ギオーム・ハーリンジャー著、栗野盛光訳
「マーケットデザイン：オークションとマッチングの理論・実践」（中央経済社発行）
付録（熊野太郎訳）

付録 B: メカニズムデザイン

付録 A では基礎的なゲーム理論を概観し、いくらかのゲームの分析方法を提供した。この章ではゲームの分析を通して、そのゲームがどうして生じたのかという問いと答えを模索する。そのためには付録 A の再確認が肝要である。この疑問を専門的に扱う分野はメカニズムデザインとして知られている。オークションとマッチング問題は実はメカニズムデザインの特殊なケースである。メカニズムデザインの主目的は特定の帰結を導くためにゲームを設計することにある。後にみるように、メカニズムデザインは不完備情報の伴う状況において特に興味深い。

B.1 準備

メカニズムデザインは個人間でどのように相互作用するルールが帰結に対してどのように影響するかを非常に普遍的または抽象的な方法を用いて理解することを目指す。メカニズムデザインの一般的な方法は、特定の目的を達成するためにルールの集合である**メカニズム**を設計することである。その設計は個人が戦略的である可能性と最終的な目的に影響を及ぼす私的情報を保持している可能性を考慮に入れて行う。

オークションを例にとると、２章では入札ルールが入札者の入札行動にどのように影響を与え、最終的にはルールが最終配分（誰が財を手に入れいくら支払うか）にどのような影響を及ぼすかを学んだ。これよりオークションは、個人とは買い手であり彼らのもつ私的情報とは彼らの評価値（または彼らが受け取る財の評価値に関するシグナル）であるメカニズムである。目的の例としては、売り手の収入の最大化と最も高い評価値をもつ買い手への財の配分の一方だけまたは両方の達成である。同様に、マッチングまたは配分問題（それぞれ９章と１１章を参照）の場合、個人をマッチさせる、または個人に財を配分するアルゴリズムは個人の行動に影響を与え、ひいては最終的な結果に影響を与える。マッチング手順は、個人の私的情報が潜在的なマッチ相手に対する選好であるメカニズムの例である。想定される目的は安定マッチングの達成などとなる。

メカニズムデザインの理論は、多くの場合非常に一般的な手法をとる。例えば買い手と売り手がいる状況を考えると、メカニズムデザインの理論は個人が関わる手順の具体的な描写なしに買い手と売り手の間の効率的な取引の問題を解決できる。この具体的な例としては、各オークション形式はそれぞれに手順が定義される（よって１つのオークション形式は１つのメカニズムの例である）が、他にもバーゲンニングなどの手順があり得る。そして買い手と売り手は最終配分を決定するためにそれらの手順に従わなければならない。するとメカニズムデザインの理論における典型的な問いとは、具体的なメカニズムを特定することなく、何かしら性質をもったメカニズムが存在するかどうかになるであろう。性質とは、例えば売り手の評価値が買い手の評価値を上回らなければ取引が行われるなどである。そのような状況は B.4.2 で分析する。

これから考察していく状況は個人が私的情報をもつ場合である。例えば、個人は自動車ディーラーで、その私的情報は販売車の質である。彼らは、これから設計していくゲームをプレイすることになる。メカニズムデザインの理論では、彼らは**エージェント**と呼ばれる。さらに**プリンシパル**と呼ばれる個人もいる。プリンシパルの役割は、エージェントがプレイするゲームを設計することである。プリンシパルの目的は、何かしらの性質を満たす帰結を得ることである。そのためにはエージェントに所定の行動を促す必要があるかもしれない。

既に触れたように、メカニズムデザイン問題の出発点は、エージェントが私的情報をもっていて、プリンシパルがエージェントにプレイさせるゲームを設計しなければならない状況である。もしプリンシパルがゲームの設計を全部自由に行えるならば、彼女はそれぞれのエージェントに対して１つの戦略しかないようなゲームをうまくを設計できる。何らかの理由でプリンシパルが戦略の集合をうまく修正できない場合でも、彼女は利得を操作することでエージェントが選択しようとする全ての戦略を支配戦略にできる。そうだとするとどこにも困難は生じない。

まず第一に、ある状況下ではプリンシパルは利得を完全に自由には設計できない。利得の古典的な操作方法は、金銭移転を導入することである。しかし真の問題は、プリンシパルの目的がエージェントのもつ私的情報に依存していて、さらにプリンパルとエージェントの目的が一致していないことにある。メカニズムは、エージェントがプリンシパルに送ることができるメッセージの集合と送られたメッセージを元にして実現される帰結を定める関数によって定義される。次の簡単な例はメカニズムがどのようなものであるかを教えてくれる。

例 B.1 ３人の子供をアリスとボブとキャロルとする。父であるデニスは夕食を準備しようとしているが、グリーン豆とレンズ豆のどちらを調理すればよいか戸惑っている。ここでは、アリスとボブとキャロルがエージェントでデニスがプリンシパルである。それぞれの子供の好みの野菜は私的情報とする。

デニスは子供達に夕食を用意しなければならないが、彼は子供たちに野菜を摂らせたいので、子供たちの好きな野菜を知る必要がある。

例として以下の２つのメカニズムが考えられる。

メカニズムその１

- 子供たちはそれぞれがレンズ豆とグリーン豆のどちらが好きかをデニスに伝える。よって「レンズ豆」と「グリーン豆」がそれぞれの子供の送ることができるメッセージである。
- デニスは、メッセージを受けて多数が好む方を調理する。帰結関数はこれより「レンズ豆」のメッセージが２つ以上であるときにレンズ豆を調理し、それ以外のときはグリーン豆を調理する。

メカニズムその２

- 子供たちはそれぞれが１から１０のうちから１つの数をデニスに伝える。よって子供たちは異なる１０のメッセージから１つを選択できる。
- デニスは３人の言ってきた数字を足し合わせる。もし和が奇数であればレンズ豆を調理し、偶数であればグリーン豆を調理する。

