メカニズムデザインにおける合理性の認識論理による分析

藤代 康誠　長谷部 浩二

メカニズムデザインは，ゲーム理論の枠組を用いて，設計者または社会全体にとって望ましい結果が生まれるような制度（メカニズム）を設計する手法である．特に，設計者がプレイヤーの行動を直接制御することなく，プレイヤーの合理的な行動によって望ましい結果が達成されるようにすることが中心的な課題である．こうした手法をもとに，これまでオークションをはじめとする数多くのメカニズムが提案されてきた．しかしながら，メカニズムが設計者の意図通りに目的を実現できるかどうかは，一般に判定が容易ではない．とりわけ，メカニズムで意図した結果が，支配戦略均衡によるものではなく，ベイジアンナッシュ均衡のようなより緩和された均衡によるものである場合，プレイヤー同士による相手の戦略に関する推論まで分析する必要があり，分析は困難なものとなる．そこで本研究は，認識論理をもとに，プレイヤーの知識状態や推論に注目したメカニズムの分析手法を提案する．特にここでは，与えられたメカニズムによって意図した結果が実現できることを，Public Announcement Logic と呼ばれる論理体系の中で証明する方法について述べる．

1　研究の背景と目的

複数の自律的主体（エージェント）によって構成されるシステムを制御する手法として，ゲーム理論が注目されている．ゲーム理論は，複数の主体（プレイヤー）の行動が相互に影響し合う状況（ゲーム）において，各プレイヤーが自分の利益を可能な限り大きくすることを目標に行動を選択するとき，どのような結果が起こるのかを分析する理論である.（ゲーム理論の詳細については文献 [1] などを参照のこと.）

与えられた状況から起こり得る結果を予測するのがゲーム理論の主要な目的であるのに対し，起こってほしい結果を実現するルールを求める理論はメカニズムデザインと呼ばれ，ゲーム理論の応用として位置づけられている．メカニズムデザインでは，各プレイヤーが自身の利得の最大化を目指して行動した結果が，起こってほしい結果と一致するようにゲーム（メカニズム）を設計する．メカニズムデザインは，主に経済学の分野において，望ましい社会制度の設計を目的として研究されてきたが，近年ではマルチエージェントシステムをはじめとする自律分散型のシステムの設計にも応用されている.（メカニズムデザインの詳細については，文献 [4] [5] などを参照のこと.）

メカニズムデザインが現実の問題に適用されている例としては，オークションの分析が挙げられる．例えば，英国式とよばれるオークションの方式では，商品の価格を競り上げていき，最後まで入札を降りなかった人が，最後の一人になった時点の価格で商品を落札する．この方式のオークションにおける最良の戦略は，商品の価格が，その商品に対して感じている価値を超える直前までは入札を続け，それより高い価格になった瞬間に降りる，というものである．すなわち，英国式オークションでは，商品に対する自分の評価を偽った入札をする誘因は生じず，商品は最も高い評価を付けている人に落札されるような望ましいルール設計になっていることが示される．

オークションの方式には英国式の他にも様々なもの

An Analysis of Rationality in Mechanism Design with Epistemic Logic
Kosei Fujishiro, 筑波大学大学院システム情報工学研究科, Department of Computer Science, Graduate School of SIE, University of Tsukuba.
Koji Hasebe, 筑波大学システム情報系, University of Tsukuba, Faculty of Engineering, Information and Systems, Division of Information Engineering.

