香 川 大 学 経 済 論 叢
第86巻 第1号 2013年6月　29－44

アプローチの選択を伴う 複数エージェントのモラルハザード

天 谷 研 一

1　はじめに

組織の経済学における最も重要な問題の一つは、組織の構成員に対する動機付けである。組織において動機付けが問題となる理由は、個々の組織構成員が自己の利益を求めて行動することは必ずしも組織全体としての効率的な結果には結びつかないところにある。組織の経済学では、動機付けの問題がプリンシパル・エージェント問題の枠組みの中で分析されてきた。

ほとんどの組織では、複数のエージェントが雇用されている。異なるエージェントに課される任務と任務の間の関係性は、組織によって違いがある。インセンティブの経済理論では、任務間の様々な関係性について研究が行われてきた。

第一に、異なるエージェントは独立した異なる任務に割り当てられる場合がある。例えば、企業の労働者のある者には研究開発が期待される一方で、別の労働者は新規の顧客を開拓することが求められる。このような状況における動機付けの問題は、Lazear and Rosen (1981), Mookerjee (1984), Itoh (1991) らによって研究されてきた。

第二に、複数のエージェントは共通の目的に向かって共同して働くことが期待される場合がある。これはチームワークの状況であり、ここで生じるインセンティブの問題はHolmström (1982) らによって研究されてきた。

本論文では、上記の二つとは異なる任務間の関係性について考察する。それは、複数のエージェントは共通の目的を追求するが、独立に行動するという場合である。

例として、新薬の開発を行う研究所のストーリーを考えよう。ここでは、複数のエージェント、すなわち研究員が雇われている。研究所には、新型インフル

エンザのワクチンを開発するという、共通の目的がある。エージェントである各研究員は、各々独自に研究を行うとする。研究所という組織としては、誰か一人の研究員がワクチンの開発に成功すれば十分である。すでに一人が成功している場合に追加的にもう一人が成功しても、その追加的価値は小さい。それでも、研究員の数を増やせば、少なくとも一人が成功する確率を増加させることができるので、複数のエージェントを雇うことに意味がある。

本論文では、このような組織において、どのようなインセンティブ体系が最適になるかを検討する。以下のようなモデルを考える。組織（プリンシパル）は2人の労働者（エージェント）を雇う。プリンシパルが達成したいと考えるプロジェクトが一つある。各エージェントの成果は、成功か失敗のいずれかであり、個々のエージェントの行動は自分自身の成果にのみ影響する。プリンシパルにとっては、どちらか一方のエージェントが成功すれば十分である。

プロジェクトには、複数のアプローチがあると考える。新薬開発の例では、ワクチンを開発するという一つの目標に向かうための、様々な研究方法があると考えられる。標準的なプリンシパル・エージェント問題の分析と同様に、各エージェントの成果は不確実性に晒されている。すなわち、努力をしていても、確率的に失敗することがある。本論文のモデルでは、どのアプローチがうまく行く方法であるのかが人々には分からないという点に、不確実性があるとする。もしアプローチが「良い」ものであれば、どのエージェントがそれを用いて働いているかに関係なく、成果は成功となる可能性が高い。一方、アプローチが「悪い」ものであれば、エージェントが誰かに関わりなく失敗する可能性が高い。このような状況では、プリンシパルは、二人のエージェントが異なるアプローチをとることを望む。

各エージェントはアプローチを一つ選び、そこに投入する努力水準を決定する。高い努力水準は、成功確率を高めるが、エージェントにより高い努力コストをもたらす。本論文のモデルでは、2つのアプローチが選択可能であるとする。一方のアプローチは、他方よりも事前の意味でより効率的である。任意の同一の成功確率を達成するための努力コストは、効率的なアプローチを選んだ

場合のほうが、他方を選んだ場合よりも小さい。契約は、最終的な結果のみに依存させることができる。エージェントが選択したアプローチは、観察することができないため、契約に含めることができない。もし各エージェントの賃金支払いが本人の成果のみに依存しているとすると、全てのエージェントは事前の意味で効率的なアプローチを選ぼうとする。しかし、プリンシパルは、少なくとも一つの成功を得る確率を高めるために、一方のエージェントは非効率的なアプローチをとることを望む。

