

広島大学学術情報リポジトリ

Hiroshima University Institutional Repository

Title	金銭的報酬は非金銭的動機付けにどのように作用するのか : Janssen and Mendys-Kamphorst (2004) “The Price of a Price” のレビュー
Author(s)	鵜野, 好文
Citation	広島大学経済論叢, 45 (3) : 33 - 71
Issue Date	2022-03-10
DOI	
Self DOI	10.15027/52077
URL	https://ir.lib.hiroshima-u.ac.jp/00052077
Right	Copyright (c) 2022 広島大学
Relation	



金銭的報酬は非金銭的動機付けに どのように作用するのか： Janssen and Mendys-Kamphorst (2004) “The Price of a Price” のレビュー[†]

鵜野好文

本稿では、金銭的インセンティブの導入が個々人の社会的承認欲求に及ぼす影響を考察する。社会を構成する（利他主義者および利己主義者の）すべての個人が社会的承認を重要視するとき、彼らが、貢献の社会規範を遵守するよう（支援するため）、金銭的インセンティブ（価格）を導入することは、実際には、利他主義者の貢献者数を減少させる可能性があることを明らかにする。このとき、貢献者数の低減に歯止めをかけるため、金銭的インセンティブを中断したとしても、貢献の社会規範を回復させるどころか、貢献水準をさらに低下させるかもしれない。しかしながら、貢献の社会規範が消滅した状態でも、高価格を導入することで、そして、その後、漸次的に価格を低下させていくことで自発的な貢献を回復させることが可能かもしれない。しかし、そのような金銭的インセンティブの導入は、極端にコストがかかり、しかも、それが成功するかどうかは不確かである。

JEL分類：I18; D10; Z13

キーワード：社会規範；内発的動機付け；外在的モチベーション；社会的承認欲求

1. イントロダクション

ほとんどの経済理論は、人々が、金銭的に補償されていなくても、努力をいとわないとするとき、彼らに金銭的報酬が支払われるならば、さらに懸命に努力するに違いないと説明する。ここでの基本的な仮定は、金銭的報酬が導入されても、既存の非金銭的モチベーションは影響を受けることがないとするものである。したがって、さまざまな種類の金銭的および非金銭的モチベーションは単純に加算することができることになる。この仮定は、経済学者が、非金銭的インセンティブ問題を、経済政策により影響を受けることがない外生的な私的嗜好（*exogenous personal preferences*）の問題として扱うことを可能にしている。このような方法で、経済学は、動機付け問題を、金銭的報酬制度の最適設計問題に単純化させようとしている。

この従来のアプローチは、経済人（*homo economicus*）と社会人（*homo sociologicus*）は、容易

[†] 研究プロジェクト（課題番号：26380462）への日本学術振興会の学術研究助成基金助成金の資金援助に深く感謝いたします。本レビューは、同研究プロジェクトの遂行にあたりなされた、一連の文献レビューの一環であり、主として、Maarten C. W. Janssen, and Ewa, Mendys-Kamphorst, “The Price of a Price: On the Crowding out and in of Social Norms,” *Journal of Economic Behavior & Organization*, Vol. 55, Issue 3, 2004, pp.377-395を概括したものである。

に、切り離すことができないと主張する一部の経済専門家の間に、しばしば、戸惑いを呼び起こすことになる。人々は、通常、複数のインセンティブがさまざまに組み合わせられることで（それぞれのインセンティブが）相互に強化されたり、あるいは、弱化されることで動機付けられると考えられている。そして、そのこと（異なるタイプの動機付けの相互作用の重要性）を想起させるような、（例えば、社会政策を経済的インセンティブのみを基礎として策定することに警鐘を鳴らすことになる）最初の研究（のひとつ）が、Titmuss（1970）により提示されることになる。彼は、医療目的のための血液調達システムについて、アメリカ型システム（大部分を、報酬支払い、あるいは、その他の諸手当を支払うことで獲得するシステム）およびイギリス型システム（完全に自発的行為に依存し、報酬支払いがなされないシステム）とを比較し、アメリカ型の報酬依存システムは輸血用の血液の不足、および、血液の品質の低下を招くと結論付けている。彼は、また、イギリスの献血者（貢献者）の社会的特性が米国の売血者（貢献者）のそれとは異なるものであることにも注目している。イギリスでの献血者は、英国民を代表するような人達であったが、アメリカでは、彼らは（アメリカでの売血者は）、低所得ないし低学歴の人達である傾向がみられた。それにもまして、彼の主張の核心をなすものは、自発的に採血に応じた人に報酬を支払う（売血）制度は、貢献者から利他的な献血動機を消滅させてしまうというものであった。しかも、この利他的動機の崩壊は、ほぼ永久的に続き、金銭的インセンティブを中断したとしても、少なくとも、すぐには回復しないとされている。

Titmussの研究は、すべての人に納得のいくような左証を与えるものではないが（例えば、Arrow, 1972を参照しなさい）、これらの実証研究を契機に、近年、多くの研究者が¹、異なるタイプの動機付け間の相互作用を考察することの重要性について論及するようになった。ここで重要なことは、とりわけ、金銭的インセンティブが、他のタイプの動機付け（自己実現の欲求充足および社会的欲求充足）、いわゆる、非金銭的インセンティブを「締め出す（crowding out）」ことが指摘されていることである²。しかも、それだけではなく、「利他的動機付け」は、いったん、締め出されると、金銭的インセンティブが中断された後でも、しばしば、元に戻ることはない。これらのことについては、例えば、イスラエルの託児所でなされたフィールド調査でも同様のことが観察されている（Gneezy and Rustichini, 2000a）。調査では、園児を決められた時刻どおり迎えに来る両親の（自発的）行動とこれを支援するはずのマイナスの金銭的動機付け（罰金と）の関係が明らかにされている。調査結果は、退出期限に遅れて子供を迎えにくる両親に罰金を科すことは、（意図に反して）遅刻する両親をさらに増やすことにつながるというものであ

¹ この主題に関する心理学的文献のオーバー・レビューについては、Deci (1999) を、また、実証研究のサーベイについては、Frey and Jegen (2001) を参照しなさい。具体的な経済学的実証研究については、例えば、Frey (1997)、Frey and Götte (1999)、Fehr and Gächter (2000)、および、Gneezy and Rustichini (2000a, b) を挙げるができる。

² 後に明らかになるが、金銭的報酬が、非金銭的インセンティブを締め出すというとき、その背後には、二つの効果が作用している。ひとつは、個人が、金銭的報酬の魅力に引き寄せられ、内発的動機付けを軽視する直接効果である。もうひとつは、金銭的報酬が、貢献活動に負の効果を持つ利己主義的な貢献者を惹き付け、これが社会的承認欲求（社会的報酬）を低下させる間接効果である。このことを一見したとき、金銭的報酬の導入が、表層的には、利他的動機付けを低下させる、あるいは、非金銭的インセンティブ全般を低下させるように見える。ただし、現段階では、（明示的な現象を観察する）実証研究の査証のように、「金銭的インセンティブが非金銭的動機付け（利他的動機付け）を締め出す」と叙述することにする。

た。また、罰金を科すことを止めても、園児の両親の行動が元の状態に戻ることはなかった。同様に、ボランティア活動に経済的インセンティブを導入したときの影響を調査したFrey and Götze (1999)の実証研究では、(完全に自発的な)ボランティア活動として遂行されたときの仕事量は、金銭的インセンティブが導入された後、顕著な低下を示すことが明らかにされている。

本稿で、私たちは、金銭的インセンティブと非金銭的モチベーションの相互作用の基礎となるひとつのメカニズムを考察することで、上述の実証結果の背景をなす論理を理論的に解明する。そこで、私たちは、まず初めに、なぜ、人々が、自発的に貢献の社会規範を遵守しようとするのかを考察することから始めることにする。Dickinson (1989, p.6) が指摘するように、「私たちの社会では、人々が、自発的に慎みのある望ましい行動を取ることは、称賛されることが多い... (中略)...。しかも、それらの行動が、いかなる外在的結果とも明らかに無関係というまさしくその事実こそが、多少なりとも、社会的承認/称賛 (social approval or praise) の根拠となっている」。言うならば、人々が、「社会的報酬 (social reward)」、すなわち、社会的承認ないし社会的地位を希求することに、多少なりとも、駆り立てられることが、結果として、自発的に (貢献の) 社会規範を遵守する社会状況を作り出すことになるとしている。

実際、モチベーションの源泉としての社会的認知 (social recognition) の重要性は、多くの研究が指摘しているところである。Glazer and Konrad (1996)、Harbaugh (1998a, b)、および、Andreoni and Petrie (2000) による慈善事業への寄付の研究では、一般的に、慈善団体が、寄付者のリストおよび (おおよその) 寄付額を公表することにより、(そして、寄付者が自らの寄付額を見られることを気に掛けるという事実により)、貢献額が影響を受けることが指摘されている。そして、これらの事実より、慈善活動においても社会的認知の重要性が (一定ではあるが) 存在することがわかるとしている。Fehr and Gächter (1999)、および、Andreoni and Petrie (2000) も、同様に、人々は、自らの貢献額が他の人々に知られるならば、彼らは (公共財に対して) より多くの貢献を行うとする実証結果を提示している。

このような状況の下で、金銭的インセンティブを導入することは、実証研究が指摘するように、自発的行動の弱体化を招くことになるのであろうか。すなわち、金銭的インセンティブを導入することは、それまで、いかなる外在的結果とも明らかに無関係である自発的行動の位置づけを危うくし、そして、まさしく、この事実が、社会的承認ないし社会的地位の持つ神通力を脆弱化し、その結果、利他的動機付け (を持つ利他主義者の貢献者数) を低下させることにつながるのであろうか。

私たちは、本稿で、金銭的インセンティブと非金銭的モチベーションとの相互作用、および、社会的に望ましい課業の遂行を支援するために導入された金銭的インセンティブにより引き起こされる結果を、単純なモデルにより解析する。その際、簡単化のため、モデルでは、次のことを仮定する。まず、人々が、自発的に社会的に慎みのある望ましい行動を取ることを、一般的に、「社会に貢献する」、あるいは、「貢献の社会規範を遵守する」行為として言及する。そして、さらに、個人が、このような (自発的な) 貢献活動を行うことに対して、(例えば、ボランティア活動あるいは献血活動を行うことに対して)、二つの (非金銭的) 報酬が準備されるとする。ひとつは、貢献活動それ自体の持つ自己実現の欲求充足 (内発的インセンティブ)、もうひとつは、社会的承認欲求、いわゆる、社会的報酬 (外在的インセンティブ) の獲得である。ところが、先にみたように、これらの非金銭的動機付けは、自発的行為の遂行努力の支援のため、権威当局等により導入された金銭的インセンティブにより有害な影響を受けるとされている。もし、そうで

あるならば、金銭的インセンティブは非金銭的モチベーションにどのような影響を及ぼすのであろうか。例えば、金銭的インセンティブは、非金銭的動機付け（利他的動機付け）を締め出し、モチベーションの源泉として取って代わるのであろうか。しかも、それだけではなく、利他的動機付けは、いったん、締め出されると、金銭的インセンティブが中断されたとしても、元に戻ることはないのであろうか。さらには、完全に利他的動機付けが締め出された状態において、金銭的インセンティブを呼び水として使用することで、利他的動機付けを再生させることができないのであろうか。私たちは、本稿で、これらのことを進化モデルにより考察することにする。

私たちのモデルでは、社会の全構成員を母集団としたとき、2つのパーソナリティ・タイプが存在するものとする。すなわち、利他主義者と利己主義者である³。利他主義者および利己主義者の両タイプとも、非金銭的および金銭的インセンティブにより、（貢献行動に）動機付けられる。また、両タイプとも貢献行動を行うのに同等の費用を被るとする。しかも、両タイプとも、利得が費用を上回るとき、貢献行動を選択するとする。しかしながら、両タイプが、いかなるインセンティブに動機付けられるかには、大きな違いがある。利己主義者は、外在的インセンティブ、すなわち、社会的報酬および金銭的報酬にのみ動機付けられるのに対し、利他主義者は、外在的インセンティブだけではなく、内発的インセンティブにも動機付けられるとする。すなわち、社会的報酬および金銭的報酬のほかにも、自己実現の欲求充足にも動機付けられる⁴。したがって、利他主義者は、一部、他の人を助けたいとする内発的欲求に動機付けられるため、利己主義者より、金銭的報酬から生じる効用を節約できることになる。かくして、利他主義者は、金銭的報酬、社会的報酬、および、内発的報酬の合計が彼らの貢献費用を越えるならば、貢献することを選択するが、他方、利己主義者は、社会的報酬および金銭的報酬の合計が彼らの貢献費用を補償するときだけに、貢献することを選択することになる。

これだけを見ると、利他主義者および利己主義者の動機付けについて、金銭的インセンティブと非金銭的モチベーションの間にはなんら相互作用も観察されず、従来の経済学における（金銭的）インセンティブの取り扱いと変わるところはない。そこで、私たちのモデルでは、金銭的インセンティブが非金銭的モチベーションに及ぼす直接的および間接的効果を次のように仮定する。まず、金銭的報酬が非金銭的モチベーションに作用する第一の（直接的）効果は、利他主義者の内発的動機付けに関わるものである。自己実現の欲求充足と金銭的報酬が同時に存在するとき、それらの価値は単純加算ではなく、相互作用し合った結果、それぞれのインセンティブに（相対的な）価値評価がなされ、その相対価値に見合うウェイト付けがなされた上で加算されるとする⁵。したがって、このことは、金銭的報酬の価値評価は、両タイプについて、同じではなくなるという結果を招くことになる。さらには、金銭的報酬の導入が、利他主義者の総モチベーションを増加させることにつながるかどうかは、報酬水準（価格規模）に大きく依存することになる。これらのことは、後に明らかになるが、利他主義者の貢献者数（均衡）に決定的な影響を及ぼすことになる。

³ 社会に利他主義タイプが存在することは、進化論的研究および実証研究の両方の研究成果より正当化することができる。他者に配慮することが、なぜ、あるいは、どうして、進化論的淘汰に耐えうるのかについては、例えば、Bester and Güth (1998) を参照しなさい。Frank (1987)、Stewart (1992)、Güth and Kliemt (2000)、および、Bar-Gill and Fershtman (2001) は、利己主義者および利他主義者が混在する人口構成が進化論的安定につながる可能性があることを示唆している。

⁴ Harbaugh (1998b) は、実証研究において、利他主義者は、慈善事業への寄付は、一部、利他主義それ自体により（内発的インセンティブにより）動機付けられていると結論付けている。

金銭的インセンティブが非金銭的モチベーションに作用する第二の（間接的な）効果は、社会的報酬に関わるものである。社会的承認に由来する（社会的）報酬には貢献の社会規範の形成・確立・維持・消滅にまつわるメカニズムが働いている。すなわち、貢献の社会規範の遵守に積極的な個人（利他主義者）が多ければ多いほど社会的承認欲求は満たされ、そこから得られる欲求充足（社会的報酬）も増大することになる。逆に、貢献の社会規範の遵守に消極的な個人（利己主義者）が増すにつれ、社会的承認欲求は満たされなくなり、その結果、欲求充足（社会的報酬）も低下することになる。すなわち、利他主義者は他の貢献者に正の外部性を、そして、利己主義者は他の貢献者に負の外部性をもたらすとする。私たちは、利他主義者の貢献者のみが社会的報酬について正の外部性を持つとしているので、社会的報酬は、少なくとも、幾人かの利他主義者が貢献するときだけに、（正の値となり）成立すると仮定する。したがって、社会的報酬（の水準）は、社会規範の形成・確立・維持・消滅のメカニズムと同様、利他主義者の貢献者数および利己主義者の貢献者数の規模に依存して決まることになる。ここで、注意すべきことは、金銭的インセンティブ（金銭的報酬）の導入が、利他主義者および利己主義者の貢献者数にどのような影響を及ぼすかにより、社会的報酬に由来する（非金銭的）インセンティブの強弱が定まることである。以上のような前提の下に、貢献の社会規範に関する社会変動の力学を考察する。このとき、私たちは進化論的モデルを応用することにする。モデルでは、利他主義者および利己主義者の各タイプは、2つの選択可能な戦略のうち一つを選ぶことができるとする。すなわち、「貢献する」、あるいは、「貢献しない」のいずれかである。個人が最適な戦略選択を行うとすれば、ほとんどの場合、他の個人が「前期」と同じ行動を選択することを前提に、当期において、最大期待利益をもたらす戦略を選択しようとする。しかしながら、最適な戦略選択を行う個人であっても、それぞれの時点で、わずかながら、最大期待利益に反する意思決定を行う可能性があるとする⁶。この枠組みを用いることで（ほとんどの場合、最適な戦略選択を行うが、わずかながら、最適な戦略選択に反する意思決定を行う可能性があるとする）で、私たちは、中期均衡と長期均衡とを区別することができる。中期的には、社会変動の力学は、最適反応としての決定

⁵ この事例については、Holmström and Milgrom (1991) を参照しなさい。例えば、クラスのインストラクターは、通常、学生に、標準的な基礎知識および技能を教えることだけでなく、同時に、学生の好奇心、創造的思考およびコミュニケーション能力等の多次元的能力開発にまで関心を払うことが知られている。したがって、インストラクターに、固定給を支払うならば、内発的動機付けが鼓舞され、そして、学生の基本的能力を含む、多次元的能力開発に腐心することが期待されるであろう。このような状況下で、インストラクターからさらに努力を誘発するよう、（観察可能な評価測度を持つ）インセンティブ給が支払われると、（すなわち、固定給+インセンティブ給が支払われると）、インストラクターは、管理者より評価を獲得しやすい学生の知的能力開発に彼らの努力を傾注する傾向がみられる。この背景には、固定給の下で作用する内発的モチベーションが、一部、金銭的インセンティブの導入により浸食されたためと考えられる。

⁶ この特定化は、（主流の）進化ゲーム理論の理論的文献に従っていることに注意しなさい（Weibull, 1995、あるいは、Young, 1998を参照しなさい）。これらの文献では、選好が所与とされている。そして、進化（ないし、学習）過程は、戦略選択に基づいて生じるものであり、選好（利他主義者あるいは利己主義者）のタイプに応じて生じるものではないとされている。私たちのモデルにおいても、両タイプのプレーヤーの選好は所与とされている。しかしながら、間接的進化アプローチにおいては、進化過程は、他より高い客観的な利得をもたらすような主観的効用関数の選択に準じることになる（例えば、Bester and Güth, 1998、あるいは、Güth and Kliemt, 2000を参照しなさい）。私たちのモデルは、客観的利得となるいかなる明確な対象もないので、間接的進化アプローチに従うのは困難である。

論的力学により導かれることになる。このとき、もし、複数の均衡が存在するならば、初期状態に応じて、いずれかが一意的均衡となる可能性がある。さらに、(超)長期的には、人々は、しばしば、非合理的に行動するという事実が、重要となる。このことにより、一般的に(複数の中期均衡が存在するときでさえ)、長期的には、優勢な一意的均衡が存在し、しかも、確率的に安定的な均衡となる。