B.2 モデル

エージェントの集合を $N = \{i_1, i_2, \cdots, i_n\}$ をする。達成される可能性がある決定の集まりを D とする。オークションでは、決定は財を誰に配分するかであったし、マッチング問題では決定は単に誰が誰とマッチするかを定めたマッチングであった。金銭移転の概念は後述するので、オークションを考える際に必要となる価格が決定には含まれないことに留意せよ。

各個人 $i \in N$ は何かしらの私的情報 θ_i をもち、i がもつ可能性がある全ての情報の集合を Θ_i とする。メカニズムデザインの理論では、個人のもつ情報は彼女の利得関数または決定上の選好と関連していて、ほとん

どの場合、効用関数 $u_i : D \times \Theta_i \to \mathbb{R}$ で表現される。個人の情報 $theta_i$ は個人の**タイプ**とも呼ばれ、付録 A のベイジアンゲームの定義でみたシグナルの概念に似ている。この設定の下で、個人 i が決定 d とタイプ θ_i から得られる利得は $u_i(d, \theta_i)$ となる。

例 B.2 オークションの環境では、買い手のタイプは評価値である。利得関数は、

$$u_i(d,\theta_i) = \begin{cases} \theta_i & d \text{ が「買い手 } i \text{ が財を手に入れる」であるとき} \\ 0 & d \text{ が「買い手 } i \text{ が財を手に入れられない」であとき} \end{cases}$$

となる。オークションの章では、買い手 i の評価値は v_i としていた。繰り返しになるが、モデルの描写は利得の譲渡 （買い手が支払う価格など）を未だ含んでいないことに注意せよ。

ここでの環境設定は、各個人の決定 d から得る効用は彼女のタイプのみに依存しており他の個人のタイプには一切依存しない。このような効用をもつ場合を私的評価値の場合という。

注意 B.1 マッチングや配分問題の基本的分析では、個人の厚生を数字で表現した効用関数とは対照的に、個人は結果に対する順序のみで表される選好関係をもつとされる。そのような環境では、個人の厚生に関係する情報は序数的で、２つの対象の良し悪しのみを比べることができる。その場合、個人のタイプとは単に選好順序そのものである。例えば３つの財 a, b, c があるとき、選好順序 $P_i : a, b, c$ と $P'_i : c, a, b$ は異なるタイプに対応している。しかしながらエージェントの選好の表現方法にかかわらず、彼らが結果に「どの程度満足しているか」がタイプに依存しているという直観に違いはない。

B.2.1 メカニズム

メカニズムは**メッセージ空間**と**帰結関数**の２つの要素から定義される。プリンシパルに設計されるメカニズムにおいて、エージェントはプリンシパルにメッセージを送らなければならない。プリンシパルは、受けとったメッセージを元に帰結関数を使って決定を遂行する。

厳密なメカニズムの定義を与える前に、**移転関数**という３つの目の要素を追加する。売り手と買い手の関係やオークションのような状況では、決定には価格や税などの金銭移転が伴う。そのような問題は、個人のメッセージのベクトルから各エージェントに対して１つの数を対応させる写像である移転関数も必要となる。各エージェントに対応するその数は、正であればエージェントがプリンシパルから受け取る、負であればエージェントがプリンシパルに支払う利得の譲渡を表す。よって帰結関数は決定関数と移転関数の組み合わせとなる。

定義 B.1 メカニズムは以下に定義される M と g のペア (M, g) である。

- $M = M_1 \times M_2 \times \cdots \times M_n$ をメッセージ空間とする。各個人 $i \in N$ について、M_i は個人 i が送ることができるメッセージの集合とする。
- g をメッセージのベクトルから決定と移転への写像 $g : M \to D \times \mathbb{R}^n$ とし、帰結関数とする。

多くの場合において、決定と移転を分離することは有用であることが知られている。帰結関数 g をエージェントのメッセージに依存する決定関数 f と移転関数 t の組み合わせとして定義する。移転関数 t はそれぞれのエージェントへの移転関数のベクトル $t = (t_1, t_2, \cdots, t_n)$ であることに注意せよ。

メカニズム (M, g) は以下のように働く。ここで帰結関数 $g = (f, t)$ は決定関数と移転関数によって構成される。最初に、各エージェント $i \in N$ はメッセージ m_i を選択する。プリンシパルはそのメッセージのベク

トル $m = (m_1, \cdots, m_n)$ を受け取る。次に、プリンシパルは決定関数を用いて決定を実現させるとともに移転を計算する。つまり決定は $d = f(m)$ で、移転は $(t_1(m), t_2(m), \cdots, t_n(m))$ である。タイプが θ_i であるエージェント i の最終的な利得は関数 v_i、

$$v_i(\theta_i, m_i, d, t) = u_i(d, \theta_i) + t_i(m) \tag{B.1}$$

で与えられる。

例 B.3 オークションは、プリンシパルが売り手でエージェントが買い手という、典型的なメカニズムの例である。買い手のタイプは支払ってもよいと考える価格の最大値で、それは私的情報である。

売り手の問題は、買い手が売り手に支払ってもよい最高価格を伝えるオークションの設計である。売り手のメッセージは入札額で、帰結関数は単にオークションの勝者（一般に最高入札者）と買い手の支払額を伝える関数である。

しかし、メカニズムデザインの理論は複数の買い手を相手にする売り手の問題をより幅広い視点から分析することが可能である。帰結関数はオークション形式を具体的に明示する必要はない。この抽象的ではあるが一般性を持った分析方法の利点は、オークションがそもそも売り手にとって例えば収入を最大化するための最もよい方法かどうかを判断できることにある。

メカニズムを次のより「簡潔」な方法で考えてみることは有益である。

- 各エージェントは送りたいメッセージを送る。
- 各エージェントの利得はエージェントから送られたメッセージの組み合わせに依存して式 (B.1) のように決める。

