

# 情報のコピーコストと情報カスケード転移

## Copy cost and information cascade transition

守 真太郎  
Shintaro Mori

北里大学理学部物理学科  
School of Science, Kitasato Univ.  
mori@sci.kitasato-u.ac.jp

久門 正人  
Masato Hisakado

スタンダード&プアーズ  
Standard and Poor's  
masato\_hisakado@standardandpoors.com

高橋 泰城  
Taiki Takahashi

北海道大学文学研究科  
Department of Behavioral Science, Hokkaido University  
taikitakahashi@gmail.com

キーワード: 情報カスケード、相転移、コピー、コスト、社会的学習

### 概要

情報を持たない個体は、他個体の示す情報をコピーすることが最適戦略である。こうした個体をハーダーと呼び、集団内のハーダーの比率がある閾値を越えると情報カスケード転移が起こり、集団のほとんどの個体が間違った選択をすることがある。では、ハーダーのコピーに対してコストをかけると、そうした状況を防ぐことは可能なのか？我々は、60名の被験者の列に対する二択のクイズを用いた情報カスケードの実験を行った。ここでは、ハーダーとは正解を知らない被験者である。被験者は、自分の知識だけで回答する以外に、(1)各選択肢を選んだ回答者数、(2)正答の場合のリターンの倍率を表すオッズ、の二つのケースで回答する。オッズは、回答者数の逆数に比例し、多数派の選択肢のコピーに対する一種のコストとみなせる。(1)では、ハーダーは多数派の選択肢を選ぶ傾向が強くデジタル的(=多数派を必ず選択)である。(2)では、各選択肢の人数が拮抗しオッズが2倍前後の場合、ハーダーは各選択肢の人数に比例した確率で選択しアナログ的である。これは、ゲーム論的に最適な混合戦略をハーダーが集団でほぼ実現していることを意味する。ハーダーの振る舞いをもとに、系をシミュレートする確率モデルを構成し解析した。その結果、どちらの場合での正解を知らない被験者の比率がある閾値を越えると相転移を起し、多数派の選択が有限の確率で誤る相に変化する。その閾値は、前者で80.0%、後者で94.8%と大きく異なり、この差がシステム全体とハーダー個体のパフォーマンスに大きく影響する。

## 1. 社会的学習と情報カスケード

社会的な学習を通じて生存に必要な情報を獲得してきたのはヒトに限った話ではない。ここで、社会的な学習とは、他の個体との相互作用による知識獲得のことであり、自分よりうまく行動している個体があればその行動を真似る個体レベルのものから、うまくいく行動を規範として集団全体の知識として共有する文化的なものまで、様々なレベルで実装されている。確かに、社会的な学習は、各個体がトライ&エラーで環境と相互作用することにより知識を獲得する非社会的な学習に比べ、個体の払うコストやリスクは低い。しかし、他個体の振る舞いなどをから学習して得た知識は、他個体が間違うリスク、知識が古くて間違うリスクなど、まったくリスクがないわけではない。また、集団の個体すべてが社会的学習のみを行い、環境からの情報がまったく入らなくなると、環境の変化に対応できなくなる[Rendell2010]。社会的な学習はこのようなりスクを伴うが、ヒトに限らず動物世界に

おいても見られる普遍的な情報収集の方法であり、自然界での長い適応過程の結果として成立したものであろう[Rendell2011]。

この社会的学習による集団行動の人間社会における顕著な例が情報カスケードと呼ばれるものである。情報カスケードは、集団の一人一人が順番に選択を行い、他個体の情報を参照できず、選択しか参照できない場合に起きるハード(群れ)現象である[Bikhchandani1992, Anderson1997, Kubler2004]。例えば、選択肢AとBの二択のクイズに次々と人が回答する状況を考える。回答者は、正解を知っているか、正解を知らないかに分けられる。正解を知る回答者はもちろん正解を選び、他の回答者の回答には影響されないで独立投票者と呼ぶ。ここで、回答者を投票者と言っているのは、二択の問題での選択が選択肢への投票とみなせるからである。一方、正解を知らない回答者は、他に何も情報がなければ、ランダムに選択するしかない。けれど、自分より前に回答した回答者の情報があればどうするだろうか？例えば、各選択肢

を選んだ人数が分かっているならどうするのか？その場合、多数派の選択肢を選ぶのが合理的であり、実際の選択も多数派を選ぶ傾向が強い。正解を知らない回答者は、多数派の選択にハードしコピーする傾向があるので、一般にハーダーと呼ぶ。また、必ず多数派を選ぶハーダーをデジタルハーダーと呼ぶ [Hisakado2011]。ハーダーは単に多数派をコピーするだけなので、情報量はゼロである。ハーダーの問題は、多数派の選択がたまたま間違っている場合に、その間違いを増幅してしまうことである。それでも独立投票者の比率が高ければ問題はない。ハーダーが間違った情報を蓄積しすぎる前に独立投票者が系の状態を修正し、集団全体としては正しい選択肢を選ぶことができる。けれど、その比率が非常に低い場合、ハーダーによる間違った情報の蓄積を修正できず、情報カスケード転移が起きる [Mori2012, Hisakado2012]。