がある．例えば，英国式のように入札状況を公開しながら進行する方式だけでなく，入札状況を非公開にする方式も存在する．これらはそれぞれ公開入札方式・封印入札方式として区別されている．このように，現実の問題にメカニズムデザインを適用する際には，メカニズムを実行するための手続き（プロトコル）としてどのような方法を採用するかを考慮する必要がある．特に，プロトコルごとの情報公開の方法の違いは，プロトコル実行中のプレイヤーの知識に変化をもたらし，意思決定に影響を与える可能性がある．しかしながら，既存のメカニズムデザインの枠組では，そのようなプレイヤーの知識の変化による影響が十分に考慮されていない．

そこで，本研究では，メカニズムを実行するためのプロトコルにおける情報公開の方法の違いが最終的な結果に及ぼす影響を，認識論理を用いて分析する．特にここでは Public Announcement Logic と呼ばれる論理を用い，プレイヤー間の知識に関する推論を定式化することにより，最終的な結果を考察する．プレイヤー間の知識に関する推論とは，具体的には，「プレイヤー 2 は『プレイヤー 1 は ϕ を知っている』ということを知っている」といった形の文で表されるような，知識の入れ子構造をもった事実に関する推論のことである．知識に関する推論は，相手の思考を読み，最良の意思決定をするために必要不可欠である．認識論理では，$K_i\phi$ という記法によって，「プレイヤー i は ϕ を知っている」という内容を表すことができる．この記法を用いると，前述の「プレイヤー 2 は『プレイヤー 1 は ϕ を知っている』ということを知っている」という内容は $K_2K_1\phi$ と簡潔に表すことができる．このように，認識論理を用いることで，知識に関する推論の形式的な分析が可能になる．

本稿では，例として投票のメカニズムとそれに対するいくつかのプロトコルを考え，プレイヤーの知識の変化やそれに伴う意思決定をどのような方針で定式化するのかを示す．

2 対象とするメカニズム

本稿では，以下のメカニズムを用いて提案手法を説明する．ここでは比較的単純なメカニズムを対象とするが，本手法はより複雑なメカニズムに対しても適用可能である．

ここで扱うメカニズムは 2 人の候補者のうち 1 人を投票によって選択するというものである．投票する権利を持った人（プレイヤー）は 3 人おり，各人は 2 人の候補者のうちどちらか一方を支持しているが，他のプレイヤーがどちらの候補者を支持しているかをあらかじめ知ることはできないとする．各プレイヤーは何らかのプロトコルにしたがってどちらの候補者に票を入れるのかを決定する．メカニズムおよびプロトコルの設計者の目標は，各プレイヤーに正直に投票をさせ，もっとも多くのプレイヤーに支持されている候補者を選択することである．

この状況設定は形式的には以下の要素によって定義される．

- プレイヤーの集合 $N = \{1, 2, 3\}$
- 候補者の集合 $X = \{a, b\}$
- 各プレイヤー i のタイプの集合 $\Theta_i = \{a, b\}, \Theta = \Theta_1 \times \Theta_2 \times \Theta_3$
- 選択ルール $f : \Theta \to X$
  ただし，任意の $\theta \in \Theta$ について，
  $$f(\theta) = \begin{cases} a & |\{i \in N \mid \theta_i = a\}| \geq 2 \text{ のとき} \\ b & \text{上記以外のとき} \end{cases}$$
- メカニズム $\Gamma = \langle \Theta, f \rangle$
- 各プレイヤー i の利得関数 $v_i : \Theta \times \Theta_i \to \mathbb{R}$
  ただし，任意の $\hat{\theta} \in \Theta$ と任意の $\theta_i \in \Theta_i$ について，
  $$v_i(\hat{\theta}, \theta_i) = u_i(f(\hat{\theta}), \theta_i) + w_i(\hat{\theta}) \tag{1}$$
  $$u_i(x, \theta_i) = \begin{cases} 1 & x = \theta_i \text{ のとき} \\ 0 & \text{上記以外のとき} \end{cases} \tag{2}$$
  $$w_i(\hat{\theta}) = \begin{cases} 1 & \hat{\theta}_i = f(\hat{\theta}) \text{ のとき} \\ 0 & \text{上記以外のとき} \end{cases} \tag{3}$$

