

囚人のディレンマ・「誰が猫の頸に鈴をつけるのか」

田 中 宏

- (一) 序論
- (二) 囚人のディレンマの一般的定式化
- (三) ユニオン・ゲームとその政治的帰結
- (四) 「鼠の会議」の定式化とその帰結
- (五) 結論

(一) 序論

本稿の目的はイソップ寓話の中の鼠の会議・「誰が猫の頸に鈴をつけるのか」を囚人のディレンマのひとつのモデルとして定式化し、その政治的帰結を追究することである。囚人のディレンマのモデルとしてはユニオン・ゲームやコモنز・ディレンマのモデルが有名であつて、かつて筆者もその政治的含意を考察したことがあるが、

それとの対比において別の新たなモデルを提示し、その政治的帰結の差異を示したいとおもう。

囚人のディレンマ、別名ソーシャル・ディレンマは俗に「総論賛成・各論反対」といわれるものである。利益を共有する一群の人々をグループ（集団）というが、その共通の利益がグループの各成員のフリー・ライディングの結果、実現せずに終わる状況を囚人のディレンマと呼ぶ。その共通の利益を実現するには各成員が合意の上で協力を互いに強制し合う (mutual coercion mutually agreed upon) という手法がとられる。村の共有地の存続や「万人の万人に対する闘争状態」(ホッブス) はともに囚人のディレンマの一例であり、それから脱却するためにとられるのもかかる手法である。

さて、いかなるグループといえどもこの手法をとるかぎりそれは政治体 (polity) といわれる。そしてこの手法そのものが政治権力の行使にほかならない。この手法のメカニズムの究明こそ政治学のひとつの主要な課題であるが、囚人のディレンマのモデルの相異によって、その解決のための手法のメカニズムにどのような差異が生ずるか。この点を究明することが本稿の課題である。

まず囚人のディレンマの一般的な定式化を説明し、次にその特殊ケースとしてのユニオン・ゲームについて述べ、次いで鼠の会議について論ずる。

(二) 囚人のディレンマの一般的定式化

いまグループの構成員—すべての点で彼等は同質である—の数を n とする。⁽²⁾ 任意の一人の構成員について考えてみる。彼が公益の実現に協力するときの効用を c 、非協力のそれを d とすると、彼は $c \vee d$ なら協力を、 $c \wedge$

α なら非協力を選擇する。もとより c や d の値は彼以外の成員が何人協力するかに依存する。いま彼が自分以外の m 人 ($0 \leq m \leq n-1$) が協力すると予想すれば c は $c(m+1)$ 、 d は $d(m)$ と表示できる。 c や d の括弧内の数字は協力する人数である。さて、(イ)として成員全員が共同歩調をとる場合を想定しよう。すなわち彼が協力するならば他のすべての成員も協力を、また彼が非協力であるならば他のすべての成員も非協力を選擇する。とすれば、彼が協力するときの効用は $c(n)$ 、非協力であるときの効用は $d(0)$ である。前者は公益が成立したときの効用であり、後者はそれが実現されない情況の下での効用である。前者は後者より大であるから、ここに

$$c(n) > d(0)$$

(1)

が成り立つ。これがいわゆる「総論賛成」に対応する。ところで(ロ)として、上の記述とは異なつて各成員が個別にバラバラに意思決定をする場合を考えよう。例えば m 人が協力し、残りが非協力である場合である。このとき彼は非協力を選擇する。すなわち

$$c(m+1) < d(m) \quad 0 \leq m \leq n-1$$

(2)

が成立し、これがいわゆる「各論反対」に対応する。(1)と(2)が同時に成立するとき、これを囚人のディレンマ、あるいはソーシャル・ディレンマという。(1)と(2)は、各成員が集団の一員として追究する目標と個別に追究する目標とが相反すること、つまり集団合理性と個別合理性が乖離することと言い換えてもよい。さて現実には(ロ)が成り立つから(2)が成立し、各人は非協力を選擇する結果、最終的には各人の効用は $d(0)$ になる。かりに $c(m+1) \geq d(m)$ であれば、各人は協力するから、その効用は最終的には $c(n)$ となる。つまり公益を享受できる。

