

## 4枚カード問題からわかること ——裏切り者検知・予防措置・利他者検知——

小 林 佳世子

### 要 約

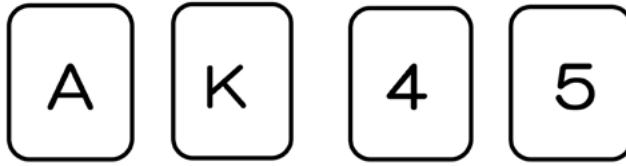
4枚カード問題とは、ウェイソン選択課題（Wason selection task）ともいわれ、イギリスの認知心理学者 Wason が考案した条件付き推論問題であり、「PならばQである」という形式の、条件付きカードに違反している可能性のあるものを見つけるという問題となる。問題の形式をほんの少し変えるだけで、その正答率が大きく変化することでも知られている。本稿では、この4枚カード問題からみえてくるヒトのもつ認知バイアスとして、裏切り者検知、予防措置、利他者検知についてサーベイをする。

またヒトの心は、適応上重要であった個々の問題の解決に特化したモジュール群から構成されるという、心の領域特殊性（domain specificity）という考え方が近年主張されるようになっていく。本稿でもその立場を支持し、4枚カード問題からみえてきたこれらの力は、領域一般的なものというよりは、進化の中で培った個別領域的な認知能力であることを議論する。

### 1章 はじめに 4枚カード問題とは

4枚カード問題とは、ウェイソン選択課題（Wason selection task）ともいわれ、イギリスの認知心理学者 P. C. Wason が考案した条件付き推論問題である（Wason 1968）。これは、「もしPならば、Qである」という形の条件付きルールに、違反している可能性のあるものを見つけださせるという、一種の論理パズルといってもよいであろう。

ウェイソンが考案した元の問題は、特に複雑な推論が必要とされるわけでもないにもかかわらず、その正答率は一般に10%にも満たないことが知られている。一方で、問題の形式をほんのわずかに変えるだけで、その正答率が大きく変化することなどから、認知科学などの分野で非常によく使われる推論問題の1つともなっている。具体的な例で考えてみよう。



ルール：「母音の裏には、偶数の数字が書いてある」

図1-1：4枚カード問題（抽象的な問題）

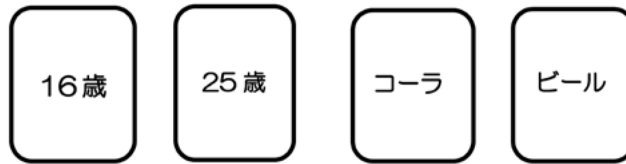
最初に4枚カード問題の基本形として、事実の状況を述べているだけの選択問題（抽象的な問題）を考える。そこでは、カードの表にはアルファベット、裏には数字が書いてある4枚のカードが並べてあり、2枚が表、2枚が裏になっている（図1-1）。表のカードには「A」と「K」と書かれており、裏のカードには、「4」と「5」と書かれている。ここでのルールは、「母音の裏には、偶数の数字が書いてある」である。このルールが守られているのかどうかを確かめるために、カードをめくってその裏を確認しないといけないのはいったいどのカードかというのが、この4枚カード問題となる。もちろん余計なカードを裏返すことは許されず、必要最小限のカードしかその裏をみることはできない。

これは、「pならばq」という命題に反しているものを見つける問題であり、4枚のカードは、それぞれ「pである」「pでない」「qである」「qでない」に対応している。

形式論理学の考え方を使えば答えはすぐにわかるとおり、「pならばq」という命題の正しさを確認するためには、「pである」と「qでない」の2枚のカードの裏を調べる必要がある。したがって、答えは、「pである（母音である）」＝「A」と、「qでない（偶数でない）」＝「5」の2枚である。

しかしながら、実際にはこの種の問題の正答率は決して高くはなく、実験をすると、その正答率はせいぜい5%～30%程度であり（Cosmides *et al.* 2010）、多くの場合には10%にも満たない程度であることが幅広く知られている（Newstead & Evans (Eds.) 2013）。

しかしながらこの問題に文脈を与えると、ときに驚くほどにその正答率が改善することが知られている。たとえばカードの表に年齢を、裏には飲んでいる飲み物を書き、ルールは、「アルコールを飲むならば、20歳以上でなければならない」とする（図1-2）。先の問題とこの問題は、もちろん論理的にはまったく同じである。しかし問題の形式をこのように変えてやることで、正答率はおおむね65%～80%程度へと跳ね上がり、先の例とは比べ物にならないほど高くなることが多くの実験から繰り返し示されている（Cosmides *et al.* 2010）。日本においても、同様の傾向が追試されている（長谷川・平石2000）。



ルール：「アルコールを飲むならば、20歳以上でなければならない」

図1-2：4枚カード問題（アルコールの問題）

この問題は以前から知られており、当初は抽象性などの問題とされ、課題をより現実的で具体的なものにすることで、その正答率が上がる可能性などが議論された（Wason & Shapiro 1971, Johnson-Laird *et al.* 1972）。しかし後述の事例をみてもわかるとおり、単にルールを現実的なものにしたり具体的なものにしたりするだけでは、必ずしも正答率が上がるとは限らないことは今日幅広く知られている。

その中でCosmidesらは、正答率の高い問題が、社会契約問題（social contract reasoning）における「裏切り者検知（cheater detection）」という形式になっていることに着目した。4枚カード問題に社会的な交換を埋め込むと、「利益を得るならば、対価を払わなければならない」あるいは「利益を得るならば、その必要条件を満たしていなければならない」という形をとることになる。この形式下で、「利益を得ていながら、代償を払っていない者」、つまり「裏切り者」を探すという状況のとき、その正答率が跳ね上がることに着目したのである。そしてCosmidesは、その自身の博士論文で、ヒトは裏切り者を検知する能力にたけていることを主張した（Cosmides 1989）。つまり、利益を得ているにもかかわらず対価を払っていない「裏切り者」を探す「裏切り者発見モジュール」が、ヒトの脳には組み込まれていることをCosmidesらは主張したのである（Cosmides and Tooby 1989, Cosmides and Tooby 1992, Cosmides and Tooby 2005）。もちろんこのような裏切り者を探す能力は、フリーライダーを素早く探し出すことに貢献し、集団秩序の維持にとって非常に重要な能力であるといえる。