このような進化論的モデル手法を用いることで、私たちは、本稿において、次のような主要な結論を得ることになる。

- (i) 金銭的インセンティブが不在のとき、すべての利他主義者が貢献し、しかも、いかなる利己主義者も貢献しない、あるいは、わずかな利己主義者しか貢献しない長期均衡が生じる可能性がある。それだけではなく、初期状態において、誰も貢献しないときでさえ、この長期均衡が生じる可能性がある。
- (ii) 貢献行動を増加させるために金銭的インセンティブを導入することは、結果として、貢献の社会規範を崩壊させることになり、中期的には、逆効果となる可能性がある。
- (iii) 金銭的インセンティブが貢献水準に及ぼす効果は、非単調であるかもしれない。すなわち、少規模、中規模、あるいは、大規模の金銭的インセンティブのいずれかが導入されるとき、それぞれ、貢献行動は増加し、また、減少し、そして、最後の場合には、再び、増加する。
- (iv) 金銭的報酬の導入は、短期的および中期的には、貢献行動を増加させるとしても、長期的には、逆効果となるかもしれない。
- (v) いったん、社会規範が締め出されたとしても、金銭的報酬の支援を得て、社会規範は、しばしば、再生させることができるかもしれない(金銭的インセンティブには社会規範の押し込み効果があるかもしれない)。しかしながら、社会的報酬が当初の水準に戻るかどうかは不確実である。
- (vi) 金銭的インセンティブの押し込み効果なしに、社会規範を再生するには、長期の期間を要することになる。

私たちは、本稿で、これらの結論が導かれる論理を考察するが、ここで、これらの結果がなぜ生じるのかを、直観的に、簡潔に説明しておくことにする。まず、(i) 初期状態において、貢献行動に対しなんの金銭的報酬も支払われない場合から始める。私たちは、金銭的インセンティブが不在のとき、とりわけ、2つのタイプの中期均衡に関心がある。すなわち、それらは、すべての利他主義者が貢献行動を選択する均衡、あるいは、誰もまったく貢献行動を選択しない均衡である。前者の場合、ほぼすべての個人が貢献するため、貢献の社会規範が遵守されると確信でき、(しかも、利他主義者にとっては、彼ら自身の) 内発的動機付けと相まって、貢献行動で被る費用を十分に補償することを確信できる。他方、後者の場合、誰も自発的貢献を供給する状況にないので、結果として、貢献の社会規範が存在しなくなる。しかしながら、このように、貢献の社会規範が消滅した状況でも、ある一定の条件の下では、貢献するという社会規範をともなう均衡こそが、確率的に安定的(均衡)である場合がある。このとき、いかなる金銭的インセンティブも十分に長い期間供給されず、しかも、初期状態で、まったく誰も貢献することがなかったとしても、すべての利他主義者は、最終的に(長期的に)、貢献行動を選択することが均衡となる。

私たちは、(ii) すべての利他主義者が貢献するという初期状態の下で、貢献行動に対し、金

銭的報酬を導入した場合、ある一定の条件の下では、いわゆる、社会規範の締め出しが生じることを知っている。価格（／金銭的報酬）の導入は、一方で、貢献する利他主義者の効用を増加（あるいは、減少）させる。そして、他方で、価格導入は、また、利己主義者の貢献を誘発させることになり、ひいては、利他主義者の（社会的報酬の）インセンティブを低下させることになろう。利己主義者により生み出された負の外部性が大きければ、内発的動機付け、価格および社会的報酬の合計はあまりにも低くなるので、利他主義者に貢献を促すことができなくなるかもしれない。これは、（社会規範の）締め出し効果とよばれるものである。（iii）一般的には、価格が貢献に及ぼす効果は非単調である。私たちの結果は、この点において、Gneezy and Rustichini(2000b)の実証結果と類似している。導入された価格が非常に低いならば、（両タイプの貢献者数に及ぼす効果は軽微で、それにとまなう）社会的報酬の増減もわずかであり、利他主義者は貢献を継続し続けるであろう。逆に、価格が非常に高いならば、（両タイプの貢献者数に及ぼす効果は相当なものとなるが、他方）、それは社会的報酬の変動を補填するのに十分な水準であるため、利他主義者は貢献し続けることになる。ところが、価格が中程度に高いならば、（両タイプの貢献者数に及ぼす効果はそれなりの規模であるにもかかわらず、しかしながら）、社会的報酬の変動を補填するのに十分でないため、利他主義者は貢献することを抑制し、その結果、社会的報酬はゼロにまで低下することになろう。また、いかなる利他主義者も貢献しない状況では、社会的報酬が不在となるので、中期的には、利己主義者も貢献するのを止めることになる。かくして、最終的に、すべての貢献者数はゼロにまで低下する。

（iv）利他主義者が、中期的に、貢献を停止せず、そして、中期的に貢献の供給量が増加するときでさえ、実際には、長期的には、貢献は減少していくかもしれない。これは、金銭的報酬が導入された後、貢献者をともなう均衡が、確率的安定性の特性を失うとき、生じることになる。この結果の背後にある直感は、利己主義的な貢献者の数が増加するとき、社会的報酬が低下するため、その結果、貢献から得られる利他主義者の効用が減少し、しかも、この減少は、金銭的インセンティブにより補償されることはないということである。利他主義者が貢献することから獲得できる効用が小さくなるならば、一部の利他主義者が、貢献から得られる効用がゼロ以下に低下すると誤認し、貢献を停止してしまうことは十分に考えられる。したがって、他のすべての利他主義者も、また、貢献するのを止めることになるであろう。

私たちは、次に、（v）権威当局が、意図したことと逆の効果が生じることに気づき、報酬を中断するとき、なにが起こるのかを分析する。社会的報酬は締め出されないが、しかし、大幅に弱体化させられるならば、したがって、利他主義者に貢献を促すためには金銭的報酬が必要とされる場合、かくして、価格を中断させることは、貢献者をともなわない均衡を生じさせることになる。他方、社会規範が金銭的インセンティブにより、さほど弱体化されていないならば、利他主義者は、価格が中断された後でさえ、貢献し続けるであろう。

（vi）利他主義者の誰もが貢献せず、しかし、金銭的報酬が不在の中で、貢献者をともなう均衡が、確率的に安定しているならば、社会規範は、長期的に、再生させることが可能である。しかし、それには非常に長い時間が必要となるであろう。このとき、権威当局は、利他主義的な貢献者および利己主義的な貢献者の両方を惹き付けるような非常に高い価格を導入することで、社会規範の再生をはかろうとする。ただし、利他主義者の存在は他のタイプの貢献者の効用の源泉となるため、この価格は、その後、漸次、低下させられなければならない。このように、社会的報酬の「押し込み（効果）」をはかることは、常に、可能である。しかしながら、金銭的インセ

ンタイプが不在の下で、貢献を保証するのに十分な水準にまで、社会的報酬を引き上げるのはより困難である。正確には、金銭的報酬が社会規範を即座に締め出してしまいうような、まさに最も有害なケースでは、十分に大きな金銭的報酬を導入したとしても、その押し込み効果は、非常に限定的であるかもしれない。それは、うまくいくかもしれないが、しかし、非常に慎重に、価格数列を選択することが必要となり、しかも、最終的な結果は不確実である。このように、社会規範を崩壊させることは、それを再生させることより、はるかに容易であり、さらに、急激になすことができる。その上、社会的報酬の押し込み効果を作用させるには、初めに、非常に高い価格を設定する必要がある、それには、あまりにも費用がかかりすぎ、したがって、実際に、それを実施するのは不可能であるかもしれない。

本稿の残りの部分の構成は、以下のとおりである。第2節では貢献の社会規範の形成・確立・維持・消滅のメカニズムを解析するための進化論的モデルを紹介する。第3節では自発的な貢献システムの下で生じる均衡のタイプを分析する。そして、貢献の社会規範が、長期的に、確率的に安定した均衡として現れる条件が明らかにされる。第4節では金銭的報酬の導入がもたらす社会的報酬の締め出し効果を考察する。第5節で、私たちは、金銭的報酬を中断した場合の結果を検証する。第6節では、金銭的インセンティブの導入がもたらす社会的報酬の押し込み効果について議論する。そして、第7節では、私たちが使用したモデルの基礎となる仮定の果たす役割について議論することで結論とする。

2. モデル

モデルでは、社会は2つのタイプの個人から構成されているとする。ここでは、それを、(参照しやすいように)単純に利他主義者と利己主義者と称することにする。そして、利他主義者と利己主義者の総数を、それぞれ、 \bar{N}_a と \bar{N}_e で表し、さらに、 t 期において、「貢献する」を選択する各タイプの人数を、それぞれ、 N_a^t と N_e^t で表す。さらに、各タイプの個人は、貢献の意思決定に際して、貢献する、あるいは、貢献しない、のいずれかを選択するだけで、貢献量等に関しては事前に設定された水準を維持するものとする。(t 期における)各タイプの効用は、貢献しないを選択したとき、ゼロに標準化されるとする。また、貢献するを選択したとき、利他主義者および利己主義者の効用 u_a^t 、 u_e^t は、それぞれ、次のように表されるとする。

$$(1) \quad \begin{aligned} u_a^t &= \theta_a a + (1 - \theta_a) p + s(N_a^t, N_e^t) - c && \text{for 利他主義者の効用} \\ u_e^t &= p + s(N_a^t, N_e^t) - c && \text{for 利己主義者の効用} \end{aligned}$$

ただし、 $0 < \theta_a \leq 1$ である。効用関数からわかるように、両タイプの個人は、社会的報酬に対し同等の評価を持つとする。私たちは、それを $s^t \equiv s(N_a^t, N_e^t)$ で表す。また、両タイプとも、貢献するとき、貢献(の参加)費用 c を負担するとする。しかしながら、利他主義者および利己主義者は、金銭的インセンティブ(の導入)から正の効用を得るが、金銭的インセンティブおよび自己実現の欲求充足の主観的価値評価には差異があるとする。すなわち、利他主義者は、自己実現の欲求充足および金銭的インセンティブに、それぞれ、 $0 < \theta_a \leq 1$ および $(1 - \theta_a)$ の(相対的)評価ウェイトを置くが、他方、利己主義者は、自己実現の欲求充足および金銭的インセンティブに、それぞれ、ゼロおよび1の(相対的)評価ウェイトを置くとする⁷。

利他主義者および利己主義者の違いのもうひとつは、両タイプの個人が、貢献活動を行うとき、他の個人に対する外部効果が異なって現れることである。先に示したように、社会的報酬は、利他主義者の貢献者数および利己主義者の貢献者数に依存する。すなわち、社会的報酬 s^t は、両タイプの貢献者数 N_a^t と N_e^t の関数 $s^t = s(N_a^t, N_e^t) \geq 0$ として表せる。私たちは、利他主義者の貢献者がいないならば、社会的報酬はゼロとなり、そして、利他主義者の貢献者数が増えるにともない、それは増加するとする。さらに、社会的報酬は、利己主義者の貢献者数が増えるにつれ、減少すると仮定している。したがって、貢献する利他主義者は、他の個人に対し、正の外部性を創造し、そして、利己主義者は負の外部性を創造することになる。公式的には、 $s(0, N_e^t) = 0$ であり、また、 $s(N_a^t, N_e^t) > 0$ を満たすすべての (N_a^t, N_e^t) について、 $\partial s^t / \partial N_a^t > 0$ 、 $\partial s^t / \partial N_e^t < 0$ であるといえる。

ここで、社会的報酬関数 (social reward function) の解釈、とりわけ、その源泉について議論しておくことは意味があろう。社会的報酬は、本来、ある特定の社会規範 (／集団規範) に従う倶楽部、協会等の大きなソサエティーにメンバーとして属し、しかも、「少なくとも部分的に利他主義的な理由でソサエティーに貢献する」とき、メンバーは社会的承認／称賛を得ることになり、その結果、社会的承認欲求を満たすことになる。したがって、社会的報酬は、外在的インセンティブであるが非金銭的報酬であるといえる。このとき、社会的報酬の大きさは2つの要因に依存していると仮定される。そのひとつは、どれくらい多くの人たちが利他的動機付けにより駆り立てられ、あるいは、利己的動機付けにより駆り立てられているかということである^{8 9}。もうひとつの要因は、また、利他主義的な貢献度がどれくらい高いかどうかである¹⁰。しかしながら、私たちは、利他主義者は貢献するあるいは貢献しないの選択のみを行い、貢献量の水準の選択は行わないと仮定し、第二の要因は無視している。

社会変動のダイナミクスを考察するため、私たちは、無作為に抽出されたある特定の個人が、各期に、貢献するあるいは貢献しない、をどのように選択するかを考える。このとき、この個

⁷ 利己主義者に、 θ_a に類似するパラメータ θ_e を導入することで、結果に影響を及ぼさずに、正の内発的動機付けを与えることは可能であろう。しかしながら、私たちは、利他主義者および利己主義者の貢献が社会的報酬 (以下で、定義される) に及ぼす影響を明確に区別しているので、両タイプの動機付けに明確な差異があることは重要である。したがって、利他主義者だけが十分に大きな内発的動機付け (自己実現の欲求充足) を持つとすることは (θ_a が十分に大きい値となることは)、そのことを保証することになる。

⁸ これに関する別のモデルとして、Lindbeck et al. (1999) を挙げるができる。そこでも、社会的報酬は、社会規範を遵守する個人が増加するとともに増加するとされている。

⁹ 社会的報酬は、社会的承認欲求の充足と同等であるかもしれない。人は、倶楽部、協会、社交界等のさまざまなソサエティーに入会することで社会的承認欲求を満たそうとする。このとき、各個人が支出する費用 c は倶楽部、協会等への入場料 (／入会金) と考えることができる。利己主義者は、入会金を支払い、ソサエティーに入会するが、倶楽部ないし協会の名声の確立・維持のための行動規範を遵守することをせず (慎みのある社会的に望ましい行動を取ることをせず)、倶楽部ないし協会の名声・称賛の恩恵にのみ預かろうとする (他の倶楽部員、協会員にイージー・ライドすることで社会的承認欲求を満たそうとする)。したがって、利他主義的会員は他の会員に対し正の外部性を創造するが、利己主義的会員は負の外部性を創造することになる。

¹⁰ この仮定に関する幾つかの実験的な証拠は、Fehr and Gächter (1999) の実証結果の中にみられる。彼らは、他の個人の貢献度 (寄付額) が高いならば、このとき、貢献することは、より高い社会的称賛 (／承認) を得ることになることを明らかにしている (Andreoni and Petrie, 2000を参照しなさい)。

人の意思決定は、前期 (t 期) において、貢献することから得られる効用 u^t を観察することでなされる。すなわち、無作為に抽出されたこの個人の最適行動は、当該個人のタイプの前期の効用が、ゼロより大きいならば、当期 ($t + 1$ 期) において貢献するを選択することである。したがって、 N_i^t 、ただし、 $i = a, e$ 、のダイナミクスは、次のように表される。

$$(2) \quad \begin{array}{l} \text{If } u_i^t > 0 \quad \text{and } N_i^t < \bar{N}_i \quad \text{then } N_i^{t+1} = N_i^t + 1 \quad \text{or } N_i^{t+1} = N_i^t, N_i^t - 1 \\ \text{If } u_i^t < 0 \quad \text{and } N_i^t > 0 \quad \text{then } N_i^{t+1} = N_i^t - 1 \quad \text{or } N_i^{t+1} = N_i^t, N_i^t + 1 \end{array}$$

その他の場合には、 N_i^t は変化しない。この動学ルールに従うと、(中期) 均衡は、 $u_i^t = 0$ 、ただし、 $i = e, a$ 、のとき、または、 $u_i^t > 0$ かつ $N_i = \bar{N}_i$ のとき、さらに、 $u_i^t < 0$ かつ $N_i = 0$ のときに達成されるといえる。私たちは、このとき、均衡における利他主義者および利己主義者の貢献者数を、それぞれ、 N_a および N_e で表すとす。

上述の (中期的な) 動学ルールは、各個人は、常に、他の個人の直近の行動に対し、最適反応となるよう自らの行動を選択するという仮定に基づいている。確かに、誤りを犯す確率が十分に低いならば、社会システムの短期的進化は、先のような決定論的最適 (反応) 動学により、ほぼ確実に支配されるであろう。この短期的な動学結果として生じる均衡は、中期均衡と呼ばれている。しかも、複数の中期均衡が存在するとき、いずれが支配的となるかは、初期状態に依存する。それは、ある領域集合が、ある吸引力の盆地/窪地 (basin of attraction) を構成するとき、しかも、(社会システムの) 初期状態が、その吸引力の盆地に位置するとき、確率 1 で、ある特定の中期均衡に収束することになるからである。

しかしながら、私たちは、1 節で言及したように、動学ルールが確率論的であるとしているので、小さな確率 ε ではあるが、エージェントは、(最適行動ではない) 「間違っ」行動を選択する可能性がある。すなわち、(最適行動が) 貢献すべきとき、エージェントが貢献しないを選択し、また、逆に、そうすべきでないとき、貢献するを選択する。公式的にいえば、例えば、利他主義者が、貢献を選択するときの効用が正のとき、多くの場合、利他主義者の貢献者数が一人だけ増加する可能性が最も高いが、しかし、実際には、一人だけ減少するか、または、変化しない確率がわずかながら存在することになる。

このような確率的動学ルールの下では、いかなる均衡にも正の確率で到達する可能性がある。しかしながら、長期的には、当該システムが、確率的に安定的に最もよく出現するある状態が存在する。複数の均衡が存在するとき、私たちは、その状態を、均衡の半径 (radius) および均衡の共半径 (coradius) の概念を使用して、確率的に安定性を決定することができる (Ellison, 2000 を参照しなさい)。その手順は以下の通りである。まず、最初に、(複数の均衡の) それぞれに関する吸引力の窪地を見つける (一定の動学ルールが作用するそれぞれの均衡を含む領域を明らかにする)。次に、それぞれの均衡を含む吸引力の窪地から逸脱するのに必要な最小誤差として定義される均衡の半径を求める。この均衡の半径を、この均衡の吸引力の盆地に到達するのに必要な最小誤差のうち、最大の (必要) 誤差を持つ均衡の共半径と比較する。当該均衡の半径がこの均衡の共半径を超えるならば、確率的に安定した状態はすべてこの均衡に含まれる。また、そのような均衡が一つしかないならば、それは一意的な確率的に安定均衡といえる。直観的にいえば、均衡の半径は、当該均衡から逸脱するのがどれくらい難しいのかを示すものであり、他方、均衡の共半径は、他の均衡から当該均衡へ到達するのがどれくらい難しいのかを表すものである。したがって、当該均衡へ到達することよりも、当該均衡から逸脱することの方が困難である

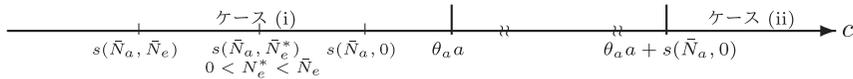
均衡は、確率的に安定しているといえる。私たちは、ある均衡を逸脱するのに最大誤差を要する均衡を長期均衡と定義し、中期均衡と区別することにする。

3. 自発的貢献と均衡

私たちは、金銭的インセンティブが貢献行動にまったく関与しないところから分析を始める。このとき、非金銭的インセンティブ、すなわち、自己実現の欲求充足（内発的動機付け）および社会的報酬のみがエージェントのモチベーションを誘発する源泉となる。したがって、利他主義者および利己主義者の両タイプの効用は、それぞれ、次のように表されることになる¹¹。

$$(3) \quad \begin{aligned} u_a^t &= \theta_a a + s(N_a^t, N_e^t) - c \\ u_e^t &= s(N_a^t, N_e^t) - c \end{aligned}$$

非金銭的インセンティブしか作用しない状況の下では、すなわち、(2) 式によって与えられた決定論的動学ルールの下では、パラメータ値、および、初期状態に応じて、複数の（中期）均衡が生じる可能性がある。そこで、結果0において、私たちは、直感的理解のため、まず、極端なパラメータ値を仮定することで、一意的均衡が生じる二つのケースを考察することから始める。



