

行動メカニズム・デザインに関する概説

川越 敏司

本稿ではまず、真の選好を表明することが支配戦略である耐戦略的メカニズムに関する実験研究を概観する。実験研究の多くでは支配戦略からの逸脱が見られるが、それは支配戦略と同じ結果をもたらす他の戦略が存在するために、すべての戦略の組み合わせを見通すことができない限り支配戦略が発見しづらいことが原因と考えられる。こうした限定合理的なプレーヤーを前提としたメカニズム設計の可能性について、行動メカニズム・デザインにおける最近の成果を紹介し、それらの分析結果を評価する。

キーワード：メカニズム・デザイン、公共財、オークション、マッチング、実験経済学

1. 耐戦略的メカニズムに関する実験研究とその課題

本稿では、メカニズム・デザインに関する実験を通じて見出されたバイアスや行動上の法則を基に、既存のメカニズムの性能を再検討したり、そうしたバイアスを念頭に置いたメカニズム設計を行ったりする行動メカニズム・デザインについて、最近の成果を解説する（実験経済学の方法論やその成果については川越 [1, 2] を参照のこと）。以下で取り上げるのは、公共財供給やオークション、マッチングに関する実験研究とその行動ゲーム理論的分析である。ただし、ここでの議論は、いずれも耐戦略性 (strategy-proofness) を満たすメカニズムに限定する。耐戦略性とは、各プレーヤーにとって自身の真の選好を表明することが支配戦略になっているということを意味する。ここで、プレーヤー i にとって、他のプレーヤーが選ぶ任意の戦略プロファイル s_{-i} に対して、戦略 s_i が常に最適反応であるとき、 s_i は支配戦略であるという。つまり、他のプレーヤーが選ぶ戦略の組み合わせに関係なく最適な戦略のことである。なお、メカニズム・デザインにおける表明原理 (revelation principle) によれば、耐戦略的メカニズムは、もしそれが設計可能であれば、各プレーヤーが可能な範囲から自らの選好を当局に報告する直接メカニズム (direct mechanism) に限定することができる（詳細は川越 [3], 第 1 章参照）。

1.1 公共財供給問題における耐戦略的メカニズムの実験

プレーヤーの選好は私的財 x_i と公共財 y に関して

$u(x_i, y) = x_i + v(y)$ のように準線形 (quasi linear) の効用関数で表せるものとする。ここでは単純化のために、公共財供給水準 y は 1 か 0 かの離散的な値を取るものとする。つまり、住宅地に住民の協力で公園や図書館の分室を設立するといった、ある一定規模の公共財を供給するかしらないか、二者択一の選択をする場合を考える。さらに、公共財の生産費用は 0 に基準化しておく。プレーヤー i にとって公共財からの便益は $v_i(y) = v_i y$ であるとする（ただし、 $v_i(0) = 0$ ）。この場合、 v_i は定数で、これをプレーヤー i の評価値と呼ぶ。

こうした準線形の効用関数の下では、公共財の効率的な供給を保証するボーエン＝サミュエルソン条件を満たす誘因両立的メカニズムを設計することが可能であることが知られている。その一つは、ピボタル・メカニズム (pivotal mechanism) と呼ばれるものである（川越 [3], 第 2 章参照）。この場合、ピボタル・メカニズムとは、次のようなルールのメカニズムになる。

各プレーヤーは公共財への需要（支払い意志額） w_i を表明し、もしその総和 $\sum_i w_i$ が生産費用=0 を超えたならば公共財が生産され、そうでない場合は生産されない。すなわち、

$$y = \begin{cases} 1 & \text{if } \sum_i w_i \geq 0 \\ 0 & \text{if } \sum_i w_i < 0 \end{cases}$$

さらに、各プレーヤー i は以下のような追加費用（クランク税ともいう） t_i を徴収される。

$$t_i = \begin{cases} \left| \sum_{j \neq i} w_j \right| & \text{if } (\sum_i w_i) \left(\sum_{j \neq i} w_j \right) < 0 \\ 0 & \text{それ以外} \end{cases}$$

かわごえ としじ

公立はこだて未来大学システム情報科学部

〒041-8655 北海道函館市亀田中野町 116-2

kawagoe@fun.ac.jp

ここで、 $(\sum_i w_i) (\sum_{j \neq i} w_j) < 0$ ということは、すなわち、(1) 全員の公共財への需要表明額の合計 $\sum_i w_i$ の符号がプラスの、プレーヤー i 以外の公共財への需要表明額の合計 $\sum_{j \neq i} w_j$ の符号がマイナスのとき、あるいは逆に、(2) 前者がマイナスで後者がプラスのとき、そのときにのみクラーク税が徴収される仕組みになっている。つまり、たとえば (1) の場合、プレーヤー i 以外のプレーヤーの間で決定した場合は公共財が供給されないはずだったのに、プレーヤー i が加わったことでその決定が覆り、公共財を供給することになったので、そのことで他のプレーヤーが被った損失をプレーヤー i は支払わなければならない、ということである。