本論文の分析において、最適契約は、次のような非対称的な構造を持つことが示される。一方のエージェントの賃金は自分の成果のみに依存し、もう一人のエージェントの賃金は両エージェントの成果に依存する。

この結論の直観は以下のとおりである。最適契約では、一方のエージェント（一般性を失うことなく、エージェント１と呼ぶ）に効率的なアプローチを取らせ、もう一方のエージェント（エージェント２）には非効率的なアプローチを取らせる。エージェント１の賃金は、本人の成果のみに依存させる。このような賃金体系のもとでは、エージェント１は実際に効率的なアプローチを取るインセンティブを持つ。エージェント２が非効率的なアプローチを選ぶよう動機づけるためには、賃金は本人の成果だけではなく別のエージェントの成果にも依存させなければならない。プリンシパルは、逸脱して効率的なアプローチをとるインセンティブをエージェント２に持たせないという誘因整合性条件のもとで最適化問題を解くことになる。エージェント２が逸脱して効率的なアプローチをとると、エージェント１も同じアプローチをとっているので、成果は一致する可能性が高くなる。従って、エージェント２に正しいインセンティブを与えるためには、エージェント１とは異なる結果を出していることに対して報酬を与えるべきである。よって、エージェント２が成功することによるボーナスは、エージェント１が失敗している場合に、より大きくなる。さらには、エージェント１が成功している場合にはエージェント２の成功をむしろ罰することが最適になる場合もある。

以下では、第２節でモデルを提示する。第３節はベンチマークとなるケースを

いくつか分析する。第 4 節において主要な結論を示す。

2 モデル

プリンシパルと、2 人のエージェント $n = 1, 2$ の関係を考える。エージェントは、各々別々に仕事を行う。エージェント n の仕事の結果（もしくは成果）を y_n で表す。これは、2 通りの値 $y_n \in \{S, F\}$ をとる。エージェント n について、$y_n = S$ は結果が成功であることを、$y_n = F$ は結果が失敗であることを示す。2 人のエージェントは対称的であること、すなわち、選好や技術的環境に相違がないことを仮定する。2 人のエージェントの成果がともに F であるとき、プリンシパルの収入は 0 である。少なくとも 1 人のエージェントの成果が S である場合には、プリンシパルは R の収入を得る。2 人ともに成果が S である場合、プリンシパルの収入はやはり R である。すなわち、すでに一つの成功がある場合、もう一つ目の成功は何ら追加的価値をもたない。

仕事を行うための 2 つの異なるアプローチ $k = 1, 2$ がある。各エージェントは、どちらかのアプローチを選ぶ。エージェント n がとるアプローチを k_n で表す。エージェントはさらに努力水準を選ぶ。より高い努力水準は成功確率を高める。ここでは努力水準を成功確率そのもので測ることにし、エージェント n の努力水準、すなわち成功確率を e_n で表す。従って、エージェント n の意思決定は $a_n = (k_n, e_n)$ を選ぶことである。ここで、$k_n \in \{1, 2\}$ と $e_n \in [0, 1)$ が成り立つ。両エージェントの意思決定のベクトルを $a = (a_1, a_2)$ とする。

努力のコストは両エージェントに共通で、関数 $G(a_n) = G(k_n, e_n)$ で表される。各 $k = 1, 2$ について、$g_k(e_n) = G(k_n, e_n)$ とする。すなわち、$g_k(e_n)$ は、アプローチ k を取り、成功確率が e_n であるような努力水準を選んだ場合の努力コストである。以下では、簡潔化のため、誤解が生じない範囲内で、添え字 n を省略し e と書いて e_n を表すことがある。

仮定 1 費用関数 $g_k(e)$ は *2* 回微分可能であり、

(i) $k = 1, 2$ について、$g_k(0) = 0$, すべての $e \in (0, 1)$ に対して $g'_k(e) > 0$,

$g'_k(0)=0$, $\lim_{e\to 1} g'_k(e)=+\infty$, すべての $e\in[0,1)$ に対して $g''_k(e)>0$ が成り立つ。