4枚のカードに社会契約を組み込むと、カードの表には利益を得たかどうか、裏にはその代償を払ったかが書かれることになる。具体的には、4枚のカードにはそれぞれ、「利益を得た者」「利益を得ていない者」「代償を払ったもの」「代償を払っていない者」が示されることとなる。

社会契約問題における裏切り者とは、先にもふれたように「利益を得ているにもかかわらず、対価を支払っていない者」である。したがって4枚カード問題において確認すべきカードは、「利益を得た者」が代償を払っているかという点と、「代償を払っていない者」が利益を得ていないかどうかとなる。したがって、確認すべきカードは「pである」と「qでない」となることから、この回答は論理的解答とも一致することがわかる。

その後、裏切り者検知以外にも、予防措置問題および利他主義者検知の少なくとも3種類の設定で、正答率が高くなることが議論されている。本稿ではこれらの問題をサーベイするとともに、社会契約仮説と並んで提示されてきた「記憶手掛かり仮説」と「実用的スキーマ仮説」の2つの仮説、およびそれらへのCosmidesらの反論についてもあわせて議論することで、ヒトは進化の中で出会う重要な個別の問題それぞれに適した個別領域的な推論能力をもつことを示す。

## 2章 社会契約仮説とその他の仮説

本章では、Cosmidesらの社会契約仮説とそれに対立する代表的な仮説についてサーベイをし、それらの仮説に対するCosmidesらの反論についてまとめることで、社会契約仮説の優位性を示す。まず2-1節では、問題のなじみの問題としての「記憶手掛かり仮説」について、2-2節では、義務や許可の場面での認知能力の問題とした「実用的推論スキーマ仮説」についてそれぞれ議論する。

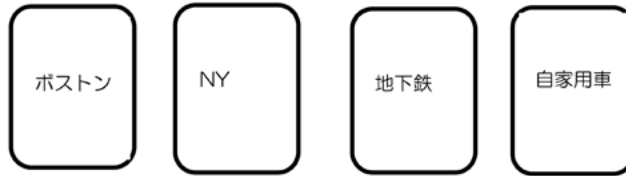
### 2-1 記憶手掛かり仮説

1章で議論したとおり、4枚カードの「アルコールの問題」ではその正答率が著しく跳ね上がることが知られている。文脈に応じてその正答率が大きく変化するのはなぜかという点に注目が集まる中で、まず考えられた仮説の1つが、「記憶手掛かり仮説 (availability hypothesis)」であった。

この説明によれば、社会契約の場面で4枚カード問題の正答率が高くなるのは、多くの人が日常的に社会契約の場面を経験しているためであり、そうした問題や状況に「なじみ」があるからであるとされる。抽象的な形式の問題は、実際の生活の中で経験をすることはほとんどない。しかし、提示されたルールや状況が被験者にとって慣れ親しんだものであるときには、ルールに反した事例 (PかつQでない) を思い出しやすくなり、その結果として正答率が上がるのだという考え方である (Griggs & Cox 1982, Manktelow & Evans 1979)。

直観的には説得力のある仮説ではあるが、Cosmidesらは、この仮説に対して以下の事例を用いて反論した。具体的でなじみがあると考えられるが、社会契約にはなっていない問題を考え、そこではその正答率が社会契約仮説ほどには高くないことを示したのである (Cosmides 1989)。

具体的には、図2-1のとおり、ここでのルールは、「ボストンに行くならば、地下鉄を使う」というものである。カードの表には行き先が、裏には交通手段が書いてある。2回行われた実験の参加者は、ハーバード大学の学部生各24名である。



ルール：「ボストンに行くならば、地下鉄を使う」

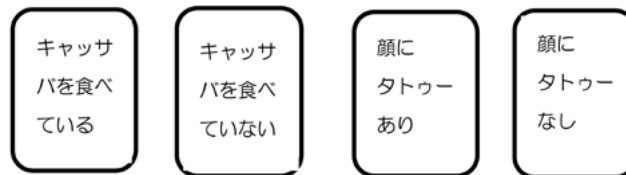
図2-1：4枚カード問題（具体的でなじみのある問題）

ルールをみるとすぐにわかるとおり、これはとても具体的な問題であるうえに、状況としてもなじみがあると考えられる。しかしながら、「利益」や「対価」といった概念が入っておらず、もちろん社会契約とはなっていない。ここでの問題の答えは「ボストン」と「自家用車」であるが、2回行われた実験の正答率はそれぞれ46%と38%と、抽象的な問題である数字とアルファベットの問題よりは高かったものの、社会契約の問題よりは大幅に低いことが示されている。

さらに Cosmides らは、非常に巧みな実験を行っている。まったく同じ問題を、社会契約の文脈のときとそうでないときで、結果を比較したのである。

ここでは、キャッサバができる地域に住んでいる人は、顔にタトゥーをいれる習慣があると説明される。キャッサバがとれる地域の人しかこれを食べることができないため、キャッサバを食べる地域の人にはタトゥーがあることになる。つまりここでのルールは、「キャッサバを食べるなら、顔にタトゥーがある」となる。

図2-3では、4枚のカード表にはキャッサバを食べているかどうか、裏には顔にタトゥーがあるかどうか書かれている。



ルール：「キャッサバを食べるなら、顔にタトゥーがある」

図2-2：4枚カード問題（具体的だがなじみのない問題）

これらの問題は、なじみはないとはいえ具体的な問題であるといえるが、この種の課題の正答率はわずか20%程度にとどまり、やはりあまり高くはないことが示されている（Cosmides & Tooby 1989）。ここでの文脈は、ある特定の地域の人がある特定の食べ物を食べるという状況の記述でしかなく、利益や代価といった概念はいっていないので、もちろん社会契約とはなっていない。

しかしここで、状況の説明だけを少し変え、社会契約の文脈として説明を行う。具体的には、キャッサバはとてもおいしくて栄養価もある魅力的な食べ物だが、希少性

の高いものでもあり、忠誠の証しである顔にタトゥーをいれた者しか食べることが許されていないと説明される。カードもルールもまったく同じであり、具体的とはいえなじみのない状況であることに変わりはないが、状況の説明を社会契約の形に変えただけのこのバージョンでは、正答率は一気に跳ね上がり、70%を超えることが示された。こうした結果から、ヒトには、裏切り者がいるかもしれないという可能性に直面したときに働くメカニズムが備わっているのだと、Cosmidesらは主張したのである。

