社会的ワクチン：メタ規範ゲームにおける裏切りの効果

山本仁志*[1]　岡田勇*[2]
*[1] 立正大学・経営学部・准教授・hitoshi@ris.ac.jp
*[2] 創価大学・経営学部・准教授・okada@soka.ac.jp

1. はじめに

集団における規範維持のモデルとして良く知られたメタ規範ゲーム[1]は、n 人囚人のジレンマの拡張モデルとして、国際問題における協調問題など中央集権的でない集団においていかに規範を維持するかを検討する上で優れたモデルである。たとえば、Heckathorn[2] や Horne and Cutlip[3] は，心理学的な実験を行いメタ規範が存在することを示している。進化論的な分析によっても、規範ゲームでは協調は維持されないが、メタ規範を導入することで協調が維持されることが知られている。しかし近年、メタ規範がシミュレーションの世代数に対して脆弱であるとの指摘がなされている。織田[4]では，初期の懲罰確率によってはメタ規範においても協調が成立しないと述べている．さらに，Galan and Izquierdo[5] は Axelrod [1] をコンピュータシミュレーションと数理解析で精査した結果，メタ規範が協調を安定させるパラメータ空間は限定的であることを指摘した．我々は、様々なシミュレーション条件におけるメタ規範の成立条件を精査し、脆弱性のメカニズムを探る。また、我々は新たに，集団に少数の常に非協調行動をとるエージェントを導入することで頑健に協調が維持されることを発見した．我々はこの効果を社会的ワクチン効果と呼ぶ．
Lotem et.al.[6]は、1 対 1 の間接互恵性ゲームにおいてイメージスコアリングを導入した環境下では、常に裏切る phenotypic defectors を導入することで、同様のアイデアを提案している。本研究では、メタ規範ゲームを用いることで、懲罰と裏切りの効果の比較ができるほか、ｎ対ｎの相互作用のある社会を記述できることにより影響範囲の大きさの議論ができるほか、さまざまな領域に適用範囲が広くなる。

2. モデル

ここで、Axelrod の規範ゲーム・メタ規範ゲームを整理し，社会的ワクチンの導入をおこなう。

規範ゲームは n 人囚人のジレンマゲームの拡張としてとらえることができる．N 人のエージェントで構成される集団を考える．エージェント i は裏切るか協調するかの二つの行為を選択することができる．裏切る確率を Bi（大胆さ）で表現する．i が裏切ると，i は T(=3)の利得を得ることができる．残りの(N-1)人エージェントは H(=-1)の利得を得る．i が協調すれば，すべてのエージェントの利得は 0 である．

ここまでは n 人囚人のジレンマゲームであるが，規範ゲームでは，このあと(N-1)人のエージェントに懲罰のチャンスがある．エージェント j は確率 s で i の裏切りを発見する．発見しなかった場合，なにも起こらず i,j いずれの利得も変化しない．j が i の裏切りを発見した場合，j は自身の持つ復讐度 Vj の確率によって i を罰する．j が i を罰した場合，i は P(=-9)の利得を j は E(=-2)の利得を得る．罰しなかった場合，i,j の利得に変化はない．

ここまでが規範ゲームである．メタ規範とは，エージェント j が i の裏切りを発見し，更に j が i を罰しなかったことをエージェント k が発見したときに k が j を罰するという構造を導入したものであるこのとき，k が j を罰すれば，j は P(=-9)の利得，k は E(=-2)の利得を得る．

社会的ワクチンは、常に(B,V)=(1,0)の戦略をとるエージェントをさす。集団内で少数の社会的ワクチンエージェントが存在することが、規範の維持にどのような効果をもたらすかを検討することが本稿の目的である。また、社会的ワクチンエージェントの戦略を(B,V)=(0,1)とすることで、純然たる規範維持のための監視者が存在することの効果を観察できるなど、メタ規範に社会的ワクチンを適用することの応用範囲は広い。

3. メタ規範の脆弱性

進化シミュレーションの分析により、メタ規範の脆弱性に関して以下のことが明らかになった。

集団の規模 N を 20 から 100 まで変化させ, 更に世代数も 100 から 100,000 まで変化させた実験を行った. 実験は 50 回の試行を行い，最終世代の大胆さ B の平均値を観察した。復讐度 V は，B と強い負の相関があり B の値を観察することで V の挙動もわかるため，本論文では大胆さ(B)のみを観察する.

規範ゲームにおいて，世代数が増えるとほぼ裏切り支配になることがわかる．これはもともと規範ゲームが懲罰に対するフリーライドを容易にしている構造のため，長期的には裏切りが優位になるためである.

また，集団の規模が大きくなると協調が維持されやすくなっている．これは，集団の規模が大きくなることで，裏切りが発見される回数も増え，裏切ることで得られる利得より裏切りを発見されて集中的に罰せられることで裏切りが不利になるためである．ただしこれは大規模な集団での完全な相互監視を意味しており，現実的には非常に厳しい制約であると考えられる.

メタ規範ゲームにおいては，ほぼ協調が支配的となるが，N=20 において世代数を長くすることで規範が崩壊していることが分かる．集団規模が少し大きくなると協調が安定していることは，先に述べたように完全な相互監視がメタ規範のレベルで徹底しているために，非常に厳しい監視社会となり協調が維持されている.

続いて，集団規模は基本である N=20 に固定し，突然変異率を変化させた実験を行う．その結果、突然変率が 0%と 5%以上のときに協調が成立していることが観察される.