(ii) すべての $e\in(0,1)$ に対して $g_1(e)<g_2(e)$ と $g'_1(e)<g'_2(e)$ が成り立つ。

仮定のうち、(i) は、費用関数が成功確率に関する増加関数であり凸関数であることを述べている。また、ファーストベストとセカンドベストは内点解を持つことを保証している。一方、(ii) は、アプローチ１はアプローチ２よりも費用、限界費用ともに低いことを仮定している。

２人のエージェントが異なるアプローチをとった場合には、両者の成果は互いに独立である。一方、２人のエージェントが同じアプローチをとった場合には、次の意味で「完全に相関」している[1]。

第一に、両エージェントの努力水準が同一の場合、両エージェントの成果は必ず一致する。すなわち、２人ともに成功するか、もしくは２人ともに失敗する。第二に、２人の努力水準が異なる場合には、努力水準が低いエージェントが成功するならば、努力水準が高いエージェントも必ず成功する。反対に、努力水準が高いエージェントが失敗するならば、努力水準が低いエージェントも必ず失敗する。数学的には、以下のようになる。もし $k_n=k_m$ かつ $e_n\geq e_m$ ならば、確率 e_m で $y_n=y_m=S$、確率 e_n-e_m で $y_n=S$ かつ $y_m=F$、確率 $1-e_n$ で $y_n=y_M=F$ となる。

この仮定の背後にある考え方は、次のとおりである。努力水準を一定としたとき、成果は不確実である。すなわち、成功するかもしれないし失敗するかもしれない。ここでは、不確実性が純粋にアプローチの質に関するものであるような状況を考えている。個々のアプローチは、正しい、すなわちうまく行くやり方かもしれないし、間違った、すなわちうまく行かないやり方かも知れない。しかし、実際にそのアプローチを使って仕事をしてみない限り、うまくいくものであるかどうかは分からない。人々は、アプローチ１のほうが正しいやり方である可能性が高いという共通の予想を持っている。個々のエージェントに特有の不確実性は存在しない。二人のエージェントが同一のアプローチをとって

[1]統計学の通常の意味での完全相関とは異なる。

いるならば、より高い努力水準を選んでいるエージェントが、必ずより高い成果をあげる。

エージェントによる意思決定が $a = (a_1, a_2) = ((k_1, e_1), (k_2, e_2))$ であるとき、少なくとも一方のエージェントが成功する確率は

$$P(a) = \begin{cases} e_1 + e_2 - e_1 e_2 & \text{if} \quad k_1 \neq k_2 \\ \max\{e_1, e_2\} & \text{if} \quad k_1 = k_2 \end{cases}.$$

となる。プリンシパルがエージェント n に支払う賃金を w_n とする。プリンシパルはリスク中立的であると仮定する。従って、プリンシパルの期待効用は $P(a) \cdot R - E[w_1 + w_2]$ である。

エージェント n の効用は $u(w_n) - G(k_n, e_n)$ で与えられる。ここで、$u(w_n)$ は賃金効用を、$G(k_n, e_n)$ は努力コストである。本論文の主要な関心は、エージェントがリスク回避的である場合、すなわち $u(\cdot)$ が凹関数である場合である。ただし、ベンチマークとしてリスク中立的な場合も考察する。

プリンシパルとエージェントの関係は以下のように進行する。まず、ステージ1において、プリンシパルは各エージェントに契約、すなわち賃金体系を提示する。プリンシパルはそれぞれのエージェントに異なる賃金体系を提示することができる。提示された契約は両方のエージェントに観察される。ステージ2において、エージェントは同時に契約を受諾するか拒否するかを決定する。少なくとも一方のエージェントが契約を拒否した場合、関係はここで終了し、両エージェントは外部効用 U_0 を受け取る。一般性を失うことなく、$U_0 = 0$ とおく。両エージェントが契約を受諾した場合にはステージ3に進み、エージェントは同時にアプローチと努力水準を決定する。ステージ4において、結果が実現し、契約に従って賃金が支払われる。

ここでは、一方のエージェントが契約を拒否した場合には、もう一人のエージェントとも関係は継続しないと仮定している。もちろん、一人のエージェントだけが受諾すれば、そのエージェントのみが雇われて仕事をすると考えることもできる。このようなケースを考えても結果は変わらない。ここでは議論の単純化のために上記のような仮定をおくことにする。