(三) ユニオン・ゲームとその政治的帰結

以下は囚人のディレンマの具体例である。 $c(m+1)$ や $d(m)$ が具体的にどのような特殊な関数であるかという点について考える。まずメシック・オールソン・ドーズのユニオン・ゲーム⁽³⁾について説明する。題材は変えているが、骨子は変らない。

村人の間で自分達の漁の安全のために燈台が必要だということになった。皆口々に賛成をと考えたが、その建設の負担に話が及ぶと口が重くなった。もとより一人だけの力で燈台の建設は不可能である。皆の協力が必要である。協力とは等分の労力を皆で提供し合うことである。この応分の負担をしても燈台は建設する価値があると皆は認めた。しかし、不思議なことに燈台を建設する場所に約束どおりに労力の提供に來た者は一人もいなかった。いまでもその村には燈台がないままである。なぜなのであろうか。

村人が提供するのは労力のみであった。この労力を他の用途に転用したならばなにかの効用が得られたはずである。この効用の大きさを k （正の定数）とする。この労力を燈台の建設に振り向けることはこの効用 k を失うことであって、これが村人各自の負担する費用となる。他方、燈台が建設された場合、それが各人にもたらす効用の大きさを b （正の定数）としよう。村人の数を n とすると、各村人に見てみると自分以外の m 人が協力すると予想したときの彼の効用 $c(m+1)$ は具体的には

$$c(m+1)=[(m+1/n)] \cdot b - k$$

で示される。右辺第一項は b の効用とそれが実現する確率——村人全体のうちの何人が協力するかの比率——の積であって、 b の期待値である。また第二項は労力提供の費用である。右辺はまた

$$m[1/n]b + [1/n]b - k$$

と書くことができる。これは m 人の協力者が b の実現の確率をそれぞれ $(\frac{1}{n})$ ずつ高めてくれること、その合計が $m[1/n]b$ として各人に恩恵となって与えられること、他方彼自身の貢献は b の実現の確率を $(\frac{1}{n})$ だけ高めただけ——それを期待値 $[1/n]b$ で示す——である。それには k だけの犠牲が必要であることを示している。他方、各人が非協力に踏み切ったときの効用 $d(m)$ は、

$$d(m) = [m/n]b$$

で示される。これはまさに m 人の協力者の余澤そのものが自己の効用となっていることを示す。以上を念頭においてみる。

$$c(n) = b - k$$

$$d(0) = 0$$

になる。 $c(n)$ とは皆が協力の結果、燈台が設置されるが、そこからの効用 b とそのために各人が払う犠牲 k との差である。他方、全員が非協力なら燈台は建設されず、そのときの効用はゼロである。さて囚人のディレンマが成立するための必要・十分条件は明らかに、

$$[b/n] > k > b$$

である。各人がバラバラに意思決定をするときの燈台実現への自己の貢献よりも自己犠牲の方が大であるが、その犠牲よりも皆が協力して燈台が建設されたときの効用が大である、というものである。非協力の結果、最終的には各人の効用はゼロとなる。かりに $[b/n] \equiv k$ ならば皆は協力するから、燈台は設置され、各人の効用は最終的には $b - k$ となる。

さて、囚人のディレンマから脱却するにはどうしたらよいか。各人が内心では燈台の建設を希望しながら、しかし非協力を決め込むため燈台は実現されないままとなっている。燈台の設置が急務であるとするならば、どうしたらよいか。この場合には村人が協力し合うよう相互に強制し合うことである。非協力者に対して罰を加えて、その試みが割りに合わぬようにすること、これを予め皆で約束することである。その約束とは以下の如きものである。