可能ならハーダーには情報をあたえないほうがいい。この場合、ハーダーはランダムに選択し、間違った選択、正しい選択がキャンセルし、独立投票者の選択だけが残る。独立投票者の比率が低くても、十分多数の回答者が選択すれば、最終的には正しい選択肢が多数派に選ばれることになる。しかし、ハーダー個体の観点からこの方法の採用は難しい。では、ハーダーが多数派を選択するインセンティブを減らすにはどうすればいいのか。そこで考えられるのが、競馬などで用いられるオッズである。競馬の単勝馬券はレースの1着の馬をあてるゲームである。強い馬は人気があり得票率が高く、弱い馬は人気がなく得票率が低い。けれど、人々の選択は最も強い馬に集中することはなく、出走馬すべてに分散する。その理由は、競馬でのリターンは得票率の逆数に比例するオッズで決まるからである。勝率とオッズのバランスを考えて人々は投票し、結果、馬の得票率と勝率はほとんど一致する。この事実を競馬市場の効率性と呼ぶ [Ziemba2008]。

では、上記の二択のクイズの状況で、各選択肢の人数の代わりにオッズを与えたとき、ハーダーはどう選ぶだろうか？正答の場合のリターンはオッズに比例するとする。各選択肢の選択者数の情報を与え、正答でのリターンが定数の場合、ハーダーは多数派をコピーするのが最適戦略であった。一方、正答に対するリターンがオッズで決まる場合、多数派の選択肢のオッズが小さいため、自分のもつ知識によっては少数派を選ぶインセンティブが働く。つまり、オッズは多数派の選択肢のコピーに対するコストとして機能する。ゲーム論を用いてオッズを与えたときのハーダーの最適戦略を導いてみる。二つの選択肢 A と B のオッズが  $O_A, O_B$  とする。A が正解の確率  $\alpha$ 、B が正解の確率  $1 - \alpha$  はハーダーにとって不明である。A を選択する確率  $P_A$ 、B を選択する確率  $P_B$  の混合戦略のリターンは、 $\alpha P_A O_A + (1 - \alpha) P_B O_B$  となる。ハーダーには  $\alpha$  が不明なので、最適な混合戦略は  $\alpha$  依存性をキャンセルする  $P_A O_A = P_B O_B$  を満たすものとなる。キャンセルしない場合、 $\alpha$  の値によってはリターンがかなり

悪くなるため、最小のリターンを最大にするマクシミン戦略では、 $P_A O_A = P_B O_B$  となる。 $O_A, O_B$  は、選択肢 A, B の得票率  $x_A, x_B$  の逆数なので、 $P_A = x_A, P_B = x_B$  という関係が成立する。このような、選択の確率が選択肢の得票率に比例するハーダーをアナログハーダーと呼ぶ [Mori2010]。アナログハーダーの存在する系は、何も情報があたえられなかったハーダーの場合と同じく、ハーダーの寄与はキャンセルする。ハーダーの比率がある閾値を越えると、得票率の収束がゆっくりになる相への相転移が起きるが、独立投票者の情報をかき乱すことはなく、情報カスケード転移は起きない。独立投票者が正解を知り、かつ正の比率で存在するなら、投票数  $T$  が無限大の極限で正解の選択肢の得票率は 100% に収束する [Hisakado2010]。

本研究では、オッズの情報をあたえたとき、ハーダーがどのような選択をするのか実験により明らかにする。それをもとに、ハーダーの比率の変化に対し、システム全体のパフォーマンス、ハーダーの個体としてのパフォーマンスを解析し、選択者数の情報を与えた場合と比較する。論文の構成は以下の通りである。セクション 2 では、実験の設定を説明する。我々の情報カスケード実験では、二択のクイズを用いた。被験者は、自分の知識のみで答える場合に加え、(1) 各選択肢を選んだ人数の情報がある場合（コピーコストゼロ）、(2) 各選択肢のオッズが与えられている場合（コピーコストあり）の 2 つのケースで回答する。次のセクション 3 では、実験結果について述べる。コピーコストがゼロの場合、ハーダーは多数派を選ぶデジタル性が強い。コピーにコストが存在する場合、ハーダーはアナログ的に多数派をコピーし、最適戦略を集団で実現する。セクション 4 では、セクション 3 で導いたハーダーのコピー確率をもとに、系をシミュレートする確率モデルを導入し、情報カスケード転移を議論する。また、ハーダーの比率を変化させたときの多数派が間違える確率、ハーダーの正答率に注目し、コピーコストゼロとありのシステムの優劣を議論する。セクション 5 は、まとめと今後の課題について触れる。

## 2. 二択のクイズによる情報カスケード実験

A か B のどちらが正しいかという二択のクイズの正解を被験者に選択してもらった実験を行った。この実験では、正解を知る回答者（自己情報が完全）と、正解を知らず可能なら他の回答者の回答を参考にすることが合理的な回答者（自己情報がゼロまたは不完全）に分けられる。前者を独立投票者、後者は、他の回答者の多数派の選択にハードしコピーする可能性があるハーダーと呼ぶ。被験者が独立投票者かハーダーかは、もちろん問題ごとに異なる。問題が難しい場合、回答を知る被験者は少なくなり、独立投票者が減って、ハーダーが増えることになる。被験者は最初に他の被験者の回答に関する情報を