プリンシパルは、エージェントが選択したアプローチも努力水準も観察できない。できることは、報酬を成果に依存させることである。エージェントへの賃金支払いは、そのエージェント自身の成果だけでなく、もう一人のエージェントの成果にも依存させることができる。エージェント n への賃金スケジュールは、$w^n = (w^n_{SS}, w^n_{SF}, w^n_{FS}, w^n_{FF})$ で表される。ここで、w^n_{ij} は、エージェント 1 の成果が i、エージェント 2 の成果が j である場合のエージェント n への賃金支払い額である。

成果が ij であるときのエージェント n の賃金効用を $v^n_{ij} = u(w^n_{ij})$ とする。また、$h(\cdot)$ を、$u(\cdot)$ の逆関数とする。プリンシパルは $w^n = (w^n_{SS}, w^n_{SF}, w^n_{FS}, w^n_{FF})$ を選択するが、これは $v^n = (v^n_{SS}, v^n_{SF}, v^n_{FS}, v^n_{FF})$ を選択することと同義である。以下では分析上の扱いやすさの理由から、w^n ではなく v^n を選択変数とする。エージェント n に賃金効用 v^n_{ij} を与えるために必要な賃金支払いは、$w^n_{ij} = h(v^n_{ij})$ となる。

3　ベンチマーク

3.1　最善：行動が立証可能な場合

エージェントの行動が観察・立証可能な場合には、報酬を行動に依存させることができる。この場合、プリンシパルはいかなるアプローチと努力水準をも実現できる。実現しようとする行動から逸れた場合にエージェントを十分大きく罰する賃金体系にすれば良い。すなわち、プリンシパルがエージェント n に行動 a^*_n を選ばせようとする場合、エージェントが実際に行動 a^*_n を選んだ場合には成果に応じて報酬 w^n_{ij} を与え、それ以外の行動を選んだ場合には極めて低い報酬を与える。均衡経路上ではエージェントは逸脱しないので、罰されることは無い。プリンシパルはリスク中立的でエージェントはリスク回避的であるので、最適契約では、エージェントは固定賃金を受け取る。すなわち、各 n について、実際に a^*_n が選ばれ、結果が ij であった場合の賃金効用を v^n_{ij} とおくと、あらゆる ij に対して $v^n_{ij} = v_n$ となる。このとき、最適な賃金契約は、以下の問題の解

となる。

$$\max_{a_1,a_2,v_1,v_2} P(a)\cdot R - h(v_1) - h(v_2)$$

subject to

$$(IR)\ \ v_n - G(a_n) \geq 0,\ \ n = 1, 2.$$

ここで、(IR) は個人合理性条件である。明らかに、2 人のエージェントに同じアプローチをとらせることは最適にならない。一般性を失うことなく、最適契約ではエージェント 1 にアプローチ 1 を、エージェント 2 にアプローチ 2 を取らせるとしよう。すると、個人合理性条件は等号で成立するので、$v_1 = g_1(e_1)$, $v_2 = g_2(e_2)$ となる。これより、最適契約問題は以下のようになる。

$$\max_{e_1,e_2} R(e_1 + e_2 - e_1e_2) - h(g_1(e_1)) - h(g_2(e_2)).$$

一階の条件は

$$R(1 - e_2) - h'(g_1(e_1))g'(e_1) = 0$$
$$R(1 - e_1) - h'(g_2(e_2))g'(e_2) = 0$$

である。これより、次の結果が容易に分かる（証明は省略する）。

補題 1 最適な努力水準は $e_1^* > e_2^* > 0$ を満たす。

努力の限界費用は $e_n = 0$ において 0 であるので、最適な努力水準は正の値をとる。

3.2　リスク中立的なエージェント

次に、行動は観察可能でないが、エージェントがリスク中立的な場合を考えよう。エージェントの賃金効用は $u(w) = w$ である。

全ての参加者の効用の合計を、社会的余剰と定義する。この場合、ファーストベストのエージェントの決定 a^* は、社会的余剰を最大化する。従って、ファー

ストベストにおいては、一般性を失うことなく $k_1^* = 1,\ k_2^* = 2$ であり、かつ、(e_1^*, e_2^*) は次の問題の解である。

$$\max_{e_1,e_2} R(e_1 + e_2 - e_1 e_2) - g_1(e_1) - g_2(e_2).$$

一階の条件は

$$R(1 - e_2^*) - g_1'(e_1^*) = 0$$
$$R(1 - e_1^*) - g_1'(e_2^*) = 0.$$

で与えられる。

標準的なインセンティブの理論におけるものと同様に、エージェントがリスク中立的である場合には、エージェントの取り分を総余剰から定額を増減したものにすることによって、ファーストベストを達成することができる。すなわち、最適契約は、$n = 1, 2$ について