## 2-2 実用的推論スキーマ仮説

しかし実は、Cosmidesらの社会契約仮説がだされた時点で、Cosmidesらの説にはすでに1つ大きな問題が存在していた。それは、社会契約ではない文脈で、その正答率が大幅に改善する問題が知られていたのである（Cheng & Holyoak 1989）。今日、Cosmidesらが「危機管理理論（hazard management theory）」（Cosmides *et al.* 2005）とよんでいるものであり、予防措置問題ともいわれる。

予防措置問題とは、「危険な活動に従事するならば、予防的措置がとられなければならない」あるいは「もし危険がせまったら、自分を守る行動をとらなければならない」といった問題である。こうした問題の正答率は80%前後となることが知られ、社会契約と並んで非常に高いことがすでに知られていた。しかしすぐにわかるとおり、こうしたルールの侵犯自体は「裏切り」を意味しない。つまり社会契約仮説が当てはまらないにもかかわらず、その正答率が跳ね上がる状況がすでに発見されていたのである。

こうした状況から、有力な代替仮説として、ヒトは裏切り者を検出する能力をもつのではなくて、何をしなくてはならないか、あるいは何をしてもよいかという、「義務や許可に関する推論能力（deontic reasoning）」をもつのではないかという仮説が提示された（Cheng & Holyoak 1985, Cheng *et al.* 1986）。それが、「実用的推論スキーマ仮説（pragmatic reasoning schema hypothesis）」である。この仮説は「前提条件yが満たされたのなら、ある行為xを行ってよい」という許可のルールが与えられたときに、「許可スキーマ（permission schema）」が働き「pである」と「qでない」という正答を導くことができるようになるという考え方である。

すぐにわかるとおり、「利益を得ること」を「行為xを行うこと」に、「代償を払うこと」を「前提yを満たすこと」に読みかえることで、社会契約もまた許可のルールの1つと考えることができる。同様に、「危険な活動を行うこと」を「行為xを行うこと」に、「予防的な措置をとること」を「前提yを満たすこと」に読みかえることで、予防措置のルールもまた許可のルールの1つと考えることができる。したがってこの考え方にたつと、ヒトは社会契約と予防措置の両者を含んだ、「義務」や「許可」の文脈での統一的な推論能力にたけていると考えることができる。つまり、ヒトは個別領域的な力をもつのではなく、むしろ領域一般的（領域普遍的）な認知能力をもつと



いう立場を支持する考え方となる。

これを支持する仮説として、Cumminsは、社会的地位という視点を導入すると、4枚カードにおける正答率が大きく改善することを示す例を提示している(Cummins 1999)。

私たちヒトという生き物を含む霊長類の多くは、「順位構造」をもつ社会で暮らしている。こうした社会においては、各個体は、常に「他者と自分の相対的な社会的地位」の関係を意識したうえでの意思決定が必要である。相対的に地位の高い者はルールに従う「義務」は弱いが、相対的に地位の高い者はその「義務」が強くなると考えられ、自分と他者との相対的な社会的地位の高さに応じて、適切なふるまいが変化すると考えられよう。Cumminsは、この点に着目したのである。

そこでCumminsは、4枚カード問題に相対的な社会的地位を導入し、回答者の視点をヒエラルキー内の高位置に置く場合と、ヒエラルキー内の低位置に置く場合とで、その回答が変化することを示した。そこでは、自分よりも社会的な地位の低い者、つまり「目下」の人間による意図的な裏切りは、自分より社会的地位の高い者、つまり「目上」の人間による裏切りよりも見つけやすかったのである。実験の参加者は、カリフォルニア州立大学サクラメント校とカリフォルニア大学デイヴィス校の、160人の学生である。

ここでの実験では、目上の者同士や目下の者同士など同じ立場の者同士、あるいは目下の者が目上の者の裏切りを見つけるときには、その正答率は15%～20%程度と、あまり高くはないにもかかわらず、目上の者が目下の者の裏切りを見つけるときに限って、その正答率が65%と一気に引きあがることが示された(図2-3)。これは、そうでないときとくらべて3倍以上もの高い正答率である。つまり、自分より地位の低い者の恣意的な義務や許可への違反には敏感であるが、自分より高い地位の者による違

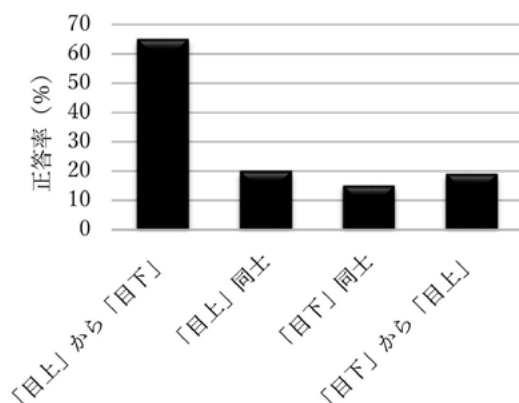


図2-3：立場ごとの正答率

出典：Cummins(1999)より

反には必ずしもそうでもないことが示されたのである。

こうした実用的推論スキーマ仮説に対して、Cosmidesは、社会契約の文脈としてとらえ高い正答率が得られたケースであっても、その同じ問題を、社会契約でもなく予防措置でもないが「許可」を与える文脈として考えると、その正答率があまり高くはならなかったことを示し、社会契約仮説の優位性を主張している（Cosmides 1989）。

具体的には、図2-2と基本的には同じ状況を考え、キャッサバができる地域に住んでいる人は、顔にタトゥーをいれる習慣があるという先と同じ説明をされる。ただしここでは、キャッサバはごく普通の食べ物であり、またタトゥーをいれることも、この地域の人間にとってはごく当たり前のことで特段のコストのかかることではないとされる。また資源保護のために、キャッサバができる地域の人しかこれを食べることが許されていないといわれる。つまり、キャッサバができる地域の人しかこれを食べることができないので、キャッサバを食べる人の顔にはタトゥーがあることになる。つまり、ここでのルールもカードも図2-2のときとまったく同じままで、状況の説明だけを変更したのである。

この文脈は、キャッサバの価値や顔の入れ墨の意味を変えた結果として、「許可」の文脈ではあるが、社会契約にはなっていない。すると、社会契約としてとらえたときの正答率は80%であったにもかかわらず、許可の文脈であるここでの正答率はその半分程度の44%にとどまり、社会契約仮説を支持する結果となった。