結果0：パラメータ値の範囲

結果0. $p = 0$ および $c < \theta_a a$ 、あるいは、 $p = 0$ および $\theta_a a + s(\bar{N}_a, 0) < c$ であるとするならば、それぞれのケースについて、一意的均衡が存在する。

- (i) $p = 0$ および $c < \theta_a a$ であるならば、一意的均衡 $(N_a, N_e) = (\bar{N}_a, N_e^t)$ 、 $N_e^t = \bar{N}_e$ 、 $N_e^* = 0$ 、が存在する。ただし、 N_e^* は $u_e = s(\bar{N}, N_e^*) - c = 0$ を満たす N_e の値である。
- (ii) $p = 0$ および $\theta_a a + s(\bar{N}_a, 0) < c$ であるならば、一意的均衡 $(N_a, N_e) = (0, 0)$ が存在する。

証明. 極端なパラメータ値の仮定の下で、一意的均衡が存在することを確かめる。

- (i) パラメータ値の範囲が $c < \theta_a a$ のケースをみてみる。このとき、利他主義者の自己実現の欲求充足（内発的動機付け）は貢献費用より大きいため、すべての利他主義者は、（他の個人の貢献の意思決定がどうであれ）、貢献することを選択する。結果として、すべての利他主義者が貢献する一意的均衡 E_a が存在することになる。ただし、利己主義者の貢献者数は、社会的報酬が貢献費用を補償するかどうかに依存して決まることになる。すなわち、 $s(\bar{N}_a, \bar{N}_e) > c$ のとき、すべての利己主義者が貢献し、また、 $s(\bar{N}_a,$

¹¹ 金銭的報酬のインセンティブ効果が存在しないため、利他主義者の自己実現の欲求充足は、金銭的インセンティブからなんの影響も受けない。このとき、本来、 $\theta_a = 1$ とするのが適切かもしれない。しかしながら、元来、 θ_a は1に近い値であり、結論にはなんら影響を及ぼさないので、効用関数の一貫性をより重視し、ここでは、 θ_a 、ただし、 $0 < \theta_a \leq 1$ 、を用いて表記している。

0) $< c$ のとき、利己主義者の誰も貢献せず、さらに、その中間の値 $s(\bar{N}_a, N_e^*) = c$ をとるとき、一部の利己主義者 N_e^* が貢献することになる。

- (ii) 一意的均衡を生じるもうひとつのケースは、パラメータ値の範囲が $c > \theta_a a + s(\bar{N}_a, 0)$ のケースである。利他主義者は、このとき、獲得可能な最大の非金銭的報酬を得たとしても、貢献に要する費用を補填することができない。したがって、いかなる利他主義者も貢献を選択することはない。(また、利己主義者の効用は、常に、利他主義者の効用よりも小さいので ($u_e < u_a < 0$)、彼らの誰も貢献することはない)。結果として、利他主義者および利己主義者の誰も貢献しない一意的均衡 E_n が存在することになる。

□

一意的均衡を生じる二つのケースに対応する社会動学を図1に示している。図の表記、無差別曲線の形状および動学ルール等の詳細については、結果1の議論の際に触れることにする。ここでは、直感的な理解が可能な程度に説明しておくことにする。利他主義者および利己主義者の無差別曲線は、それぞれ、 $u_a = 0$ および $u_e = 0$ で表されている。(3) 式より明らかに、 $u_a > u_e$ であるので、利他主義者の無差別曲線は、利己主義者のそれより左側に位置する。また、いずれの無差別曲線の場合も、南東方向(右側)で正となり、その北西方向(左側)で負となる。したがって、(2) 式に示したように、動学ルールは、利他主義者の無差別曲線の右側で N_a 次元の右押し、左側で N_a 次元の左押し圧力が働き、また、利己主義者の無差別曲線の上方で N_e 次元の下押し、下方で N_e 次元の上押し圧力が働く。動学ルールは、 N_a および N_e の二次元となるので、 $\pm 45^\circ$ の小さな矢印 (\Rightarrow) で表されることになる。この動学ルールに従えば、利他主義者の無差別曲線の右側の領域は、($u_a > 0$ であるので、この領域のすべてにおいて)、利他主義者の貢献者を吸引する盆地となり、図1(a)のように、すべての利他主義者が貢献する均衡 E_a が生じることになる。他方、これとは逆に、利他主義者の無差別曲線の左側の領域は、($u_a < 0$ であるの

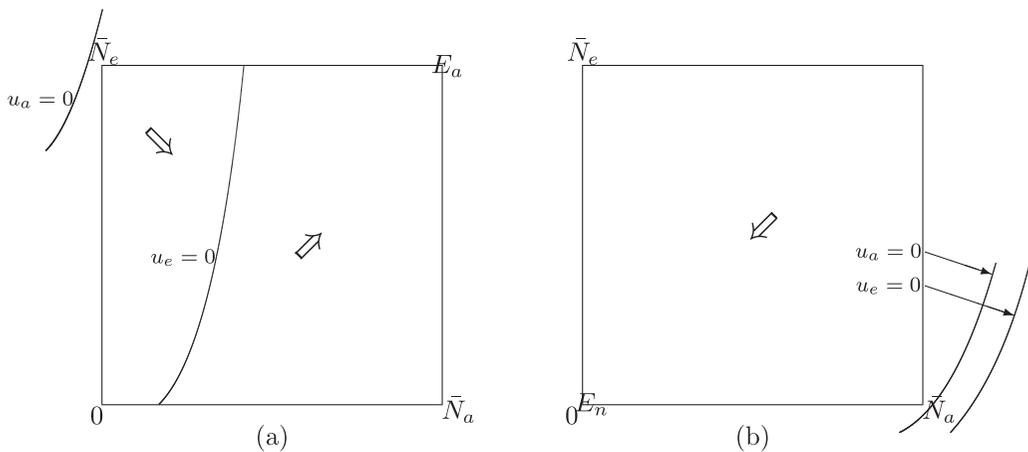


図 (a) : $p = 0$ and $c < \theta_a a$;
 図 (b) : $p = 0$ and $\theta_a a + s(\bar{N}_a, 0) < c$

図 1. 内発的動機付けと一意的均衡

で、この領域のすべてにおいて)、利他主義者の非貢献者を吸引する盆地となり、図1 (b) のように、利他主義者が貢献しない均衡 E_n が生じることになる。

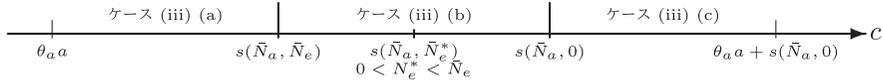
先の二つのケースは、スペクトルの両極端をなすパラメータ値について、(中期の) 一意的均衡が起こりうることを示すものである。これに対し、次に考察する結果1は、スペクトルの中間の値をとるパラメータ値について、複数均衡が生じる興味深いケースを記述したものである。中間のパラメータ値 $\theta_a a < c < \theta_a a + s(\bar{N}_a, 0)$ を仮定したとき、利他主義者が貢献するには、内発的動機付けだけでは十分ではないが ($\theta_a a < c$)、社会的報酬が追加されるならば ($c < \theta_a a + s(\bar{N}_a, 0)$)、貢献行動を誘発するかもしれない (図「結果1：パラメータ値の範囲」を参照しなさい)。したがって、このパラメータ値の範囲については、誰も貢献しない均衡、また、一部の利他主義者が貢献する均衡、さらに、すべての利他主義者が貢献する均衡の3つの可能性が考えられる。ただし、このとき、利己主義者の貢献者数は、貢献に要する費用を補償する社会的報酬の水準に依存することになる。

結果1. $p = 0$ および $\theta_a a < c < \theta_a a + s(\bar{N}_a, 0)$ であるとするならば、次の3つのタイプの均衡が存在する。

- (i) $N_a = N_e = 0$
- (ii) $N_a = N_a^*$ 、および、 $N_e = 0$ 、ただし、 N_a^* は $\theta_a a + s(N_a^*, 0) - c = 0$ を満たす N_a の値である。
- (iii) $N_a = \bar{N}_a$ 、および、
 - (a) $N_e = \bar{N}_e$ if $c < s(\bar{N}_a, \bar{N}_e)$
 - (b) $N_e = N_e^*$ if $s(\bar{N}_a, \bar{N}_e) < c < s(\bar{N}_a, 0)$ 、ただし、(結果0と同様、ただし、値域は異なる)、 N_e^* は $s(\bar{N}_a, N_e^*) - c = 0$ を満たす N_e の値である。
 - (c) $N_e = 0$ If $c > s(\bar{N}_a, 0)$

証明. (i) – (iii) に示した3つのケースでは、均衡は、それぞれ、 $N_a = 0$ 、 $N_a = \bar{N}_a$ 、あるいは、($u_a = 0$ が満たされる) $0 < N_a^* < \bar{N}_a$ とされている。そこで、これらの3つの均衡の可能性について考察する。ただし、 u_a 、 u_e はそれぞれの均衡での効用を表すとする。

- (i) $N_a = 0$ ならば (このとき、社会的報酬 $s(0, N_e^t)$ はゼロであり、しかも)、仮定より $0 < \theta_a a < c$ であるので $u_e < u_a < 0$ である。したがって、 $N_e = 0$ でなければならない。かくして、均衡において、 $N_a = N_e = 0$ であることは容易にわかる。
- (ii) $u_a = 0$ を満たす $0 < N_a^t < \bar{N}_a$ が存在すると仮定する。また、任意の N_a^t および N_e^t について、 $u_e^t < u_a^t$ が常に成り立つ。このとき、 $u_a = 0$ を満たす $0 < N_a^t < \bar{N}_a$ について、 $u_e^t < u_a^t = 0$ が成り立つことを意味する。したがって、 $u_a = 0$ および $u_e < 0$ を満たす唯一可能な均衡 (状況) は、 $N_a = N_a^*$ および $N_e = 0$ であることと同等である。ただし、 N_a^* は $u_a = \theta_a a + s(N_a^*, 0) - c = 0$ を満たさなければならない。
- (iii) $N_a = \bar{N}_a$ ならば、それぞれのパラメータ値に応じて、 N_e について、次の3つのケース (a) – (c) が生じる可能性がある。(図「結果1；パラメータ値の範囲」は、それぞれのケースについて、対応するパラメータ値の範囲を示したものである)。



結果 1 : パラメータ値の範囲

- (a) $(\theta_a a <) c < s(\bar{N}_a, \bar{N}_e)$ のとき、すべての $N_e^t \leq \bar{N}_e$ について、利己主義者が貢献することから得られる効用は正である。したがって、均衡において、利己主義者の貢献者数は $N_e = \bar{N}_e$ となる。
- (b) $s(\bar{N}_a, \bar{N}_e) < c < s(\bar{N}_a, 0)$ のとき、 $N_e^t = 0$ であれば、社会的報酬は貢献費用を上回ることになり、利己主義者は、(次期以降に)、貢献することを選択することになる。しかしながら、利己主義者の貢献者数が $N_e^t = \bar{N}_e$ にまで達すれば、利己主義者の効用は負となる。したがって、中間のある値 $0 < N_e^* < \bar{N}_e$ において、利己主義者が貢献することおよび貢献しないことが無差別となる。かくして、均衡において、(均衡の条件は $u_e^t = 0$ であるので)、利己主義者の貢献者数は $N_e^t = N_e^*$ 、ただし、 $u_e = s(\bar{N}_a, N_e^*) - c = 0$ 、となる。
- (c) $s(\bar{N}_a, 0) < c < \theta_a a + s(\bar{N}_a, 0)$ のとき、社会的報酬が最大のときでさえ ($s(\bar{N}_a, 0)$ のとき最大)、利己主義者が貢献することから得られる効用は負となる。したがって、均衡において、利己主義者の貢献者数は $N_e = 0$ となる。

N_e が先の (iii) (a) - (c) に示した値をとるとき、 $u_a \geq 0$ であることを明らかにすることが、まだ、残されている。3つの考えられる可能なケースは、 $N_e = \bar{N}_e, u_e = 0$ (を満たす $N_e = N_e^*$ 、ただし、 $0 < N_e^* < \bar{N}_e$)、あるいは、 $N_e = 0$ である。最初の2つのケースでは、 $u_e \geq 0$ であるので、 $u_a > u_e \geq 0$ を意味する。最後のケースでは、均衡において $N_e = 0$ および $N_a = \bar{N}_a$ が満たされている。このとき、仮定より、 $u_a = \theta_a a + s(\bar{N}_a, 0) - c \geq 0$ が成り立つことになる。これは、すべての利他主義者は、均衡で、貢献することを意味する¹²。□

結果 1 から、すべての利他主義者が貢献することを最低限保証するパラメータ値 $\theta_a a < c < \theta_a a + s(\bar{N}_a, 0)$ の下では、中期的には (決定論的動学ルールに従うと)、複数均衡が存在することがわかる。しかしながら、この場合でも、初期状態に応じて、それらの均衡 (E_a あるいは E_n) のいずれかが、一意的均衡となる。私たちは、これらのことをさらに直感的に理解できるよう、次に、結果 1 のケース (その一部) を図示することにする。

結果 1 でみた (自発的貢献の下での) 社会動学の一例を図 2 に示している。図の横軸および縦

¹² 結果 1 において、私たちは、内点均衡、 $0 < N_a < \bar{N}_a$ および $0 < N_e < \bar{N}_e$ の場合に表れる整数問題を無視した。これらは (N_a および N_e は)、整数でなければならないので、結果の中で記述された条件を満たす N_a^* および N_e^* が存在しないことも起こりうる。 $u_a = \theta_a a + s(N_a^*, 0) - c = 0$ を満たす整数 N_a^* が存在しないならば、利他主義者が貢献すること、および、貢献しないことが無差別となるような利他主義者の貢献者数 (整数値) が存在しないことになり、したがって、 $N_a = N_a^*$ の内点均衡は存在しないことを意味する。他方、 $u_e = s(\bar{N}_a, N_e^*) - c = 0$ を満たす整数 N_e^* が存在しないならば、均衡は、 N_e^* よりも小さい最大整数 N_e と N_e^* よりも大きい最小整数 N_e との間を揺れ動くことになる。このことは、また、本稿の他のすべての内点均衡についても適用されることになる。

軸は、それぞれ、利他主義者および利己主義者の貢献者数を表し、その範囲は、ゼロからそれぞれのタイプの総人数 \bar{N}_a および \bar{N}_e までである。二つの（実線の）曲線は、それぞれ、利他主義者および利己主義者が貢献することおよび貢献しないことが無差別となるような (N_a^t, N_e^t) の組み合わせの集合（無差別曲線）を表している。効用（関数）の主要因を構成する社会的報酬は、 N_a^t の増加とともに増加し、そして、 N_e^t の増加とともに減少するので、無差別曲線は右上がりとなる¹³。利他主義者および利己主義者の無差別曲線は、それぞれ、 $u_a = 0$ および $u_e = 0$ で表されている。利他主義者が貢献することから得られる効用は、その無差別曲線 ($u_a = 0$) の右側の領域で（利他主義者の貢献者数が増加するので）、正であり、そして、その左側の領域で（利他主

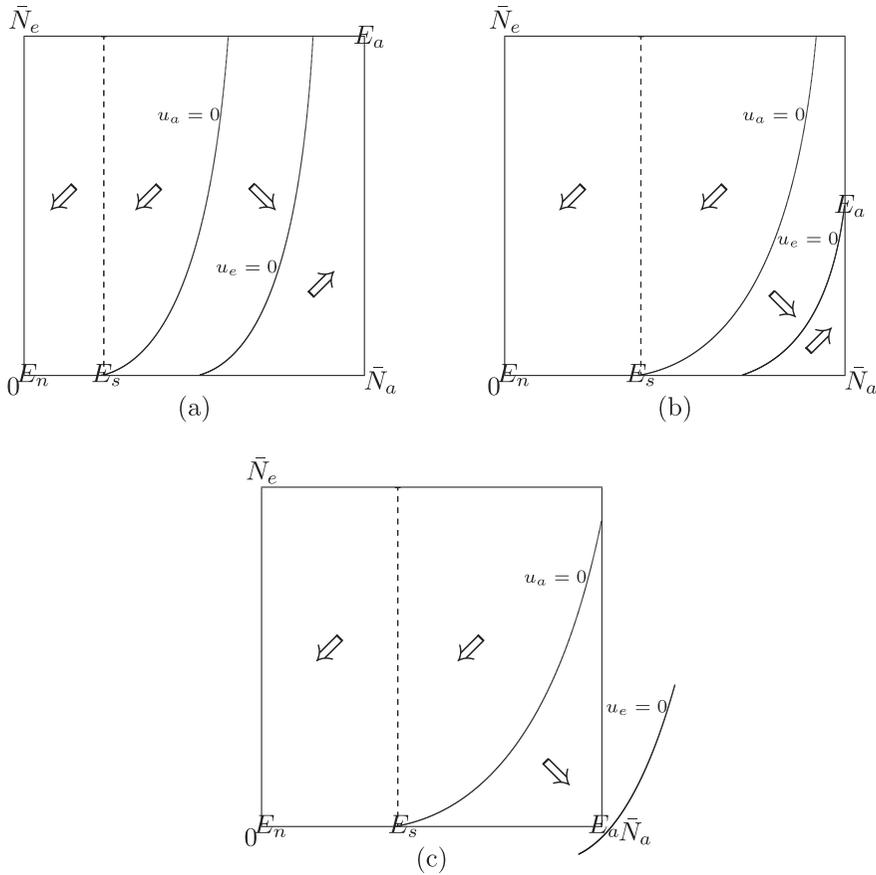


図 (a) : ケース (iii) (a) $\theta_a a < c < s(\bar{N}_a, \bar{N}_e)$;
 図 (b) : ケース (iii) (b) $s(\bar{N}_a, \bar{N}_e) < c < s(\bar{N}_a, 0)$;
 図 (c) : ケース (iii) (c) $s(\bar{N}_a, 0) < c < \theta_a a + s(\bar{N}_a, 0)$

図 2. 内発的動機付けと中期の動学

¹³ 無差別曲線が $u_i = 0$ 、ただし、 $i = a, e$ 、を満たすためには、（無差別曲線上の (N_a^t, N_e^t) の組み合わせについて）、社会的報酬 $s(N_a^t, N_e^t)$ は一定でなければならない。社会的報酬は N_a^t の増加とともに増加し、そして、 N_e^t の増加とともに減少するので、 N_a が増加したとき、 N_e を増加させることで、社会的報酬 $s(N_a^t, N_e^t)$ を一定に保つことができる。したがって、無差別曲線は右上がりの曲線となる。無差別曲線の正確な形状は、社会的報酬関数の形状に依存することになる。

義者の貢献者数が減少するので)、負である。したがって、 N_a^t は無差別曲線の右側で増加し(右押し圧力が働き)、また、その左側で減少する(左押し圧力が働くことになる)。その結果、 N_a は無差別曲線上から逸脱し端点($N_a = 0$ あるいは $N_a = \bar{N}_a$)へ向けて発散していくことになる。(これは、利他主義者が生み出す正の外部性が臨界量効果(critical mass effect)を誘発することにより引き起こされるものである)。これに対し、利己主義者が貢献することから得られる効用は、その無差別曲線($u_e = 0$)の下方で(利己主義者の貢献者数が減少するので)、正であり、そして、その上方で(利己主義者の貢献者数が増加するので)、負である。したがって、 N_e^t は無差別曲線の下方で増加し(上押し圧力が働き)、また、その上方で減少する(下押し圧力が働く)。その結果、 N_e は無差別曲線上から逸脱することなく、無差別曲線上の内点均衡へ収束していく。(これは、利己主義者が生み出す負の外部性の作用によるものである)。