上記の定義において，タイプは各プレイヤーがどちらの候補者を支持しているのかを表す．選択ルールは，プレイヤーの真のタイプに対する望ましい候補者の選択を表す．メカニズムとしては，各プレイヤーがどちらかの候補者に投票し，より多くの票数を獲得し

た候補者が選択される方法を採用している．ただし，各プレイヤーが自分のタイプに正直に投票するとは限らない．プレイヤーは，意思決定の時点で自分のタイプと異なる候補者に投票した方が利得が高くなることを知っている場合には，自分のタイプを偽った投票をする．各プレイヤーの利得は，(i) 自分のタイプと同じ候補者が選択されたか，(ii) 自分の投票と同じ候補者が選択されたか，という 2 つの基準で定まるものとしている．式 (1) において $\hat{\theta}$ は各プレイヤーの投票を，θ_i はプレイヤー i 自身の真のタイプを表しており，式 (2)，(3) はそれぞれ基準 (i)，(ii) に対応している．

3 論理体系

本研究では，Plaza による Public Announcement Logic（PAL）[2] [3] を論理体系として用いる．PAL は，通常の命題論理に，知識を表す様相演算子と，プレイヤーに対する情報の伝達を表す様相演算子を加えた論理体系である．

原子命題としては，次のものを定義する．

- 任意の $i \in N$，$x \in X$ について，$decl(i) = x$
- 任意の $i \in N$，$x \in X$ について，$eval(i) = x$
- 任意の $i \in N$ について，$decl(i) =?$

ここで，$decl(i) = x$ は「プレイヤー i は候補者 x に投票する」という内容を表し，$eval(i) = x$ は「プレイヤー i は真のタイプにおいて候補者 x を支持する」という内容を表す．また，$decl(i) =?$ はプレイヤー i が支配戦略を持たないことを表す． さらに,「プレイヤー i は真のタイプにおいて支持する候補者に投票する戦略を取る」という内容を表す論理式 $decl(i) = T$ を $(eval(i) = a \rightarrow decl(i) = a) \wedge (eval(i) = b \rightarrow decl(i) = b)$ で定義する．なお，ここでは可読性を考慮し，述語論理の記法を用いて説明しているが，これらの概念はすべて命題論理の原始命題として記述することが可能である．

非論理公理としては，次のものを定義する．ただし，プレイヤー i は真のタイプにおいて候補者 x を支持しているとする．

(a) $eval(i) = x$

(b) $K_i eval(i) = x$

(c) $\neg K_j K_i eval(i) = x$（ただし，$i \neq j$）

(d) 任意のプレイヤーの列 $i_1, i_2, \ldots, i_n$ について，

$$K_{i_1} K_{i_2} \ldots K_{i_n} decl(i) =?$$

(e) 互いに異なる任意の $i, j, k \in N$ について，

$$decl(i) =? \wedge decl(j) =? \rightarrow decl(k) = T$$

(f) 互いに異なる任意の $i, j, k \in N$ について，

$$decl(i) = T \wedge decl(j) = T \rightarrow decl(k) = T$$

上記の定義において，(a) はゲームの設定を表し，(b)，(c) はタイプに関するベイジアンゲームの仮定を表す．また，(d) はプレイヤー全員の利得関数が共有知識であることを表す．(e) は相手の戦略が予想できない場合に，仮に相手の戦略がランダムであると仮定したときの最適反応が自分のタイプの通りに投票することであることを反映している．(f) は相手が正直に投票する戦略を取ったときの最適反応が，正直に投票する戦略であることを表している．(d)，(e)，(f) は対象とするメカニズムによっては成り立たない場合もあるが，今回対象としているメカニズムでは成り立っている．

4 PAL を用いたプロトコルの分析

ここまでの定式化によって，単純なプロトコルの性質を示すことができる．ここでは，対象とするプロトコルとして，一切情報を公開せず投票を実施することを考える．このプロトコルではベイジアンゲームの仮定が満たされており，全員が正直に投票をするという結果がベイジアンナッシュ均衡として達成されると考えられるが，実際にこの性質をここまでの定式化によって証明することができる．