この簡潔な描写は付録 A でみたメッセージが戦略の役割をしエージェントのタイプがシグナルの役割をするベイジアンゲームの描写と似通っており、メカニズムはベイジアンゲーム以外の何ものでもないと言えそうである。しかし、そこには決定的な違いがある。ゲームでは、エージェントと戦略、利得関数があった。もし利得を変えれば、それは異なるゲームであった。対してメカニズムでは、利得が存在しない。B.1 の定義より、メカニズムにはメッセージと、決定と移転からなる帰結のみがある。一旦、利得関数 $u_i(d, t)$ が定義されると、それはゲームとなる。ここで注意しなければならないのは、メカニズムを変えることなく利得関数を変えることができることである。よって、メカニズムは非常に多くのゲームを同時に描写している。このことからしばしばメカニズムは**ゲーム形式**と呼ばれる。

B.2.2 社会選択関数の遂行

ここでまでの内容を整理すると、次のような状況となる。プリンシパルは帰結関数を設計し、エージェントが選ぶことのできるメッセージの集合を選択する。これよりエージェントは、選んだメッセージが帰結に影響することを考慮に入れて、送るメッセージを選ばなければならない。この文脈におけるプリンシパルの目的は何であろうか？既に触れたように、プリンシパルの望む帰結は一般にエージェントのタイプに依存する。つまり、エージェントのタイプの組み合わせに対して、プリンシパルにとって理想的な帰結が存在する。正確には、プリンシパルは関数

$$\gamma : \Theta \to D \times \mathbb{R}^n$$

を**遂行 (implement)** するという。この関数 γ を**社会選択関数 (social choice function)** と呼び、それぞれのタイプに対して決定と移転を規定する。

例 B.4 例 B.1 の状況で、社会選択関数のひとつは、

- 3人の子供達が皆レンズ豆を好むのであればレンズ豆を調理し、
- それ以外の場合はグリーン豆を調理する、

と考えられる。

よって社会選択関数はエージェントのタイプ（この場合は子供の好み）に対して定義されるもので、子供達が伝えることでに定義されるわけではない。

プリンパルの問題は次の 2 つである。1 つは、目的となる社会選択関数である。それはエージェントのあらゆるタイプの組み合わせに対してプリンシパルが達成したい決定を描写する。違う言い方をすると、もしもプリンシパルがエージェントのタイプを知っていたならば選ぶであろう決定である。もう 1 つは、メカニズムと、エージェントの行動である。エージェントのタイプ、エージェントが選択可能なメッセージ、決定関数を所与とすると、エージェントは特定のメッセージを選択する。それはつまり決定を選択することになる。プリンシパルの問題は結局、

$$g(m_1, m_2, \cdots, m_n) = \gamma(\theta_1, \theta_2, \cdots, \theta_n)$$

とすることができるかどうかである。つまり帰結関数 g による決定と移転が社会選択関数 γ で与えられる目的と一致しているかどうかである。メカニズムデザインの言葉を使うと、式 (B.2) が成り立つとき**メカニズムは社会選択関数 γ を遂行する**という。

例 B.5 プリンシパルである売り手は財を n 人の潜在的な買い手に売りたいと考えている。エージェントである各買い手には支払ってもよい最高価格がある。この最高価格はエージェントのタイプである。この場合、結果は、誰が最終的な買い手（決定）でそれぞれのエージェントがいくらを支払うかである。一般には、財を手に入れる買い手のみが 0 でない価格を支払うが、くじ引きのように全ての個人がいくらかを支払う状況も考えることができる。

$\theta_1, \theta_2, \cdots, \theta_n$ をエージェントのタイプとすると、売り手の 1 つの目的は以下のような社会選択関数となるかもしれない。

$$\gamma(\theta_1, \theta_2, \cdots, \theta_n) = \begin{cases} i \text{ が財を } \theta_i \text{ で買う} & \text{もし } \theta_i = \max\{\theta_1, \cdots, \theta_n\} \\ i \text{ は財を得られず、何も支払わない} & \text{もし } \theta_i < \max\{\theta_1, \cdots, \theta_n\} \end{cases}$$

言い換えると、関数 γ は最も高い額を支払える（最も高いタイプ）エージェントが財を手に入れ、そのタイプ通りに支払うことを規定する。最も高いタイプ以外のエージェントは何も払わず、何も手に入れない。

これよりプリンシパルの目的は、エージェントが行動の結果、最も高いタイプが常に財を手に入れられるメカニズムを設計することとなる。

価格（入札額）をメッセージの集合とするのは直観的で分かりやすい。つまりメカニズムは、それぞれのエージェントに価格を表明させることで構成され、最も高い価格を伝えてきたエージェントにその価格を支払ってもらい財を与えることになる。

この場合、メカニズムと社会選択関数 γ は同じものとなる。しかしこのメカニズムはいつでもうまく働くであろうか？容易にうまくいかないことがわかる。アリスとボブの二人だけがエージェントとする。アリスは 5 ドル、ボブは 10 ドルまでなら支払ってもよいと考えている。アリスとボブはそれぞれ 5 ドル、10 ドルというメッセージを送るだろうか？彼らがそうしないことは容易にわかる。もしそのようなメッセージが送られ

ると、売り手は嬉しいだろうが、それはボブの利益に反する。もし 6 ドルとメッセージを送れば、それでもアリスのメッセージの方が低いので、彼は財を手に入れられ、支払いも低くなるからである。

この簡単な例でもわかるように、正しくメカニズムを設計する問題は見た目以上に難しい。もしメカニズムと社会選択関数が同じ場合、エージェントはプリンシパルの望む帰結とは異なる帰結が遂行されてしまう行動を選んでしまうかもしれない。

