をもたらす場合とそうでない場合で、インセンティブ・システムの設計方法がまったく異なることがわかる。たとえば、官僚的リーダーシップが有効でない組織において、官僚的リーダーシップを発揮しようという非金銭的動機づけが大きいマネージャーには、業績給導入が効果的となるが、研究開発組織などにおいて、研究者のもつ個人目的が組織に対して正の利益をもたらすようなときは、そのような個人目的への非金銭的動機づけが大きい研究者へは、業績給導入は効果的でないといえる。

本研究では、エージェントの行動が組織に対して負の利益をもたらすこともありうるという、より包括的なモラル・ハザード概念を表現しうるモデルを提案し、個人目的のもつ組織への生産性の値の正負によって、つまり、個人目的のための努力が組織のためにプラスになるのか直接マイナスにはたらくのかの違いによって、ISの設計方法がまったく異なってくることを示

した。今後は、様々なケースのモラル・ハザードを表現するため、特に非金銭的効用関数の設定を変化させながら分析をすすめる予定である。今回の分析に当たっては、関数の設定方法に結果が依存しているという面もある。特に、 $e_i$ と $e_o$ の相互作用はないという仮定は本論文の結論を導くのに不可欠なものである。ただし、その他の設定は、あくまで解析的な分析を容易にするためのもので、たとえば、成果関数の線形性をゆるめたり、職務の遂行による不効用を下に凸な別の関数におきかえたりしても、同様な結論に達するものと思われる。本モデルは、分析の容易性という観点から、直観に反しない範囲で最適な関数形を選択したつもりである。

**謝辞** 最後に、貴重なご助言を賜りました査読者の先生方に、感謝の言葉を述べさせて頂きたいと思います。どうも有り難うございました。

#### 参考文献

- [1] 谷内正文, 「エージェント・モデルについて」, オペレーションズ・リサーチ, 1983年11月号, pp. 558-564.
- [2] Spremann, K., 'Agent and Principal', in "Agency Theory, Information, and Incentives", Springer-Verlag, 1987.
- [3] 森俊治他, 『研究開発管理の理論と体系』, 丸善, 1978.
- [4] 桑田耕太郎, 田尾雅夫, 「組織論」, 有斐閣, 1998.
- [5] 松村良平, 中野文平, 猪原健弘, 高橋真吾, 「複数の目的をもった労働者に対するインセンティブ・システムの設計方法に関する分析」, オペレーションズ・リサーチ, 1999年11月号, pp. 621-626.
- [6] Holmstrom, B. and Milgrom, P., "Multitask Principal-Agent Analysis: Incentive Contracts, Asset Ownership, and Job Design", Journal of Law, Economics and Organization, 7 Special Issue, pp. 24-52, 1991.
- [7] 伊丹敬之, 『マネジメント・コントロールの理論』, 岩波書店, 1986.