図に示したように、枠内は、利他主義者および利己主義者の無差別曲線により3分割され¹⁴、そして、(2)式に示した)それぞれの領域におけるそれぞれのタイプの貢献者数の増減の組み合わせを、動学ルールとして、小さな矢印(\Rightarrow)で表している。ただし、この動学ルールは、 N_a および N_e の二次元で、しかも、それぞれの次元を1増減の測度で計測しているため、すべての矢印は $\pm 45^\circ$ で描写される。私たちは、この動学ルールに従う動学経路(初期状態から均衡へのパス)をたどることで、社会システムが、最終的に到達する均衡を知ることができる。具体的には、二つの無差別曲線により分割されたそれぞれの領域から始まる動学経路をたどることで、社会システムが新たに到達する均衡を知ることができる。このとき、私たちは、図の枠内の領域を、3つの異なる均衡に到達する3つの領域に区分することができる。それは、利他主義者の無差別曲線の右側の領域、利他主義者の無差別曲線と破線に囲まれた領域(図2を参照)、および、破線の左側の領域である¹⁵。初期状態がこれらのいずれの領域に位置するかで、動学経路は異なる軌跡をたどることになり、最終的に、それぞれの均衡、すなわち、すべての利他主義者が貢献する均衡、一部の利他主義者が貢献する均衡、および、いずれの利他主義者も貢献しない均衡に到達することになる。図では、これらの均衡を、それぞれ、 E_a 、 E_s 、および、 E_n で表している。私たちは、次に、いずれの領域からパスが始まるかに依存して、社会システムがどのような均衡にいたるのかを、図2に沿って、少し詳しくみていくことにする。

- (i) 安定均衡 E_a : 社会システムが、初期状態において、利他主義者の無差別曲線の右側の領域に位置するとき($u_a > 0$ 、および、 $u_e \geq 0$)、中期的には、すべての利他主義者が貢献する均衡 E_a へ向うことになる。したがって、この領域は、すべての利他主義者が貢献する均衡の吸引力の盆地であるといえる。他方、均衡での利己主義者の貢献者

¹⁴ このとき、利他主義者の無差別曲線 $u_a = 0$ は、利己主義者の無差別曲線 $u_e = 0$ の左側に位置していることに注意しなさい。利他主義者の内発的インセンティブは常に正であるので ($\theta_a a > 0$)、任意の $N_e = N_e^t$ について、 $u_a = \theta_a a + s(N_a^t, N_e^t) - c = 0$ を満たす $N_a = N_a^t$ 、および、 $u_e = s(N_a^t, N_e^t) - c = 0$ を満たす $N_a = N_e^t$ は、 $N_a^t < N_e^t$ でなければならない。したがって、利他主義者の無差別曲線 $u_a = 0$ は利己主義者の無差別曲線 $u_e = 0$ の左側に位置することになる。

¹⁵ 破線は、利他主義者の無差別曲線と横軸との交点からの垂線である。自発的貢献のとき、利他主義者および利己主義者の無差別曲線は横軸 ($N_e = 0$) とどの位置で交差するのだろうか。利他主義者および利己主義者の無差別曲線が横軸と交差する点での、それぞれのタイプの効用は、 $u_a = \theta_a a + s(N_a^t, 0) - c = 0$ 、および、 $u_e = s(N_a^t, 0) - c = 0$ である。このとき、パラメータ値の範囲は $0 < \theta_a a < c$ であるので、 $s(N_a^t, 0) > s(N_a^t, 0) > 0$ でなければならない。すなわち、 $N_e^t > N_a^t > 0$ となる。したがって、利他主義者および利己主義者の無差別曲線は、常に、原点より右側で横軸と交差することになる。

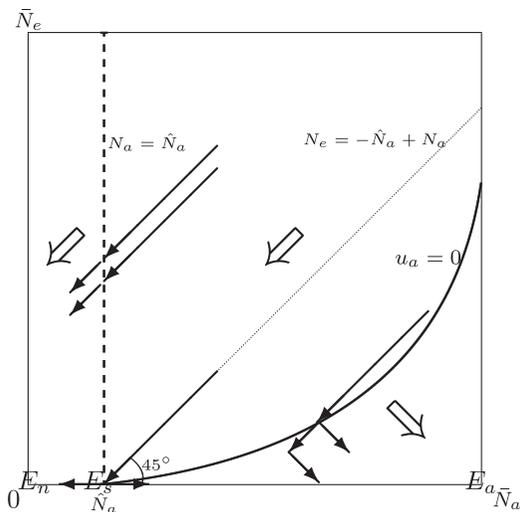
数は、利己主義者の無差別曲線が直線 $N_a = \bar{N}_a$ の左側に位置するとき (図 2 (a))、 $N_e = \bar{N}_e$ となる。また、利己主義者の無差別曲線が直線 $N_a = \bar{N}_a$ と交差するとき (図 2 (b))、交点の値 $0 < N_e^* < \bar{N}_e$ が貢献者数となる。さらに、利己主義者の無差別曲線が枠外に位置するとき (図 2 (c))、貢献者数はゼロとなる。これは、結果 1 (iii) を表したものである。

- (ii) 安定均衡 E_n : 社会システムが、初期状態において、破線 $N_a^t = N_a^*$ の左側の領域に位置するとき ($u_a < 0$ 、および、 $u_e < 0$)、中期的には、利他主義者の誰も貢献しない均衡 E_n へ向うことになる。したがって、この領域は、誰も貢献しない均衡の吸引力の盆地であるといえる。他方、均衡での利己主義者の貢献者数は、社会的報酬が利他主義者の貢献者数の減少と共に減少していくので、結果として、利己主義者の貢献者数もゼロとなる (図 2 (a) - (c))。これは、結果 1 (i) を表したものである。
- (iii) 確率的安定均衡 : 最後に、社会システムが、初期状態において、先の領域以外にあるとき (破線および利他主義者の無差別曲線に囲まれた領域にあるとき) ($u_a < 0$ 、および、 $u_e < 0$)、均衡 E_a あるいは E_n のいずれかが正の確率で生じることになる。より詳しく言えば、図 3 に示したように、 $(\hat{N}_a, 0)$ に端を発する 45° 線の上方の領域から動学経路が始まるならば、図 3 において、太い矢印 (→) で示したように、中期の動学経路は、破線を越えて、その後、均衡 E_n にいたるパスで表せる。他方、動学経路が 45° 線の下方の領域から始まるならば、太い矢印 (→) で示したように、中期の動学経路は、利他主義者の無差別曲線を越え、その後、均衡 E_a にいたるパスで表せることになる。したがって、破線と利他主義者の無差別曲線に囲まれた領域では、均衡 E_a あるいは E_n のいずれかが正の確率で生じることになる¹⁶。

さらに、特殊ケースがある。それは、社会システムが、初期状態において、図 3 に示した 45° 線上 $N_e = -\hat{N}_a + N_a$ に位置するとき、均衡 E_s に到達することである。ところが、これは安定均衡とはならない。このとき、 $N_e = 0$ であるので、($u_e < 0$ となる) E_s では、下押し圧力はなくなる。しかしながら、 $0 < \hat{N}_a < \bar{N}_a$ であるので、($u_a = 0$ となる) E_s では、右押し圧力あるいは左押し圧力が働くことになる。したがって、 E_s は不安定となり、最終的に、安定均衡 E_a あるいは E_n のいずれかが正の確率で生じることになる。これは、結果 1 (ii) を表したものである。

これまで、自発的貢献の中期動学 (決定論的動学) をみてきたが、ここで、長期動学、すなわち、長期均衡について触れておくことにする。長期的には、人々が、彼らの利益に反するような偶発的誤りを犯すとき、初期状態とは関係なく、いかなる均衡も正の確率で生じる可能性がある。とりわけ、初期状態において、貢献レベルがゼロであったとしても、貢献者が存在する均衡が生じる可能性があり、しかも、それは長期的に持続するかもしれない。しかしながら、そのためには、貢献者がいる均衡が確率的に安定していることが求められる。第 2 節で説明した半径 - 共半径法 (radius-coradius method) に従えば、 E_a から利他主義者の無差別曲線までの距離が (誤差の数という意味で)、 E_n からの距離よりも大きいならば、すべての利他主義者が貢献する均衡

¹⁶ 図 2 に示したように、破線 $N_a^t = \hat{N}_a$ および無差別曲線 $u_a = 0$ で挟まれた領域では、決定論的動学ルールに従うと、(一見すると)、均衡 E_n に到達するように見える。しかしながら、初期状態が、図 3 に示したように、 $(\hat{N}_a, 0)$ に端を発する 45° 線の下方の領域にあるならば、太い矢印 (→) が示すように無差別曲線を越えて貢献者数が増加することになり、均衡 E_a に到達することになる。



破線 $N_a = \hat{N}_a$ の右側かつ無差別曲線 $u_a = 0$ の左側の領域において、45°線の上側では（無差別曲線から離れているところでは）、破線を越えて、均衡 E_n に到達する。他方、45°線の下側では（無差別曲線の近傍では）、太い矢印（→）で示したように、無差別曲線を越えて、均衡 E_n に到達する。さらに、不安定均衡 E_a においては、太い矢印（→）で示したように、右押しあるいは左押し圧力が働き、最終的に、正の確率で、安定均衡 E_a あるいは E_n が生じる。

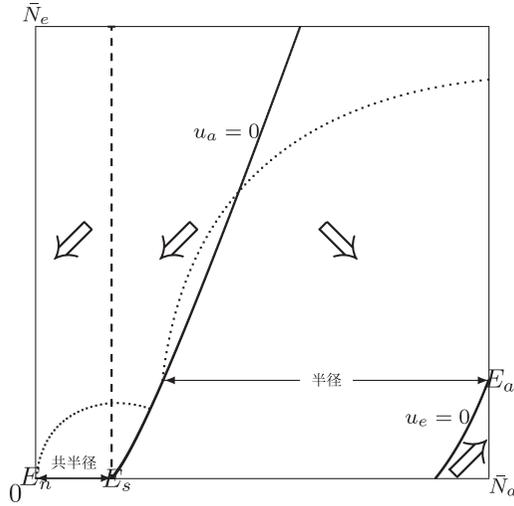
図3. 中期の動学メカニズム

は、確率的に安定していることを容易に示すことができる。この条件は、 $\theta_a a$ が相対的に大きいならば、すなわち、均衡 E_s において N_a の値が小さいならば、また、均衡 E_a において N_e の値が小さいならば、満たされる可能性が最も高くなる。言い換えれば、もし、利他主義者の利他的動機付けが大きく、しかも、ほとんどの利己主義者が、利他主義者によって創造される正の外部性により貢献行動を誘発されないのであれば、自発的貢献が最も生じやすくなる（図4を参照しなさい）。

本稿の残りの部分では、私たちは、(初期状態として)、複数の中期均衡が存在し、しかも、「すべての利他主義者が貢献する均衡が確率的に安定している」ケースに焦点を当て考察する。このケースでは、長期的には、貢献するという社会規範が出現しているが、しかしながら、また、金銭的報酬の導入によってそれが締め出される可能性があることを明らかにする。それ以外のパラメータ値については、結果は明白であり、したがって、それほど興味深いものではない。

4. 金銭的報酬の導入

本節では、非金銭的インセンティブしか作用していない状況において、金銭的報酬を導入したときの効果を考えることにする。私たちは、権威当局が、(ボランティア活動等の) 貢献活動への参加者数（あるいは、個々人の貢献水準）を増加させるため、金銭的インセンティブ $p > 0$ を導入することをよく観察する。これらの施策の背後にある基本的なアイデアは、さまざまな種類の金銭的および非金銭的モチベーションは単純に加算することができるというものである。経



$\theta_a a$ が $s(N_a^t, N_e^t)$ より非常に大きいならば、二つの無差別曲線は離れた位置関係となる。すなわち、均衡 E_s において N_e の値が小さく、しかも、また、均衡 E_a において N_e の値が小さくなるため、均衡 E_a は確率的に安定している。

図 4. 長期の動学メカニズム

経済学は、このような方法で、動機付け問題を、金銭的報酬制度の最適設計問題に単純化させようとしてきた。しかしながら、近年、非金銭的インセンティブしか作用していない状況に、金銭的報酬を持ち込むことは、既存の（非金銭的）モチベーションに大きな影響を及ぼすことが次第に明らかになってきている。そこで、私たちは、本節で、このことについて考察してみることにする。

私たちは、（非金銭的インセンティブしか作用していない状況で、しかも）、貢献者が存在する均衡が確率的に安定している状態から始める。具体的に、（金銭的報酬が導入される時点での）初期状態として、（結果 1 (iii)） (\bar{N}_a, \bar{N}_e) 、ただし、 $0 \leq N_e^t \leq \bar{N}_e$ 、を考える。そして、この初期状態において、金銭的報酬を導入したときの中期動学（エージェントが最適行動を選択する決定論的動学）がどのようなになるのかを考察する。結果 2 および結果 3 は、2つの異なるパラメータ値について（結果 2 では、小規模あるいは大規模の金銭的報酬が導入されたとき、そして、結果 3 では、中規模の金銭的報酬が導入されたとき）、どのような新たな中期均衡が生じるのかを記述したものである。なお、このとき、 $p > 0$ であるので、両タイプの貢献者の効用関数は (1) 式のようになることに注意しなさい。

$$\begin{aligned} u_a^t &= \theta_a a + (1 - \theta_a)p + s(N_a^t, N_e^t) - c && \text{for 利他主義者の効用} \\ u_e^t &= p + s(N_a^t, N_e^t) - c && \text{for 利己主義者の効用} \end{aligned}$$

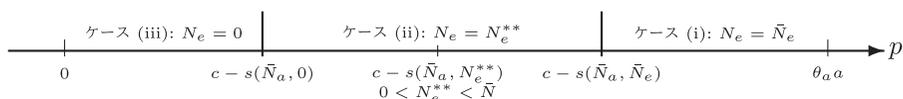
結果 2. $\theta_a a < c < \theta_a a + s(\bar{N}_a, \bar{N}_e)$ であると仮定する。このとき、金銭的報酬 $p > 0$ が導入されるとき、（初期状態において、 $u_a = \theta_a a + s(N_a^0, N_e^0)$ 、ただし、 $N_a^0 = \bar{N}_a$ 、は、新たな中期均衡において、 $u_a = \theta_a a + (1 - \theta)p + s(N_a, N_e)$ となり）、新たな中期均衡では、 $N_a = \bar{N}_a$ であり、しかも、 N_e について、それぞれ、次のことが満たされる。

$$(i) N_e = \bar{N}_e \quad \text{if } p > c - s(\bar{N}_a, \bar{N}_e)$$

- (ii) $N_e = N_e^{**}$ if $c - s(\bar{N}_a, 0) < p < c - s(\bar{N}_a, \bar{N}_e)$
 ただし、 N_e^{**} は $p + s(\bar{N}_a, N_e^{**}) - c = 0$ を満たす N_e の値である。
- (iii) $N_e = 0$ if $p < c - s(\bar{N}_a, 0)$

証明. 私たちは、まず、新たな（中期）均衡において、 $N_a = \bar{N}_a$ であると仮定する。このとき、 N_e が、(i) – (iii) のそれぞれのパラメータ値について、それぞれ、 $N_e = \bar{N}_e$ 、 $N_e = N_e^{**}$ 、および、 $N_e = 0$ の値をとることを明らかにしなければならない。さらに、また、新たな中期均衡、 (\bar{N}_a, \bar{N}_e) 、 (\bar{N}_a, N_e^{**}) および $(\bar{N}_a, 0)$ の下で、 $u_a > 0$ が保証されることを明らかにしなければならない。（すなわち、 $N_a = \bar{N}_a$ の仮定の下で、 $u_a(\bar{N}_a, N_e^t) > 0$ 、ただし、 $N_e^t = \bar{N}_e, N_e^{**}, 0$ 、が保証されることを明らかにしなければならない）。

私たちは、まず、前半の部分を証明することから始める。中期均衡において、 $N_a = \bar{N}_a$ であると仮定する。このとき、 N_e について、(i) – (iii) に示した三つのケースが成り立つことを明らかにする。（図「結果2：パラメータ値の範囲」は、それぞれのケースについて、パラメータ値の範囲を示したものである）。



結果2：パラメータ値の範囲

- (i) $p > c - s(\bar{N}_a, \bar{N}_e)$ および $N_a = \bar{N}_a$ ならば、 $u_e = p + s(\bar{N}_a, \bar{N}_e) - c > 0$ であるので、利己主義者は、いかなる $0 \leq N_e^t \leq \bar{N}_e$ についても、正の効用 ($u_e^t > 0$) を得る。したがって、均衡において、すべての利己主義者が貢献することになる ($N_e = \bar{N}_e$)。
- (ii) $c - s(\bar{N}_a, 0) < p < c - s(\bar{N}_a, \bar{N}_e)$ および $N_a = \bar{N}_a$ ならば、 $u_e = p + s(\bar{N}_a, \bar{N}_e) - c < 0$ および $u_e = p + s(\bar{N}_a, 0) - c > 0$ であるので、利己主義者の貢献者数が $N_e = \bar{N}_e$ ならば、負の効用を、そして、また、利己主義者の貢献者数がゼロの近傍まで低下すると、正の効用を得る。したがって、その中間のある貢献者数 $0 < N_e^{**} < \bar{N}_e$ について、 $p + s(\bar{N}_a, N_e^{**}) = c$ となる。かくして、均衡において、利己主義者の貢献者数は $N_e = N_e^{**}$ である。ただし、 N_e^{**} は $u_e = p + s(\bar{N}_a, N_e^{**}) - c = 0$ を満たす N_e の値である。
- (iii) (最大の社会的報酬 $s(\bar{N}_a, 0)$ のときでさえ)、 $p < c - s(\bar{N}_a, 0)$ および $N_a = \bar{N}_a$ ならば、 $u_e = p + s(\bar{N}_a, 0) - c < 0$ であるので、利己主義者は、いかなる $0 \leq N_e^t \leq \bar{N}_e$ についても、負の効用を得る。したがって、均衡において、利己主義者の誰も貢献することはない ($N_e = 0$)。

私たちは、さらに、後半の部分、すなわち、仮定したパラメータ値の範囲の下で、 $u_a > 0$ が保証されることを証明しなければならない。仮定より、(最小の社会的報酬 $s(\bar{N}_a, \bar{N}_e)$ のときでさえ)、 $c < \theta_a a + s(\bar{N}_a, \bar{N}_e)$ であるので、すべての $0 \leq N_e^t \leq \bar{N}_e$ について、 $c < \theta_a a + s(\bar{N}_a, \bar{N}_e) \leq \theta_a a + s(\bar{N}_a, N_e^t)$ である。したがって、 $u_a^t = \theta_a a + (1 - \theta_a)p + s(\bar{N}_a, N_e^t) - c > 0$ が保証される。（これで、証明は完結するが、補足として次のことを指摘しておく）。ただし、このとき、注意しなければならないのは、 $p > a$ であるならば、 $p > \theta_a a + (1 - \theta_a)p$ であるので、 $p + s(\bar{N}_a, N_e^t) - c = u_e > u_a = \theta_a a + (1 - \theta_a)p + s(\bar{N}_a, N_e^t) - c > 0$ であり、そして、 $p < a$ であるならば、 $p < \theta_a a$