最終的に各プレーヤー i の利得は、私的財からの効用と公共財から受ける便益の合計からクラーク税を差し引いた

$$u_i(x_i, y) = x_i + v_i y - t_i$$

となる。

このピボタル・メカニズムは耐戦略性を満たすので、各プレーヤー i にとって、表明するその需要額を公共財に対する真の評価値に等しくすること、すなわち、 $w_i = v_i$ が支配戦略になる（その証明は川越 [3]、第2章ウェブ付録参照）。

ピボタル・メカニズムに関する実験研究には、Attiyeh et al. [4]、Kawagoe and Mori [5]、Cason et al. [6] といったものがあるが、いずれの実験でも、支配戦略がプレーされる割合はかなり低いものであることが報告されている（公共財供給における耐戦略的メカニズムに関する実験研究は、Chen [7] にサーベイされている）。

その理由として、Kawagoe and Mori [5] は、ピボタル・メカニズムにおいて真の評価値を表明することが弱支配戦略 (weakly dominant strategy) であること、つまり、他のプレーヤーが選ぶ戦略の組み合わせがある一定の範囲に含まれる場合、真の評価値を表明することと利得が等しい他の戦略が存在するということを指摘している。つまり、あるプレーヤーが他のプレーヤーが選ぶ戦略が上記の範囲であるという「間違っただ」信念 (belief) をもってしまった場合、そのプレーヤーにとっては真の評価値以外の値を表明することも最適反応になってしまい、支配戦略からの逸脱が発生するということである。Kawagoe and Mori [5] はまた、公共財の供給水準が離散的な値である限り、こうした弱支配戦略にまつわる問題は解消できないことも

示している¹。支配戦略は他のプレーヤーが選ぶどの戦略の組み合わせに対しても最適な戦略であるため、最も「自明な」選択と考えられるが、それがプレーされないのはゲーム理論において深刻な問題だと考えられている。

1.2 オークションにおける耐戦略的メカニズムの実験

先ほどの公共財供給の場合と同様に、プレーヤーの選好が準線形である場合には、私的価値・2位価格オークション (second-price auction) は耐戦略性を満たしている。いま、一つの財が2位価格オークションで販売されるものとし、各プレーヤー i はその財に対して v_i の評価値をもっているとする。各プレーヤー i は同時に封入入札で入札価格 b_i を提出し、最も高い価格を提出したプレーヤーが落札し、2番目に高い価格を支払う。落札者以外のプレーヤーは財を得られない代わりに、何も支払う必要がない。この2位価格オークションは耐戦略性を満たすので、各プレーヤー i にとって、入札価格を財に対する真の評価値に等しくすること、すなわち、 $b_i = v_i$ が支配戦略になる（その証明は川越 [3]、第3章参照）。

しかし、2位価格オークションに関する実験研究では、先ほどのピボタル・メカニズムの場合と同様に、支配戦略がプレーされる割合は低いことが報告されている（オークション実験に関する初期のサーベイ論文としてはKagel [8] がある）。

その理由としては、先ほどのピボタル・メカニズムにおけるのと同様に、2位価格オークションでも真の評価値を表明することが弱支配戦略 (weakly dominant strategy) であること、つまり、他のプレーヤーが選ぶ戦略の組み合わせがある一定の範囲に含まれると予想される場合、真の評価値を表明することと利得が等しい他の戦略が存在するためである。

いま、評価値が v_i であるプレーヤー i が2位価格オークションに参加しており、プレーヤー i の入札価格を除いた中で最も高い入札価格を $b_j = \max \{b_k\}_{k \neq i}$ とする。もし、プレーヤー i が落札者ならば、 b_j がプレーヤー i の支払うべき2位の価格となる。ここで、 $v_i > b_j$ だとしてみよう。横軸にプレーヤー i の入札価格 b_i 、縦軸にプレーヤー i の利得 u_i を示したのが

¹ Cason et al. [6] は、支配戦略遂行とナッシュ遂行を同時に満たす安全遂行 (secure implementation) が可能であるための必要十分条件 (rectangular property) を示しているが、公共財の供給水準が離散的な値の場合、この条件は満たされない。