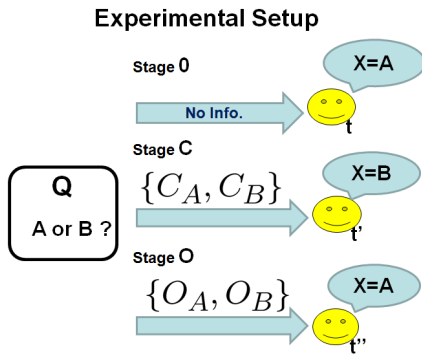


図 1 実験の概念図。二択の問題に対し、被験者は Stage 0 では自分の知識で回答 ( $r=0$ )。  $t$  は、その問題に回答した順番である。 Stage C では、自分より以前に回答した被験者のうち、A が何人、B が何人という分布  $\{C_A, C_B\}$  を参考に回答 ( $r=C$ )。被験者は  $t'$  番目に回答。同じ問題  $i$  であっても、各ステージで実験サーバーがランダムに問題を割り振るので被験者の回答の順番はステージ間で一般に異なる。 Stage O では、各選択肢の倍率  $\{O_A, O_B\}$  を参考に回答 ( $r=O$ )。被験者は  $t''$  番目にその問題に回答。倍率は、Stage O での A と B の選択者数から決定。

まったく得ずに回答する。この情報から、各問題についてハーダーの比率  $p$  を推定することができる。つまり、2 択のクイズの情報カスケード実験ではクイズの難易度を変えることによりハーダーの比率  $p$  をコントロール可能である [Mori2012]。

図 1 が実験の概念図である。被験者は、まず Stage 0 で自分の知識だけで回答する。問題番号を  $i$  で表すとし、問題  $i$  に対して  $t$  番目に回答したとき、回答が正解なら  $X(i, t|0) = 1$ 、間違いなら  $X(i, t|0) = 0$  とする。次の Stage C では、自分より前に回答した被験者すべての回答情報も参考にして回答する。  $t'$  番目に回答するなら、その回答を  $X(i, t'|C)$  と記す。被験者が得る情報は、  $s < t'$  の被験者の回答の時系列情報  $\{X(i, s|C)\}$  から、それぞれの選択肢の選択者数  $C_1(i, t' - 1|C) = \sum_{s=1}^{t'-1} X(i, s|C)$  と  $C_0(i, \infty, t' - 1|C) = (t' - 1) - C_1(i, t' - 1|C)$  に要約したものである。もちろん、選択肢 A、B の選択数  $\{C_A, C_B\}$  と正解、不正解の選択数  $\{C_1(i, t' - 1|C), C_0(i, t' - 1|C)\}$  の対応関係は、正解を知らない被験者には分からない。最後の Stage O では、二択の選択肢に対して、人気のある選択肢のオッズは小さく、人気のない選択肢はオッズを大きく設定した情報を参考に回答する。回答の順番が  $t''$  番目なら、その回答を  $X(i, t''|O)$  と書く。被験者の参考にするオッズ  $\{O_A, O_B\}$  は、Stage O での選択肢 A、B の選択者数  $\{C_A, C_B\}$  から、  $O_A = \frac{t''}{C_A + 1}$ 、  $O_B = \frac{t''}{C_B + 1}$  とする。これは、その選択肢を選んだ人数で  $t''$  を割ったものであり、競馬のオッズと同様のシステムである。競馬のオッズは投票を締め切るまで確定しないのに対し、この実験では回答時点での倍率で報酬を計算する。実験では、倍率の最小単位を 0.1 とし、少数 2 桁は四捨五入とした。

[Mori2012] との違いは第 1 に Stage O の倍率の場合の

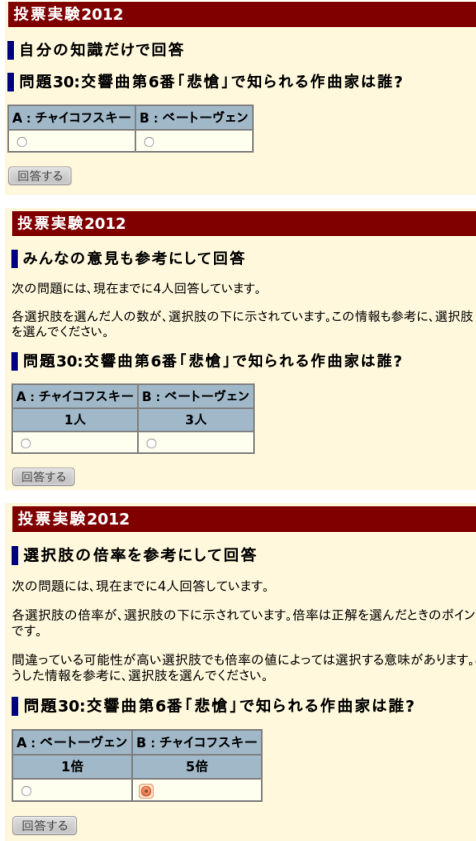


図 2 二択のクイズの情報カスケード実験での回答画面。被験者はまず Stage 0 で「自分の知識だけで回答」で回答する。次に Stage C で「みんなの意見も参考にして回答」で回答する。最後に、Stage O で「選択肢の倍率を参考にして回答」で回答する。