+ $(1 - \theta_a)p$ であるので、 $p + s(\bar{N}_a, N_e^t) - c = u_e < u_a = \theta_a a + (1 - \theta_a)p + s(\bar{N}_a, N_e^t) - c$ 、ただし、 $u_a > u_e \geq 0$ あるいは $u_a > 0 \geq u_e$ 、となることである。□

結果2は、初期状態の利他主義者の貢献者数になんら影響を及ぼさないパラメータ値の条件を明らかにするものである。すなわち、パラメータ値の範囲を $\theta_a a < c < \theta_a a + s(\bar{N}_a, \bar{N}_e)$ と仮定したとき、利己主義者がいかなる行動を選択しようとも、利他主義者は、決して、貢献を思いとどまることがない。これに対し、利己主義者の貢献者数は、 (\bar{N}_a, N_e^0) の初期状態からスタートしたとすれば、利己主義者がさらに貢献するのは、少なくとも、 $c < p + s(\bar{N}_a, N_e^t)$ 、ただし、 $N_e^t < N_e^t \leq \bar{N}_e$ 、が満たされるときだけである。私たちは、 $p > 0$ を導入するとき（増加させるとき）、 $(N_e$ が増加し、ひいては)、 s^t を減少させるため、 p および N_e^t の適切な組み合わせを模索しなければならない。このとき、適切な組み合わせのひとつは、 p の増加を最小限にし、 s^t の低下を抑制する方法である（結果2(ii)）。もうひとつは、 p を大きく増加させ、 s^t の低下分を補填する方法である（結果2(i)）。私たちは、これらの二つの価格導入戦略に関して、前者は金銭的報酬 p が小規模ケース、および、後者は大規模ケースと定義することができるかもしれない。

結果2でみた価格導入後の動学経路の一例を図5に示している。図において、(破線の)曲線は、(価格導入前の)利他主義者および利己主義者の古い無差別曲線 $u_a^0 = 0$ および $u_e^0 = 0$ を表し、また、小さな矢印 (\Rightarrow) は古い無差別曲線の下での動学ルール、そして、 E_a^0 は (すべての利他主義者が貢献する) 古い均衡を表している。さらに、(実線の)曲線は、価格導入後の利他主義者および利己主義者の新たな無差別曲線 $u_a = 0$ および $u_e = 0$ を表し、また、大きな矢印 (\triangleright) は新たな無差別曲線の下での動学ルール、そして、 E_a 、 E_s 、および、 E_n は新たな均衡を表している。さらに、太い矢印 (\rightarrow) で、(価格導入後に生じる)、旧均衡 E_a^0 から新均衡 E_a への動学経路を示している。したがって、私たちは、この太い矢印をたどることで、同じ (古い均衡を持つ) 社会システムが、異なる価格水準の導入を契機として、異なる動学経路を経て、新たな (異なる) 均衡を持つ社会システムにいたる過程を観察することができる。

動学経路 (初期状態から新たな均衡へのパス) をみていく前に、価格導入に関わる両タイプの無差別曲線のシフトの仕方について、簡単に触れておくことにする。まず、価格導入により生じる変化は、初期状態 E_a^0 が、新たな無差別曲線の右側の領域に位置するようになることである。さらに、詳しくいえば、私たちは、パラメータ値の範囲を $\theta_a a < c < \theta_a a + s(\bar{N}_a, \bar{N}_e)$ と仮定しているので、(すなわち、 $u_a(\bar{N}_a, \bar{N}_e) > 0$ となるので)、利他主義者の無差別曲線は、導入される価格水準に関わらず、常に、座標 (\bar{N}_a, \bar{N}_e) の左側に位置する (図5)。これに対し、利己主義者の無差別曲線は、導入される価格水準に依存し、その位置関係が定まることである。まず、大規模な金銭的報酬 $p > a > 0$ が導入されるとき、利己主義者の無差別曲線は利他主義者の無差別曲線の左側に位置することになる ($u_e > u_a$)。そして、また、小規模の金銭的報酬 $0 < p < a$ が導入されるとき、価格導入前と変わらず、利他主義者の無差別曲線の右側に位置する ($u_a > u_e$)¹⁷。ただし、このとき、 $p + s(\bar{N}_a, N_e^0) = u_e(\bar{N}_a, N_e^0) > u_e^0 = s(\bar{N}_a, N_e^0)$ となるので、利己主義者の新

¹⁷ 内発的インセンティブ a を分類基準とするのは、この価格水準を境に ($p \geq a$)、二つの無差別曲線の位置関係が逆転するからである。 $p > a$ を仮定した場合、 $\theta_a a + (1 - \theta_a)p < p$ であるので、任意の $N_e = N_e^t$ について、 $u_a = \theta_a a + (1 - \theta_a)p + s(N_a^t, N_e^t) = 0$ を満たす $N_a = N_a^t$ 、および、 $u_e = p + s(N_a^t, N_e^t) = 0$ を満たす $N_a = N_a^t$ は、 $N_a^t > N_e^t$ でなければならない。したがって、 $u_a = 0$ は $u_e = 0$ の右側に位置することになる。逆に、 $p < a$ を仮定した場合、 $u_a = 0$ は $u_e = 0$ の左側に位置することになる。

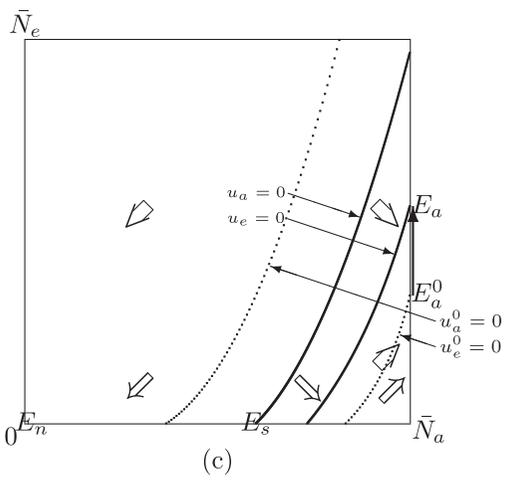
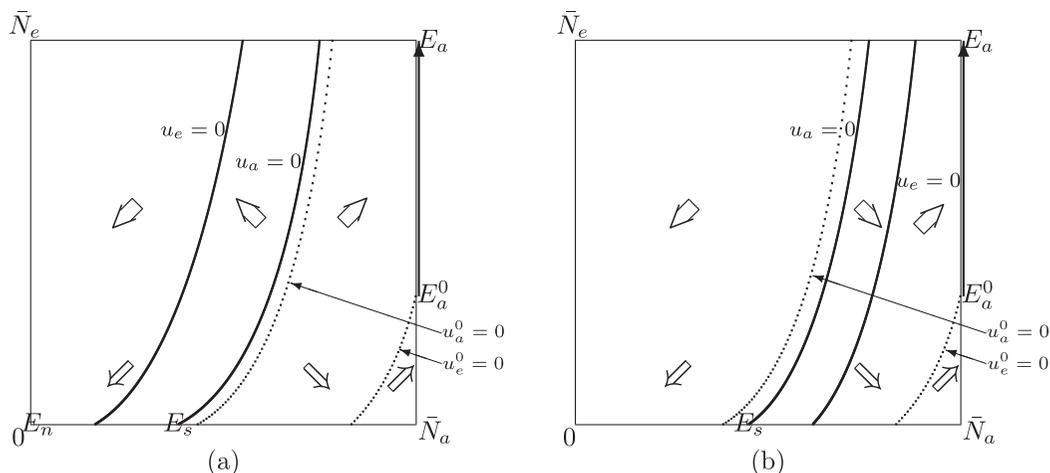


図 (a) : $\theta_a a < c < \theta_a a + s(\bar{N}_a, \bar{N}_e)$ および $p > a$;
 図 (b) : $\theta_a a < c < \theta_a a + s(\bar{N}_a, \bar{N}_e)$ および $p < a$;
 図 (c) : $\theta_a a < c < \theta_a a + s(\bar{N}_a, \bar{N}_e)$ および $p < a$

図 5. 小規模および大規模の金銭的報酬の導入と中期の動学

たな無差別曲線は、利他主義者の無差別曲線と初期状態との間に位置することになる。私たちは、導入価格の分類基準 $p \geq a$ による小規模および大規模に沿って、次に、社会システムの動学経路をみていくことにする。

(i) 大規模な金銭的報酬を導入したときの中期動学 大規模価格 $0 < a < p$ が導入される
 とき、古い無差別曲線の位置関係は新たな無差別曲線では、逆転することになる ($u_e > u_a$)。しかも、パラメータ値の範囲を $\theta_a a < c < \theta_a a + s(\bar{N}_a, \bar{N}_e)$ と仮定している
 ので、 $u_e(\bar{N}_a, \bar{N}_e) > u_a(\bar{N}_a, \bar{N}_e) > 0$ となり、初期状態の社会システム E_a^0 および座標 (\bar{N}_a, \bar{N}_e) は、利他主義者および利己主義者の無差別曲線の右側の領域に位置することになる。したがって、決定論的動学ルールに従えば、中期的には、すべての利他主義者
 および利己主義者が貢献する新たな均衡 E_a が生じることになる (図 5 (a))。

(ii) 小規模の金銭的報酬を導入したときの中期動学 小規模価格 $a > p > 0$ が導入される
 とき、旧い無差別曲線の位置関係は新たな無差別曲線においても維持されることになる
 ($u_a > u_e$)。しかしながら、パラメータ値の範囲を $\theta_a a < c < \theta_a a + s(\bar{N}_a, \bar{N}_e)$ と仮定
 しているの、(大規模価格のときと同様)、初期状態の社会システム E_a^0 は、利他主義
 者の無差別曲線の右側の領域に位置することになる。したがって、このとき、中期的
 には、すべての利他主義者が貢献する均衡が維持されることになる。これに対し、利
 己主義者の貢献者数は、次の二つの場合が考えられる。

(a) $u_e(\bar{N}_a, \bar{N}_e) > 0$: このとき、座標 (\bar{N}_a, \bar{N}_e) および初期状態 E_a^0 は、利己主義者
 の無差別曲線の右側の領域に位置する。したがって、新たな均衡での利己主義者
 の貢献者数は $N_e = \bar{N}_e$ となる (図 5 (b))。

(b) $u_e(\bar{N}_a, \bar{N}_e) < 0 < u_e(\bar{N}_a, N_e^0)$: このとき、利己主義者の無差別曲線は、座標 $(\bar{N}_a,$
 $\bar{N}_e)$ と初期状態 E_a^0 の中間の値 (\bar{N}_a, N_e^{**}) で、直線 $N_a = \bar{N}_a$ と交差する。した
 がって、新たな均衡での利己主義者の貢献者数は $N_e = N_e^{**}$ 、 $N_e^0 < N_e^{**} < \bar{N}_e$ 、と
 なる (図 5 (c))。

結果として、価格の導入は、いずれの場合も、利他主義者の貢献者数を $N_a = \bar{N}_a$ に維持したま
 ま、利己主義者の貢献者数を増加させることになる。

私たちが、結果 2 から言えることは、利他主義者の間で貢献の社会規範が浸透しているとき、
 内発的動機付けに届かない規模、あるいは、逆に、内発的動機付けを凌駕する規模で金銭的報酬
 が導入されるとき、貢献の社会規範は弱化するのではないということである。この結論は、価格
 の導入が、利他主義者の貢献者のモチベーションを低下させるのではないという意味で、伝統的
 経済学の報酬インセンティブの基本的アイディア (さまざまな種類の金銭的および非金銭的モチ
 ベーションは単純に加算することができるとする仮定) となら矛盾するものではない。

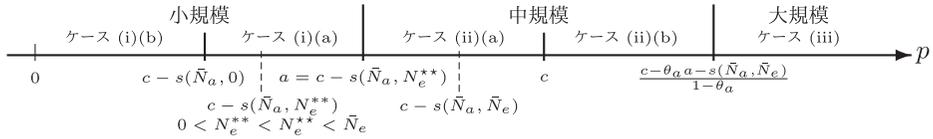
最後に、金銭的報酬の規模の定義について、整理しておくことにする。私たちは、内発的動機
 付けに届かない規模で金銭的報酬が導入されるとき、小規模、そして、また、内発的動機付けを
 凌駕する規模で金銭的報酬が導入されるとき、大規模と定義している。小規模価格の特質は、両
 タイプの無差別曲線の位置関係を変えずに、利己主義者の無差別曲線を北西方向へシフト
 させることである。また、大規模価格の特質は、ひとつは、両タイプの無差別曲線の位置関係
 を逆転させる程、利己主義者の無差別曲線を北西方向へシフトさせることである。さらに、もう
 ひとつは、座標 (\bar{N}_a, \bar{N}_e) の左側に利己主義者の無差別曲線をシフトさせるほど価格規模 $p > p^*$
 が大きいことである。ただし、このとき、価格 p^* は、公式的には、 $u_a = \theta_a a + (1 - \theta_a)p + s(\bar{N}_a,$
 $\bar{N}_e) - c = 0$ を満たす p の値である。すなわち、 $p^* = [c - \theta_a a - s(\bar{N}_a, \bar{N}_e)] / (1 - \theta_a)$ である。(当然、
 小規模と大規模の中間の価格規模 $a < p < p^*$ が存在する。私たちは、これを中規模と定義する)。
 ただし、このとき、価格規模は、あくまで、内発的動機付け a を基準とした「相対的」規模を表
 している。したがって、 a が絶対的に大きな値であるとき、 $a > p > 0$ であっても、 p の絶対的規
 模は大きいかもしれない。しかしながら、このときでも、 $a > p > 0$ ならば、金銭的報酬は (相
 対的に) 小規模と規定していることに注意しなさい。

私たちは、次に、結果 3 で、中規模の金銭的報酬の導入が、(結果 2 とは逆に)、中期的には、
 貢献者を減少させるかもしれないことを考察する。結果 3 では、金銭的報酬が小規模、中規模お
 よび大規模の場合における貢献パラメータ値を比較することで、このことが生じる (中規模の金
 銭的報酬の下で、貢献の社会規範が消滅する) 可能性があることを考察する。

結果 3. $a + s(\bar{N}_a, \bar{N}_e) < c < \theta_a a + s(\bar{N}_a, 0)$ であると仮定する。次のような（小規模、中規模および大規模の）金銭的報酬 $p > 0$ が導入されると、新たな中期均衡は、それぞれ、次のようである。

- (i) $N_a = \bar{N}_a$ if $p < a$ であり、しかも、
 (a) $N_e = N_e^{**}$ if $(a >)p > c - s(\bar{N}_a, 0)$ である。ただし、 N_e^{**} は、 $c - s(\bar{N}_a, 0) < p < c - s(\bar{N}_a, \bar{N}_e)$ について、 $p + s(\bar{N}_a, N_e^{**}) - c = 0$ を満たす N_e の値である。あるいは、
 (b) $N_e = 0$ if $p < c - s(\bar{N}_a, 0)$ である。
 (ii) $N_a = 0$ if $a < p < [c - \theta_a a - s(\bar{N}_a, \bar{N}_e)] / (1 - \theta_a)$ であり、しかも
 (a) $N_e = 0$ if $p < c$ である。あるいは、
 (b) $N_e = \bar{N}_e$ if $p > c$ である。
 (iii) $N_a = \bar{N}_a$ and $N_e = \bar{N}_e$ if $p > [c - \theta_a a - s(\bar{N}_a, \bar{N}_e)] / (1 - \theta_a) > a$ である。

証明. (i) – (iii) に示した三つのケース、小規模、中規模、および、大規模の金銭的報酬について、それぞれ、次のような貢献パラメータ値 N_a^t および N_e^t が成り立つことを明らかにする¹⁸。



結果 3 : パラメータ値の範囲

- (i) $p < a$ であると仮定する。このとき、 $p < \theta_a a + (1 - \theta_a)p$ であるので、いかなる N_a^t および N_e^t についても、 $p + s(N_a^t, N_e^t) - c = u_e^t < u_a^t = \theta_a a + (1 - \theta_a)p + s(N_a^t, N_e^t) - c$ である。また、(パラメータ値の仮定より、 $a + s(\bar{N}_a, \bar{N}_e) < c$ であるので)、 (\bar{N}_a, \bar{N}_e) のとき、 $u_e < u_a = \theta_a a + (1 - \theta_a)p + s(\bar{N}_a, \bar{N}_e) - c < a + s(\bar{N}_a, \bar{N}_e) - c < 0$ であるため、利他主義者および利己主義者の効用は負であることを意味する。(したがって、 (\bar{N}_a, \bar{N}_e) は均衡とはなり得ない)。次に、 $(\bar{N}_a, 0)$ のとき、利他主義者および利己主義者の効用はどのようになるのかをみってみる。

¹⁸ 結果 3 の証明に直接に関連するものではないが、 $[c - \theta_a a - s(\bar{N}_a, \bar{N}_e)] / (1 - \theta_a) > c - s(\bar{N}_a, \bar{N}_e)$ の大小関係を確認しておく。ただし、仮定より、 $a + s(\bar{N}_a, \bar{N}_e) < c$ であることに注意しなさい。

$$\begin{aligned} & \frac{c - \theta_a a - s(\bar{N}_a, \bar{N}_e)}{1 - \theta_a} - [c - s(\bar{N}_a, \bar{N}_e)] \\ &= c - \theta_a a - s(\bar{N}_a, \bar{N}_e) - (1 - \theta_a)[c - s(\bar{N}_a, \bar{N}_e)] \\ &= c - \theta_a a - s(\bar{N}_a, \bar{N}_e) - [c - s(\bar{N}_a, \bar{N}_e)] + \theta_a [c - s(\bar{N}_a, \bar{N}_e)] \\ &= -\theta_a a + \theta_a [c - s(\bar{N}_a, \bar{N}_e)] \\ &= \theta_a [c - a - s(\bar{N}_a, \bar{N}_e)] > 0 \end{aligned}$$

さらに言えば、仮定 $a + s(\bar{N}_a, \bar{N}_e) < c$ より、 $a < c - s(\bar{N}_a, \bar{N}_e)$ であるので、最終的に、 $a < c - s(\bar{N}_a, \bar{N}_e) < [c - \theta_a a - s(\bar{N}_a, \bar{N}_e)] / (1 - \theta_a)$ といえる。(これは、図「結果 3 : パラメータ値の範囲」に示したとおりである)。

ケース (a) では、 $p > c - s(\bar{N}_a, 0)$ を仮定しているので、正の金銭的報酬を導入したとき、(利己主義者の効用は $u_e = p + s(\bar{N}_a, 0) - c > 0$ となるため)、次期以降、利他主義者は貢献することを選択することになる。しかしながら、また、 $u_e = p + s(\bar{N}_a, \bar{N}_e) - c < 0$ であるので、利己主義者の貢献者数が増加し、 $N_e = \bar{N}_e$ に達するとき、利己主義者の効用は負となる。かくして、この中間のある値について、 $u_e = p + s(\bar{N}_a, N_e^{**}) - c = 0$ を満たす N_e の値が存在する。すなわち、均衡において、利己主義者の貢献者数は $N_e = N_e^{**}$ となる。このとき、 $0 < N_e^{**} < N_e^* < \bar{N}_e$ となることに注意しなさい。ただし、 $N_e = N_e^{**}$ は $a + s(\bar{N}_a, N_e^{**}) - c = 0$ を満たす N_e の値である¹⁹。(すなわち、均衡では、 $(N_a, N_e) = (\bar{N}_a, N_e^{**})$ となる)。