村人は互いにその持てる財を取引して生計を立てている。罰とは各村人が非協力者の村人との取引を一定期間はしないということである。その結果、罰をこうむる村人の生計は苦しくなる。その期間が永くなければなるほど罰は重くなる。しかし処罰はそれに加わる村人にとっても負担である。それというのも非協力者との取引停止は取引に伴う効用の断念を意味するからである。したがって村人の中には処罰に加わらない者も出てくる。このことを防止するため処罰に加わらない者も「非協力者」として処罰される。村人には結局のところ燈台の建設の応分の負担とそれをのがれようとする者の処罰に加わること、つまり処罰に伴う負担をすることの二点が求められる。少なくともそのいずれか一方を怠れば「非協力者」とされ処罰される。

さて、この約定をどう定式化するか。まず協力ということを燈台の建設と処罰への参加の両面での負担であるとし、非協力を少なくともこのいずれか一方の負担の回避であるとする。任意の個々の村人を B、その他の(□)一人からなる村人の結託を A とする。村人が相互に強制し合う⁽⁴⁾ことは、A が、B がもし非協力ならば、B に対し処罰あるいはその威嚇をするということである。ここに A も B も n 通りあることに注意したい。A の B に対する処罰は取引停止であるが、それは一ヶ月の期間を単位として行なわれるとする。期間が一ヶ月から二ヶ月、さらに三ヶ月と延長されるにつれ刑罰は重くなる。それに応じて B の効用は低下する。刑罰の重さを期間 x で示

し、 x が1なら一ヶ月の期間、 x が2なら二ヶ月の期間というように表示する。B にとつての刑罰は x に正比例するものとし、比例定数を効用 v (正の定数) とすると、それは効用 $\leq x$ の減少として示される。他方処罰は各人にBとの取引がなされたなら得られたはずの効用の断念を意味する。断念される効用は処罰の重さ x に正比例するとすれば、それを πx (π は正の定数) と書く。⁽⁵⁾ここに

$$v > \pi > 0$$

でなくてはならない。もし u が v よりも大であれば、B は非協力による効用の増加分 u が同じ非協力による効用の減少分 v を上回ることから非協力を選択するからである。以上からBが協力するときの効用は