回答である。また、[Mori2012] では参照する人数を徐々に増やすことでの選択の変化を追跡したが、今回の Stage 0 から Stage C では参照人数が 0 から過去の回答者全員と一度に増やしている。このことにより、ハーダーのデジタル的な振る舞いがさらに顕著になると期待される。[Mori2012] では、一つの問題に対しさまざまな参照人数での情報をもとに回答してから次の問題に回答するとしたが、今回の実験では Stage 0, C, O と分け、各条件での回答間の相関を減らす工夫を行っている。つまり、Stage 0 で  $r=0$  の条件ですべての問題に回答し、Stage C に移って  $r=C$  の条件ですべてに回答し、最後に Stage O で倍率 ( $r=O$ ) をもとにすべての問題に回答する、とした。各ステージですべての問題に答えている間に、前ステージでの自分の回答と現ステージでの回答をどう組み合わせるか、といった工夫が入り込むのを極力排除している。特に、Stage C と Stage O の間は、Stage C で多数派に賭けたから、Stage O では倍率の大きい方を選ぶといった戦略が可能なのでこうしている。各ステージで実験サーバーはランダムに問題を被験者に割り振るが、その際他の回答者が回答している問題は除外する。

クイズで用いる問題は [Mori2012] と同じ 120 問。実験は 2012 年 6 月 13、14 日に北大社会科学実験センター

で実施し、集団実験室のパーティションに被験者に入ってもらい、備え付けのノートPCのWEBブラウザで実験サーバーに接続し、二択の問題に回答。他の回答者の選択状況や倍率はPC画面に表示。実験サーバーのプログラムはPHPで記述。実験両日とも5セッション実施し、1セッションあたり10名から13名の被験者が実験に参加。13日には計57名、14日には計63名の被験者が実験に参加し、平均60名でとなった( $T_{avg} = 60$ )。13日の被験者集団をグループA、14日の集団をグループBとし、各グループに対し同じ実験を実施した。被験者は集団実験室に入室後、パーティションに入る。実験者はスライドを用いて、実験の趣旨、実際の二択の問題での回答の様子、報酬について説明。報酬は、参加費として500円、通常の正答に対して2円、選択肢に倍率Oがある場合は、正答の場合O円とした。選択肢A、Bの回答が割れて半々のとき、倍率はほぼ2倍となるので、その正解は他の場合での正解と同じく2円となる。多数派の選択肢は倍率が2倍より小さくなり正解でも報酬は2円を切る。少数派の選択肢は状況によって非常に大きな倍率になりうる。実験時間は説明時間も含め講義の1コマを限度としたが、大体の学生が60分程度で回答を終えた。120問の問題を前半60問、後半60問に分け、その間に5分の休憩をもうけた。図2は、各ステージでの被験者のブラウザに表示される情報を示したものである。

### 3. 実験結果：マイクロとマクロ

約60名の被験者が120問の二択に、ステージ0,C,Oの3回回答している。それをA、Bの2グループで実施し、240問に対する3ステージでの時系列 $\{X(i, t|r)\}$ が得られたことになる。ここで、 $i \in \{1, 2, \dots, 240\}$ は問題をラベルし120問2グループの240問に対し、Aグループを最初の120、Bグループを後半の120とする。 $r$ は、各ステージをラベルし、 $r \in \{0, C, O\}$ 。回答の順番を示す $t$ の範囲は、Aグループでは1から57、Bグループでは1から63で、 $t \in \{1, 2, \dots, T_i\}$ で表す。

#### 3.1 正答率の分布のマクロな変化

興味ある量は正答率 $Z(i, t|r)$ であり、 $\frac{1}{t} \sum_{s=1}^t X(i, s|r)$ と定義する。まず、最終正答率 $Z(i, T_i|r)$ の分布に着目する。表1に、各ステージ $r$ での正答率の分布 $N(r)$ を示した。3列目に $r=0$ 、5列目に $r=C$ 、6列目に $r=O$ の分布を示している。まず、 $r=0$ では、正答率の中心が60%の瓶No.7にピークをもつ分布となっているが、 $r=C(O)$ の正答率の分布は10(30)%とほぼ100(90)%の2ヶ所にピークを持つことが分かる。この正答率の分布のマクロな変化は、ハーダーの選択の変化により引き起こされている。ここで、クイズ $i$ に対するハーダーの比率 $p(i)$ の推定法について述べる。独立投票者は正解を知り、その正答率は100%である。一方、ハーダーは正解

表1 最終正答率の分布とハーダーの平均比率。 $r \in \{0, C, O\}$ に対する240個のサンプル $\{Z(i, T_i|r)\}$ を2列目に示した条件下で11個の瓶に分類。 $r=0$ での正答率の分布 $N(0)$ を3列目に、各瓶内のハーダーの比率 $p(i)$ の平均値 $p_{avg}$ を4列目に、 $r=C$ と $r=O$ での正答率の分布 $N(C), N(O)$ を5、6列目に示した。

No.	$Z(i, T_i r)$	$N(0)$	$p_{avg}[\%]$	$N(C)$	$N(O)$
1	< 0.05	0	NA	5	0
2	0.05 ~ 0.15	3	NA	33	7
3	0.15 ~ 0.25	5	NA	28	25
4	0.25 ~ 0.35	18	NA	9	30
5	0.35 ~ 0.45	35	NA	5	13
6	0.45 ~ 0.55	38	97.5	5	13
7	0.55 ~ 0.65	57	78.3	5	14
8	0.65 ~ 0.75	29	60.6	7	19
9	0.75 ~ 0.85	41	40.6	17	44
10	0.85 ~ 0.95	11	21.3	57	62
11	$\geq 0.95$	3	6.6	69	13
Total		240		240	240