証明の概略は次の通りである．以下では i, j, k は互いに異なるプレイヤーであるとする．

1. 公理 (d)，(e) により $K_i decl(j) =? \wedge K_i decl(k) =?$ から $K_i decl(i) = T$ を導くことができる．
2. 公理 (d)，(f) により $K_i K_i decl(i) =? \wedge K_i K_i decl(k) =?$ から $K_i K_i decl(i) = T$ を，$K_j K_i decl(j) =? \wedge K_j K_i decl(k) =?$ から $K_j K_i decl(i) = T$ を，$K_k K_i decl(j) =? \wedge K_k K_i decl(k) =?$ から $K_k K_i decl(i) = T$ をそれぞれ導くことができる．
3. 2. で導いた式に対して，2. と同様の操作を続け

ることで，任意のプレイヤーの列 $i_1, i_2, \ldots, i_n$ について，$K_{i_1} K_{i_2} \ldots K_{i_n} decl(i) = T$ を導くことができる．

したがって，各プレイヤー i について，$decl(i) = T$ が共有知識となることが証明できる．これはゲームの結果として全員が正直に投票することを意味している．

以上で述べたプロトコルでは，プレイヤーに対して一切情報を公開しないことを仮定したが，プロトコルの実行過程においてより多くの情報が公開される状況を考えると，一般に $decl(i) = T$ は共有知識とならない．例えば，次のようなプロトコルを考えることができる．

Step 1. 各プレイヤーが第 2 章で述べた投票を行う．

Step 2. プロトコルの調停者からプレイヤーに対して投票結果に関する情報が知らされる．(ここでは，投票の結果選ばれた候補者の名前だけが知らされる場合や，各人が誰に投票したかまで知らされる場合など，様々なプロトコルが考えられる．)

Step 3. Step 1 に戻り，各プレイヤーはこれまでに得た情報をもとに再度投票を行う．全員の投票が変化しなくなるまでこれを繰り返す．

こうしたプロトコルにおいては，先に考えたプロトコルと比べ，各プレイヤーはより多くの知識を得ることができるため、正直な投票以外の戦略が最適反応となる場合がある．本提案手法は，こうした微妙なプロトコルの違いによる共有知識の成立の有無について分析することを目的としている．

5　結論と今後の課題

本稿では，メカニズムのプロトコルが最終的な結果に及ぼす影響について，認識論理を用いて分析する手法を提案した．さらに，単純な投票のメカニズムとプロトコルに対して提案手法を適用し，最終的な結果が正直な投票となることを確認した．今回対象としたプロトコルは情報の公開がないものであったが，今後は提案手法を拡張し，様々な情報公開の方法を取るプロトコルについて分析できるようにすることを目指す．

参 考 文 献

[1] Osborne, M. J. and Rubinstein, A.: *A Course in Game Theory*, The MIT Press, 1994.

[2] Plaza, J.: Logics of public communications, *Synthese*, Vol. 158, No. 2(2007), pp. 165–179.

[3] van Ditmarsch, H., van der Hoek, W., and Kooi, B.: *Dynamic Epistemic Logic*, Synthese Library, Springer Netherlands, 2007.

[4] 横尾真, 岩崎敦, 櫻井祐子, 岡本吉央: 『計算機科学者のためのゲーム理論入門』シリーズ第 3 回メカニズムデザイン (基礎編), コンピュータ ソフトウェア, Vol. 29, No. 4(2012), pp. 4_15–4_31.

[5] 横尾真, 岩崎敦, 櫻井祐子, 岡本吉央: 『計算機科学者のためのゲーム理論入門』シリーズ第 4 回メカニズムデザイン (応用編), コンピュータ ソフトウェア, Vol. 30, No. 1(2013), pp. 1_34–1_52.