$$w_{FF}^n = w_0^n,\ \ w_{SS}^n = w_{SF}^n = w_{FS}^n = w_0^n + R$$

を見たし、w_0^n は個人合理性条件が等式で満たされる水準に選ばれた値となる。

3.3 立証可能なアプローチ

もう一つの興味深いベンチマークは、各エージェントが選ぶアプローチは観察・立証可能だが、努力水準は観察できない場合である。

一般性を失うことなく、最適契約ではエージェント 1 にアプローチ 1 を、エージェント 2 にアプローチ 2 を選ばせるものとしよう。

Grossman and Hart (1983) の方法に従い、最適契約を 2 段階で求める。第一に、与えられた意思決定の組み合わせ $a = (a_1, a_2)$ に対して、それを最小費用で実現する賃金体系を求める。第二に、最適な意思決定の組み合わせを探す。

プリンシパルは、エージェント $n(= 1, 2)$ に、行動 $a_n = (k_n, e_n) = (n, e_n)$ を選ばせようとしているとしよう。2 人のエージェントが異なるアプローチをとる場合、成果は互いに独立である。よって、一方のエージェントの成果は、他のエー

ジェントの努力水準に関して情報を含まない。Holmström (1979) の十分統計量に関する結果に従えば、最適契約における各エージェントの報酬は自分自身の成果のみに依存することになる。従って、プリンシパルにとっての最適契約を考える上での第一ステップは、各エージェントについて別々に、一定の努力水準を選ばせるための費用最小化問題を考えることである。ここで、$v_S^1 = v_{SS}^1 = v_{SF}^1$, $v_F^1 = v_{FS}^1 = v_{FF}^1$, $v_S^2 = v_{SS}^1 = v_{FS}^1$, $v_F^2 = v_{SF}^2 = v_{FF}^1$ とおく。

エージェント n に対する費用最小化問題は以下のように与えられる。

$$\min_{v_S^n, v_F^n} e_n h(v_S^n) + (1-e_n) h(v_F^n)$$

subject to

$$\text{(IC)} \;\; e_n v_s^n + (1-e_n) v_F^n - g_n(e_n) \geq e_n' v_s^n + (1-e_n') v_F^n - g_n(e_n') \forall e_n',$$

$$\text{(IR)} \;\; e_n v_s^n + (1-e_n) v_F^n - g_n(e_n) \geq 0.$$

ここで、(IC) は誘因整合性条件、(IR) は個人合理性条件である。

First order approach (Mirrlees (1975), Rogerson (1985)) に従い、誘因整合性 (IC) 条件は、以下の一階の誘因整合性 (FOIC) 条件に書き換えることができる。

$$\text{(FOIC)} \;\; v_S^n - v_F^n = g_n'(e_n)$$

また、個人合理性 (IR) 条件が等号で成立することは容易に確認できる。従って、

$$e_n v_s^n + (1-e_n) v_F^n - g_n(e_n) = 0$$

となる。これら 2 つの条件より、以下を得る。

$$v_S^n = g_n(e_n) + (1-e_n) g_n'(e_n)$$

$$v_F^n = g_n(e_n) - e_n g_n'(e_n).$$

エージェント n にアプローチ n と努力水準 e_n を選ばせるための期待賃金支払額は

$$\begin{aligned} C_n(e_n) &= e_n h(v_S^n) + (1-e_n) h(v_F^n) \\ &= e_n h(g_n(e_n) + (1-e_n) g_n'(e_n)) + (1-e_n) h(g_n(e_n) - e_n g_n'(e_n)) \end{aligned}$$