を知らず、正答率は50%とする。ハーダーの比率 $p$ 、独立投票者の比率 $1-p$ での正答率 $Z$ の期待値 $E(Z)$ は $E(Z) = (1-p) + \frac{1}{2}p = 1 - \frac{1}{2}p$ となる。この関係式を解くことで、正答率 $Z$ のクイズでのハーダーの比率 $p$ の最尤推定値は $p = 2(1-Z)$ となる。クイズ $i$ での $p(i)$ を $p(i) = 2(1 - Z(i, T_i|0))$ と推定する。表1の4列目に、各瓶のクイズに対する $p(i)$ の平均値 $p_{avg}$ を示した。クイズの正答率が50%に近づくと、ハーダーの比率は100%に近づく。表1を見ると $Z(i, T_i|0)$ が、0.5よりもずっと小さな値をとっている問題がある。二択のクイズの場合、選択肢にバイアスがなければ、正答率の最小値は50%である。50%より極端に小さな正答率は、その問題の選択肢にバイアスが存在することを意味する。システムサイズが有限であることからくるゆらぎ $\frac{1}{2\sqrt{T_i}} \simeq 6.5\%$ を考慮して、 $Z(i, T_i|0) < 45\%$ のクイズ $i$ のデータは以下の解析では用いないとする。240問のうち、61問のサンプルを除外し、179問分のデータが残った。また、 $Z(i, T_i|0) < 50\%$ では $p(i)$ は100%を越えるが、これもゆらぎから来るエラーとし $p(i) = 100\%$ とする。 $p(i)$ の推定誤差は、正答率の推定の誤差の2倍となり、13%程度である。

ハーダーの回答の変化の様子を見るため、x軸に $r=0$ の正答率、y軸に $r=C$ と $O$ の正答率をプロットしたものが図3である。x軸に垂直な線は、表1での瓶の境界線を表している。ハーダーが選択を変えないときに、散布図は対角線に乗るはずだが、図はそれを否定する。x軸に示した $Z(i, T_i|0)$ が65%以上の場合、ハーダーの比率 $p$ は70%以下であり、このとき、 $Z(i, T_i|C)$ はほぼ100%近くに収束している。一方、 $Z(i, T_i|0)$ が65%未満、45%以上の領域では、 $p$ が70%から100%となり、 $Z(i, T_i|C)$ は90%以上のところと10%から20%のところの二ヶ所に収束しているように見える。これが、表1

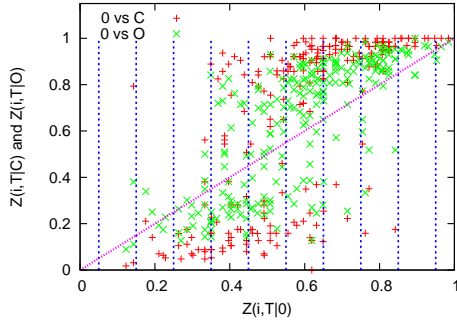


図 3 正答率  $Z(i, T_i|r)$  の散布図。x 軸に  $Z(i, T_i|0)$ 、y 軸に  $Z(i, T_i|C)$  (赤+) と  $Z(i, T_i|O)$  (緑×) をプロットした。x 軸に垂直な青点線は、表 1 での瓶の境界線を表している。対角線は、他の被験者の選択情報の影響がない場合の参照のためのもの。

での 2 つのピークに対応するものであり、 $p$  が高いところで正答率が 2 つのピークを持つことが分かる。一方、 $r = O$  の場合、 $Z(i, T_i|O)$  の分布は  $Z(i, T_i|0)$  が 65% 以上の場合、 $r = C$  の場合と同様に高い正答率に集中するが、その値は 80% から 90% と  $r = \infty$  に比べて低い。また、 $Z(i, T_i|0)$  が 65% 未満、45% 以上の領域では 60% から 90% と 20% から 40% 前後の広い領域に分布する。 $r = \infty$  の分布より正答率は対角線側に寄っていて収束は明確ではない。

### 3.2 ハーダーのコピー確率

ハーダーの比率  $p$  と  $r = 0$  と  $r = C, O$  での正答率の分布の比較から、ハーダーが他の回答者の回答情報をもとに選択を変えていると推測される。 $r = C$  の場合は選択の変化は一目瞭然で、正答率も二つのピークに収束しているが、 $r = O$  の場合は  $r = C$  と比較して、収束が弱く変化も小さい。このように、コストの有無でハーダーのコピーに明確な違いがある。そこで、ハーダーが他の回答者の情報をどのようにコピーするのかを調べ、コピーコストの有無の影響を明かにする。まず、問題  $i$  に対する選択の時系列  $\{X(i, t|r)\}$  から、正解を知る独立投票者の寄与を差し引いて、ハーダーの情報を求めたい。独立投票者の比率は、 $1 - p(i)$  であり、 $X(i, t|r)$  から、 $1 - p(i)$  を引けば、比率  $p(i)$  のハーダーの寄与となる。問題  $i$  に対する  $t + 1$  番目の被験者が参照した情報は  $\{C_1(i, t|r), C_0(i, t|r)\}$  である。被験者がハーダーの場合、その正答率は  $t$  と  $C_1(i, t|r)$  の値  $n_1 \in \{0, 1, \dots, t\}$  のみに依存すると考えられ、 $q_h(t, n_1|r)$  と書くことができる。すると、被験者の正答率  $q(t, n_1|r)$  は、