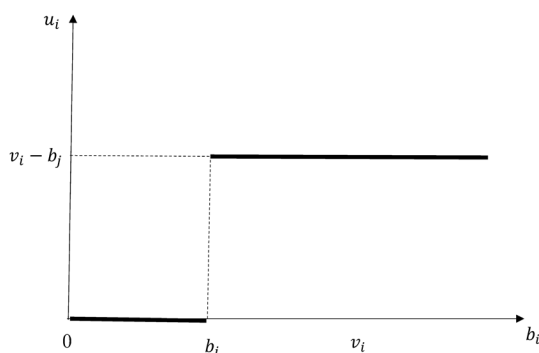


図1 2位価格オークションの利得構造

図1である。この図1からわかるとおり、プレーヤー i にとっては、 b_j 以上の価格 b_i はどれも同じ利得 $v_i - b_j$ になる。このように、2位価格オークションにおいても、各プレーヤー i にとっては、真の評価値を入札することとそれ以外の戦略とが無差別になるという弱支配戦略の問題が生じ得て、それが実験結果にも反映されているのだと考えられる²。

1.3 マッチング理論における誘因両立的メカニズムの実験

先ほどの公共財供給やオークションのように、プレーヤーに金銭的移転 (monetary transfer) が生じる場合以外にも耐戦略的なメカニズムが設計可能である。その一つがマッチング問題における受入保留 (deferred acceptance: DA) 方式である。

マッチング問題とは、たとえば、男女のような二つのグループが互いに相手側に対する希望順位 (選好) をもっていて、その順位に基づいて双方の組み合わせ (誰と誰がカップルになるか) を決定する問題である (ほかに、人とモノとの間のマッチング問題もある)。こうしてできる組み合わせのことをマッチングといい、こうしたマッチングを生み出すルールをマッチング方式という。ここでは、男女のカップルを形成する安定結婚問題 (stable marriage problem) を考えよう。

いま、それぞれ n 人ずつの女性と男性がいて³、女性の集合を W とし、そのそれぞれのメンバーを $w_1, w_2, \dots, w_n$ とする。また、男性の集合を M とし、そのそれぞれのメンバーを $m_1, m_2, \dots, m_n$ とする。また、女性 w_i が n 人の男性たちに対してもつ希望順

位 (選好) を一つのベクトルにまとめたものを**選好プロファイル** P という。男性 m_i が n 人の女性たちに対してもつ選好とその選好プロファイル Q についても同様に定義される。こうした男女の選好プロファイルに基づいて、どの男女をカップルにするかを定めるマッチング方式を $g(P, Q)$ とし、こうしたマッチング方式 g によって生み出された結果をマッチング $\mu = g(P, Q)$ と表す。

マッチング μ は、たとえば、1行目に男性を順に並べ、2行目にそれぞれの男性とマッチした女性を対応した位置に並べた行列として表現できる。男女それぞれ3名の場合、たとえば、以下のマッチング μ は、男性 m_1 と女性 w_2 、男性 m_2 と女性 w_1 、男性 m_3 と女性 w_3 がそれぞれマッチしたことを表している。

$$\mu = \begin{pmatrix} m_1 & m_2 & m_3 \\ w_2 & w_1 & w_3 \end{pmatrix}$$

なお、マッチする相手がいない場合、形式的に自分自身とマッチすると考える。

あるマッチング方式によって生み出された結果をどう評価するかについては、いくつかの基準がある。まず、マッチング μ が**個人合理性** (individual rationality) を満たすとは、マッチングに参加する前の状態よりは、マッチング結果が少なくとも良いものになっていることをいう。また、個人合理性を満たすマッチング μ において、ブロッキング・ペアが存在しないとき、 μ を**安定マッチング** (stable matching) という。ここで、ブロッキング・ペアとは、女性 w_i がマッチング μ でマッチした相手 $\mu(w_i)$ よりも違う男性 $m_j \neq \mu(w_i)$ を好んでおり、かつその男性 m_j はマッチング μ でマッチした相手 $\mu(m_j)$ よりもその女性 w_i を好んでいる、そのような男女のペア (w_i, m_j) のことである。

この安定結婚問題における DA 方式は、女性からプロポーズする場合、次のようなマッチング方式になる。

ステップ1. 女性は第1希望の男性にプロポーズする。男性は、プロポーズしてきた女性のうち最も希望順位が高い女性を仮受け入れする。すべての女性がいずれかの男性に仮受け入れされるか、まだプロポーズしていない男性がもはやいなければ、この時点で仮受け入れされているカップルを確定して終了。まだ仮受け入れされておらず、まだプロポーズしていない男性がいる女性が存在すれば次のステップに進む。