しかし、突然変異率が 5%のときに協調が達成されているように観察されるが，時系列の推移を観察してみると非常にランダムな世界であることが分かる．メタ規範において大胆さ(B)は 0.15 から 0.5 までの間を不規則に推移する．突然変異率が 0%のときは，いったん戦略が安定すると変化する要因がないため最終的な結果は安定的となる．規範ゲームにおいても，初期にいったん協調的となると，早い世代で戦略が一様になるため裏切りの侵入がなく安定的な結果となる．ただし，一様になる直前に交叉で裏切りが発生して裏切りが支配的になることも低い頻度では観察される. その結果大胆さの平均値は 0.2 前後で安定している.

4. 社会的ワクチンの効果

前節までの結果をまとめると以下のように整理できる．Axelrod[1]の結果は、集団規模=20,世代数 100 期では，「規範ゲーム＝３パターンの結果が生じる」「メタ規範＝協調が達成される」を示している。しかし，規範における３パターン（裏切り，中間，協調）はすべて裏切りへの過程にすぎず，世代数を伸ばすことで裏切り支配となってしまう．つまり規範ゲームでは集団規模を増やさない限り裏切りに収束する．メタ規範ゲーム（Agent=20）も，超長期では裏切りになる．突然変異率が 0%や 5%では平均的な裏切率は抑制されたが，0%は進化ゲームとしては不自然であるし，5%の状態はランダム性が大きい.

我々は，頑健に協調を維持するための方策として「社会的ワクチン」の導入を提案する．ワクチンとは一般的に弱毒化した病原体を接種することで抗体をつくり病原体への感染を予防することをいう．社会的ワクチンとは，集団の中にごく少数の常に裏切り行為をとるエージェントが存在することで，集団全体の規範を高く維持することができる効果をいう.

規範ゲームにおいては，裏切りが支配的となっているが，メタ規範ゲームにおいては世代数を変化させても協調が安定的に維持されることが分かった.

メタ規範が崩壊する理由は，協調達成時に復讐度の低いエージェントが侵入してきても，裏切り行為がないため，それを発見できないため復讐度の低いエージェントが広まってしまうためであった．しかし，ワクチンエージェントがいることで，復讐度の低いエージェントは発見されやすくなり，集団全体の復讐度が下がることを防ぐことができる.

続いて，突然変異率と世代を変化させた．突然変異率に対しても頑健に協調が維持されていることが分かった.

更に、社会的ワクチンの戦略(B,V)を、(0,0),(1,0),(0,1),(1,1)にそれぞれ設定し規範の安定性を調べた結果、次にようになった（表１）。

表１：社会的ワクチンの戦略によるシミュレーション結果

戦略(B,V)	意味	結果
(0,0)	常に協調し、懲罰に対してフリーライドする	協調は崩壊する
(1,0)	常に裏切り、懲罰に対してフリーライドする	協調は維持される
(0,1)	常に協調し、懲罰に対して厳格	協調は維持される
(1,1)	常に裏切り、懲罰に対して厳格	協調は維持される

実験の結果、常に協調するが裏切りに対しても寛容なエージェントを導入する以外の形式では協調が安定となった。それぞれの場合の ESS がどこに存在するかを調べてみると、戦略(0,0)の際には、集団全体の挙動はメタ規範ゲームとほぼ同様となり、協調は崩壊する。(1,0)および(1,1)の場合は、動学的にも協調の安定が唯一の安定点となり、協調は頑健に安定する。一方、(0,1)の場合には、動学的には、協調戦略および非協調戦略の２点に安定点があり、不安定となるが非協調への推移が極めて稀な条件でしか起こらないため、結果として協調は維持される。

5. まとめ

メタ規範ゲームは安定的な協調の維持にメタ規範が有効であるという有益な知見が発表されて以来，多くの研究がその安定性を前提とした研究が多くなされている．しかし一方で，メタ規範が協調を安定させるパラメータ空間は限定的であることが指摘されている．我々は，メタ規範が協調を安定させる条件を探るためにシミュレーション実験をおこなった．その結果多くのパラメータ環境において，協調が崩壊することを示した．また，我々は従来協調が崩壊するといわれているパラメータ空間においても協調が頑健に維持されるための方策として「社会的ワクチン」の導入を提案した．社会的ワクチンを導入することでメタ規範における超長期および様々な突然変率における安定達成を可能とした．

文　　献

[1] Axelrod, R.M., An Evolutionary Approach to Norms, American Political Science Review, 80 (4), 1095-1111, 1986.

[2] Heckathorn, D.D., Collective Sanctions and Compliance Norms: A Formal Theory of Group-Mediated Social Control, American Sociological Review, 55(3), 366-384, 1990.

[3] Horne, C., and A. Cutlip, Sanctioning Costs and Norm Enforcement: An Experimental Test, Rationality and Society 14(285), DOI: 10.1177/1043463102014003002, 2002

[4] Galan, J.M. and L.R. Izquierdo, Appearances Can Be Deceiving: Lessons Learned Re-Implementing Axelrod's 'Evolutionary Approach to Norms', Journal of Artificial Societies and Social Simulation 8(3), http://jasss.soc.surrey.ac.uk/8/3/2.html, 2005.

[5] 織田輝哉，秩序問題への進化論的アプローチ-メタ規範ゲームの展開-，理論と方法, 5(1), 81-99, 1990.

[6] Lotem,A., Fishman M. A. & Stone,L., "Evolution of cooperation between individuals", Nature 400, 226-227,1999.