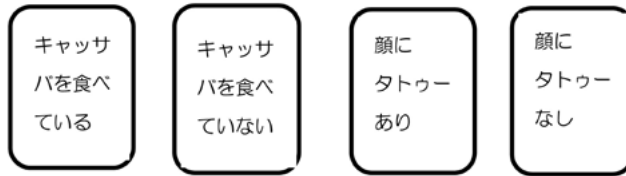
さらにCosmidesは、社会契約問題が「何らかの形で論理的思考を助けている」のであって、必ずしも裏切り検知を行っていない可能性を排除するために、次のような実験も行っている。

これまでの実験では、「利益を得るならば (p)、代償を払う (q)」というルールで実験が行われていた。そこでは、ヒトは、「利益を得ている (p)」にもかかわらず「代償を払っていない者 (非q)」つまり「裏切り者」を探していた。

そこでCosmidesは、この前段と後段をいれかえ、「代償を払うなら (p)、利益を得る (q)」というルールに変更をした。たとえば、「顔にタトゥーがある (p) なら、キャッサバを食べる (q)」というルールである（図2-4）。

もし、社会契約の文脈が何らかの形で論理的思考を助けているのならば、「pである」と「qでない」のカードを選ぶはずである。つまり「顔にタトゥーあり (p)」と、「キャッサバ食べていない (qでない)」の2枚のカードである。しかし、もし本当に「裏切り者検知」を行っているのであれば、やはり「利益を得ている (q)」にもかかわらず、「対価を払っていない (pでない)」という者を選ぶはずである。つまり、選ぶカードは「キャッサバ食べている (q)」と「顔にタトゥーなし (pでない)」の2枚である。これらのカードは、もちろん論理的思考によって得られる解答とはまったく逆となる。





ルール：「顔にタトゥーがあるなら、キャッサバを食べる」

図2-4：4枚カード（前段と後段をいれかえる）

実験の結果は、「pでない」と「q」を選ぶものが60%～80%となり、Cosmidesらの社会契約仮説を強く支持するものとなった。こうした結果は、その後ドイツや日本でも追試され、同様の結果を得ている（長谷川・平石 2000）。

このように、社会契約仮説に対して複数の代替仮説が考えられてきたが、Cosmidesらはそれらの仮説に対して反論を行い、社会契約仮説の優位性を示してきた。

さらに次の章では、4枚カード問題からみえてくる第2の力としての予防措置問題と、第3の力としての利他主義者検知について考え、さらにこの「予防措置問題」を支えるヒトの危機管理の力は、社会契約とは独立の個別領域的な力として考えられることを議論する。

### 3章 予防措置問題と社会契約

本章では、4枚カード問題からみえてくる、ヒトのもつ第2の力としての「予防措置」にかかわる問題についてサーベイを行う。

ここでは、ヒトは裏切り者から身を守る社会契約の問題と並んで、危険な活動から身を守るための予防措置に関する推論能力をヒトはもつことを議論する。また心は、適応上重要であった個々の問題の解決に特化したモジュール群から構成されるという、心の「領域特殊性（domain specificity）」が近年主張されるようになっている（Pinker 1997, Tooby & Cosmides 1992）。ここでの議論も、社会契約と危険管理とのそれぞれの問題に特化した、「個別の推論能力」をヒトはもつという考え方を支持するものである。

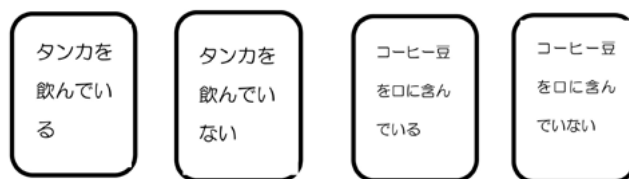
#### 3-1 予防措置問題

ではまず最初に、予防措置問題とはどのようなものを理解するために、具体的な事例をみしてみる（Fiddick 2004）。

ここでは、以下の仮想的な状況を考える。タンカという宗教的な飲み物は毒が含まれており、その毒を中和するために、タンカを飲むときにはコーヒー豆を口に含んでいなければいけないという状況を考える。つまりここでのルールは、「タンカを飲む

4枚カード問題からわかること

ならば、「コーヒー豆を口に含む」となる（図3-1）。



ルール：「タンカを飲むならば、コーヒー豆を口に含む」

図3-1：予防措置問題

これは、タンカという毒を含んだ飲み物を飲むならば、毒を中和するためにコーヒー豆をあらかじめ口に含ませておかねばならないので、まさに予防措置の形式をとっている。このケースでの正答率は90%を超えており、非常に高い正答率となっていた。

2章で議論したとおり、社会契約と予防措置は、どちらも「許可」や「義務」への違反として統一的にとらえることも可能である。そうした仮説に対し、CosmidesやCosmidesの学生であったFiddickらは、さまざまな視点からの反論を行っている。たとえばFiddickは、社会契約の場面では「意図」の有無でその正答率が大きく変化するが、予防措置の問題では意図の有無による変化は起きないことを示した。FiddickやCosmidesらは、これは、社会契約と予防措置が異なる領域の力である証拠の1つであると主張している。

そこでまず3-2節では、社会的交換の文脈では、偶然の逸脱よりも意図的な逸脱の方をより高率で見つけることを示し、ついで3-3節では、予防措置問題では、こうした「意図」の有無による結果の差がみられなかったことを示す。

### 3-2 意図的な裏切り者

まず、社会契約の文脈であっても、意図の有無でその結果が変わることを示す(Fiddick 2004)。ここでの実験参加者は、ハーバード大学人類学専攻の学部生89名である。

ここでの問題はこどものスイミングのクラスの登録で、もしこどもを上級クラスに登録するならば、50ドルの追加費用を払わなければならないとする。つまりここでのルールは、「上級クラスに登録するならば、追加費用を払う」となる。図3-2では、カードの表にはこどもが登録されたクラスが、裏には追加費用を払ったかどうか書かれている。



ルール：「上級クラスに登録するならば、追加費用を払う」  
図3-2：4枚カード問題（スイミングと追加費用の問題）

ここでは、祖母が登録の手続きをするが、アルツハイマーという病気のため、“うっかりと”間違ったクラスに登録してしまうことがあると説明される。するとこのときの正答率は、45%ほどにとどまった。もちろんこの場合の正解は、「上級クラス」と「追加費用を払っていない」である。

しかしここで状況の説明を変え、今度は父親がその登録をするが、この父親は、“意図的に”追加費用を支払わないまま上級クラスに登録しようとしている状況だと説明される。するとこのケースにおける正答率は、80%を超えることが示された。