ステップ k ($k \geq 2$). ステップ $k-1$ 終了時にまだ男性に仮受け入れされておらず、まだプロポーズ

² 一方、Nishimura et al. [9] は、自分が落札できないなら他人の支払額が高くなるようにしようといういじわるの動機から、2位価格オークションにおける入札額が評価値とは異なることを説明しようとしている。

³ 一般的には、男性と女性の数は同じでなくてもよい。その場合、カップルになれない人が必ず存在することになる。

してない男性がいる女性は、まだプロポーズしていない男性のうち最も希望順位の高い男性にプロポーズする。男性は、プロポーズしてきた女性とその直前のステップ $k-1$ で仮受け入れした女性のうち最も希望順位が高い女性を新たに仮受け入れし直す。すべての女性がいずれかの男性に仮受け入れされるか、まだプロポーズしていない男性がもはやいなければ、この時点で仮受け入れされているカップルを確定して終了。まだ仮受け入れされておらず、まだプロポーズしてない男性がいる女性が存在すれば次のステップに進む。

この DA 方式は必ず有限のステップで終了する。また、どのような選好プロファイルに対しても必ず女性最適の安定マッチングを生み出す。また、女性から男性にプロポーズする場合、プロポーズする側の女性には選好を偽るインセンティブはなく、耐戦略性を満たしている。

一方、女性が第 1 希望から順にプロポーズしていき、男性に受け入れられた時点でマッチングが確定するような方式を即時受入 (immediately acceptance: IA) 方式という。IA 方式は一般に耐戦略性を満たしておらず、また安定マッチングを導くとは限らない。

これまでの DA 方式に関する実験研究では真の選好が表明される割合は 4 割から 8 割とかなりばらつきがある (DA 方式を含むマッチング理論の実験に関しては Hakimov and Kübler [10] がサーベイしている)。また、そうした実験では、自分の真の希望順位では 2 位以下の相手を 1 位に指名するといった行動が観察されることが示されている。

Kawagoe et al. [11] では、こうした支配戦略からの逸脱を詳しく調べ、プロポーズする側は自分の選好順位ではなく、相手側が自分を何位に評価しているかに注視し、自分を高く評価してくれている相手の順に希望順位を決めるというスキッピング・ダウン戦略 (skipping-down strategy) に従っていることを明らかにしている。たとえば、ある女性 w_i は真の選好では男性 m_i を男性 m_j より好んでいるが、男性 m_j の方が男性 m_i よりも女性 w_i を高く評価している場合に、この女性 w_i が男性 m_j を男性 m_i よりも高くするような希望順位を提出するなら、この女性はスキッピング・ダウン戦略に従っていることになる。

また、ちょうど公共財供給やオークションにおける弱支配戦略の場合と同じように、このスキッピング・ダウン戦略は、真の選好を表明する場合と同じマッチング結果を導く場合がある。そのため、DA 方式におい

ても支配戦略からの逸脱が発生するのだと考えられる。

2. 行動メカニズム・デザインの成果

1 節では、支配戦略と同じ結果をもたらす他の戦略があるために、実験において支配戦略からの逸脱が生じている可能性があることを示した。本節ではそうした結果を受けて、必ずしも支配戦略をプレーするわけではないプレーヤーがいることを前提とした行動メカニズム・デザインにおける最近の成果の一部を紹介する。

2.1 自明に耐戦略的なメカニズム

Li [12] は、自明に耐戦略的 (obviously strategy-proof: OSP) なメカニズムという考え方を定式化した。ある展開形ゲームにおいて、ある戦略 S_i が自明な支配戦略 (obviously dominant) であるのは、任意の戦略 S'_i に対して、任意の情報集合以降で S_i と S'_i が異なるパスをたどるとき、 S'_i がもたらす可能な最善の結果が S_i がもたらす可能な最悪の結果よりもよいものではないときをいう。そして、あるメカニズムが自明に耐戦略的であるのは、その均衡が自明な支配戦略によって構成されているときをいう。