となる。

第2ステップは、上記で導出した期待賃金支払額を用いて、プリンシパルにとって最適な努力水準を求めることである。両エージェントが努力水準 e_n $(n = 1, 2)$ を選択するときに少なくとも1人のエージェントが成功する確率は $P(a) = e_1 + e_2 - e_1 e_2$ であるので、プリンシパルの期待効用は $R(e_1+e_2-e_1e_2)-C_1(e_1)-C_2(e_2)$ である。従って、第2ステップは以下の問題を解くことになる。

$$\max_{e_1, e_2} R(e_1 + e_2 - e_1 e_2) - C_1(e_1) - C_2(e_2)$$

これを解くと、以下が容易に示される（証明は省略）。

補題 2 最適解においては $e_1^* > e_2^* > 0$ が成り立つ。

4　エージェントがリスク回避的である場合の最適契約

本節では、本論文の主要な関心の対象、すなわち、エージェントがリスク回避的で、行動がアプローチ、努力水準ともに観察不可能である場合の契約について考える。

以下の結論が得られる。

定理 1 最適契約で選ばれる行動は、$a_1 = (k_1, e_1) = (1, e_1)$, $a_2 = (k_2, e_2) = (2, e_2)$, $e_1 > 0$, $e_2 > 0$ を満たすとする。このとき、最適な賃金体系は以下を満たす。

(i) エージェント1 *:* $w_{FS}^1 = w_{FF}^1 < w_{SS}^1 = w_{SF}^1$

(ii) エージェント2 *:*

- $w_{FF}^2 < w_{SF}^2 \leq w_{SS}^2 < w_{FS}^2$ または
- $w_{FF}^2 < w_{FS}^2$ かつ $w_{SS}^2 < w_{SF}^2$

プリンシパルは、エージェント1が効率的なアプローチを、エージェント2が非効率的なアプローチをとることを望む。定理の (i) は、エージェント1の賃

金は、本人の成果のみに依存することを示している。一方、(ii) は、エージェント２の賃金は本人の成果だけでなく、エージェント１の成果にも依存することを示している。エージェント２の賃金の形状には二つの可能性がある。第一のケースでは、エージェント１の成否にかかわらずエージェント２は自分自身の成功に対してボーナスを受け取るが、ボーナスはエージェント１が失敗した場合により大きい。第二のケースでは、エージェント１が成功している場合にはエージェント２は自分が成功することによってむしろ賃金が低くなる。

複数のエージェントが選択する努力水準がそれぞれ自分自身の成果に影響を与える状況におけるモラルハザードの状況で、異なるエージェントの間の成果が正の相関を持つ場合については、先行研究において、「相対業績評価」が望ましいという結果が知られている (Lazear and Rosen (1981), Mookherjee (1984))。各エージェントの成果が成功または失敗であるような場合にこのロジックをあてはめると、エージェントの賃金は、自分が失敗し、もう一人のエージェントが成功している場合に最も低くなる。この結果の直観は以下の通りである。あるエージェントが失敗しているとすると、それは怠けていたことが理由かもしれないし、あるいは単に運が悪かっただけかもしれない。結果に正の相関がある状況では、一人のエージェントが運が悪いゆえに失敗している場合には、他のエージェントも同様に悪い運に直面して結果的に失敗している可能性が高い。インセンティブとリスク分担の観点からは、そのような結果に対してエージェントを罰することは好ましくない。一方、他のエージェントが成功しているにもかかわらず本人が失敗した場合には、それは怠けたことのシグナルである可能性が高いので、罰せられるべきである。

本論文のモデルにおいては、エージェント２の賃金は他のエージェントの成果に依存するが、その依存のしかたは上記の「相対業績評価」の場合とは異なる。ここでの問題は、3.3 節で議論したベンチマークケースに、追加的な誘因整合性条件が加わったものである。その条件とは、各エージェントは別のアプローチに逸脱する誘因を持たないというものである。従って、3.3 節でとりあげた問題を、ここでの問題に対する「緩和された問題 (relaxed problem)」と見なすこと

ができる。すなわち、緩和された問題の解が追加的な制約を満たすならば、それはそのままもとの問題に対する解となる。そうでない場合には、異なる解を探すことになる。

もしエージェントの賃金が本人の成果のみに依存するならば、そのエージェントにとってアプローチ２をとることは決して最適にならない。あるエージェントがアプローチ２で努力水準 e を選んでいるとしよう。もし、ここからアプローチ１に変更して同じ e を選べば、受け取る賃金の分布は変わらないまま、努力のコストを減らすことができる。従って、エージェント１にとっては、緩和された問題の解は、追加的な制約を満たす。一方、エージェント２にとっては、緩和された問題の解は、追加的な制約を満たしていない。