アリスとボブはそれぞれ 5 ドル、10 ドルを支払ってもよいと考えている時、売り手はボブに 10 ドルで財を売りたい。これが社会選択関数の帰結である。しかしアリスとボブは異なる価格、例えば 3 ドルと 6 ドル、を伝えるかもしれないし、それは「ボブに 6 ドルで財を売る」という帰結になる。

例 B.5 におけるエージェントの戦略的行動は、プリンシパルにとっての最適なメカニズムが一般には遂行したい社会選択関数と一致しないことを示唆する。もしエージェントが戦略的でないならば、プリンシパルは遂行したい社会選択関数をそのままメカニズムとすれば十分である。しかしながら、エージェントが戦略的である場合、プリンシパルはエージェントの戦略的行動を考慮しなければプリンシパルの導きたい帰結に対応するメッセージを引き出せない。

ゲームとして考えるならば、いくつかのゲーム理論の概念がメカニズムにおけるエージェントの行動を「予測する」ことに使うことができる。事実、一旦エージェントの利得関数が定まればベイジアンゲームが規定されるので、様々な解概念によってゲームを分析することができる。

B.2.3 直接メカニズムと間接メカニズム

ここまではメッセージの性質を詳細にしてこなかった。メカニズムデザインの問題を定義する過程においてメッセージはどのようなものでもよい。それはメッセージがエージェントのタイプと直接関係付いていなくてもよいということである。例 B.1 のメカニズム 2 のメッセージはそのような例の 1 つである。

メカニズムデザインの文献では、メカニズムを直接メカニズムと間接メカニズムの 2 種類に分けて考える。**直接メカニズム**とはエージェントのメッセージをタイプとするメカニズムである。プリンシパルはエージェントに単にタイプを表明してもらう。エージェントのタイプに依存するどのような社会選択関数 γ も、メッセージ空間は $\Theta = \Theta_1 \times \Theta_2 \times \cdots \times \Theta_n$ で与えられ、帰結関数は社会選択関数 γ そのものであることより、それは直接メカニズムとなる。

もっとも例 B.5 でみたように、直接メカニズムは必ずしもエージェントに正直に彼らのタイプを表明させることを導けるわけではない。対して、間接メカニズムではメッセージは何でもよく、タイプで構成されるわけではない。例 B.1 では、メカニズム 2 が間接メカニズムで、メカニズム 1 が直接メカニズムである。

B.3 支配戦略遂行

メカニズムを設計する際、エージェントの行動を予測するための概念が必要である。最も納得できる概念は、支配戦略の概念である。(M, g) は $g = (f, t)$ を帰結関数とするメカニズムとし、全てのエージェント $i \in N$ について u_i を i のタイプに依存する効用関数とする。

メカニズムデザインの文脈ではメッセージに対応する戦略 m_i がタイプ θ_i の**支配戦略**であるとは、全てのメッセージ $m_{-i} = (m_1, \cdots, m_{i-1}, m_{i+1}, \cdots, m_n)$ と全ての戦略 m'_i について

$$u_i(f(m_i, m_{-i}), \theta_i) + t_i(m_i, m_{-i}) \geq u_i(f(m'_i, m_{-i}), \theta_i) + t_i(m'_i, m_{-i}) \tag{B.3}$$

となることである。

この支配戦略の考え方は、ゲームで既に見た支配戦略の考え方と同じである。（付録 A の定義 A.7 を参照。）

定義 B.2 メカニズム (M, g) が社会選択関数 γ を支配戦略で遂行するとは、全てのエージェント i について、

- ある関数 $h_i : \Theta_i \to M_i$ が存在して全てのタイプ θ_i について $h_i(\theta_i)$ が支配戦略となり、かつ
- 全ての $\theta \in \Theta$ について

$$g(h(\theta)) = \gamma(\theta)$$

となることである。

定義 B.2 は次のように読める。社会選択関数 γ をメカニズム (M, g) を用いて遂行したい。このメカニズムにおいてエージェントはメッセージを申告するが、特に支配戦略を申告することを求める。もし全てのタイプについて支配戦略があれば、定義 B.2 の h_i という関数が存在する。この関数は単にどのメッセージが各タイプの下で支配戦略であるかを示す。最後に、エージェントが支配戦略をプレイすれば、プリンシパルがあたかもタイプを知っているときに遂行したい帰結 $\gamma(\theta)$ を実際に実現する帰結として得られる。

ここまでのところ非常に一般的な分析方法を構築してきたが、一方でエージェントのタイプが、他方でエージェントが送るべきメッセージがあるため、操作が少し面倒になり始める。特定の社会選択関数を遂行するメカニズムが存在するかどうかを知りたければ、複雑なメッセージ空間を考える必要があるかもしれないし、それに依存する定義 B.2 にある関数 h の探索にうんざりしてしまうかもしれない。直接メカニズムの場合は、各エージェント i に対する関数 h_i の存在を気にする必要はない。定義 B.2 はより簡素にできる。

定義 B.3 直接メカニズム (Θ, γ) において、社会選択関数 γ が**支配戦略誘因両立**であるとは全てのエージェント $i \in N$ と、全ての $\theta_i \in \Theta_i$ について、タイプが θ_i であるとき θ_i が支配戦略となることである。

メカニズムデザインの文献では、社会選択関数が支配戦略誘因両立であることを<u>耐戦略的 (strategy-proof)</u> であると呼ぶ。直接メカニズムにおいてエージェントはタイプを表明するため、全ての $i \in N$ について、全ての $\theta_i \in \Theta_i$ を $h_i(\theta_i) = \theta_i$ とすればよく、定義 B.2 と B.3 は同値となる事が容易に分かる。

B.3.1 表明原理 (Revelation principal)

メカニズムの分析は困難な作業になる場合がある。特定の社会選択関数を遂行するメカニズムを見つけようとするとき、最初に直接メカニズムを使うか間接メカニズムを使うかを決めなければならない。メッセージ集合の定義を免れることから、直接メカニズムは明らかに好ましくみえる。しかし場合によっては間接メカニズムを考える必要がある。幸運にも既存研究の素晴らしい結果によってそのような作業は驚くほど簡単になる。