$$q(t, n_1|r) = (1 - p(i)) \cdot 1 + p(i) \cdot q_h(t, n_1|r) \quad (1)$$

となる。ハーダーの正答率  $q_h(t, n_1|r)$  を次の条件付き平均値として評価する [Mori2012]。

$$\sum_i \frac{\{(X(i, t+1|r) - (1 - p(i)))/p(i)\}}{\#\{C_1(i, t|r) = n_1\}} \delta_{C_1(i, t|r), n_1} \quad (2)$$

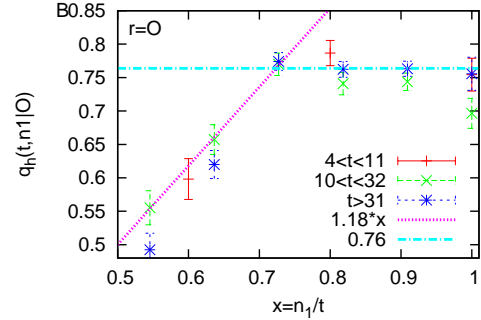
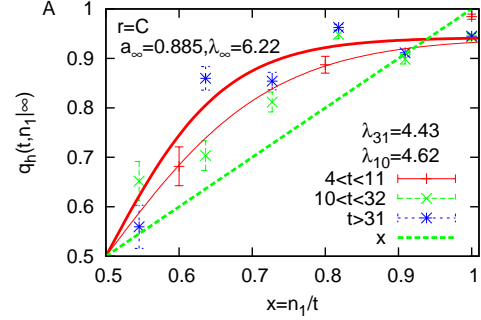


図 4  $r = C, O$  でのハーダーのコピー確率  $q_h(t, n_1)$  を  $n_1/t$  に対してプロットした。(A)  $r = C$ , (B)  $r = O$ 。データは、 $5 \leq t \leq 10$ 、 $11 \leq t \leq 31$  と  $t \geq 32$  の被験者列の前半、中盤、後半に分けている。また、各場合、 $n_1/t$  に対するデータは、整数  $k$  に対する  $k/5, k/11, k/11$  にもっとも近い点のデータ全体でまとめている。参考のため、 $q_h(t, n_1|C) = n_1/t$  と (A) では 3 のフィットの結果も示している。

平均値の計算での和  $\sum_i$  は  $C_1(i, t|r) = n_1$  の問題  $i$  に限ることをクロネッカーのデルタ  $\delta_{C_1(i, t|r), n_1}$  で示している。分母の  $\#\{C_1(i, t|r) = n_1\}$  はその条件を満たす問題数を表す。 $q_h(t, n_1|r)$  は、過去  $t$  人中  $n_1$  人が正解のときのハーダーの正解率であるが、ハーダーは選択肢のどちらが正解かは不明であり、 $q_h(t, n_1) = 1 - q_h(t, t - n_1)$  という対称性を持つ。平均の計算ではこの対称性を仮定した。図 4 は  $q_h(t, n_1)$  を  $n_1/t$  に対してプロットしたものである。対称性から  $n_1/t \geq 1/2$  のみ表示している。

まず、 $r = C$  の場合、 $q_h(t, n_1)$  は  $n_1/t = 1/2$  での傾きが 2 ~ 2.8 程度の  $n_1/t$  の単調増加関数となる。被験者列の前半、中盤では傾きは約 2 で、後半になると約 2.8 になり、参照する被験者数  $t$  が多いとコピーの応答感度が高まり、よりデジタル的になることが分かる。コピー確率は、 $n_1/t = 1$  の近くでも 1 には収束せず、95% 程度となる。これらの性質は、過去の我々の実験結果とも一致している [Mori2012]。以前の結果との差は、予想されたとおり、コピーの応答感度が高くなることである。次に、 $r = O$  の場合のコピー確率の振る舞いを見てみる。 $r = C$  の場合と異なり、 $n_1/t = 0.5$  での傾きは被験者列の前半、中盤、後半のいずれでも 1.0 ~ 1.2 となっている。 $q_h \geq 0.76$  では、 $n_1/t$  依存性は消え、 $q_h(t, n_1)$  はほぼ定数となる。つまり、オッズが 2 倍をはさんで、 $4/3 (\approx 1/0.76)$  倍から  $4 (\approx 1/(1.0 - 0.76))$  倍の領域では、ハーダーはアナログ

ハーダー ( $q_h(x|O) = x$ ) として振る舞うことが分かる。その範囲外では倍率にはほとんど応答しなくなり、高い倍率では約 1/8 が少数派を、約 7/8 が多数派を選ぶ。