$$[(m+1)/n] b - k - \pi x$$

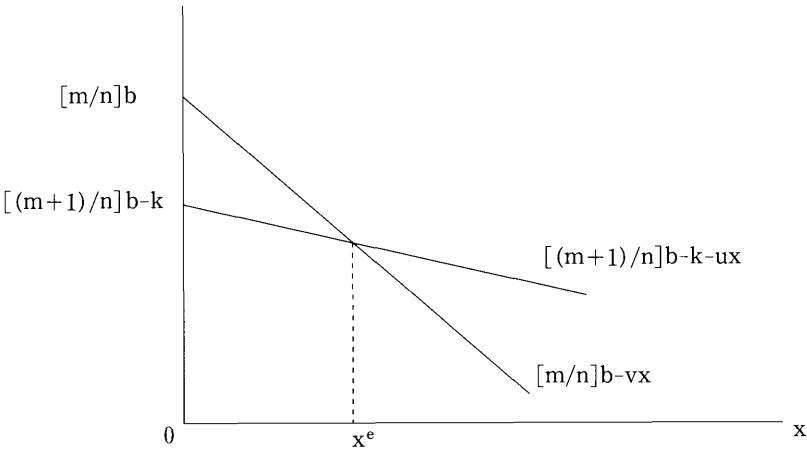
他方、B が非協力であるときの効用は

$$[m/n] b - vx$$

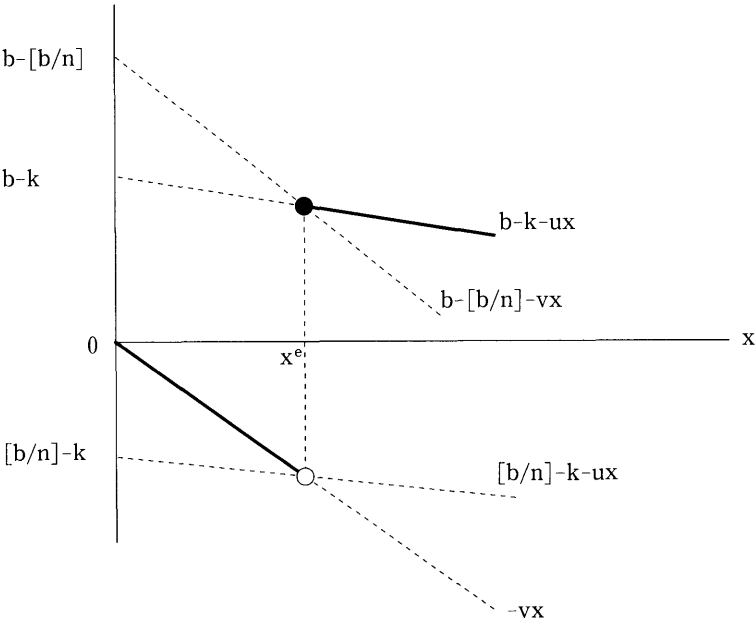
で示される。前者が後者より小さければBは非協力、さもなければ協力を選擇する。両者を均等にする x の値を x^e とすると、 $x \equiv x^e$ ならばBは協力を、 $0 \equiv x \wedge x \neq x^e$ ならば非協力を選擇する。ここに $x^e \equiv [1/(v-\pi)] [k - (b/c)]$ である。図解すると第1図、第2図のようになる。なお、 x^e が m の値から独立である点に注意されたい。

第四節の「鼠の会議」のモデルでは x^e が m に依存していることが示され、この対比がポイントになるからである。第1図は、 m 人の協力者を予想したときBが刑罰の重さ x の値如何で協力したり非協力になったりすることを示している。 x^e より x が大なれば協力、未滿ならば非協力である。第2図は第1図について、 m が $0, 1, 2, \dots, (n-1)$ の値を逐次とったときの詳細を図解したものである。いま x が x^e 以上に設定されるとBは協力する。すなわち、すべての村人は協力するから燈台は実現し、Bの最終的効用は、

第 1 図



第 2 図



$b-k-ux.$ $x \geq x^e.$
 となる。他方、 x が x^e 未満であれば、各村人は非協力に走るから燈台は実現しない。そのときの各村人の最終的な効用は、

$$-vx, \quad 0 \leq x < x^e$$

となる。これらは第2図において太い実線で示されている。問題はAが刑罰の重さしたがって、 x をどの水準に定めるかである。AはBの厚生を最大化するための組織体であるから、Aの目的関数を $\pi(x)$ とすると、これを最大にするように x を決定すると考えて差支えない。すなわち、

$$\text{Max}_x a(x)$$

である。この

$$a(x) = \begin{cases} b-k-ux & x \leq x^e \\ -vx & 0 \leq x < x^e \end{cases}$$

である。第2図から明らかなように

$$b-k-ux^e \geq 0 \iff x = x^e$$

$$b-k-ux^e < 0 \iff x = 0$$

である。

$x = x^e$ ならば各村人は協力するから燈台は設置され、各村人の最終的な効用は $b-k-ux^e$ となる。他方、 $x = 0$ ならば刑罰は科されず、各村人は非協力となって、そのため燈台は設置されず、各村人の最終的な効用はゼロとな

る。別言すれば刑罰がゼロの下では各村人は非協力となってAの設立そのものが不可能ということ、つまり政治体の設立が割りに合わないこと、したがって政治権力の行使 (mutual coercion mutually agreed upon) が各村人にとって余りに高くつきすぎて実行不可能ということである。