結果 B.1（表明原理） 社会選択関数 γ を支配戦略で遂行するメカニズムを (M, g) とする。すると直接メカニズム (Θ, γ) は支配戦略誘因両立性を満たす。

この結果は次のように説明される。特定の社会選択関数 f が遂行可能かどうかを知りたいとする。そのためには本来、<u>全ての</u>直接メカニズムと間接メカニズムを考慮する必要があるが、もしその中に遂行可能なメカニズムが存在するならば、表明原理の出番となる。つまり社会選択関数 f それ自体をメカニズムとする事ができ、その均衡においてエージェントは自らのタイプを正直に表明する。（正直にタイプを表明することが 1

つの均衡となる。均衡はそれ以外にも存在する可能性があることに注意せよ。）

表明原理の背後には比較的分かり易い直観がある。(M, g) は M を全てのエージェントのメッセージ空間とし g を帰結関数とするメカニズムとし、(M, g) は社会選択関数 γ を支配戦略で遂行するとする。すると、それぞれのエージェント i についてタイプが θ_i であるとき $h_i(\theta_i)$ が支配戦略となる関数 h_i が存在する。よってエージェントのタイプが $\theta = (\theta_1, \cdots, \theta_n)$ であるとき、彼らは $(h_1(\theta_1), \cdots, h_n(\theta_n))$ というメッセージを送り、その下での関数 g が導く帰結は、 (M, g) が γ を遂行するので θ の下での社会選択関数 γ の帰結と等しくなり以下が成り立つ。

$$g(h_1(\theta_1), \cdots, h_n(\theta_n)) = \gamma(\theta_1, \cdots, \theta_n)$$

表明原理の「仕掛け」は、エージェントにメカニズム (M, g) をプレイすることを要求する代わりにタイプそのものを表明するよう要求し、それらの代わりに関数 $(h_1, \cdots, h_n)$ に従うメッセージを作り出し、帰結関数 g にそのメッセージを送ることにある。メッセージ $h_i(\theta_i)$ を送ることが支配戦略であるとき、θ_i そのものを送ることも支配戦略となることは容易に確認できる。θ_i というメッセージを受け取ったプリンシパルがそれを $h_i(\theta_i)$ というメッセージに置き換えるからである。間接メカニズムにおける誘因が直接メカニズムの誘因と一致していることは、もし直接メカニズムにおいてタイプ θ_i のエージェント i がタイプを θ'_i と偽ることで得をするならば、プリンシパルはメッセージ $h_i(\theta'_i)$ を用いて帰結関数 g を計算するので、間接メカニズムにおいても $h_i(\theta'_i)$ を送ることで同様に得をすることができることから分かる。

B.3.2 ギバード＝サタスウェイト定理 (Gibbard-Satterthwaite theorem)

これまでにみてきたモデルと簡単な分析結果は価格のある交換の問題に着想を得ている。それらの問題は、決定とは別にエージェントとプリンシパルとの間の金銭移転を含んでいる。しかしそのような金銭移転が不可能である、または避けるべきである状況も多くある。マッチングと配分問題に関する章はそのような状況を扱っている。

まず始めに、これまでの金銭移転のあるモデルはほとんどそのまま金銭移転が不可能である場合に書き換えられる。プリンシパルが受け取る全てのメッセージに対して移転を 0 とすれば十分である。

移転を含まず、実現可能な決定の数が有限であるとき、エージェントの選好は効用関数ではなく単に順序で表現される事が多い。つまり、各エージェントは 9 章で議論したような実現可能な決定に対して「個人的なランキング」をもっている。この文脈においてメカニズムデザインの理論は何を示唆するだろうか？答えはむしろ否定的である。それが何を意味するかを述べる前にもう 1 つ定義が必要である。決定関数 f が独裁的 (dictatorial)であるとは、あるエージェント i が一人だけ存在して全てのエージェントがどのようなタイプであったとしても、f が選択する決定が常に i の最も好ましい決定と一致することである。言い換えると、決定関数が独裁的であるとき、遂行される決定は一人の独裁者 (dictator)と呼ばれるエージェントにのみ依存する。次の結果は社会選択関数の遂行に関する基本結果である。

結果 B.2 (Gibbard と Satterthwaite) 有限な決定の集合を D とし、決定関数 f がエージェントのタイプに依存して少なくとも 3 つの異なる決定を選択し、各エージェントのタイプが決定上の全ての選好で構成されるとする。このとき決定関数 f が支配戦略誘因両立であることと f が独裁的であることは同値である。

結果 B.2 はよいニュースではないようにみえる。この結果は支配戦略誘因両立的な唯一の決定関数は独裁的であることを提示している。ここで「提示している」というのは、結果 B.2 の表面的な解釈による。どのよ

うな結果であれ、そこには何かしらの仮定がある。ここでの重要な仮定はエージェントの選好に制限がなく、どのような選好順序も可能なことである。これは次のことを意味する。4 つの異なる決定 d_1, d_2, d_3, d_4 が実現可能とすると、それらに対して $4! = 24$ の異なる順序の付け方があり、それぞれの順序がタイプに対応する。選好、順序またはタイプのいくつかが無関係である状況が幸運にもある。例えば、エージェントが d_3 よりも d_2 を d_2 よりも d_1 を好む場合または d_1 よりも d_2 を d_2 よりも d_3 を好む場合があるとき、d_2 よりも d_3 を d_3 よりも d_1 を好むことはあり得ない。

結果 B.2 の 2 つ目の制約的な仮定は、移転の欠如である。よってこの結果は、社会選択関数を支配戦略で遂行するためには 0 でない移転が必要となることを示唆している。B.3.3 は移転があるときそのようなメカニズムが存在することを扱う。

B.3.3 ヴィックリー＝クラーク＝グローブス・メカニズム (Vickrey-Clarke-Groves)