他方、ケース (b) では、 $p < c - s(\bar{N}_a, 0)$ を仮定しているので、正の金銭的報酬を導入しても、まだ、利己主義者の効用は負のままである ($u_e = p + s(\bar{N}_a, 0) - c < 0$)。したがって、価格導入以前と同様、利己主義者は全く貢献しないであろう。このとき、利他主義者の効用は、(仮定 $\theta_a a + s(\bar{N}_a, 0) > c$ より)、 $u_a = \theta_a a + (1 - \theta_a)p + s(\bar{N}_a, 0) - c > \theta_a a + s(\bar{N}_a, 0) - c > 0$ であるので、貢献し続けるであろう。すなわち、均衡では、 $(N_a, N_e) = (\bar{N}_a, 0)$ となる。

金銭的報酬の導入が小規模のとき ($a > p > 0$)、利他主義者の貢献者数は $N_a = \bar{N}_a$ が維持されることになる。

- (ii) $0 < a < p < [c - \theta_a a - s(\bar{N}_a, \bar{N}_e)] / (1 - \theta_a)$ であると仮定する。このとき、 $p > \theta_a a + (1 - \theta_a)p$ であるので、いかなる N_a^t および N_e^t についても、 $p + s(N_a^t, N_e^t) - c = u_e^t > u_a^t = \theta_a a + (1 - \theta_a)p + s(N_a^t, N_e^t) - c$ が成り立つ。ここで、仮定 ($N_a = 0$ かつ $u_a < 0$) とは逆に、いかなる $\bar{N}_a \geq N_a^t > 0$ についても、 $u_a \geq 0$ が成り立つとする。(すなわち、任意の $N_a^t > 0$ について、社会的報酬の最小値 $s(N_a^t, \bar{N}_e)$ のときにも $u_a \geq 0$ が成り立つと仮定する)。これは、結果として、 $u_e > u_a \geq 0$ 、かつ、 $N_a = \bar{N}_a$ を意味する。すなわち、 $u_e > u_a = \theta_a a + (1 - \theta_a)p + s(\bar{N}_a, \bar{N}_e) - c \geq 0$ である。しかし、仮定 $p < [c - \theta_a a - s(\bar{N}_a, \bar{N}_e)] / (1 - \theta_a)$ より、 $u_a = \theta_a a + (1 - \theta_a)p + s(\bar{N}_a, \bar{N}_e) - c < 0$ である。これは、利他主義者の効用が負であるのに、その貢献者数が $N_a > 0$ とする仮定は矛盾する。したがって、均衡において、 $N_a = 0$ でなければならない。

$N_a = 0$ のとき、いかなる N_e^t についても、 $s(0, N_e^t) = 0$ であるので、 $u_e = p + s(0, N_e^t) - c = p - c$ となる。ケース (a) では、 $p > c$ を仮定しているので、いかなる N_e^t についても、 $u_e = p - c > 0$ であるので、均衡において、 $N_e = \bar{N}_e$ となる。また、ケース (b) では、 $p < c$ を仮定しているので、いかなる N_e^t についても、 $u_e = p - c < 0$ であるので、均衡において、 $N_e = 0$ となる。

金銭的報酬の導入が中規模のとき ($a < p < [c - \theta_a a - s(\bar{N}_a, \bar{N}_e)] / (1 - \theta_a)$ のとき)、利他主義者の貢献者数は $N_a = 0$ まで低下することになる。

- (iii) $p > [c - \theta_a a - s(\bar{N}_a, \bar{N}_e)] / (1 - \theta_a) > a$ であると仮定する。このとき、利他主義者の効

¹⁹ 仮定 $a + s(\bar{N}_a, \bar{N}_e) < c < \theta_a a + s(\bar{N}_a, 0) < a + s(\bar{N}_a, 0)$ より、 $c - s(\bar{N}_a, 0) < a < c - s(\bar{N}_a, \bar{N}_e)$ であるので、 $a = c - s(\bar{N}_a, N_e^{**})$ を満たす N_e の値が存在する。また、 $p + s(\bar{N}_a, \bar{N}_e) < c < p + s(\bar{N}_a, 0)$ であるので、 $c - s(\bar{N}_a, 0) < p < c - s(\bar{N}_a, \bar{N}_e)$ である。かくして、 $p = c - s(\bar{N}_a, N_e^{**})$ を満たす N_e の値が存在する。このとき、 $p < a$ であるので、 $N_e^{**} < N_e^*$ でなければならない。(図「結果3：パラメータ値の範囲」を参照しなさい)。

用 $u_a = \theta_a a + (1 - \theta_a)p + s(\bar{N}_a, \bar{N}_e) - c$ は、 $N_a = \bar{N}_a$ について、利己主義者の貢献者数が $N_e = \bar{N}_e$ のときでさえ（社会的報酬が最小値 $s(\bar{N}_a, \bar{N}_e)$ のときでさえ）、 $u_a > 0$ である。したがって、 $N_a = \bar{N}_a$ のとき、すべての $0 \leq N_e \leq \bar{N}_e$ について、利他主義者の効用は正である。また、仮定 $p > a$ より、 $(p > \theta_a a + (1 - \theta_a)p)$ であるので、いかなる (N_a^t, N_e^t) についても、 $u_e > u_a$ である。結果として、 $N_a = \bar{N}_a$ のとき、すべての N_e について、 $u_e > u_a > 0$ であることが明らかである。このことは、均衡において、 $N_e = \bar{N}_e$ となることを意味する。

金銭的報酬の導入が大規模のとき ($p > [c - \theta_a a - s(\bar{N}_a, \bar{N})]/(1 - \theta_a) > a > 0$ のとき)、利他主義者の貢献者数は $N_a = \bar{N}_a$ が維持されることになる。 □

結果3に示された状況は、小規模および大規模の価格（金銭的報酬）の導入は、利他主義者および利己主義者の貢献意欲を強化させるが、しかし、同時に、中規模の価格の導入は、利他主義者の貢献意欲を喪失させるかもしれないということである。したがって、結果3は、また、価格の導入が総貢献者数に及ぼす効果は非単調であることを示すものでもある。すなわち、価格がスペクトルの中間に設定されるとき、貢献者数の大幅な低下、いわゆる、締め出し効果（crowding out effect）が生じるかもしれないということである。このとき、利他主義者の貢献者が完全に排除されるならば、社会的報酬はゼロまで低下し、利己主義者に残された唯一の貢献インセンティブは金銭的報酬だけとなる。しかも、金銭的報酬が貢献費用を越えないならば ($p < c$)、利己主義者も、また、貢献することを中断することになる。結果として、総貢献者数はゼロにまで低下することになろう。これは、考えられる最悪の結果である（図6(b)を参照しなさい）。他方、かろうじて、金銭的報酬が貢献費用を越えるならば ($p > c$)、貢献者は、確実に、利己主義者のみとなる（図6(c)を参照しなさい）。このとき、結局のところ、総貢献者数の増減は、人口に占める利己主義者の総数に依存することになる。

私たちは、ここで、結果3に示した中期の動学（その一部）を、社会システムの動学経路として図示することで、社会動学のメカニズムをより明らかにする。図6は、結果3の(i)(a)（小規模の金銭的報酬）、(ii)(a)(b)（中規模の金銭的報酬）および(iii)（大規模の金銭的報酬）の3つのケースを取り上げ描写している。図6では、これまでと同様、 E_a^0 は、価格導入前の（すべての利他主義者が貢献する）旧い均衡を表している。また、 $u_a = 0$ および $u_e = 0$ は、それぞれ、利他主義者および利己主義者の価格導入後の新たな無差別曲線、そして、大きな矢印 (\rhd) は新たな無差別曲線の下での動学ルール、さらに、 E は新たな均衡を表している。社会システムの動学経路（旧い均衡 E_a^0 から新たな均衡 E へのパス）は、太い矢印 (\rightarrow) で示されている。私たちは、この太い矢印をたどることにより、（旧い均衡に位置する）社会システムが、異なる価格水準の導入を契機として、異なる動学経路を経て、異なる（新たな均衡を持つ）社会システムにいたる過程を観察することができる。ここでは、金銭的報酬が中規模の場合（図4(b)(c)）に限定し、社会システムの動学経路を、さらに、詳しくみていくことにする。

- (i) 中規模の金銭的報酬を導入したときの中期動学 中規模価格 $a < p < p^*$ が導入される
とき、旧い無差別曲線の位置関係は、新たな無差別曲線の下では逆転することになる
($u_e > u_a$)。また、中規模価格パラメータの範囲を $a < p < p^* = [c - \theta_a a - s(\bar{N}_a, \bar{N}_e)]/$

($1 - \theta_a$)と仮定しているの、利他主義者の新たな無差別曲線は座標 (\bar{N}_a, \bar{N}_e) の下方に位置する ($u_a(\bar{N}_a, \bar{N}_e) < 0$)。さらに、利他主義者の無差別曲線は、初期状態において、古い均衡 E_a^0 の上方にあるので ($u_a^0(\bar{N}_a, N_e^0) > 0$)、価格導入後も古い均衡の上方に位置する ($u_a(\bar{N}_a, N_e^0) > 0$)。これらのことより、初期状態 (古い均衡) は、利他主義者 (および利己主義者) の新たな無差別曲線の右側の領域に位置するため、利他主義者の無差別曲線と直線 $N_a = \bar{N}_a$ との交点 (\bar{N}_a, N_e^{**}) が (不安定) 均衡となる (図6 (b) (c))。ところが、利他主義者の無差別曲線上では、(無差別曲線上の左側では $u_a < 0$ であるので左押し圧力が、そして、右側では $u_a > 0$ であるので右押し圧力が働くため)、内点均衡とはならず、無差別曲線上から逸脱し端点均衡へ向かうことになる。この場合、 $N_a = \bar{N}_a$ であるため、 N_a 次元の左押し圧力が働き、利他主義者の無差別曲線を越え左側の領域に侵入する。利他主義者の無差別曲線の左側の領域では $u_a < 0$ であるので、利他主義者の貢献者数はゼロまで低下していくことになる。他方、利己主義者の貢献者数については、次の二つのことが考えられる。

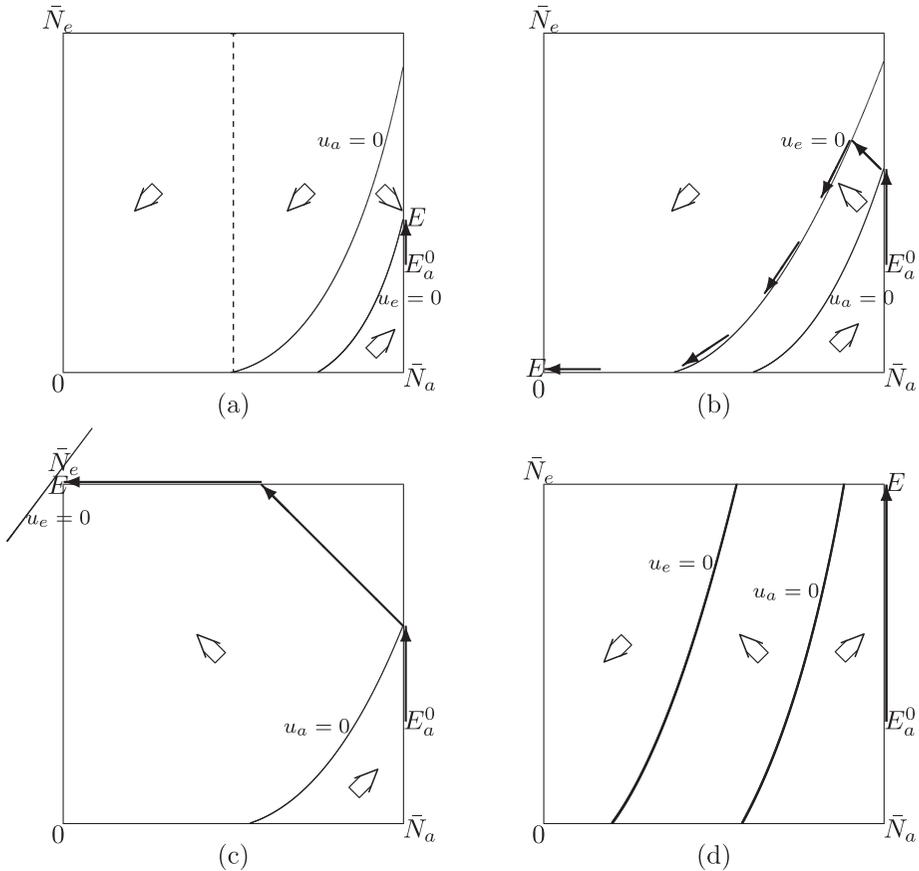


図 (a) : ケース (i) (a) $c - s(\bar{N}_a, 0) < p < a$;
 図 (b) : ケース (ii) (a) $a < p < [c - s(\bar{N}_a, \bar{N}_e) - \theta_a a]/(1 - \theta_a)$ and $p < c$;
 図 (c) : ケース (ii) (b) $a < p < [c - s(\bar{N}_a, \bar{N}_e) - \theta_a a]/(1 - \theta_a)$ and $p > c$;
 図 (d) : ケース (iii) $a < [c - s(\bar{N}_a, \bar{N}_e) - \theta_a a]/(1 - \theta_a) < p$

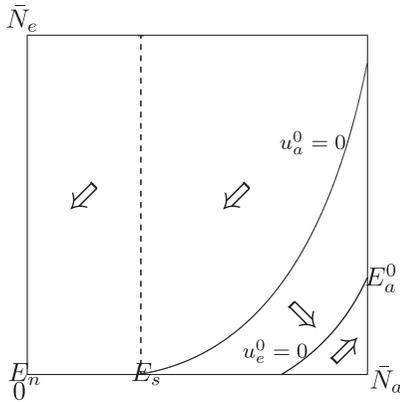
図 6. 中規模の金銭的報酬の導入と中期の動学

- (a) 利他主義者の貢献者数がゼロまで低下すると（社会的報酬もゼロとなるので）、利己主義者に残された唯一の貢献インセンティブは金銭的報酬だけとなる。したがって、金銭的報酬が貢献費用を補填しないならば ($p < c$)、利己主義者も、また、貢献することを中断することになる。結果として、総貢献者数はゼロまで低下することになる（図 6 (b)）。すなわち、中期均衡 E は $(N_a, N_e) = (0, 0)$ となる。
- (b) 他方、かろうじて、金銭的報酬が貢献費用を補填するならば ($p > c$)、貢献者は、確実に、利己主義者のみとなる（図 6 (c)）。すなわち、中期均衡 E は $(N_a, N_e) = (0, \bar{N}_e)$ となる。

この結論は、価格の導入が、利他主義者の貢献者のモチベーションを低下させることがあるという意味で、伝統的経済学の報酬インセンティブの基本的アイデアと矛盾するものとなっている。

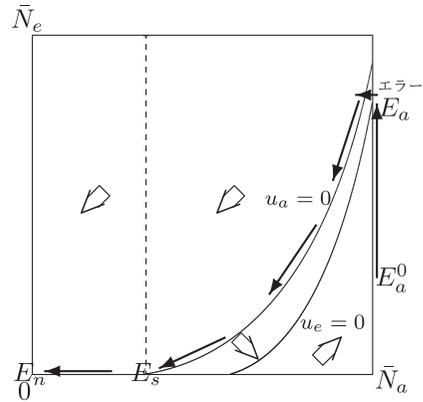
これまで、金銭的報酬を導入したときの中期動学をみてきたが、次に、長期動学、すなわち、長期均衡を考えることにする。中期的には、図 6 (b)–(d) に示したように、中規模および大規模の金銭的報酬 $p > a$ が導入されたとき、利他主義者の誰も貢献しない均衡 ($N_a = 0$)、あるいは、すべての利他主義者が貢献する均衡 ($N_a = \bar{N}_a$) のいずれかが一意的均衡 E として表れる。しかも、また、これらの均衡は、長期均衡においても一意的均衡として表れることになる。これに対し、小規模の金銭的報酬 $p < a$ が導入されたとき、図 7 に示したように、 E_a および E_n の両方が長期均衡として表れることになる。これは、まさに、3 節で示したように、 E_a の確率的安定性は、それぞれの均衡から利他主義者の無差別曲線までの最短距離との比較により（均衡からどれほどのエラー回数で利他主義者の無差別曲線に到達できるかという意味で）、確定されることになる。価格の導入前と導入後の状況を比較したとき、貢献者が存在する新しい均衡は、確率的安定性の性質を失うかあるいは以前のまま維持されるかもしれない。（古い均衡において）、 $0 < N_e < \bar{N}_e$ を所与とすると、金銭的報酬の導入は、いかなる (N_a^t, N_e^t) についても、貢献する個人の効用を増加させ、（とりわけ）利己主義者の無差別曲線を（大幅に）左方にシフトさせる。このことは、利他主義者の貢献者が存在する均衡の確率的安定性に（どちらかと言えば）負の効果を持つ。しかも、この価格は、また、 N_e を増加させるかもしれない。これは、社会的報酬を低下させ、そして、利他主義者の効用を低下させる。したがって、 E_a の確率的安定性に負の効果を持つ。価格の長期的な正味効果は、パラメータ値、および、報酬関数の形状に依存することになる。

例えば、小規模の金銭的報酬 $p < a$ を導入したときの均衡の確率的安定性の一例を、図 7 に示しておくことにする。この例では、 $\theta_a = 1$ 、しかも、貢献者が存在する（古い）均衡 (\bar{N}_a, N_e^t) 、 $0 < N_e^t < \bar{N}_e$ を初期状態と仮定することで、価格導入後に新たな均衡が不安定となることを示している。このとき、金銭的報酬 $p > 0$ は、利他主義者の効用にいかなる直接的効果を持つものではない（ $(1 - \theta_a)p = 0$ ）。他方、価格導入は、利己主義者の貢献者数の増加を招くため、利他主義者に対し負の間接的效果を持つ。まず、価格導入前の状況は、図 7 (a) に示している。このとき、均衡 E_a^0 は安定均衡である（均衡から利他主義者の無差別曲線までの最短距離が大きい）。これに対し、価格導入後の状況は、図 7 (b) に示している。このとき、 $(\theta_a = 1$ であるので)、利己主義者の無差別曲線のみが左方へシフトするため、新たな均衡 E_a と利他主義者の無差別曲線が近接することになり、エラーに対し不安定となる。すなわち、この均衡では、エージェントが最適行動の選択を誤ることで、内点均衡 E_a を逸脱し、端点均衡 E_n へ発散していく可能性がある。



(a) : 価格導入前

図 (a) : $c - s(\bar{N}_a, 0) < 0 < a$;
 価格導入以前では、均衡 E_n および E_a が長期均衡として表れる可能性がある。これは、均衡から利他主義者の無差別曲線までの距離が離れているため、エージェントが誤って決定論的最適行動に反した行動を選択したとしても、均衡から逸脱する可能性が小さいからである。この場合、 E_a は確率的安定的である。



(b) : 価格導入後

図 (b) : $c - s(\bar{N}_a, 0) < 0 < a$ and $p < a$;
 $\theta = 1$ のとき、利他主義者の無差別曲線は (小規模の) p 値によって影響を受けない。他方、利己主義者の無差別曲線は北西方向へシフトする。このとき、二つの無差別曲線はより近接するので、貢献に関する誤った選択がなされるとき、利他主義者の無差別曲線を越える可能性が生じ、その結果、貢献者が減少していく可能性がある。