Li [12] による自明な支配戦略という概念は、実験の被験者が示しがちな認知的な制約がある意味で取り入れたものになっている。前節で述べたように、2 位価格オークションやピボタル・メカニズムのような耐戦略的なメカニズムにおいては、真実表明が弱支配戦略になっているために、他のプレーヤーが選ぶ戦略プロファイル次第では、真実表明とそれ以外の戦略がもたらす利得が同じになる場合がある。そのため、プレーヤーは、真実表明が弱支配戦略であることを知るためには、他のプレーヤーが選ぶすべての戦略プロファイルの下での結果を検討しなければならない。しかし、実際には、実験の被験者は、他のプレーヤーが選ぶ戦略プロファイルの一部だけに注視して戦略の良し悪しを判断しがちである。そこで、Li [12] は、プレーヤーはある戦略 S_i と別の戦略 S'_i が異なるパスを通る時点に着目し、そのときに S_i における最悪の結果が S'_i における最善の結果を下回らないような「自明な」な場合にのみ、それが支配戦略だと認識しうるのだと考えたのである。

2 位価格オークションは耐戦略的であるが OSP ではないことは、次の例によってわかる。

いま、ある財を \$10 と評価しているプレーヤーがいるとする。ここで、真の評価値を表明するという支配戦略に従って \$10 を入札することと、\$11 を入札する

ことを考える。入札する前の情報集合から考えると、これら二つの戦略は異なるパスを通る戦略といえる。 $\$11$ を入札した場合、この戦略がもたらす最善の結果は財を $\$10$ より低い価格（たとえば、 $\$8$ ）で落札するというものであり、 $\$10$ を入札した場合、この戦略がもたらす最悪の結果は財を手に入れられず 0 の利得になることである。したがって、 $\$10$ を入札するという支配戦略は自明な支配戦略にはなっていない。

Ashlagi and Gonczarowski [13] は OSP 概念をマッチング理論に適用し、一般的には安定マッチングを生み出すようななどのようなメカニズムも OSP を満たさないことを示している。

いま、3人の女性 a, b, c と 3人の男性 x, y, z によるマッチングを考える。すでに見たように、女性プロポーズ DA 方式は女性最適の安定マッチングを生み出し、かつ女性側には耐戦略性が満たされている。ここで、男性の選好が以下のようなものとする。

$x: abc$
 $y: bca$
 $z: cab$

ここで、たとえば、 abc は男性 x にとって a が 1 位、 b が 2 位、 c が 3 位であるという選好を意味する。なお、男性も女性も誰ともマッチしないよりは誰かとマッチした方がよいと思っているとする。

女性の選好については、以下のいずれかを考える。

$a: zyx \text{ or } yxz$
 $b: xzy \text{ or } zyx$
 $c: yxz \text{ or } xzy$

ここで、たとえば、 zyx は女性 a にとって z が 1 位、 y が 2 位、 x が 3 位であるという選好を意味する。なお、Li [12] の「剪定原理 (pruning principle)」によれば、このように限定的な選好ドメインで OSP を満たさないメカニズムは、一般の選好ドメインでも OSP を満たさない。

さて、いま女性 a が選好を表明するという時点を考える。ここで、女性 a は真の選好 yxz を表明するか、虚偽の選好 zyx を表明するかを検討しているものとする。真の選好 yxz を表明した場合、女性 a にとって最悪の結果は、女性 b が zyx 、女性 c が yxz を表明した場合で、このとき DA 方式の下では以下のマッチング μ が実現する。

$$\mu = \begin{pmatrix} a & b & c \\ x & z & y \end{pmatrix}$$

女性 a が虚偽の選好 zyx を表明した場合、女性 a にとって最善の結果は、女性 b が xzy 、女性 c が xzy を表明した場合で、このとき DA 方式の下では以下のマッチング ν が実現する。

$$\nu = \begin{pmatrix} a & b & c \\ y & x & z \end{pmatrix}$$

このように、女性 a は真の選好 yxz を表明した場合には最悪で男性 x にマッチすることになるが、虚偽の選好 zyx を表明した場合には最善で男性 y にマッチすることになり、真の選好 yxz の下では後者の方が望ましく、DA 方式の下で真の選好を表明するという戦略は自明な支配戦略ではないことがわかる。

この例からもわかるとおり、選好（優先順位）プロフィールが循環的であると、メカニズムは OSP を満たさない⁴。Ashlagi and Gonczarowski [13] は、選好プロフィールが非循環的 [14] であることが OSP を満たすための十分条件であるが、必要条件ではないことを示している。Thomas [15] は、選好プロフィールが循環的な場合であっても OSP を満たすメカニズムが存在する例を示しており、選好プロフィールが非循環性を弱めた概念である限定循環性 (limited cyclicity) を満たすことが、DA 方式が OSP であるための必要十分条件であることを示している⁵。