こうした結果から、ヒトは、「結果的な裏切り」ではなく、「意図的な裏切り」に敏感であることが示唆される。同様の結果はほかにもあり、そこでは、意図的な逸脱の場合の正答率は68%だったが、意図のない失敗の場合には、その正答率はわずか27%にとどまっていた（Cosmides &Tobby 2005）。

なお、「意図のないうっかりの失敗」の場合には正答率が高くはならないというこの結果は、「なじみ」が問題であるとした「記憶手掛かり仮説」に対して、さらに不利な証拠ともなっていることがわかる。

### 3-3 予防措置問題と意図

では次に、予防措置問題における意図の有無による変化を考えてみる。ここでの例もまた、こどものスイミングのクラスの登録の問題である（Fiddick 2004）。ここでは、上級クラスは泳ぎに慣れていない初心者には危険なので、泳ぎに熟達したこどもしかそのクラスには登録できないと説明される。つまりここでのルールは、「上級クラスに登録するならば、泳ぎに熟達していなければならない」である（図3-3）。

この予防措置の状況で、意図のあるバージョンと意図のないバージョンを考える。意図のないうっかりミスのバージョンでは、やはり祖母はアルツハイマーのため、うっかりと間違って登録をする可能性があるとして説明される。意図のあるバージョンでは、今度のお父さんは、こどもがまだ初心者であっても無理をしてでも上級クラスにいたいと考えていると説明される。

両者ともに、先の意図的な裏切り者を探すケースと、形式もルールも非常によく似ている。しかし今回の予防措置問題の状況では、意図のないうっかりミスのものも意

4枚カード問題からわかること



ルール：「上級クラスに登録するならば、泳ぎに熟達していなければならない」

図3-3：4枚カード問題（スイミングと上級クラスの問題）

図的なものも、どちらもその正答率は80%前後と両者が並んで非常に高い率となった。

3-2節で議論したとおり、社会契約の状況では意図的な裏切りのみの正答率が高かったが、よく似た状況を予防措置としてとらえると、意図の有無にかかわらずその正答率が高くなることが示されたのである。「意図的な裏切り者」は典型的なフリーライダーだと考えられるが、「うっかりの失敗」は必ずしもフリーライダーではない。また一方で、危険な状況は、それがうっかりであろうと意図的であろうと、どちらも危険であることには変わりがない。その両者を、脳は無意識に区別しているというのである。

さらにFiddickらは、fMRIを用いて、社会契約問題と予防措置問題では、脳の別々の部位が働いていることを示した（Fiddick *et al.* 2005）。またさらにStoneらは、脳に局所的に損傷を負った患者に社会契約問題と予防措置問題を解いてもらったところ、予防措置問題では正常な推論でき、その正答率は70%となったが、社会契約問題では著しくその推論能力が劣り、わずか39%にとどまったことを示している（Stone *et al.* 2002）。

またFiddickは、反復プライミング法という手法を使い、この両者の推論を切り離すことができることを示している（Fiddick 2011）。そこでは、社会契約問題を先に解いていた場合には、予防措置問題を先に解いていた場合と比較して、より正しく社会契約の問題の推論ができていた。一方で、予防措置問題を先に解いていた場合には、社会契約の問題を先に解いていた場合と比較して、予防措置問題の方が正しい推論ができていた。

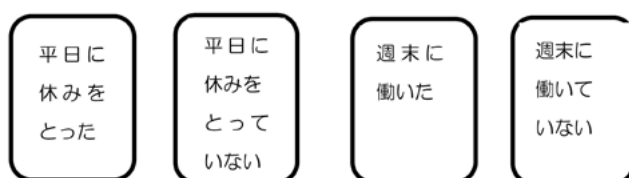
さらに、その両者では関連する「感情」が異なることも示されている（Fiddick 2004）。ドイツ人の被験者を対象とした実験で、裏切り者検知の文脈では“怒り”の感情と関連していたのに対し、予防措置問題は“恐怖”の感情と関連していることが示されたのである。この結果は、オーストラリア、イギリス、インド、日本、シンガポールの参加者で追試され（Fiddick 2011）、さらに4歳児であっても同様の結果が得られたとしている（Pooley & Fiddick 2010）。これらの結果は、社会契約と予防措置問題では、その両者は抽象的には同じ論理構造であっても、その処理をする脳の領域は異なっていることを強く示唆している。

こうした一連の研究を受け、CosmidesやFiddickらは、ヒトは、「義務」や「許可」に関する領域一般的な能力をもつのではなく、進化の中の重要な課題である、「裏切り者から搾取されないこと」と「危険な活動から身を守ること」の、それぞれに対応する「個別領域的な力」を身に付けてきたのだと主張している。裏切り者から搾取されないことと危険な活動から身を守ることは、進化の中の淘汰圧がかなり違うと考えられることを思えば、それぞれが進化の中で別個に刻まれてきた認知能力だとする考え方はごく自然なものであると考えられる。

### 3-4 視点変化効果

さらにGigerentzerとHugは、同様の4枚カード問題で、立場によって裏切り行為が変わる状況において、その立場に応じて「相手が社会契約を守っているかどうか」を確認するということを示し、Cosmidesらを支持する結論を導いた（Gigerentzer & Hug 1995）。ここでの実験参加者は、ドイツのコンスタンツ大学の、19歳から35歳の93名の学生である（図3-4）。

そこでは、まず自分自身を小さな工場のオーナーで、4人の従業員を雇っていると考えるように説明される。この工場では、平日に休みをとったら、週末には働かなければならないというルールがある。ここにそれぞれの従業員の働き方が記録されたカードがあり、表には平日に休みをとったかどうか、裏には週末に働いたかどうかがかかれていいる。ここでのルールは、「平日に休みをとるなら、週末は仕事をする」であり、このルールに反した者を探すにはどのカードをめくればよいかと質問される。



ルール「平日に休みをとるなら、週末は仕事をする」

図3-4：裏切り者検知（視点変化効果）

この状況でよく選ばれたカードは、「平日に休みをとった」と「週末に働いていない」の2枚であった。しかし、まったく同じカードと同じルールのまま、今度は労働組合の代表として考えるように指示をされると、よく選ばれるカードは先とはまったく逆となり、「平日に休みをとっていない」と「週末に働いた」の2枚が選ばれることとなった。立場を少し変えるだけで、答えがまったく逆になってしまったというこの結果は、視点変化効果ともいわれる。