(四) 「鼠の会議」の定式化とその帰結

鼠の会議・「誰が猫の頸に鈴をつけるのか」は周知のイソップの寓話であるが、大要次のようなものである。

天敵の猫にいかに対処すべきかについて鼠が会議を開いた。席上、ある鼠が「猫の頸に鈴をつけられ、その襲来を事前に察知でき、被害を未然に防止できる」と提言し、皆々これを妙案として賛意を表した。しかるに別の鼠がいうには「では、誰が猫の頸に鈴をつけるのか」と。これに対して一人として「我こそ」と発言する者はない。よって「そのたびの議定事終らで皆々退散しぬ」⁽⁶⁾という結果となった。

この話では誰か一疋が身を挺して猫の頸に鈴をつけられ、それによって猫の襲来をすべての鼠に知らせることになり皆々身の安全をはかることができる。つまり一疋の犠牲的貢献に他の鼠が只乗りできるということである。任意の一疋の鼠が猫の頸に鈴をつけるのに協力するときの彼の効用をc、非協力のそれをdとする。いま彼のほかに協力者がゼロであるとする、協力の効用 Ω は $\Omega = c$ である。ここにbは彼が猫の頸に鈴をつけることに伴う危険を示す効用による安全の効用であり、その恩恵は他の成員にもひとしく及ぶ。kとは鈴をつけることに伴う危険を示す効用表示の指標である。これは彼のみの負担である。bもkもともに正の定数とする。さらに協力者が自分の他に一疋いるとすると、自分か彼かのどちらか一疋がクジ引きによってその危険な行為を引き受けるから、bから差し

引かれる効用はクジに当たる確率 $[1/2]$ と k との積である。すなわち $c(2) = b - [1/2]k$ となる。一般に自分以外の m 正が協力すると予想されるとき、協力の効用 $c(m+1)$ は

$$c(m+1) = b - [k/(m+1)] \quad 0 \leq m \leq n-1$$

で示される。他方、協力者ゼロのときの非協力の効用は $d(0)$ 。協力者が一正のときの非協力の効用は $d(1)$ で、これは他の鼠の貢献の余澤であるから、 b の値をとる。協力者が二正のときの非協力の効用 $d(2)$ も b の値である。これは誰か一正が事にあたれば足りるからである。一般に自分以外の m 正が協力すると予想したときの彼の非協力の効用 $d(m)$ は

$$d(m) = b \quad 1 \leq m \leq n-1$$

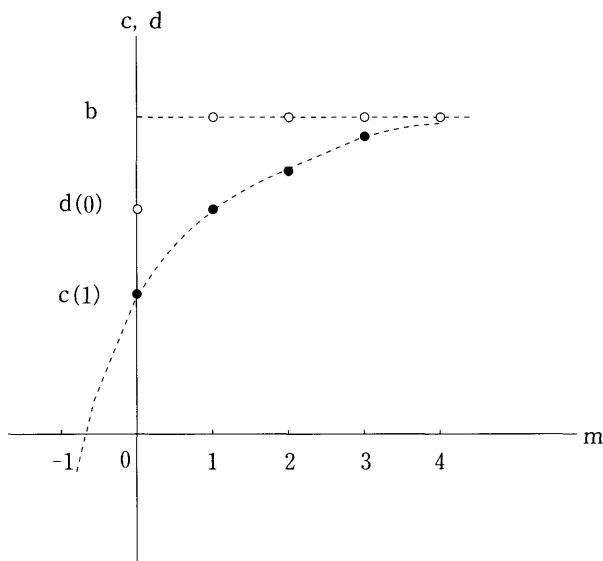
で示される。以上を詳しく書くと、

$$\begin{array}{lll} m=0 & c(1) = b-k & d(0) \\ m=1 & c(2) = b - [1/2]k & d(1) = b \\ m=2 & c(3) = b - [1/3]k & d(2) = b \\ \vdots & \vdots & \vdots \\ m=n-1 & c(n) = b - [1/n]k & d(n-1) = b \end{array}$$