2.2 レベル K 理論に基づくメカニズム・デザイン

Li [12] の OSP は、プレーヤーがすべての戦略の組み合わせを見通すことができないという認知上の制約を考慮したものであったが、レベル K 理論 (level-k theory) では、相手プレーヤーの合理性に関する信念の形成に制約があるものとする。つまり、レベル K 理論では、各プレーヤーは利己的に自らの利得最大化

⁴ 男性 x は女性 c より女性 b 、女性 b より女性 a を好んでいて、男性 y は女性 a より女性 c を好んでいるといった女性 a, b, c と男性 x, y が存在するとき、その選好プロフィールは循環的であるという。ある選好プロフィールが循環的ではないとき、非循環的という。ちなみに、Ergin [14] は、選好プロフィールが非循環的であることが、あるメカニズムが（提携）耐戦略的であり、かつパレート効率的であるための必要十分条件であることを示している。

⁵ DA 方式以外で耐戦略性を満たすマッチング方式にトップ・トレーディング・サイクル (top trading cycle: TTC) 方式がある。Li [12] は一般に TTC 方式が OSP を満たさないことを示している。Troyan [16] は、選好プロフィールが弱循環的 (weak acyclic) であることが、TTC 方式が OSP であるための必要十分条件であることを示している。

を目指して合理的に行動するものの、その戦略選択に当たっては、相手プレーヤーの合理性、あるいは、そこから導かれる行動に関して必ずしも正しくない予測をするものとする。レベル K 理論では、各プレーヤーをいくつかのタイプに分類する。\$L_0\$ タイプは最も naïve なプレーヤーで、通常は一樣ランダムな選択をするものと想定される（ほかには、真実表明などの「自明な」選択をするものとする場合もある）。\$L_1\$ タイプは、相手プレーヤーが \$L_0\$ タイプと想定したうえで、それに対する最適反応をプレーする。以下、\$L_k\$ タイプは、相手プレーヤーが \$L_{k-1}\$ タイプと想定したうえで、それに対する最適反応をプレーするといった形で帰納的に定義される。

レベル K 理論は数多くのゲーム理論実験に当てはめられ、その説明力は高いとされているが、Crawford and Iriberri [17] は私的価値オークションと共通価値オークションにレベル K 理論を適用し、実験で観察される規則性を再現できるかを検討しており、Crawford et al. [18] は、最適オークションの枠組みの下で、レベル K 理論に従うプレーヤーの行動を考察している⁶。

これらの研究は既存のメカニズムにレベル K 理論に従うプレーヤーが参加した場合の結果を考察したものであるが、de Clippel et al. [20] と Kneeland [21] は、こうしたレベル K 理論に従うプレーヤーたちがいる環境で設計可能になるメカニズムの特徴付けを行っている。一般に、こうしたメカニズムの設計においては、一部の特殊な環境を除いては、プレーヤーがすべてレベル K 理論に従うと想定したとしても、メカニズム設計の可能性は広がらないという結論が得られている。

Zhang [22] はマッチング問題についてレベル K 理論に基づくメカニズム・デザインを検討している⁷。Zhang [22] は DA 方式と IA 方式とを比較する際、\$L_0\$ としては真実表明を仮定している。したがって、DA 方式の下ではすべてのタイプのプレーヤーが支配戦略をプレーすることになる。そこで、IA 方式の下での結果がどうなるかであるが、以下の例を考えてみよう。3 人の女性 \$a, b, c\$ と 3 人の男性 \$x, y, z\$ がいるとする。女性の男性に対する選好と、男性の女性に対する選好は以下のとおりだとする。

$$\begin{aligned} a: xyz \quad x: cab \\ b: xyz \quad y: bac \\ c: yxz \quad z: abc \end{aligned}$$

ここで、たとえば、女性 \$a\$ について \$xyz\$ は、\$x\$ が 1 位、\$y\$ が 2 位、\$z\$ が 3 位であるという選好を意味する。なお、これらの選好は完備情報であるとし、男性も女性も誰ともマッチしないよりは誰かとマッチした方がよいと思っているとする。