図 7. 金銭的報酬導入と確率的安定性

る。これは、 $\theta_a = 1$ を仮定する特殊例であるが、すべての関連距離は確率 θ_a において連続であるので、同じ結論が、1 に近い θ_a についても保証される。したがって、価格導入により、新しい均衡が確率的に不安定となる状況は起こりそうであるといえる。

5. 金銭的報酬の中断

先の節で、金銭的インセンティブを導入することは、貢献レベルに負の影響を及ぼすかもしれないことを示した。権威当局は、このような状況に直面するとき、(金銭的報酬の導入が、貢献の社会規範を脆弱化させることを認識するとき)、以前の状況を回復するため、金銭的報酬の支払いを中断することを決意するかもしれない。本節で、権威当局が、金銭的報酬の中断の施策を実行したとき、それが意図したように機能しない可能性があることを明らかにする。

金銭的報酬の中断は、おおよそ、次のような状況を招くことが知られている。まず、金銭的報酬を中断することは、無差別曲線を元のポジションへシフトさせてしまう。そして、金銭的報酬の導入以前のように、二つの中期均衡 E_a および E_n が存在する状況へ回帰させる。このとき、貢献者が存在する均衡が確率的に安定しているならば、長期的には、それらの均衡が再現することになる。しかしながら、中期的には、社会システムは、利他主義者が貢献する均衡へ回帰することではなく、その代わりに、いかなる貢献者も存在しない均衡へ収束していく可能性がある。例えば、金銭的報酬を中断した時点で、社会システムが、誰も貢献しない均衡の吸引力の盆地に位置しているならば、誰も貢献しない均衡 E_n が生じることになる。

私たちは、本節で、権威当局による金銭的報酬の中断が、誰も貢献しない均衡 E_n を生じさせるメカニズムをさらに詳しくみていく。価格 $p > 0$ が支払われているのに、利他主義者が全く貢献しない状況があるとき、金銭的報酬を中断するならば、誰も貢献しない均衡 E_n が表れることは容易にわかる。誰も貢献しない状況にあるとき、（したがって、貢献から得られる社会的報酬が締め出されているとき）、その上、さらに、金銭的報酬が中断されるならば、貢献インセンティブが失われてしまうからである。このような状況で、権威当局が、なによりも、まず、金銭的報酬を中断する施策を実行するならば、状況を改善するどころか、さらに悪化させる可能性さえある。結果 4 はこのことを公式的に示したものである。本節で触れる結果 4 とその証明に関するすべての議論では、金銭的報酬 $p_0 > 0$ が導入されている状況で、均衡における利他主義者の貢献者数および利己主義者の貢献者数を、それぞれ、 N_a^0 および N_e^0 で表し、そして、また、この均衡における彼らの効用を、それぞれ、 u_a^0 および u_e^0 で表すとす。これは、結果 4 において仮定される初期状態を示すものである。

結果 4. $\theta_a a < c < \theta_a a + s(\bar{N}_a, 0)$ であると仮定する。しかも、（金銭的報酬が導入されているときの）初期状態において、 $p_0 > 0$ および $N_a^0 = 0$ であるとする。このとき、金銭的報酬が中断されるならば、新たな中期均衡は $N_a = 0$ および $N_e = 0$ となる。

証明.（初期状態において）、 $N_a^0 = 0$ ならば（このとき、利他主義者が貢献しないのは、 $u_a^0 < 0$ であるからである。したがって）、次のことがいえる。

$$\begin{aligned} u_a^0 &= \theta_a a + (1 - \theta_a)p_0 + s(0, N_e^t) - c \\ &= \theta_a a + (1 - \theta_a)p_0 - c < 0 \end{aligned}$$

したがって、すべての $p < p_0$ およびすべての N_e^t について、 $u_a^t < 0$ である。このとき、金銭的報酬の中断がなされるならば（ $p = 0$ のとき）、新しい均衡では、利他主義者について、 $u_a < 0$ および $N_a = 0$ となるのは明らかである。他方、利己主義者について、 $p = 0$ および $N_a^0 = 0$ のとき、次のことがいえる。

$$\begin{aligned} u_e^t &= p + s(0, N_e^t) - c \\ &= -c < 0 \end{aligned}$$

したがって、 $p = 0$ およびすべての N_e^t について、 $u_e^t < 0$ であるので、均衡において、 $(N_a, N_e) = (0, 0)$ が生じることになる。□

先のケースとは異なり、社会規範が完全に消滅していないならば（ $0 < N_a^t < \bar{N}_a$ のとき）、金銭的報酬の中断効果はいままでのものとなる。導入されている価格が小規模であり、したがって、利己主義者が（貢献に）さほど惹き付けられていないならば、あるいは、多数の利己主義者の貢献者が存在するにも関わらず、社会的報酬が十分に高水準であるならば、利他主義者が貢献するのに十分なインセンティブを準備することになる。このとき、金銭的報酬を中断したとしても、利他主義者のモチベーションは低下することなく、彼らの貢献行動になんら影響を及ぼすこともない。しかしながら、そうでない場合、利他主義者の貢献モチベーションを維持するには、金銭的インセンティブは不可欠なものとなる。

これらの結果は、公式的には、結果5のように表せる。まず、表記の説明から始めることにする。段階的な価格調整が可能なとき、長期的結果は価格調整過程に依存するので、私たちは、まず、ゼロに収束する価格数列、 $p_0 > p_1 > \dots > p_k > \dots > p_K = 0$ を定義する。ただし、 p_0 は初期状態での価格である。価格が、直ちに中断されるならば、 $p_K = 0$ 、ただし、 $K = 1$ 、と表されることになる。また、価格が段階的に低下させられるとき、 $p_k = 0$ 、ただし、 $1 < k \leq K$ 、と表されることになる。分析を容易にするため、それぞれの間接価格 p_k 、ただし、 $k = 1, \dots, K$ 、は前回の価格変更後の調整過程が完了し、中間均衡に達した後にのみ導入されるものとする。そして、これらの中間均衡を、 $(N_a^1, N_e^1), (N_a^2, N_e^2), \dots, (N_a^K, N_e^K)$ で表すとする。

これらの表記を用い、結果4を再述すると次のように表される。結果4は、価格が一気にゼロまで下げられた場合であるが、価格が段階的に下げられるときにも保証される。初期状態が $p_0 > 0$ および $N_a^0 = 0$ のとき、すなわち、利他主義者の誰も貢献しないとき、必然的に、 $u_a^0 = \theta_a a + (1 - \theta_a)p_0 + s(0, N_e^0) - c < 0$ となるので、すべての $p_k < p_0$ およびすべての N_e^t について、 $u_a^t < 0$ が成り立つことになる。したがって、(価格が)ゼロに収束するすべての価格数列について(価格が一気にゼロまで低下する場合も含む)、新たな均衡において、(利他主義者について)、 $u_a < 0$ および $N_a = 0$ となることは明らかである。他方、利己主義者について、初期状態 $p_0 > 0$ および $N_a^0 = 0$ において、 $u_e^0 = p_0 + s(0, N_e^0) - c = p_0 - c$ であるので、価格数列 p_k がゼロに向かって収束していくとき、新たな均衡でも、 $u_e < 0$ および $N_e = 0$ となる。すなわち、(初期状態が $p_0 > 0$ および $N_a^0 = 0$ のとき)、価格がゼロに収束するすべての(価格)数列について、長期均衡 $(N_a, N_e) = (0, 0)$ が生じることになる。

これに対し、次に示す結果5は、金銭的報酬が段階的に低下させられるとき、(価格導入以前の)初期状態が、パラメータ値 $\theta_a a < c < \theta_a a + s(\bar{N}_a, 0)$ の下で、均衡 $N_a^0 = \bar{N}_a$ を持つとき、すべての利他主義者が貢献する中期均衡(このとき、2つの均衡パスが考えられる)が生じる可能性があることを説明している。この結果は、あるパラメータ値について、金銭的報酬を段階的に減少させることで、価格導入以前の均衡が回復することを示唆している。

結果5. $\theta_a a < c < \theta_a a + s(\bar{N}_a, 0)$ 、 $p_0 > 0$ 、および、 $N_a^0 = \bar{N}_a$ であると仮定する。 $p_0 < a$ あるいは $c < a + s(\bar{N}_a, \bar{N}_e)$ であるならば、また、そのときにだけ、最終的な中間均衡 $N_a^K = \bar{N}_a$ を確率1で生じさせる価格数列 $p_0 > p_1 > \dots > p_k > \dots > p_K = 0$ が存在する。

証明. 結果4の証明から明らかなように、中間均衡において、 $N_a^k = 0$ ならば、それは、また、(金銭的報酬を漸次的に減少させたとき)、それに続くすべての中間均衡においても同等である。したがって、最終的な中間均衡において、 $N_a^K = \bar{N}_a$ であるためには、すべての中間均衡において、 $N_a^k = \bar{N}_a$ が保証されなければならない。もちろん、 $c < s(\bar{N}_a, \bar{N}_e)$ であれば、そのとき、即座に、価格を中断させることができ、そして、このとき、最終的な中間均衡において、 $N_a^K = \bar{N}_a$ 、 $N_e^K = \bar{N}_e$ となることは容易にわかる。したがって、ここでは、 $c > s(\bar{N}_a, \bar{N}_e)$ のケースを考えることにする。

私たちは、まず、 $c > s(\bar{N}_a, \bar{N}_e)$ であると仮定する。もし、 $N_a^k = \bar{N}_a$ が確率1で起こるならば、中間価格のすべての(当期および次期の)組み合わせ p_{k-1} および p_k について、当期の中間均衡 (\bar{N}_a, N_e^{k-1}) は、次期の中間均衡 (\bar{N}_a, N_e^k) の吸引力の盆地に位置しなければならない。ただし、ここで、 $N_e^k \neq N_e^{k-1}$ が生じるような価格変化に限定して考えてもなんの問題もないであろう。

これにより、私たちは、すべての中間均衡 (\bar{N}_a, N_e^{k-1}) について、 $u_a^{k,k-1}(\bar{N}_a, N_e^{k-1} | p_k) > 0 \geq u_e^{k,k-1}(\bar{N}_a, N_e^{k-1} | p_k)$ を満たす価格数列 $\{p_k\}$ が存在することを明らかにしなければならない。すなわち、価格数列 $\{p_k\}$ について、貢献することから得られる利他主義者の効用が正であり、例えば、次のようであり、

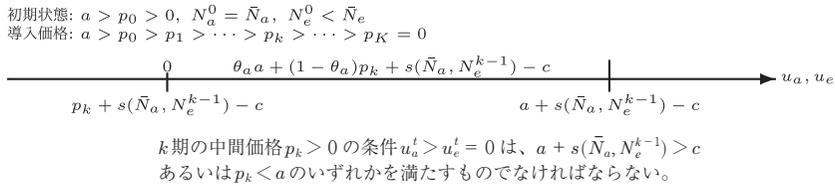
$$u_a^{k,k-1} \equiv \theta_a a + (1 - \theta_a) p_k + s(\bar{N}_a, N_e^{k-1}) - c > 0$$

そして、また、貢献することから得られる利己主義者の効用が、例えば、次に示すように、非正でなければならない。

$$u_e^{k,k-1} \equiv p_k + s(\bar{N}_a, N_e^{k-1}) - c \leq 0$$

これは、権威当局が、それぞれの中間価格の設定段階において、幾人かの利己主義者が貢献することを諦める程に十分に低い、しかし、同時に、また、利他主義者が貢献を維持し続ける程に十分に高い価格を選択しなければならないことを意味している。このとき、問題は、どのような条件の下で、 $u_a^{k,k-1} > 0 \geq u_e^{k,k-1}$ を満たす価格数列が存在するのかである。

このとき、2つのケースが考察されなければならない。第1のケースは、初期状態が $(N_a^0 = \bar{N}_a$ および $N_e^0 < \bar{N}_e$ のような場合である。この初期状態が示唆する要件は、(図7(a)のように)、 $u_a(\bar{N}_a, N_e^0) > u_e(\bar{N}_a, N_e^0) = 0$ である。すなわち、初期状態の価格は、 $p_0 < a$ である ($p_0 < \theta_a a + (1 - \theta_a) p_0 < a$)。したがって、すべての価格 $p_k < p_0 < a$ について、 $u_a^{k,k-1} > u_e^{k,k-1}$ であるといえる。さらに、また、私たちは、すべての p_k について、 $\theta_a a + (1 - \theta_a) p_k + s(\bar{N}_a, N_e^{k-1}) - c = u_a^{k,k-1} > u_e^{k,k-1} = p_k + s(\bar{N}_a, N_e^{k-1}) - c = 0$ となるよう価格を調整することが可能である。また、 $c < \theta_a a + s(\bar{N}_a, 0)$ (および $p_K = 0$) と仮定しているので、結果として、最終的な中間均衡において、 $N_a^K = \bar{N}_a$ および $N_e^K = 0$ が確率1で生じるような価格数列 $\{p_k\}$ が存在することになる。



結果 5 : k 期の中間価格の設定

2番目のケースは、初期状態が $(N_a^0 = \bar{N}_a$ および $N_e^0 = \bar{N}_e$ のような場合である。この初期状態 $(N_a^0 = \bar{N}_a$ および $N_e^0 = \bar{N}_e$) が示唆する要件は、(初期状態において)、すべての個人の効用が正でなければならないことである。すなわち、次のことが満たされなければならない。

$$u_a^0 = \theta_a a + (1 - \theta_a) p_0 + s(\bar{N}_a, \bar{N}_e) - c > 0$$

$$u_e^0 = p_0 + s(\bar{N}_a, \bar{N}_e) - c > 0$$

このとき、 $u_a^0 > u_e^0 > 0$ あるいは $u_e^0 > u_a^0 > 0$ のいずれかの場合が考えられる。このような状態にあるとき、1期目の価格 $p_1 > 0$ は、利他主義者および利己主義者の効用が次のような大小関係を満たすよう選択されなければならない。

$$\theta_a a + (1 - \theta_a) p_1 + s(\bar{N}_a, \bar{N}_e) - c = u_a^{1,0} > u_e^{1,0} = p_1 + s(\bar{N}_a, \bar{N}_e) - c = 0$$

$c > s(\bar{N}_a, \bar{N}_e)$ を所与とするとき、 $u_a^{1,0} > u_e^{1,0} = 0$ であるための価格条件は、 $a + s(\bar{N}_a, \bar{N}_e) > c$ 、あるいは、 $a > p_1$ でなければならない²⁰。そして、その後の価格経路は、先と同様である。すなわち、すべての価格 $p_k < p_1$ は、 $u_a^{k,k-1} > u_e^{k,k-1} = 0$ を満たすように選択され、そして、このとき、中間均衡および最終的な中間均衡は、 $N_a^1 = \dots = N_a^k = \dots = N_a^K = \bar{N}_a$ および $\bar{N}_e > N_e^1 > \dots > N_e^k > \dots > N_e^K = 0$ を満たすことになる。結果として、最終的な中間均衡 $(N_a^K, N_e^K) = (\bar{N}_a, 0)$ が確率 1 で生じるような価格数列が存在するといえる。

□

結果 5 から明らかのように、例えば、 $\theta_a a < p_0 < a$ の場合には、完全にしかも即座に社会的報酬を中断するのではなく、むしろ、段階的に価格を低下させていくことで、最終的に、価格を完全に中断した後でも、社会規範が消滅することはなく、しかも、貢献者数が正であり続ける状況が生じる可能性がある。このことを保証するためには、社会システムが貢献者が存在する均衡の吸引力の盆地に位置するよう、権威当局は価格数列を選択しなければならない。言い換えれば、権威当局は、常に、微妙な価格、すなわち、幾人かの利己主義者が貢献することを諦める程に十分に低い、しかし、同時に、また、利他主義者が貢献し続ける程に十分に高い価格を選択しなければならない。このとき、利己主義者の貢献者数が漸次的に減少し、同時に、社会的報酬が増加していくため、価格は利他主義者の動機付けの源泉として、次第に重要ではなくなり、その結果、さらに価格を低下させても利他主義者のインセンティブを下げることはない。

このように、価格導入により、貢献の社会規範が（完全には締め出されたわけではないが、しかし）、相当に弱体化させられているとき、価格を段階的に引き下げることで、回復させることは可能であるかもしれない。しかし、これは、常に、真であるというわけではない。4 節でみた「締め出し効果」を思い起こしなさい。結果 3 (ii) でみたように、パラメータ値が $c > a + s(\bar{N}_a, \bar{N}_e)$ であるならば、中規模価格 $a < p < p^* (= [c - \theta_a a - s(\bar{N}_a, \bar{N}_e)] / (1 - \theta_a))$ を導入することは、利他主義者の誰も貢献しない均衡 ($N_a = 0$) に落ち着くことを知っている。中規模価格については、これが唯一の均衡であり、価格をゼロからこの水準まで引き上げても、あるいは、高価格レベルからこの水準まで引き下げても、必ず、この結果が生じることになる。

6. 社会規範の押し込み

権威当局による金銭的報酬の導入、あるいは、また、その後の中断により、社会規範が消滅したとする。私たちが、3 節でみたように、貢献者が存在する均衡が確率的に安定しているならば、金銭的報酬を中断したとしても、長期的には、社会規範は再生される可能性がある。しかしながら、(利他主義者の貢献者がおのずから戻り)、社会的報酬が自発的に再生するのを待つのではなく、新たな金銭的報酬を呼び水として（導入することで）、社会的報酬を押し込む（再生させる）

²⁰ $a + s(\bar{N}_a, \bar{N}_e) - c > 0 (= p_1 + s(\bar{N}_a, \bar{N}_e) - c)$ は、 $a + s(\bar{N}_a, \bar{N}_e) - c > \theta_a a + (1 - \theta_a)p_1 + s(\bar{N}_a, \bar{N}_e) - c > p_1 + s(\bar{N}_a, \bar{N}_e) - c = 0$ 、あるいは、 $a > p_1$ と同じ意味である。 $p_1 + s(\bar{N}_a, \bar{N}_e) - c = 0$ と $a + s(\bar{N}_a, \bar{N}_e) - c$ の加重平均は $\theta_a a + (1 - \theta_a)p_1 + s(\bar{N}_a, \bar{N}_e) - c$ である。この関係が成立するためには、 $a + s(\bar{N}_a, \bar{N}_e) - c > 0 = p_1 + s(\bar{N}_a, \bar{N}_e) - c$ あるいは $a > p_1$ でなければならない。(図「結果 5 : k 期の中間価格の設定」を参照しなさい。)

ことが可能である。しかし、それには、利他主義者が金銭的インセンティブを嫌悪せず、しかも、それが利他主義者の貢献モチベーションの源泉となるほど（すなわち、社会的報酬を代替するほど）十分に高い水準でなければならない。さらに、価格導入後、ゼロに向けて、順次、価格を低下させていくにあたって、利他主義者の貢献意欲を削ぐことなく、しかも、同時に、幾人かの利己主義者に貢献を諦めさせるような価格範疇が、常に、存在しなければならない。