ということになる。ここに $d(0)$ は b よりも小さいと考えるのが自然である。前者は猫の頸に鈴がついていないときの彼の効用、後者は猫の頸に鈴がついているときの効用だからである。

さて、囚人のディレンマが成立するための必要十分条件は(1)と(2)が同時に成り立つことであつた。いま $d(0)$ が正であるとすると、上記が成り立つための必要十分条件は

第 3 図



$$c(1) > d(0) < c(n)$$

である。以上を図示すれば第3図のようになる。

この囚人のディレンマから脱却するには政治権力を使用して各鼠をして協力せしめる以外に方法はない。これは、任意の一足の鼠をB、その他の $(n-1)$ 足の鼠からなるひとつの結託をAとするとBが非協力ならばAがBを処罰するということである。協力する鼠の数を m と n と $n-m$ と、前と同様にBの非協力に対する刑罰を $-ux$ 、また処罰に伴う負担を $-vx$ とすると、Bの協力の効用は $b - [k/(m+1)] - ux$ 、非協力の効用は $b - vx$ になる。そこで、

$$b - [k/(m+1)] - ux = b - vx \quad (1 \leq m \leq n-1)$$

を満たす x の値 x^e は

$$x^e(m) = k / [(m+1)(v-u)]$$

となって、 m に依存する。つまり m が増加するにつれて x^e は小さくなる。ただし、 m が0のときの x^e は

$$x^e(0) = [d(0) - c(1)] / (v-u) \quad \text{or} \quad [d(0) - b + k] / (v-u)$$

となつて、この値と

$$x^e(1) = k / [2(v-u)]$$

この値の大小は定めがたい。

さて、 $x \parallel x^e(m)$ であれば各鼠は協力し、その中の一疋の手によって猫の頸に鈴がつけられる。その結果、鼠の安全は確保される。そのときのBの最終的な効用は

$$b - [k/n] - ux$$

である。他方、 $0 \parallel x \wedge x^e(m)$ であれば、各鼠は非協力となり猫の頸に鈴はつけられないままとなって各鼠はその脅威にさらされる。そのときのBの効用は

$$d(0) - vx$$

である。さて刑罰の重さしたがって、 x を定めるのは結託Aである。Aはそれ自身の目的関数——実はそれはBの最終的な効用である——を最大にするように x の値を定めると考えよう。つまりAの行動は次のように示される。

$$\text{Max}_x a(x)$$

ただし、

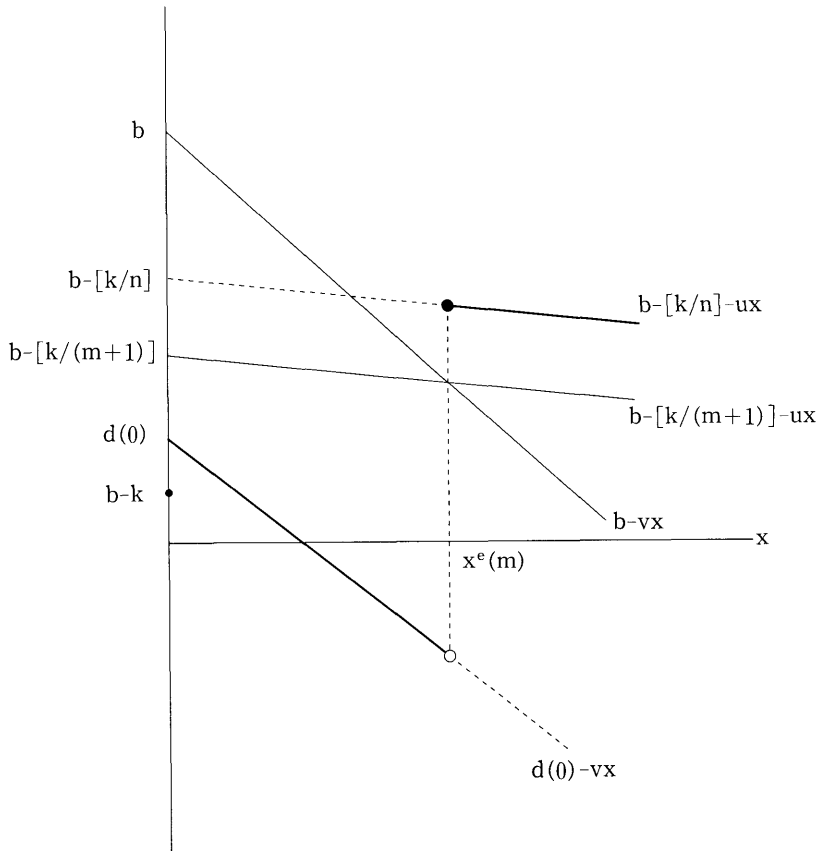
$$a(x) = \begin{cases} b - [k/n] - ux & x \geq x^e(m) \\ d(0) - vx & 0 \leq x < x^e(m) \end{cases}$$