定理の (i) は、上記の議論によって証明された。以下では、(ii) を証明する。

エージェント２に対する費用最小化問題は、次のものである。

$$\min_{v^2} e_1^* e_2^* h(v_{SS}^2) + e_1^*(1-e_2^*)h(v_{SF}^2) + (1-e_1^*)e_2^* h(v_{FS}^2) + (1-e_1^*)(1-e_2^*)h(v_{SS}^2)$$

subject to

$$\text{(IR)}\ \ e_1^* e_2^* v_{SS}^2 + e_1^*(1-e_2^*)v_{SF}^2 + (1-e_1^*)e_2^* v_{FS}^2 (1-e_1^*)(1-e_2^*)v_{FF}^2 - g_2(e_2^*) \geq 0$$

$$\text{(IC}_1\text{)}\ \ e_1^* e_2^* v_{SS}^2 + e_1^*(1-e_2^*)v_{SF}^2 + (1-e_1^*)e_2^* v_{FS}^2 (1-e_1^*)(1-e_2^*)v_{FF}^2 - g_2(e_2^*)$$
$$\geq e_1^* e_2 v_{SS}^2 + e_1^*(1-e_2)v_{SF}^2 (1-e_1^*)e_2 v_{FS}^2 (1-e_1^*)(1-e_2)v_{FF}^2 - g_1(e_1)\ \ \forall e_2$$

$$\text{(IC}_2\text{)}\ \ e_1^* e_2^* v_{SS}^2 + e_1^*(1-e_2^*)v_{SF}^2 + (1-e_1^*)e_2^* v_{FS}^2 (1-e_1^*)(1-e_2^*)v_{FF}^2 - g_2(e_2^*)$$
$$\geq e_2 v_{SS}^2 + (e_1^* - e_2)v_{SF}^2 + (1-e_1^*)v_{FF}^2 - g_1(e_2)\ \ \forall e_2 \in [0, e_1^*]$$

$$\text{(IC}_3\text{)}\ \ e_1^* e_2^* v_{SS}^2 + e_1^*(1-e_2^*)v_{SF}^2 + (1-e_1^*)e_2^* v_{FS}^2 (1-e_1^*)(1-e_2^*)v_{FF}^2 - g_2(e_2^*)$$
$$\geq e_1^* v_{SS}^2 + (e_2 - e_1^*)v_{FS}^2 + (1-e_2)v_{FF}^2 - g_1(e_2)\ \ \forall e_2 \in [e_1^*, 1]$$

ここで、(IR) は個人合理性条件である。(IC_1) から (IC_3) は、誘因整合性条件である。(IC_1) は、アプローチ２をとるが、異なる努力水準を選ぶという逸脱が起きないための条件である。(IC_2) と (IC_3) は、アプローチ１に逸脱しないため

の条件である。これは、逸脱する先の努力水準が、エージェント1が選ぶ努力水準 e_1^* よりも小さいか大きいかによって場合分けされる。

以下では、誘因整合性条件 (IC_1) から (IC_3) を書き換えて、それらの関係性を見ていこう。

まず、(IC_1) は、First Order Approach を用いて、一階の誘因整合性

$$(e_1^* v_{SS}^2 + (1-e_1^*)v_{FS}^2) - (e_1^* v_{SF}^2 + (1-e_1^*)v_{FF}^2) = g_2'(e_2^*) \tag{1}$$

と同値となる。

次に、(IC_2) を考えよう。左辺を変形すると

$$e_1^* v_{SF}^2 + (1-e_1^*)v_{FF}^2 + e_2^*\{(e_1^* v_{SS}^2 + (1-e_1^*)v_{FS}^2) - (e_1^* v_{SF}^2 + (1-e_1^*)v_{FF}^2)\}$$
$$-g_2(e_2^*)$$