エージェントとプリンシパルの間で移転が可能であるとき、支配戦略誘因両立的であるだけでなく効率的でもあるメカニズムが存在する。ここでいう効率性は、任意のタイプベクトル θ に対して選択される決定が個人の（移転を除いた）ペイオフの和を最大化することを意味する。正確には、決定関数 f が効率的であるとは、全ての $d' \in D$ と全ての $\theta \in \Theta$ について

$$\sum_{i \in N} u_i(f(\theta), \theta_i) \geq \sum_{i \in N} u_i(d', \theta_i) \tag{B.4}$$

が成り立つことである。つまり、エージェントのいかなるタイプの組 θ についても、実現される決定 $f(\theta)$ はその決定の下での各エージェントのペイオフの和が他のどのような決定 d' の下でのエージェントのペイオフの和よりも大きいか同等である。決定 d を所与として、タイプの組 θ におけるペイオフの和 $\sum_{i \in N} u_i(d, \theta_i)$ は社会厚生である。よって決定関数が効率的であるとは、全てのタイプの組 θ に対して、常に社会厚生を最大化することである。

次に移転関数を以下のように定義する。$d \in D$ を任意の決定とする。

$$t_i(\theta) = \sum_{j \neq i} u_j(d, \theta_j) - \max_{\hat{d} \in D} \sum_{j \neq i} u_j(\hat{d}, \theta_j) \tag{B.5}$$

式 (B.5) の関数は次のようなものである。決定 d についてエージェント i 以外のエージェントのペイオフの和を計算する。これは

$$\sum_{j \neq i} u_j(d, \theta_j)$$

の項である。式 (B.5) の後半はエージェント i 以外について社会厚生を最大化している。決定 $\hat{d}$ はエージェント i を除いた残りのエージェントのペイオフの和を最大化する。よって、式 (B.5) で表現される移転関数は決定 d の下でのエージェント i を除いた社会厚生と、エージェント i を除いた残りのエージェントの社会厚生を最大化した場合の値との差である。

定義 B.4 ヴィックリー＝クラーク＝グローブス (VCG) メカニズムは直接メカニズム (f, t) である。ここで f は効率的な決定関数で、移転 t は式 (B.5) の $d = f(\theta)$ として与えられる。

構成から VCG メカニズムは効率的であるが、さらに望ましい性質ももつ。

結果 B.3 ヴィックリー＝クラーク＝グローブス・メカニズムは支配戦略誘因両立的である。

B.4 ベイジアン・メカニズムデザイン

設定によっては、たとえ移転を許容しても支配戦略誘因両立的なメカニズムは見つからないかもしれない。例えば移転を実行可能にする場合または移転を予算均衡させようとする場合に該当する。

移転関数 t が実行可能であるとは、全てのタイプの組 θ に対して、個人の移転の和が 0 以下、つまり全ての $\theta \in \Theta$ について

$$\sum_{i \in N} t_i(\theta) \leq 0$$

が成り立つことである。式 (B.1) のように移転はエージェントの効用関数に含まれている。よってもし移転の和が厳密に正であると遂行するために金銭を幾らか外部から持ち込まれなければならないことを意味する。

より厳しい要求は移転の予算均衡である。これは移転が実行可能なだけでなくエージェントが払い過ぎないことを意味し、全ての $\theta \in \Theta$ に対して

$$\sum_{i \in N} t_i(\theta) = 0$$

が成り立つことである。

B.4.1 ベイジアン誘因両立性

前節では、支配戦略誘因両立的なメカニズムである VCG メカニズムを導入した。このメカニズムの問題は移転関数が予算均衡しない可能性があることである。1 つの方策は正直申告が支配戦略であることを弱め、代わりにそれが均衡戦略であることを要求することである。一旦個人の利得関数が明らかになればメカニズムは事実ベイジアンゲームなので、ベイジアン＝ナッシュ均衡の概念を使うことができる。まず直接メカニズムにおける誘因両立性を定義する。つまり $\gamma = (f, t)$ とするメカニズム (Θ, γ) を考える。ここで f は決定関数で t は移転関数である。

定義 B.5 直接メカニズム (Θ, γ) が**ベイジアン誘因両立的である**とは、任意のタイプの組 θ について、エージェントが自らのタイプを正直に表明する戦略の組がベイジアン＝ナッシュ均衡となること、すなわち全ての $i \in N$、$\theta_i \in \Theta_i$、$\theta'_i \in \Theta_i$ に対して

$$E\left[u_i(f(\theta_i, \theta_{-i}), \theta_i) + t_i(\theta_i, \theta_{-i}) | \theta_i\right] \geq E\left[u_i(f(\theta'_i, \theta_{-i}), \theta_i) + t_i(\theta'_i, \theta_{-i}) | \theta_i\right] \tag{B.6}$$

が成立することである。

式 (B.6) は付録 A の式 (A.10) にあるベイズ＝ナッシュ均衡の条件と全く同じであるが、メカニズムの概念と枠組みで書き換えたものである。間接メカニズムと直接メカニズムについて、メカニズム (M, g) が社会選択関数をベイジアン＝ナッシュ均衡で遂行するとは、(M, g) にベイジアン均衡 m が存在して全ての θ に対して $g(m(\theta)) = f(\theta)$ となることである。

表明原理は、支配戦略誘因両立的なメカニズムについては直接メカニズムのみに着目すれば十分であることを保証したが、同様の原理がベイジアン均衡による遂行でも成立する。

結果 B.4 （表明原理）メカニズム (M, g) が社会選択関数 γ をベイジアン均衡によって遂行するとする。そのとき直接メカニズム (Θ, γ) はベイジアン誘因両立的である。

B.4.2 取引：マイヤーソン＝サタスウェイトの定理 (Myerson-Satterthwaite 定理)