工場のオーナーの立場から考えると、裏切り行為は、「平日に休みをとったのに、週末に働かないこと」である。一方で働く者の立場にたてば、「平日に休みをとらず

に働いたのに、週末に休みがもらえないこと」が裏切り行為となる。立場によって「裏切り」とみなされる行為が異なるために、それぞれの立場から相手の裏切り行為を探そうとした結果、調べるカードがまったく逆になってしまったのではないかと考えられている。Gigerentzerらは、複数のパターンで同様の実験を行っているが、すべてのパターンで、相手が社会契約を守っているか、つまり「相手が裏切っていないか」を確認するという点で一致していたことが示されている。回答者の視点によって、「回答者の視点からみた裏切り者を探す」というこの結果は、Cosmidesらの説を支持するものとなっている。

### 3-5 裏切り者検知 ～ 他文化やこども、動物など

さらに、社会契約における違反者を探すという能力は、少なくとも3～4歳ごろには発揮されることも示されている。たとえば、Harrisらは、イギリスとネパールという異なる文化に属する3歳から7歳のこどもを対象とした実験を行い、おもちゃを交換すると約束をしたのにそれを破った“悪い”お友達は誰かを探させることで、すでに3～4歳のこどもが、社会契約の違反者を探す能力があることを示している (Harris *et al.* 2001)。

さらにCumminsは、3～4歳のこどもを被験者として、「単なる状況の叙述」の違反者を探すよりも、「社会規範」の違反者を探す方が得意であることを示した (Cummins1996)。

ここではまずこどもたちに、ネズミの物語を聞かせる。このネズミたちは夜になると家の裏庭に遊びに行くが、そのうちの何匹かはチューチューと鳴き声をあげるので、近所のネコに襲われてしまう。ここまです基本設定として、ここから先のストーリーは2つのグループで異なるものをこどもたちに聞かせる。一方のグループ（状況の叙述：グループ1）には、「だから夜には、チューチューなくネズミはすべて家の中にいます」と語られる。もう一方のグループ（社会規範：グループ2）には、「だから夜には、チューチューなくネズミはすべて家の中にいなければなりません」と語られる。すぐにわかるとおり、グループ1は単なる状況の叙述として、グループ2は社会規範として物語が語られるのである。

ここでこどもたちに、家の中や裏庭で遊ぶ10匹のネズミのぬいぐるみの入った家のセットが与えられる。このネズミのぬいぐるみは、おなかを握ると鳴き声をあげるものとあげないものがいた。そして、状況の叙述としてとして聞かされたグループ1のこどもたちには、説明された状況が正しいかどうかを確認させ、社会規範として聞かされたグループ2のこどもたちには、規則を破ったネズミがいなかったかどうかを探させた。

このときの正答は、鳴くネズミの数がわからない以上、グループ1もグループ2も



どちらも、裏庭にいるすべてのネズミを調べることである。実験の結果は、社会規範として聞かされたグループ2のこどもたちのほとんどは、裏庭で遊ぶネズミを調べようとした。しかし、単なる状況の叙述として聞かされたグループ1のこどもたちは、裏庭にいるネズミを調べようとせず、家の中のネズミを調べようとしていた。これは、3～4歳のこどもであっても、「社会規範」に違反している者を見つけることにだけしている可能性を意味する。いいかえればこれは、社会の違反者を探すという力が、経験や慣れとはあまり関係しないことを示唆していると考えられる。

さらに同様の傾向は、他文化でもみられることも示されている。たとえば先のHarrisらによる実験は、イギリスとネパールのこどもたちを対象として行われたものであった。また人類学者のSugiyamaは、Cosmidesらと共に、アマゾンの奥地に住むシウィア族に対して同様の実験を行っている (Sugiyama *et al.* 2002)。

シウィア族とは、アマゾンの熱帯雨林の奥地という、人里離れた未開の地で暮らす民族である (Kenrick & Griskevicius 2013)。彼らはよそ者とほとんど接触せず、交流する相手はほとんど近親者に限るという非常に閉鎖的な生活を営み、木の実や果物を採集し、槍で魚をつき、多くは吹き矢を使った狩りによってさまざまな動物を得るという伝統的な生活様式を守っているそうである。正式な学校教育はまったく受けておらず、読み書きができないとされる。

こうしたシウィア族を対象として実験を行ったところ、4枚カード問題の正答率は0%であったが、裏切り者検知の文脈ではその正答率は83%となり、これは同様の問題を与えたときのハーバードの学部生よりも1%高い正答率であったことが示された。これは、社会契約における裏切り者検知という能力は、幅広い文化にみられるということを示すだけでなく、いわゆる「教育」によって後天的に学ぶということとの関連の薄さも示唆している。

またごまかしを見破る能力は、フサオマキザルなど一部の動物にもみられるとも主張されている (Brosnan & de Waal 2003)。ここでは、小石を渡すという「仕事」をすると、報酬として、普通のご褒美であるキュウリか、よりよいご褒美であるがブドウがもらえる。しかし同じ仕事をしていても、隣の仲間がよりよい報酬を得ていると、彼らはとたんにいらだちをあらわにした。その後、多くの霊長類で同様の結果が報告されている (Brosnan 2013)。この実験結果は動物の不公平回避を示す実験であるとされるが、同時に、社会契約の中の「ごまかし」を見破る能力の表れであると考えられることができる (Gazzaniga 2008)。

こうしたさまざまな結果から、ずるを見つけごまかしを見破る能力は、裏切り者から身を守るため、進化の中でヒトの脳に組み込まれた能力なのではないかと、進化心理学者のCosmidesらは主張するようになったのである (Cosmides & Tooby 2005)。さらにその後Cosmidesらは、偶然であつたり、裏切り者に利益がなかったり、状況