に、右辺を変形すると

$$e_2(v_{SS}^2 - v_{SF}^2) + (e_1^* v_{SF}^2 + (1-e_1^*)v_{FF}^2) - g_1(e_2)$$

になる。従って、制約条件 (IC_2) は、以下と同値である。

$$e_2^* g_2'(e_2^*) - g_2(e_2^*) \geq e_2(v_{SS}^2 - v_{SF}^2) - g_1(e_2) \ \ \forall e_2 \in [0, e_1^*]$$

すなわち

$$v_{SS}^2 - v_{SF}^2 \leq \frac{e_2^* g_2'(e_2^*) - g_2(e_2^*) + g_1(e_2)}{e_2} \ \ \forall e_2 \in [0, e_1^*] \tag{2}$$

となる。

次に、(IC_3) を考えよう。左辺を変形すると

$$e_1^* v_{SS}^2 + (1-e_1^*)v_{FS}^2\} - (1-e_2^*)\{(e_1^* v_{SS}^2 + (1-e_1^*)v_{FS}^2) - (e_1^* v_{SF}^2 + (1-e_1^*)v_{FF}^2)\}$$
$$-g_2(e_2^*)$$

に、右辺を変形すると

$$(e_1^* v_{SS}^2 + (1-e_1^*)v_{FS}^2) - (1-e_2^*)(v_{FS}^2 - v_{FF}^2) - g_1(e_2)$$

になる。従って、制約条件 (IC_3) は、以下と同値である。

$$-(1-e_2^*)g_2'(e_2^*) - g_2(e_2^*) \geq -(1-e_2)(v_{FS}^2 - v_{FF}^2) - g_1(e_2) \ \ \forall e_2 \in [e_1^*, 1],$$

すなわち

$$v_{FS}^2 - v_{FF}^2 \geq \frac{(1-e_2^*)g_2'(e_2^*) - g_2(e_2^*) + g_1(e_2)}{1-e_2} \ \ \forall e_2 \in [e_1^*, 1]. \tag{3}$$

となる。

もし $e_2^* \in [0, e_1^*]$ ならば (2) より $v_{SS}^2 - v_{SF}^2 < g_2'(e_2^*)$ となる。これと (1) により、$v_{FS}^2 - v_{FF}^2 > g_2'(e_2^*)$ が成立する。一方、もし $e_2 \in [e_1^*, 1]$ ならば、(3) より $v_{FS}^2 - v_{FF}^2 > g_2'(e_2^*)$ となる。これと (1) により、$v_{SS}^2 - v_{SF}^2 < g_2'(e_2^*)$ が成立する。つまり、e_2^* の値にかかわらず、

$$v_{SS}^2 - v_{SF}^2 < g_2'(e_2^*) < v_{FS}^2 - v_{FF}^2$$

となる。$v_{SS}^2 - v_{SF}^2 \geq 0$ のときは、定理の第一のケースのように $w_{FF}^2 < w_{SF}^2 \leq w_{SS}^2 < w_{FS}^2$ が成立し、$v_{SS}^2 - v_{SF}^2 < 0$ のときは、定理の第二のケースのように $w_{FF}^2 < w_{FS}^2$ かつ $w_{SS}^2 < w_{SF}^2$ が成立する。

参考文献

[1] Grossman, S. J., and O. D. Hart (1983) "An Analysis of the Principal-Agent Problem," *Econometrica*, 51, 7-45.

[2] Holmström, B. (1979) "Moral Hazard and Observability," *Bell Journal of Economics*, 10, 74-91.

[3] Holmström, B. (1982) "Moral Hazard in Teams," *Bell Journal of Econoomics*, 13, 324-40.

[4] Itoh, H. (1991) "Incentives to Help in Multiagent Situations," *Econometrica*, 59, 611-36.

[5] Lazear, E. P., and S. Rosen (1981) "Rank-Order Tournaments as Optimum Labor Contracts," *Journal of Political Economy*, 89, 841-64.

[6] Mirrlees, J. A. (1975) "The Theory of Moral Hazard and Unobservable Behaviour, Part 1," mimeo, Oxford. (Published in 1999 in the *Review of Economic Studies*, 66,3-21.)

[7] Mookherjee, D. (1984) "Optimal Incentive Schemes with Many Agents," *Review of Economic Studies*, 51, 433-46.

[8] Rogerson, W. P. (1985) "The First-Order Approach to Principal-Agent Problems," *Econometrica*, 53, 1357-67.