である。 m は所与であるから、これを一定と考ええると、第4図より明らかなように、

$$(\alpha) \quad b - [k/n] - ux^e(m) \equiv d(0) \quad \leftrightarrow \quad x \equiv x^e(m)$$

$$(\beta) \quad b - [k/n] - ux^e(m) < d(0) \quad \leftrightarrow \quad x = 0$$

第 4 図



という結果を得る。 (α) が成り立つと刑罰 $\propto x_e(\beta)$ が科されるから各鼠は協力し、そのうちのどれか一足によって猫の頸に鈴がつけられるから鼠の安全は保障される。 (β) が成り立つと刑罰はゼロであるから各鼠は非協力となり、その結果として猫の頸に鈴はつけられないままとなる。かくて猫の脅威は継続する。ここで注意すべきことは x^e が m に依存していることである。 m が下落すると x^e は上昇するから、他の事情にして不変であるかぎり (α) から (β) の方向へと事態は変化する、ということになる。ユニオン・ゲームではこのようなケースは生じなかった。これが両者の相異点である。

(五) 結論

同じ囚人のディレンマでもユニオン・ゲームにおいては各成員をして強制的に協力させるに必要な最小限の刑罰 $(\propto x_e)$ の大きさは自発的な協力者の数 m の大きさには一切依存しない。他方、鼠の会議のゲームでは最小限の刑罰の大きさは自発的な協力者の数に依存する。自発的な協力者の数が減少すると最小限の刑罰は重くしなくてはならない。その結果として刑罰を執行することの費用が高つくようになり、ついには割が合わなくなつて刑罰の執行そのものが不可能となる。つまり政治体が解体せざるを得ないということになる。これが上記のふたつのモデルの帰結の際立った差異である。

ところで、鼠の会議のモデルの場合に自分以外の全員が協力すると予想されるならば、つまり「ヨロヨロ」であるならば、強制力行使は不要であろうか。答は否である。それというのも当の鼠そのものがフリー・ライダーを選択するからである。ここに「当の鼠」とは「任意の個々の鼠」の意であるから、すべての鼠がフリー・ライダー

1 となるということを意味する。したがって、フリー・ライダーを予防するために強制力の行使—処罰は必要なのである。ただ、他の事情にして等しいかぎり、その刑罰そのものが最小($\min$ (b - 1)) になり、刑罰の執行の費用も最小($\min$ (b - 1)) になるから、執行そのものの実行可能性 (feasibility)、つまり $b - [k/n] - ux^e$ と $d(0)$ との格差が、最大になる、ということである。