がそもそも裏切りを難しくするときには、その正答率が低くなることを示している (Cosmides *et al.* 2010)。

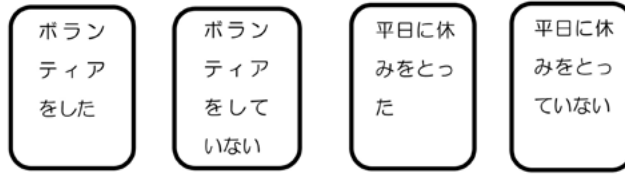
なお、社会契約仮説の優位性を主張するためには、たとえばCumminsの示した社会的地位が入った場合の事例などについても、社会契約の視点から説明ができる必要がある。先のGigerentzerらの「相手の裏切り」を調べるという結果と合わせると、ヒトは「自分に都合の悪い裏切り」に敏感であるという可能性が考えられる (小林 2021 予定)。特にヒトには損失回避の傾向あること、さらに広くいえば、ネガティブなものへの刺激を優先するネガティビティ・バイアスがあることが、さまざまな文脈で繰り返し議論されていることを思えば、自らの損失につながりやすい裏切りにこそ敏感であるという可能性は、十分にあると思われる。こうした視点で、現在検証を進めている。

### 3-6 利他主義者検知

なお、裏切り者検知と予防措置問題以外では、「利他主義者」を見つける利他主義者検知が、正答率を上げる状況の第3の候補として議論されている。本節では、4枚カード問題からみえてくる第3の力である利他主義者検知についてサーベイするとともに、この力もまた、裏切り者検知とは異なる推論能力である可能性が高いことを示す。

BrownとMooreは、4枚カード問題をもちいて利他主義者を見つけやすいという認知バイアスがあることを示した (Brown & Moore 2000)。ここでは、「もし他人のためになることをするなら、報酬をもらう」というルールで、新しい友人を探す、あるいはベビーシッターの候補を探すというような状況を想定して4枚カード問題として提示されると、正答率が上がることが示された (図3-5)。「他人のためになることをするのだが、報酬は要求しない」といった利他的な人を探すことが得意であるという可能性が示されたのである。

さらに小田らは、利他主義者は「決して裏切らない人」という意味では、裏切り者検知と同じメカニズムが働いている可能性を考慮し、裏切り者検知と利他主義者検知との間に関連があるかどうかを検証している (Oda *et al.* 2006)。被験者は、日本人の大学生である。すると、裏切り者検知と利他主義者検知の間には、何ら統計的に意味のある関係は見いだせないという結果が報告された。つまり裏切り者検知がよくできたとしても利他主義者検知がよくできるということはなく、利他主義者検知がよくできたからといって、裏切り者検知がよくできるとは限らなかったのである。これは、裏切り者検知と利他主義者検知の間もまた、独立した別のメカニズムが働いているという可能性を強く示唆している。この結果は、オーストラリアでも追試されている (Fiddick & Erlich 2010)。またこの両者は、背景にある感情面でも異なる可能性が示されている (Chang & Wilson 2004)。



ルール：「ボランティアをしたなら、平日に休みをとる」

図3-5：4枚カード問題（ボランティアの問題）

いわゆる互惠的利他主義を提唱した進化生物学者の Trivers は、裏切り行為には2つあるとしている (Trivers 1971)。1つが「明白な裏切り (gross cheating)」であり、もう1つが「微妙な裏切り (subtle cheating)」である。小田は、明白な裏切りに対して進化したものが先の裏切り者検知の力であり、「微妙な裏切り」に対応して進化したものが、この「利他主義者検知」の力なのではないかとしている (小田 2016)。最初から利他的な人とだけ付き合うことで、微妙な裏切りに会う危険性が少なくなると考えられるからである。

なおここまでの結果をまとめると、「裏切り者検知」と「予防措置」は異なる力である可能性が高く、同様に、「裏切り者検知」と「利他主義者検知」もまた異なる力である可能性が高い。しかし、「予防措置」と「利他主義者検知」の両者が異なる力であることを明示的に調べた研究はないと思われ、この点の検証は今後行われなければならないであろう。

#### 4章 適応という視点からみたヒトの認知能力

本稿では、Wason の考案した4枚カード問題からみえてくるヒトのもつ認知バイアスとして、裏切り者検知、予防措置、利他者検知の力があることを議論した。

またヒトの心は、適応上重要であった個々の問題の解決に特化したモジュール群から構成されるという、心の領域特殊性 (domain specificity) という考え方が近年主張されるようになってきている。その中で、4枚カード問題からみえてきたこれらの力は、領域一般的なものというよりは、進化の中で培った個別領域的な推論能力であることを議論してきた。

裏切り者から搾取されないようにすることと、危険な活動から身を守ることは、それぞれ進化の中の淘汰圧はかなり異なると考えられる。そうであるならば、これらの力は進化の中で重要なそれぞれの問題に対して個別に培われた個別領域的な推論の能力であると考えすることは非常に自然であるといえよう。

## 引用文献

- Brosnan, S. F. (2013) "Justice-and fairness-related behaviors in nonhuman primates," *Proceedings of the National Academy of Sciences*, 110 (Supplement 2), pp. 10416–10423
- Brosnan, S. F., & de Waal, F. B. M. (2003) "Monkeys reject unequal pay," *Nature*, 425, pp. 297–299
- Brown, W. M., & Moore, C. (2000) "Is prospective altruist-detection an evolved solution to the adaptive problem of subtle cheating in cooperative ventures? Supportive evidence using the Wason selection task," *Evolution and Human Behavior*, 21 (1), pp. 25–37
- Chang, A. & Wilson, M. (2004) "Recalling emotional experiences affects performance on reasoning problems," *Evolution and Human Behavior*, 25 (4), pp. 267–276
- Cheng, P. W., & Holyoak, K. J. (1985) "Pragmatic reasoning schemas," *Cognitive Psychology*, 17 (4), pp. 391–416
- Cheng, P. W., Holyoak, K. J., Nisbett, R. E., & Oliver, L. M. (1986) "Pragmatic versus syntactic approaches to training deductive reasoning," *Cognitive Psychology*, 18 (3), pp. 293–328
- Cheng, P. W., & Holyoak, K. J. (1989) "On the natural selection of reasoning theories," *Cognition*, 33 (3), pp. 285–313
- Cosmides, L. (1989) "The logic of social exchange; Has natural Selection Shaped with Humans Reason? Studies with the Wason Selection Task," *Cognition*, 31, pp. 187–276
- Cosmides, L., Barrett, H. C. & Tooby, J. (2010) "Adaptive specializations, social exchange, and the evolution of human intelligence," *Proceedings of the National Academy of Sciences*, 107 (Supplement 2), pp. 9007–9014
- Cosmides, L., & Tooby, J. (1989) "Evolutionary psychology and the generation of culture Part II," *Ethology and Sociobiology*, 10, pp. 51–97
- Cosmides, L., & Tooby, J. (1992) "Cognitive adaptations for social exchange," in "The adapted mind: Evolutionary psychology and the generation of culture," pp. 163–228, Oxford University Press
- Cosmides, L. and Tooby, J. (2005) "Social Exchange: The Evolutionary Design of a Neurocognitive System," in Gazzaniga, M.S. ed. The New Cognitive Neuroscience III, Cambridge, MA: pp. 1295–1308, MIT Press
- Cosmides, L., Tooby, J., Fiddick, L., & Bryant, G. A. (2005) "Detecting cheaters," *Trends in Cognitive Sciences*, 9, pp. 505–506
- Cummins, D. D. (1996) "Evidence of deontic reasoning in 3-and 4-year-old children," *Memory & Cognition*, 24 (6), pp. 823–829
- Fiddick, L. (2004) "Domains of deontic reasoning: Resolving the discrepancy between the cognitive and moral reasoning literature," *Quarterly Journal of Experimental Psychology*, Section A, 57 (3), pp. 447–474
- Fiddick L., (2011) "From social exchange to safety: Further explorations of adaptive reasoning" in Wang, X. T., (ed) "Thus spake evolutionary psychologists," pp. 136–145, Peking University Press (「ヒトの理屈はいつだって論理的? —交換と安全の論理」(2018) 平石監修・翻訳他『進化心理学を学びたいあなたへ: パイオニアからのメッセージ』pp. 147–156, 東京大学出版会)
- Fiddick, L., Spampianto, M V., & Graman, J. (2005) "Social contracts and precautions activate