$r = C$  の場合、 $q_h(t, n_1|C)$  に対し、[Mori2012] と同様に次の関数形を仮定する。

$$q_h(n_1/t|C) = \frac{1}{2} (a \tanh(\lambda(n_1/t - 1/2)) + 1) \quad (3)$$

パラメータ  $a$  は、多数派に従う人の正味の比率、 $\lambda$  はコピー応答の感度を表している。また、 $x = 1/2$  でのコピー確率の傾きから集団での感度は  $a\lambda/2$  で評価される。このモデルとデータをフィットした結果が図 4A に示してある。まず、 $a, \lambda$  を  $t \geq 32$  のデータでフィットし、その結果を、 $a_\infty, \lambda_\infty$  と書く。 $a$  の値を  $a_\infty$  に固定し、 $5 \leq t \leq 10$  と  $11 \leq t \leq 31$  のデータでフィットし、その結果を  $\lambda_{10}, \lambda_{31}$  と書くことにする。 $a_\infty$  は 0.885 となり、正味約 90% が多数派にハードすることを意味する。一方、コピーに対する感度  $\lambda$  は、 $\lambda_\infty = 6.62$  で、参照人数が少ない場合、 $\lambda_{10} = 4.62, \lambda_{31} = 4.23$  となった。ただ、フィットの結果にはばらつきがあり、参照人数  $t$  の増加に対し  $\lambda$  はより大きな値をとる可能性がある。

$r = O$  の場合、 $q_h(t, n_1|O)$  に対し次の関数形を仮定する。

$$q_h(n_1/t|O) = \lambda_O(n_1/t - 1/2) + \frac{1}{2} \quad (4)$$

ただし、 $q_h$  の上限は 0.76、下限は 0.24 とする。 $t \geq 32$  のデータでフィットした結果、 $\lambda_O = 1.18$  となった。

#### 4. 投票モデルの熱力学極限とカスケード転移の閾値

ハーダーのコピー確率のモデルをもとに、系の確率モデルを導入しマクロな性質を解析する。ハーダーの比率  $p$  の被験者列を考え、 $t$  番目の被験者の回答を確率変数  $X(t|r) \in \{0, 1\}$  で表すとす。確率過程  $\{X(t|r)\}$  は、式 (1) の条件付き確率  $q(t-1, n_1|r)$  に従う。 $q_h(t, n_1|r)$  は  $t, n_1$  に  $n_1/t$  の依存性をするとしたことから、 $q(t, n_1|r)$  の  $t, n_1$  依存性も  $q(n_1/t|r)$  と書くことができる。 $q_h(n_1/t|r)$  におけるパラメータ ( $\lambda, a$ ) は、 $r = C$  の場合、 $a = a_\infty$  と固定し、 $2 \leq t \leq 11$  では、 $\lambda_{10}$ 、 $11 \leq t \leq 32$  では  $\lambda_{10}$  と  $\lambda_{31}$  の値を線形に補間したもの、 $t \geq 33$  では  $\lambda_\infty$  を用いるとする。 $r = O$  の場合、 $\lambda = \lambda_O$  とする。

確率変数  $C_1(t|r) \equiv \sum_{s=1}^t X(s|r)$  の従う確率分布関数に対するマスター方程式は、確率分布関数を  $P(t, n_1|r) \equiv \text{Prob}(C_1(t|r) = n_1)$  と書くと、

$$P(t+1, n_1|r) = q((n_1 - 1)/t|r)P(t, n_1 - 1|r) + (1 - q(n_1/t|r))P(t, n_1|r) \quad (5)$$

である。ここで、 $P(t, n_1|r)$  は、 $n_1 < 0, n_1 > t$  に対してゼロとする。興味があるのが、正答率  $Z(t|r) = C_1(t|r)/t$  の収束の振る舞いである。情報カスケード転移は、ハー

ダーの比率  $p$  の変化に対し、 $p < p_c$  では正答率が 100% 近くのある一点に確率 100% で収束する (One Peak) 相、 $p > p_c$  では、100% 近くと 50% 以下の低い正答率の 2 点にそれぞれ有限の確率で収束する (Two peaks) 相となる相転移である [Hisakado2011, Mori2012]。前者の相は、選択の列の途中で多数派が間違っているにもかかわらず、最終的に多数派の選択肢は正しい選択肢に収束するので、自己修正的、後者の相は、間違った選択肢に多数派の選択肢が収束する確率が有限なので、非自己修正的と呼ばれる [Goree2007]。どちらの相にあるかは、 $p$  に対し、自己無撞着 (Self-Consistent, SC) 方程式  $q(x|r) = x$  が、1 個の解を持つか、3 個の解を持つかで決まる [Hisakado2012]。SC 方程式の解のうち、対角線を  $q(x|r)$  が上から下へ交叉する解が正答率  $Z(t|r)$  の収束点に対応する。