さて、鼠の会議のモデルの特質は一疋の力だけで猫の頸に鈴がつけられるということであった。これを燈台の事例に適用すると、村人が独力で燈台を建設できるということ、しかも二人以上の人手は不必要という意味で独力で建設する方が効率的であるという想定をすることになる。そして猫の頸に鈴をつけることによる安全確保の効用を燈台のもたらす効用に、猫の頸に鈴をつける負担を燈台建設の労力負担と読み替えば、鼠の会議のモデルの帰結がそのままではまるのである。つまり自発的に独力で燈台を建設すると予想される協力者の数が多くなればなるほどフリー・ライダー防止の刑罰のコストは小さくなるから刑罰の執行そのものが実行可能となつて、したがって燈台が建設されやすくなる、という帰結である。これに対しユニオン・ゲームのモデルでは刑罰のコストが予想協力者数には一切依存しないことは先述したところである。

- (1) 田中宏「国家と権力の理論」萩原・根岸・田中他著『国家の解剖学』（日本評論社、一九九四年）所収、一一八—一四四頁。田中宏「国家崩壊の理論」田中・大石編『政治・社会理論のフロンティア』（慶應義塾大学出版会、一九九八年）所収、二一—一四四頁。ただし、本稿の第二節はこれらと若干異なつた理論展開をしている。

- (2) この定式化はドーズに依拠している。R. Dawes, "Formal Models of Dilemmas in Social Decision Making," in *Human Judgement and Decision Processes*, ed. by M. F. Kaplan and S. Schwartz/Academic Press: New York,

1975), pp. 88-107.

(c) D. M. Messick, "To Join or Not to Join: An Approach to the Unionization Decision," *Organizational Behavior and Human Performance*, 10(1973), pp. 145-156. M. Olson, Jr., *The Logic of Collective Action* (Harvard University Press: Cambridge, 1968), pp. 43-52. Daves, op. cit., pp. 88-107. なおユニオン・ゲームとコモンズ・ディレンマ・ゲームは本質的には同一であるから、両者をまとめてユニオン・ゲームとして扱う。

(4) 主体Aが主体Bに権力ないし強制力行使することを、AがBの重要な与件(data)を操作し、Bの効用を減少せしめることと規定する。その際、与件操作が有効であるためには与件操作に関してAが独占的でなくてはならない。他方、政治権力とはグループにかかわるものである。n人からなるグループについて言えば、それは任意の一人の成員Bの与件をひとつの結託となった他の(5-1)人Aが操作し彼の効用を減少させること、あるいはそのようにすると威嚇することである。これはすべての個人の知力・体力はほぼ同一であるから結託こそ力の源泉であるという事実に立脚している。各メンバーの重要な与件を独占的に操作するには単一の結託が必要である。もし操作しようとする側が単一ではなくて複数の結託となるならば、それら結託間の競争によって各人の与件を操作することが不可能となるからである。また与件の操作をされる方が結託をすると、それだけ与件の操作をしづらくさせる。かくて強制力ないし権力の行使が有効でなくなる。本稿ではBにとつての与件は他の村人との取引である。なお一般に結託Aを主権者(the sovereign)、Bを被統治者(the subjects)という。そもそも主権とは一国家というグループについてであれ、それより下位のグループについてであれ一グループの構成員の力と手段を同一目的に向けること、つまり各人をして同一のストラテジーを採用させることにかかわるものである。フォレットは「多くの人々が単一の行動主体になるプロセスこそ主権が創出されるプロセスである」といっている。M. P. Follet, *The New State* (Longmans: New York, 1918/1923), p. 272. またボサンケは「政府は事実の上からも原理の上からいっても一人々(the one)に対する『その他の人々(the others)の強制としてあらわれる』と述べている。B. Bosanquet, *The Philosophical Theory of the State* (McMillan: London, 1899/1910) p. 76. これらはいずれも同じ内容を表現しようとしているのである。田中「国家崩壊の理論」三二—三三頁参照。

(5) この断念されるべき効用、つまり刑罰を執行することに伴う費用は政治体 A を機能させるための費用と解釈してもよい。

(6) 鼠の会議がイソップ寓話であるかどうかについては、小堀桂一郎『イソップ寓話―その伝承と変容―』(昭和五
十三年、中央公論社)、二〇六頁を参照のこと。なお以下の文章は筆者による要約である。