- different neurological systems: An fMRI investigation of deontic reasoning,” *NeuroImage*, 28, pp. 778–786
- Fiddick, L., & Erlich, N. (2010) “Giving it all away: Altruism and answers to the Wason selection task,” *Evolution and Human Behavior*, 31 (2), pp. 131–140
- Gazzaniga, M. S. (2008) “Human: The Science Behind What Makes Us Unique,” Ecco; 1st Edition  
(『人間らしさとはなにか? : 人間のユニークさを明かす科学の最前線』(2010) 柴田裕之訳, インターシフト)
- Gigerentzer G., & Hug, H. (1995) “Domain-specific reason: Social contracts, cheating and perspective change,” *Cognition*, 43, pp. 127–171
- Griggs, R. A., & Cox, J. R. (1982) “The elusive thematic-materials effect in Wason’s selection task,” *British Journal of Psychology*, 73 (3), pp. 407–420
- Harris, P. L., Núñez, M. & Brett, C. (2001) “Let’s swap: Early understanding of social exchange by British and Nepali children,” *Memory & Cognition*, 29 (5), pp. 757–764
- 長谷川寿一・平石界 (2000) 『進化心理学から見た心の発生』「心の比較認知科学 (ことばと心の発達) 8章」渡辺茂編, 小嶋祥三・鹿取廣人監修, ミネルヴァ書房
- Johnson-Laird, P. N., Legrenzi, P., & Legrenzi, M. S. (1972) “Reasoning and a sense of reality,” *British Journal of Psychology*, 63 (3), pp. 395–400
- Kenrick, D. K., & Griskevicius, V. (2013) “The rational animal: how evolution made us smarter than we think,” Basic Books (『きみの脳はなぜ「愚かな選択」をしてしまうのか〈意思決定の進化論〉』(2015) 熊谷淳子訳, 講談社)
- 小林佳世子 (出版予定) 『最後通牒ゲームの謎～進化心理学からみた行動ゲーム理論入門：ヒトは本当に不合理か?』
- Manktelow, K. I., & Evans, J. S B. (1979) “Facilitation of reasoning by realism: Effect or non-effect?,” *British Journal of Psychology*, 70, pp. 477–488
- Newstead, S., & Evans, J. S. B. (Eds.). (2013) “Perspectives on thinking and reasoning: Essays in Honour of Peter Wason,” Psychology Press
- Oda, R., K. Hiraishi, A. & Matsumoto-Oda, A. (2006) “Does an altruist-detection cognitive mechanism function independently of a cheater-detection cognitive mechanism? Studies using Wason selection tasks,” *Evolution and Human Behavior*, 27 (5), pp. 366–380
- 小田亮 (2016) 『道徳的行動の進化的背景』「モラルサイコロジー：心と行動から探る倫理学」太田紘史 (編・著) 春秋社
- Pinker, S. (1997) “How the mind works,” W. W. Norton & Company (『心の仕組み 上・下』2013, 椋田直子 (訳), 筑摩書房)
- Pooley, A. J., & Fiddick, L. (2010) “Social referencing “Mr. Yuk”: The use of emotion in a poison prevention program,” *Journal of Pediatric Psychology*, 35 (4), pp. 327–339
- Stone, V. E., Cosmides, L., Tooby, J., Kroll, N., & Knight, R. T. (2002) “Selective impairment of reasoning about social exchange in a patient with bilateral limbic system damage,” *Proceedings of the National Academy of Sciences*, 99, pp. 11531–11536
- Sugiyama, L. S., Tooby, J., & Cosmides, L. (2002) “Cross-cultural evidence of cognitive adaptations for social exchange among the Shiwiar of Ecuadorian Amazonia,” *Proceedings of the National Academy of Sciences*, 99 (17), pp. 11537–11542

#### 4枚カード問題からわかること

- Tooby, J., & Cosmides, L. (1992) "The psychological foundations of culture. in The adapted mind: Evolutionary psychology and the generation of culture," J. H. Barkow, L. Cosmides, & J. Tooby (eds), pp.137-162, Oxford University Press
- Trivers, R. L. (1971) "The evolution of reciprocal altruism," *Quarterly Review of Biology*, 46 (1), pp. 35-57
- Wason, P. C. (1968) "Reasoning about a rule," *Quarterly Journal of Experimental Psychology*, 20 (3), 273-281.
- Wason, P. C., & Shapiro, D. (1971) "Natural and contrived experience in a reasoning problem," *Quarterly Journal of Experimental Psychology*, 23 (1), pp. 63-71

なお本研究は、2018年度南山大学パツへ研究奨励金I-A-2の助成を受けています。

『南山経済研究』掲載論文の中で示された内容や意見は、南山大学および南山大学経済学会の公式見解を示すものではありません。また、論文に対するご意見・ご質問や、掲載ファイルに関するお問い合わせは、執筆者までお寄せ下さい。

(小林 佳世子, E-mail: kkayoko@ic.nanzan-u.ac.jp)