私たちは、ここで、幾分詳細に、（社会規範の）押し込みについて分析することにする。（導入価格が、社会規範の押し込みを可能にするには）、社会的報酬が消滅しているときでさえ、利他主義者に貢献インセンティブを供給するようなものでなければならない。かくして、社会規範の押し込みの初期段階において、価格 p_1 は、次の要件を満たすよう設定されなければならない。

$$u_a = \theta_a a + (1 - \theta_a)p_1 - c > 0$$

あるいは

$$(4) \quad p_1 > \frac{c - \theta_a a}{1 - \theta_a}$$

このような要件を満たす価格が導入されるとき、次期以降、すべての利他主義者は、決定論的最適行動に従い貢献することを選択するため、結果として、 $N_a = \bar{N}_a$ が達成されることになる。そして、社会が、このこと（ $N_a = \bar{N}_a$ ）を認知したとき、社会的報酬は正となる（ $s(\bar{N}_a, N_e^t) > s(\bar{N}_a, \bar{N}_e) > 0$ 、 $0 < N_e^t < \bar{N}_e$ ）。したがって、権威当局が、（再生した社会的報酬が補填する限度内で）、価格を、幾分、低下させたとしても、利他主義者は、 $N_a = \bar{N}_a$ を維持しながら、貢献し続けるであろう。公式的には、 $u_a(\bar{N}_a, \bar{N}_e) > 0$ を満たす価格水準であれば、利己主義者のいかなる行動選択にもかかわらず、貢献水準は、常に、 $N_a = \bar{N}_a$ で維持されることになる。

$$u_a = \theta_a a + (1 - \theta_a)p + s(\bar{N}_a, \bar{N}_e) - c \geq 0$$

あるいは

$$(5) \quad p \geq \frac{c - \theta_a a - s(\bar{N}_a, \bar{N}_e)}{1 - \theta_a}$$

ただし、 \bar{N}_e は、 $N_a = \bar{N}_a$ について、社会的報酬 $s(\bar{N}_a, N_e^t)$ が最小となる N_e^t の値である。したがって、場合によっては（ $N_e^t < \bar{N}_e$ のとき）、価格 p_1 は、さらに、低下させることが可能である²¹。このとき、私たちが最も関心の高い状況は、金銭的報酬が不要とされるほど、社会的報酬が十分に高い貢献インセンティブを持ち、結果として、 $N_a = \bar{N}_a$ が保証されるような状況である。そして、この状況が達成されるとき、まさに、社会システムにおいて、社会規範が再生されたということができる。

結果6で、社会規範が確実に押し込まれる条件を公式的に考察することにする。私たちは、すでに、（さまざまな価格水準、初期状態、および、パラメータ値の下で）、価格を中断したときの結果については、（5節の結果4および結果5において）、論及しているので、ここでは、押し込みに関連する結果の部分のみ言及することにする。ただし、先と同様に、新たな中間価格は、常

²¹ (5) 式は、価格が、いわゆる、大規模価格の範疇にあることを意味する。私たちは、中規模価格 $[c - \theta_a a - s(\bar{N}_a, \bar{N}_e)] / (1 - \theta_a) = p^* > p > a$ では、利他主義者の誰も貢献しない均衡が必ず生じることを知っている。したがって、もし、導入価格を低下させようとするならば、中規模価格を避け、大規模価格 $p > p^*$ あるいは小規模価格 $p < a$ の範疇に収まるような新たな中間価格を模索しなければならない。

に、旧い中間価格の下で調整過程が完了し、中間均衡に達した後にのみ導入されると仮定する。さらに、また、新たな中間価格が導入された後、(決定論的最適反応により)、ただちに、 $N_a = \bar{N}_a$ の事象が生じることになるので、社会規範の押し込みが正の確率で成功する蓋然性が多少なりとも存在することを想定している。結果6に、社会規範の押し込みが確実に成功するための中期動学の要件を示している。

結果6. $\theta_a a < c < \theta_a a + s(\bar{N}_a, 0)$ であるとする。また、初期状態において、 $p_0 = 0$ 、および、 $N_a^0 = N_e^0 = 0$ であるとする。このとき、 $c < a + s(\bar{N}_a, \bar{N}_e)$ であるならば、また、そのときにだけ、最終的な中間均衡において、 $N_a = \bar{N}_a$ を確率1で生じさせるような価格数列 $p_1 > \dots > p_k > \dots > p_K = 0$ 、ただし、 $p_1 > (c - \theta_a a)/(1 - \theta_a)$ 、が存在する。

証明. 私たちは、結果4 (の証明) から、(価格数列 $\{p_k\}$ について)、中間均衡が $N_a^k = 0$ であるならば、また、その後のすべての中間均衡においても、同等となることを知っている。そして、また、結果5より、(価格数列 $\{p_k\}$ について)、最終的な中間均衡が $N_a^K = \bar{N}_a$ であるならば、すべての中間均衡においても、 $N_a^k = \bar{N}_a$ でなければならないことを知っている。初期状態は、 $p_0 = 0$ 、および、 $N_a^0 = N_e^0 = 0$ であるので (このとき、 $-c = u_e^0 < u_a^0 = a - c < 0$ であるので)、ここで、私たちがすべきことは、すべての中間均衡が、 $N_a = \bar{N}_a$ となるよう、中間価格を適切に選択することである。まず、1期において、 $N_a = \bar{N}_a$ が達成されるためには、((4) 式より)、 $p_1 > (c - \theta_a a)/(1 - \theta_a)$ が満たされなければならない。また、すべての中間価格 p_k について、 $N_a = \bar{N}_a$ が達成されるためには、 $u_a^{k, k-1} > u_e^{k, k-1} = 0$ となるよう価格が設定されなければならない。そして、このような価格が存在するための条件は、(結果5より明らかのように)、 $a + s(\bar{N}_a, \bar{N}_e) > c$ あるいは $p_k < a$ である。したがって、すべての中間価格について (1期を含む)、これらの条件、すなわち、 $p_1 > (c - \theta_a a)/(1 - \theta_a)$ および $a > p_k$ 、あるいは、 $a + s(\bar{N}_a, \bar{N}_e) > a > c$ が満たされるならば²²、すべての中間均衡において、 $N_a = \bar{N}_a$ を確率1で生じさせる価格数列 $p_1 > \dots > p_k > \dots > p_K = 0$ 、ただし、 $p_1 > (c - \theta_a a)/(1 - \theta_a)$ 、が存在するといえる。□

私たちは、次に、社会規範の押し込みについての議論の一部を図示することで、押し込みの動学経路がどのような軌跡をたどるのかを確かめておくことにする。

- (i) 図8 (i): 社会規範の押し込み (過程) の初期状態は、社会システムが貢献の社会規範を喪失した状態、いわゆる、貢献者が全くいない状態 E_a^0 である。このとき、
- (ii) 図8 (ii): 権威当局は、新たな (中間) 価格 ($a > p_1 > (c - \theta_a a)/(1 - \theta_a)$) を導入

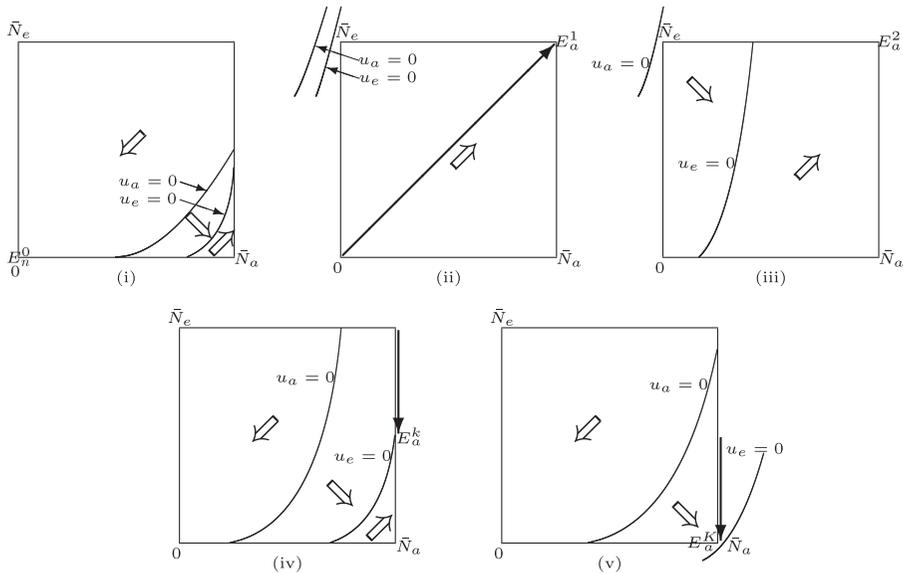
²² 1期の中間価格について、 $u_a(0, 0 | p_1) > 0$ が満たされる条件は $p_1 > (c - \theta_a a)/(1 - \theta_a)$ である。また、すべての中間価格について (1期も含む)、 $u_a^{k, k-1} > u_e^{k, k-1} = 0$ が満たされる条件は、 $a > p_k$ である。したがって、1期の中間価格の条件は次のようであればならない。

$$\begin{aligned} a > p_1 &> \frac{c - \theta_a a}{1 - \theta_a} \\ a(1 - \theta_a) &> c - \theta_a a \\ a &> c \end{aligned}$$

したがって、すべての中間価格 (1期を含む) について、条件 $p_1 > (c - \theta_a a)/(1 - \theta_a)$ および $a > p_k$ 、あるいは、 $a + s(\bar{N}_a, \bar{N}_e) > a > c$ が満たされるならば、最終的な中間均衡においても、 $N_a = \bar{N}_a$ が達成されることになる。

することで、社会システムに、一気に、社会規範 $N_a = \bar{N}_a$ を回復させようとする。導入される価格は、両タイプの無差別曲線を枠外にまでシフトさせるほどの高水準であるかもしれない。その結果、両タイプのすべての個人が貢献する均衡 E_a^1 が生じることになる。この後、

- (iii) 図8 (iii)： 権威当局は、利他主義者の貢献者数を、 $N_a = \bar{N}_a$ に維持したまま、しかも、利己主義者の貢献者数を $N_e = \bar{N}_e$ から漸次減少させる価格を模索する。権威当局は、利他主義者の無差別曲線は枠外にとどまり、しかも、利己主義者の無差別曲線はわずかに枠内にシフトするまで、中間価格を低下させる。結果として、すべての個人が貢献する状態 E_a^2 が維持されたままである。さらに、
- (iv) 図8 (iv)： 権威当局は、 $N_a = \bar{N}_a$ は維持したまま、 $N_e = \bar{N}_e$ を減少させる価格水準を模索する。そこで、権威当局は、利己主義者の無差別曲線のみが直線 $N_a = \bar{N}_a$ と交差するまで、中間価格を低下させるであろう。このとき、新たな中間均衡は E_a^k となる。この後も、また、
- (v) 図8 (v)： 権威当局は、 $N_a = \bar{N}_a$ は維持したまま、 $N_e = 0$ にまで減少させる適切な価格を模索する。そこで、利他主義者の新たな無差別曲線は古い均衡 E_a^k の左方にとどまる程度に、しかも、利己主義者の無差別曲線は枠外にシフトするまで、中間価格を低下させる。これにより、最終的な中間均衡 $(\bar{N}_a, 0)$ に到達することができる。



- 図 (i) : $u_a^0(0, 0) < 0$;
- 図 (ii) : $u_e^1(0, \bar{N}_e) > 0$;
- 図 (iii) : $u_a^2(0, \bar{N}_e) > 0, u_e^2(\bar{N}_a, \bar{N}_e) > 0 > u_e^2(0, \bar{N}_e)$;
- 図 (iv) : $u_a^k(\bar{N}_a, \bar{N}_e) > 0 > u_e^k(\bar{N}_a, N_e^{k-1})$;
- 図 (v) : $u_a^k(\bar{N}_a, N_e^{k-1}) > 0 > u_e^k(\bar{N}_a, 0)$

図 8. 社会的報酬の押し込み

金銭的報酬を導入したとき、パラメータ値が社会的報酬の締め出しを招かない条件を満たす場合のみ、社会的報酬（すなわち、社会規範）を確実に押し込むことが可能であることに注意しなさい（結果2を参照しなさい）。他方、金銭的報酬の導入が、社会的報酬の締め出しにつながるパラメータ値を持つ場合、社会的報酬の押し込みが成功するかどうかは不確実である。最終的に、このプロセスが、いかなる結果につながるかは不確実であるという意味で、金銭的インセンティブを活用して社会的報酬を再生させる可能性は限定的なものであるといえるかもしれない。

7. ディスカッションと結論

簡単な進化モデルの助けを借りて、私たちは、実証研究の結果がどのような動学メカニズムを持つのかを明らかにしようとした。例えば、多くの実証研究においてみられる結果、金銭的報酬が導入されるとき、貢献の社会規範が消滅、あるいは、脆弱化してしまうこと、そして、また、金銭的報酬が中断されるとき、貢献の社会規範は回復するどころか、（かろうじて、維持されている社会規範までも）消滅させてしまうこと、さらに、また、社会規範が消滅してしまうならば、それが再生されるまで、長い期間を要するが、しかし、また、（金銭的報酬を呼び水として用いなくても）、自発的な貢献システムが本来持つ長い歴史の結果として、社会規範が再生する可能性があること等々をモデルにより再現しようとした。

進化モデルは、実証研究のこれらの結論に対し、動学メカニズムにより一定の説明を与えている。例えば、不十分な金銭的報酬が社会規範の脆弱化を招くことについて、中規模の価格では、（利己主義者の貢献者数の増加にともなう）社会的報酬の低下分を補填できないこと、そして、また、金銭的報酬の中断が招く貢献状況の悪化についても、社会的報酬が低下している状況で、金銭的報酬を中断することは、十分な貢献インセンティブを確保できなくなる可能性があること、さらに、また、消滅した社会規範の自発的再生にしても、貢献者が存在する均衡が確率的に安定しているならば、長期的には、社会規範は再生される可能性があることを、動学メカニズムにより説明することが可能であるとしている。

私たちは、また、進化モデルによる分析から得られた知見に基づく幾つかの施策を提言している。例えば、社会規範の脆弱化がみられるとき、即座に、金銭的報酬を中断させることは、短期的あるいは中期的には、しばしば、状況をさらに悪化させるため、社会規範の自発的再生を待つことが、実現可能な選択肢でないのであれば、報酬を維持する施策は意味があるとしている。さらに、貢献の社会規範が消滅している状況において（それは、おそらく、社会的報酬が締め出されたためであるが）、社会的報酬の呼び水として価格の支援を得ることで、十分に強い社会的報酬を創出できる可能性があることを提言している。ただし、社会的規範の押し込みが成功するには、非常に高い初期価格が必要なため、場合によっては、これは利用できない可能性がある。

最後に、モデルの仮定について、幾つかの指摘をしておきたい。まず、利己主義者も（利他主義者と同様に）社会的報酬について関心を払うという仮定は、締め出しあるいは押し込み効果についての議論に関し、とりわけ、重要というわけではない。ただし、この仮定は、（利己主義者が、社会的報酬を得るためにだけ貢献することを許容することで）、分析結果をより豊かなものにしていく。これに対し、締め出しおよび押し込みの両効果は、主として、利他主義者の貢献意欲の消滅あるいは強化に依存して決まるという意味で、利他主義者の貢献モチベーションはモデ

ルにおいて重要な役割を果たすものである。さらに、利他主義者について言えば、自己実現の欲求充足 $a > 0$ は、利他主義者の特質を表すのに、それ程重要であるわけではない。すなわち、利他主義者が実際に「利他主義的」であることは、必ずしも、彼らの本質的な特質を表すものではないが、それより、むしろ、これにより、利他主義者が（利己主義者よりも）、金銭的報酬から生じる効用を節約できることが、モデルにおいて、決定的な役割を果たしている。

参考文献

- [1] Andreoni, J., and Petrie, R., “Social Motives to Giving: Can These Explain Fund Raising Institutions?” mimeo, 2000.
- [2] Arrow, K. J., “Gifts and Exchanges,” *Philosophy and Public Affairs*, Vol. 1, No. 4, 1972, pp. 343-362.
- [3] Bar-Gill, O., and Fershtman, C., “The Limit of Public Policy: Endogenous Preferences,” *SSRN Electronic Journal*, 2001.
- [4] Bester, H., and Güth, W., “Is Altruism Evolutionarily Stable?” *Journal of Economic Behavior & Organization*, Vol. 34, Issue 2, 1998, pp. 193-209.
- [5] Deci, E. L., Koestner, Richard, and Ryan, Richard M., “Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation,” *Psychological Bulletin*, Vol. 125, No. 6, 1999, pp. 627-668.
- [6] Dickinson, A., “The Detrimental Effects of Extrinsic Reinforcement on ‘Intrinsic Motivation’,” *Behavior Analyst*, Vol. 12, No. 1, 1989, pp. 1-15.
- [7] Ellison, G., “Basins of Attraction, Long-Run Stochastic Stability, and the Speed of Step-by-Step Evolution,” *Review of Economic Studies*, Vol. 67, Issue 1, 2000, pp. 17-45.
- [8] Fehr, E., and Gächter, S., “Collective Action as a Social Exchange,” *Journal of Economic Behavior & Organization*, Vol. 39, Issue 4, 1999, pp. 341-369.
- [9] Fehr, E., and Gächter, S., “Do Incentive Contracts Crowd Out Voluntary Cooperation?” Working Paper No. 34, Institute for Empirical Research in Economics, University of Zurich, 2000.
- [10] Frank, R. H., “If *Homo Economicus* Could Choose His Own Utility Function, Would He Want One with a Conscience?” *American Economic Review*, Vol. 77, No.4, 1987, pp. 593-604.
- [11] Frey, B., *Not Just for the Money: An Economic Theory of Personal Motivation*, Cheltenham, UK: Edward Elgar Publishing, 1997.
- [12] Frey, B., and Gotte, L., “Does Pay Motivate Volunteers?” Working Paper No. 7, Insitute for Empirical Research in Economics, University of Zurich, 1999.
- [13] Frey, B., and Jegen, R., “Motivation Crowding Theory: A Survey of Empirical Evidence,” *Journal of Economic Surveys*, Vol. 15, No.5, 2001, pp. 589-611.
- [14] Glazer, A., and Konrad, K., “A Signaling Explanation for Charity,” *American Economic Review*, Vol. 86, No. 4, 1996, pp. 1019-1028.
- [15] Gneezy, U., and Rustichini, A., “A Fine Is a Price,” *Journal of Legal Studies*, Vol. 29, No. 1, 2000a, pp. 1-17.

- [16] Gneezy, U., and Rustichini, A., "Pay Enough or Don't Pay at All," *Quarterly Journal of Economics*, Vol. 115, Issue 3, 2000b, pp. 791-810.
- [17] Güth, W., and Kliemt, H., "Evolutionarily Stable Co-operative Commitments," *Theory and Decision*, Vol. 49, 2000, pp. 197-221.
- [18] Harbaugh, W.T., "What Do Contributions Buy?: A Model of Philanthropy Based on Prestige and Warm Glow," *Journal of Public Economics*, Vol. 67, Issue 2, 1988a, pp. 269-284.
- [19] Harbaugh, W.T., "The Prestige Motive for Making Charitable Transfers," *American Economic Review*, Vol. 88, No. 2, 1998b, pp. 277-282.
- [20] Holmström, B., and Milgrom, P., "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics, & Organization*, Vol. 7, Special Issue, 1991, pp. 24-52.
- [21] Lindbeck, A., Nyberg, S., and Weibull, J., "Social Norms and Economic Incentives in the Welfare State," *Quarterly Journal of Economics*, Vol. 114, Issue 1, 1999, pp. 1-35.
- [22] Stewart, H., "Rationality and the Market for Human Blood," *Journal of Economic Behavior & Organization*, Vol. 19, Issue 2, 1992, pp. 125-143.
- [23] Titmuss, R. M., *The Gift Relationship: From Human Blood to Social Policy*, London, UK: George Allen and Unwin, 1970.
- [24] Weibull, J. W., *Evolutionary Game Theory*, Cambridge, London, UK: MIT Press, 1995.
- [25] Young, P., *Individual Strategy and Social Structure*, Princeton, NJ: Princeton University Press, 1998.