メカニズムデザイン理論の一般性は広範な状況を分析することを可能にし、設定はどちらかといえば抽象的なので特定の詳細から離れて状況を分析することができる。これをみるために売り手のサラと買い手のボブが取引する状況を考える。サラは自身とボブにとって価値のある財を 1 つもっているが、両者の財の評価値は異なっている。当然サラは彼女の評価値以上の価格がつけば財を売りたいと考え、ボブも彼の評価値以下の価格であれば財を買いたいと考える。

描かれた状況は非常に一般的であり、サラとボブがどのように取引を行うか詳細は与えられていない。サラが財を売ってもよいと考える価格を表明し、ボブがそれを受け入れて財を購入するか単に断るかという方法がありうる。別の方法としては、ボブが支払ってもよいと価格を表明するものである。3 つ目の方法には、片方が価格を提案し、もう片方が対案を出すといった繰り返し交渉に参加し、互いの提案が合致したところで取引が成立するというものがある。4 つ目の方法は、サラが求める最低価格とボブが支払ってもよい最高価格の間から無作為に数字を選び、その数字と同額をボブが支払うというもである。これらはサラとボブの取引の詳細の数例であり、これらのうちの 1 つの方法で取引しなければならないわけではない。

サラとボブが自らの評価値のみを知っており相手の評価値を知らない私的情報の場合、この問題はより興味深いものとなる。メカニズムデザインがどの程度有用であるかをみるために、サラとボブがいかなる取引を行うともそれはメカニズムとして描けることに注目したい。

ここで解明する疑問は「いくつかの性質を満たす取引メカニズムは存在するか？」というものである。この疑問は、メカニズムの詳細を特定化することなく単にメカニズムの存在の可否を明らかにすることなので、非常に明快である。またメカニズムに課す性質はどれも比較的容易に正当化できる。

- **事前の個人合理性**サラとボブの期待利得は非負でなければならない。
  サラとボブは取引の有無と価格に関する帰結を与えるメカニズムをプレイする。サラとボブは互いに相手の評価値を知らないので、それぞれの帰結の生起確率は各タイプの組の生起確率となるのでサラとボブの最終利得は確率的である。
  例えば、サラの評価値が 10 ドルで、価格の申告をするメカニズムを考える。帰結関数が
  - もしボブの申告価格がサラの価格よりも高ければ取引が行われ、それ以外では取引が行われない。
  - 取引が行われる場合、取引価格はサラとボブの申告価格の中間値とする。

  ボブの評価値は 0.6 の確率で 14 ドル、0.4 の確率で 8 ドルとする。またサラとボブは正直であると仮定する。するとサラにとっては、0.6 の確率で取引が行われ、価格は $(\$14 + \$10)/2 = \$12$ となる。ボブは 0.4 の確率でサラよりも低い価格を申告するので取引が行われない。よってサラの期待利得は、

  $$\text{期待収入} - \text{期待損失} = (0.6 \times \$12 + 0.4 \times \$0) - (0.6 \times \$10 + 0.4 \times \$0) = \$1.2$$

  となる。
  サラの期待利得は正なので、このメカニズムは個人合理的である。いま代わりに取引が必ず行われ価格がサラとボブの評価値の小さい方とするメカニズムを考える。この場合、サラの期待利得は $(0.6 \times \$10 + 0.4 \times \$8) - (0.6 \times \$10 + 0.4 \times \$10) = -\$0.8$ となる。彼女の期待利得は負なので、そう

するとこのメカニズムは個人合理的ではなくなる。

- **弱予算均衡**プリンシパルはメカニズムに追加的な費用を必要としない。
  例えばサラが 10 ドルと申告しボブが 8 ドルと申告したとしても、メカニズムは取引を行う決定をするとする。この場合、サラの価格よりも高い価格はなく、ボブの価格よりも低い価格は存在しない。もし取引が行われると、どのような価格であっても少なくとも 1 人は負の利得となる。メカニズムが事前の個人合理性を満たすには、サラもボブも最終的に負の利得とならないようにプリンシパルはゲームに追加的に金銭を注入する必要がでてくる。予算均衡しているメカニズムでは、プリンシパルは追加的金銭を必要としない。「弱」の意味は買い手によって支払われる価格が売り手が受け取る価格を超える可能性に触れている。それは例えば消費税が存在する場合を考慮することも可能にする。
- **ベイジアン誘因両立性**サラとボブにとって真の評価値を申告する真実表明が相手もそうしている時の最適反応になっている。
- **事後的効率性**財は最も高い評価値をもつエージェントに配分される。もしサラがボブよりも財を高く評価しているならば、彼女は財をそのまま持ち続けるべきで、そうでなければボブが財を手に入れるべきである。

ロジャー・マイヤーソンとマーク・サタスウェイトは以下の驚くべき結果を示した。

結果 B.5 (Myerson と Satterthwaite) 私的情報下の相対取引において、事前の個人合理性、予算均衡、ベイジアン誘因両立性、事後効率性を満たすメカニズムは存在しない。

言い換えると、マイヤーソン＝サタスウェイト定理は、少なくとも一方が取引の結果損をする、外部からの追加資金を入れる、または少なくとも一方が正直でない、のいずれかのリスクをとることなしに評価値が私的情報である 2 人のエージェントが効率的に 1 つの財を取引する方法がないことを示唆している。

この結果はむしろ驚異的で直観に反する。実際、そのような取引メカニズムが設計可能である状況を想像することは難しくない。これをみるために、サラの評価値は常に 10 ドルより低く、ボブの評価値は常に 10 ドルよりも高いとする。メカニズムとして

- サラとボブがタイプである評価値をどのように申告しようとも常に取引を行い、
- 取引価格は 10 ドルである

とすると、この特定の評価値をもつ状況では結果 B.5 で述べられた 4 つの性質を実際に満たす。しかし定理はいかなる状況をも考慮している！定理を理解するには、4 つの性質を満たす取引方法を設計できない状況を示せば十分である。実際、この結果は重要な性質である正直申告が均衡という誘因両立性条件に依拠している。