女性プロポーズ DA 方式の下ではすべての女性がそれぞれ真の選好を表明することになるので、そのときのマッチング \$\mu\$ は以下ようになる。

$$\mu = \begin{pmatrix} a & b & c \\ z & y & x \end{pmatrix}$$

次に IA 方式の下で、すべての女性は \$L_1\$ だとする。彼らは他の女性が \$L_0\$、つまり、真の選好を表明するという信念のもとに最適反応をプレーする。たとえば、女性 \$a\$ は、女性 \$b\$ が男性 \$x\$ を、女性 \$c\$ が男性 \$y\$ を第 1 希望にすると想定すれば、男性 \$x\$ の自分に対する希望順位が女性 \$b\$ よりも高いことから男性 \$x\$ を第 1 希望にするはずである。女性 \$b\$ は、\$L_0\$ である女性 \$a\$ が男性 \$x\$ を、女性 \$c\$ が男性 \$y\$ を第 1 希望にすると想定する。ここで、女性 \$b\$ が男性 \$x\$ を第 1 希望にすれば、男性 \$x\$ の自分に対する希望順位が女性 \$a\$ より低いため男性 \$x\$ とはマッチされない。一方、女性 \$b\$ にとっては真の選好では第 2 希望の男性 \$y\$ を第 1 希望だと（虚偽）表明すれば、女性 \$y\$ は、男性 \$y\$ の自分に対する希望順位が女性 \$c\$ より高いため男性 \$y\$ とマッチすることになる。最後に、女性 \$c\$ は、\$L_0\$ である女性 \$a\$ と \$b\$ が男性 \$x\$ を第 1 希望にすると想定すると、女性 \$c\$ にとっては男性 \$y\$ を第 1 希望にすることが最適であるが、実際には \$L_1\$ の女性 \$b\$ が男性 \$y\$ を第 1 希望にして彼とマッチされるため、女性 \$c\$ は男性 \$z\$ とマッチすることになる。よって、このときのマッチング \$\nu^1\$ は以下ようになる。

$$\nu^1 = \begin{pmatrix} a & b & c \\ x & y & z \end{pmatrix}$$

次に、すべての女性は \$L_2\$ だとすると、女性 \$a\$ は男性 \$x\$ を、女性 \$b\$ は男性 \$y\$ を、女性 \$c\$ は男性 \$x\$ を第 1 希望にすることになるので、このときのマッチング \$\nu^2\$ は以下ようになる。

$$\nu^2 = \begin{pmatrix} a & b & c \\ z & y & x \end{pmatrix}$$

⁶ Crawford [19] は、相対取引 (bilateral trade) における効率的なメカニズムの設計（不可能性）にレベル K 理論を応用し、合理的なプレーヤーを想定した場合との差異を示している。

⁷ Zhang [22] は学校選択制マッチングについて議論しているが、以下では安定結婚問題の文脈に変更して説明している。

以下同様にして、すべての女性は $L_k (k \geq 3)$ だとすると、女性 a は男性 z を、女性 b は男性 y を、女性 c は男性 x を第 1 希望にすることになるので、このときのマッチング ν^k は以下のようになる。

$$\nu^k = \begin{pmatrix} a & b & c \\ z & y & x \end{pmatrix}$$

このように、 $k \geq 2$ について、 L_k である女性に対する IA 方式のマッチング結果は DA 方式と一致することになる。なお、Zhang [22] は、IA 方式の下では、 L_k である女性は DA 方式の下でのマッチング結果より弱い意味で順位の高い男性を第 1 希望にすることや、IA 方式の下でのマッチングが DA 方式の下でのマッチングに強くパレート支配されないことなどの結果を示している。

このようにレベル K 理論に基づく行動メカニズム・デザインではいくつか興味深い結果が得られているが、Crawford and Iriberri [17] や Zhang [22] では L_0 を真実表明と仮定しているため、実験でしばしば観察される支配戦略（真実表明）からの逸脱が説明できない⁸。

3. おわりに

本稿では、耐戦略的メカニズムに関する実験研究で指摘されている支配戦略（真の選好を表明）からの逸脱について、行動メカニズム・デザインの観点からどのような理論が提案されているのかを概観した。しかし、今のところ、こうした支配戦略からの逸脱を前提とした行動メカニズム・デザインは成功していないと言わざるを得ない。本稿で検討してきた耐戦略的メカニズムでは、真実表明が弱支配戦略であるため同じ利得（結果）をもたらす他の戦略が存在することが根本的な問題としてある。このような場合、真の選好を表明することはメカニズムのルールから容易に演繹できるような「自明な」選択では決してなく、なんらかのアドバイスや追加情報がないかぎり理解されない⁹。今後の行動メカニズム・デザインの研究では、こうした

⁸ L_0 を一様ランダムと仮定したとしても、2 位価格オークションでは L_0 に対する最適反応 L_1 は真実表明となるので、自明とまではいかないにせよ、レベル K 理論では支配戦略は容易に見出せるものであることになってしまう。