サラの財に対する評価値は 0 ドルか 0.9 ドルとし、ボブの評価値は 0.1 ドルか 1 ドルとする。両者にとってそれぞれの評価値の生起確率は 0.5 とする。

予算均衡条件から買い手の支払い価格は少なくとも売り手の受け取り価格と同額以上でなければならない。簡単化のために、取引が行なわれた場合ボブの支払い価格はサラの受け取り価格と等しいと仮定する。よって消費税の存在は捨象する。

ここで価格関数 $p(v_{Sarah}, v_{Bob})$ を導入する。v_{Sarah} と v_{Bob} はそれぞれサラとボブのメカニズムに申告された評価値とする。表明原理のおかげで、ここでのメカニズムはサラとボブが評価値を表明する直接メカニズムと仮定できる。

ここでの目的はこの価格関数 $p(v_{Sarah}, v_{Bob})$ が全てのタイプ（評価値）の組み合わせに対してどのような

形状かを分析することである。サラとボブの評価値の組み合わせは全部で４つある。

ケース１ $v_{Sarah} = \$0.9$ と $v_{Bob} = \$1$
$p(\$1, \$0.9)$ の値を見つけようとするとき、個人合理性より価格は少なくとも \$0.9 でなければサラの利得は負となる。よって

$$p(\$0.9, \$1) \geq \$0.9$$

ケース２ $v_{Sarah} = \$0$ と $v_{Bob} = \$0.1$
個人合理性から価格はせいぜい \$0.1 でなければボブの利得が負となる。

$$p(\$0, \$0.1) \leq \$0.1$$

ケース３ $v_{Sarah} = \$0.9$ と $v_{Bob} = \$0.1$
サラとボブの両者の個人合理性から取引は行われない。よって価格は 0 とする。

$$p(\$0.9, \$0.1) = 0$$

ケース４ $v_{Sarah} = \$0$ と $v_{Bob} = \$1$
$p(\$0, \$1)$ の特徴づけには誘因両立性条件が関わってくる。サラの評価値が \$0 のとき、正直に申告させたい。彼女はボブの評価値を知らないので、もしボブの評価値が \$0.1 であるとき、それは 0.5 の確率で生起するが、価格はケース２よりせいぜい \$0.1 でなければならない。しかし 0.5 の確率でボブの評価値は \$1 である。この場合、サラは価格がとても高いことを望む。彼女はボブの評価値を知らないから期待利得のみが計算でき、正直申告する均衡に注目すると彼女が \$0 と報告することで得られる期待利得は少なくとも \$0.9 と報告する場合の期待利得と同等以上でなければならない。よって

$$\underbrace{\frac{1}{2} \times p(\$0, \$1) + \frac{1}{2} \times p(\$0, \$0.1)}_{\text{サラが \$0 と報告した場合の期待利得}} \geq \underbrace{\frac{1}{2} \times p(\$0.9, \$1) + \frac{1}{2} \times p(\$0.9, \$0.1)}_{\text{サラが \$0.9 と報告した場合の期待利得}} \tag{B.7}$$

が成り立たなければならない。

式 (B.7) において、サラの評価値は \$0 なので期待収入は単に期待価格であり、それぞれの価格の生起確率はボブの評価値が \$1 であるか \$0.1 であるかの生起確率と一致する。

ケース１、２、３から価格関数 $p(\cdot, \cdot)$ について分かっていることを使うと、式 (B.7) は

$$p(\$0, \$1) \geq p(\$0.9, \$1) - p(\$0, \$0.1) \Rightarrow p(\$0, \$1) \geq \$0.8 \tag{B.8}$$

と簡単にできる。

ここでボブの視点から考える。彼の評価値が \$1 であるとき、\$0.1 ではなく \$1 と申告させたいが、前述したようにボブはサラの評価値を知らないので、彼は期待利得しか計算できない。よって彼の \$1 と申告した場合の期待利得が \$0.1 と申告した場合の期待利得と同等以上であることが必要となる。ボブの利得は、財を手に入れた時にはその財の評価値または手に入れられなかった時の 0 からメカニズムによって定められた価格を引いたものとなる。つまり

$$\text{ボブの利得} = \begin{cases} \text{財の価値} - \text{価格} & \text{財を手に入れた時} \\ 0 & \text{財を手に入れられなかった時} \end{cases}$$

であるので、

$$\underbrace{\frac{1}{2}\times(1-p(\$0,\$1))+\frac{1}{2}\times(1-p(\$0.9,\$1))}_{\text{ボブが \$1 と報告した場合の期待利得}}$$
$$\geq \underbrace{\frac{1}{2}\times(1-p(\$0,\$0.1))+\frac{1}{2}\times(1-p(\$0.9,\$0.1))}_{\text{ボブが \$0.1 と報告した場合の期待利得}} \tag{B.9}$$

となる。

ケース１、２、３から価格関数 $p(\cdot,\cdot)$ について分かっていることを使うと、式 (B.9) は

$$1-p(\$0,\$1)+1-p(\$0.9,\$1)\geq p(\$0,\$0.1)\Rightarrow p(\$0,\$1)\leq \$0.2 \tag{B.10}$$

と簡単にできる。

式 (B.8) と式 (B.10) を合わせると、

$$p(\$0,\$1)\leq \$0.2 \text{ かつ } p(\$0,\$1)\geq \$0.8 \tag{B.11}$$

を得る。

式 (B.11) は明らかに成立しない。多くの人にとって、マイヤーソン＝サタスウェイト定理は直観に反するかもしれないが、通常それは誘因両立性条件を考慮に入れていないことに起因する。一旦エージェントに正直であることを要求すると、メカニズムデザイナーには最適な取引を成立させることと真の評価値を申告させるために誘因を与えることの間で葛藤が生じることとなる。