⁹ Kawagoe and Mori [5] では、すべての戦略の組み合わせの下でポータル・メカニズムにおける利得がどのように決まるのかを示した「詳細な利得表」を追加的に被験者に提示することで、真の選好を表明する割合が高まるということが報告されている。また、Masuda et al. [23] は、2 位価格オークションの実験において、被験者に事前に真の評価値を表明することが支配戦略であると告げることで、支配戦略を選択する割合が増加することを示している。

ことを念頭に置いたメカニズム設計を進めていくべきであろう。

参考文献

- [1] 川越敏司, 『実験経済学』, 東京大学出版会, 2007.
- [2] 川越敏司, 『行動ゲーム理論入門』, 第 2 版, NTT 出版, 2020.
- [3] 川越敏司, 『基礎から学ぶマーケット・デザイン』, 有斐閣, 2021.
- [4] G. Attiyeh, R. Franciosi and R. M. Isaac, “Experiments with the pivot process for providing public goods,” *Public Choice*, **102**, pp. 93–112, 2000.
- [5] T. Kawagoe and T. Mori, “Can the pivotal mechanism induce truth-telling? An experimental study,” *Public Choice*, **108**, pp. 331–354, 2001.
- [6] T. N. Cason, T. Saijo and T. Yamato, “Secure implementation experiments: Do strategy-proof mechanisms really work?” *Games and Economic Behavior*, **57**, pp. 206–235, 2006.
- [7] Y. Chen, “Incentive-compatible mechanisms for pure public goods: A survey of experimental research,” *Handbook of Experimental Economics Results*, Vol.1, C. Plott and V. Smith (eds.), Elsevier, pp. 625–643, 2008.
- [8] J. H. Kagel, “Auctions: A survey of experimental research,” *Handbook of Experimental Economics*, J. H. Kagel and A. E. Roth (eds.), Princeton University Press, pp. 501–585, 1995.
- [9] N. Nishimura, T. N. Cason, T. Saijo and Y. Ikeda, “Spite and reciprocity in auctions,” *Games*, **2**, pp. 365–411, 2011.
- [10] R. Hakimov and D. Kübler, “Experiments on centralized school choice and college admissions: A survey,” *Experimental Economics*, **24**, pp. 434–488, 2021.
- [11] T. Kawagoe, T. Matsube and H. Takizawa, “Skipping-down strategy and stability in school choice with affirmative action: Theory and experiment,” *Games and Economic Behavior*, **109**, pp. 212–239, 2018.
- [12] S. Li, “Obviously strategy-proof mechanisms,” *American Economic Review*, **107**, pp. 3257–3287, 2017.
- [13] I. Ashlagi and Y. A. Gonczarowski, “Stable matching mechanisms are not obviously strategy-proof,” *Journal of Economic Theory*, **177**, pp. 405–425, 2018.
- [14] H. I. Ergin, “Efficient resource allocation on the basis of priorities,” *Econometrica*, **70**, pp. 2489–2497, 2002.
- [15] C. Thomas, “Classification of priorities such that deferred acceptance is obviously strategyproof,” In *Proceedings of the 22nd ACM Conference on Economics and Computation*, p. 860, 2021.
- [16] P. Troyan, “Obviously strategy-proof implementation of top trading cycles,” *International Economic Review*, **60**, pp. 1–13, 2019.
- [17] V. P. Crawford and N. Iriberri, “Level-k auctions: Can a nonequilibrium model of strategic thinking explain the winner’s curse and overbidding in private-value auctions?” *Econometrica*, **75**, pp. 1721–1770, 2007.
- [18] V. P. Crawford, T. Kugler, Z. Neeman and A. Pauzner, “Behaviorally optimal auction design: Examples and observations,” *Journal of European Economic Association*, **7**, pp. 377–387, 2009.

- [19] V. P. Crawford, “Efficient mechanisms for level-k bilateral trading,” *Games and Economic Behavior*, **127**, pp. 80–101, 2021.
- [20] G. de Clippel, R. Saran and R. Serrano, “Level-k mechanism design,” *Review of Economic Studies*, **86**, pp. 1207–1227, 2019.
- [21] T. Kneeland, Mechanism design with level-k types: Theory and an application to bilateral trade, <http://hdl.handle.net/10419/157375> (2021 年 11 月 30 日閲覧).
- [22] J. Zhang, “Level-k reasoning in school choice,” *Games and Economic Behavior*, **128**, pp. 1–17, 2021.
- [23] T. Masuda, R. Mikami, T. Sakai, S. Serizawa and T. Wakayama, “The net effect of advice on strategy-proof mechanisms,” *Experimental Economics*, forthcoming, 2021.