SC 方程式の解の個数を決める上でもっとも重要なファクターはハーダーのコピー確率  $q_h(x|r)$  の  $x = 1/2$  での傾きである。傾きが 1 以下の場合、任意のハーダーの比率  $p$  に対し解の個数は一個であり、系は One Peak 相にある。この場合、ハーダーの投票は集積することなくキャンセルし、独立投票者の情報のみが集積し、 $p < 100\%$  なら多数派の選択は必ず正しい選択肢に収束する。例えば、 $q_h(x|r) = x$  のアナログハーダーモデルの場合である [Hisakado2010]。一方、傾きが 1 より大きい場合、情報カスケード転移が起きる。 $p_c < 100\%$  が存在し、 $p < p_c$  では 1 個の解、 $p > p_c$  では 3 個の解を持つ。 $p = p_c$  を境に、One Peak 相と Two Peaks 相の間の相転移が起きる。例えば、デジタルハーダーの場合、 $\theta(x)$  をヘビサイド関数として  $q_h(x|r) = \theta(x - 1/2)$  と書け、 $p_c = 50\%$  となる。このように、 $x = 1/2$  での傾きが 1 を越えるかどうかカスケード転移の存在を決定する。傾きが 1 以上でも 1 に近ければ  $p_c$  は大きく、情報カスケード転移は起こりにくい。これはハーダーのハードする傾向が弱いから、ほんのわずかの独立投票者でも多数派の間違いを修正できるからである。実験データをフィットして得られた  $q_h(x|r)$  は、 $r = C, O$  共に、 $x = 1/2$  での傾きが 1 を越える。つまり、ともにある  $p_c$  を境にして情報カスケード転移が起きる。けれど、 $r = C$  の場合、傾きは  $a_\infty \lambda_\infty / 2 = 2.75$  でデジタル性が強く  $p_c$  はデジタルハーダーの  $p_c = 50\%$  に近くなる。 $r = O$  では傾きは 1.18 でデジタル性が弱く、 $p_c$  が 100% に近くなる。

様々な  $p$  に対し、 $q(x|r)$  をプロットしたものが図 5 である。 $r = C$  の場合、SC 方程式は  $p < p_c(C) = 80.0\%$  では解が一個存在し、その  $x$  の値から正答率の収束値は 90% から 100% であることが分かる。 $p > p_c(C)$  では、解は 3 個存在し、そのうち正答率  $Z(t|C)$  の収束値に対応するのは、真ん中を除いた左右の 2 個である (図では  $p = 100\%$  が 3 個の解を持つ場合である)。右の解は、 $p < p_c(C)$  以下の解とほぼ同じ位置にあり、この点に収束するとき、被験者列のほぼ全員が正解を選ぶことになる。一方、左の解は、正答率が 10% から 25% に収束する

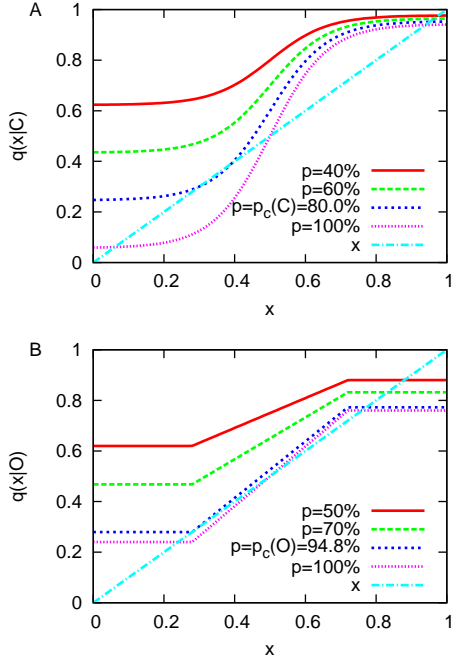


図 5  $q(x|r)$  を  $x$  に対してプロットした。(A) $r = C$ , (B) $r = O$ 。(A) $r = C$  の場合、 $p < p_c(C) = 80\%$  で SC 方程式は一個の解、 $p > p_c(C)$  で 3 個の解を持つ。(B) $r = O$  の場合、解の個数が 1 個と 3 個を分ける  $p$  の値  $p_c(O)$  は 94.8% である。

状態を表している。この結果から、 $p = p_c(C)$  で情報カスケード転移が起きていることが分かる。また、 $r = O$  の場合、SC 方程式は  $p < p_c(O) = 94.8\%$  の場合は解が一個、 $p > p_c(O)$  では解が 3 個存在する。 $p < p_c(O)$  の場合、80% から 90% に正答率は収束する。一方、 $p > p_c(O)$  の場合、右の解に収束すれば正答率は 75%、左の解に収束すれば、正答率は 25% となり、収束値は 50% に近くなる。この収束の様子の違いは、図 3 の正答率の分布の様子と定性的に一致している。

多数派が間違える確率とハーダーの比率  $p$  の関係を見るために、表 1 の  $r = 0$  の各瓶に対し、最終正答率が 50% 未満の比率を計算する。 $N(0)$  個のサンプルを  $k \in \{1, 2, \dots, N(0)\}$  でラベルし、次の平均値を計算する。

$$\sum_{k=1}^{N(0)} \frac{\theta(\frac{1}{2} - Z(k, T_k|r))}{N(0)} \quad (6)$$

図 6 は、その結果と確率モデルの計算結果をプロットしたものである。実験データは、 $r = C$  と  $r = O$  で顕著な差はないことを示す。これはシステムサイズ  $T$  が平均 60 と非常に小さいためである。 $T \rightarrow \infty$  の熱力学極限では、 $p < p_c(C)$  と  $p > p_c(C)$  では状況が全く異なる。例えば、 $p = 90\%$  の場合を考えてみる。 $T = 60$  では、 $r = C$  で約 40%、 $r = O$  で約 30% の確率で多数派が間違える。 $r = C$  の場合、 $p_c(C) = 80.0\%$  なので  $p = 90\%$  は Two-Peaks 相にあり、 $T = 10^3, 10^4$  とシステムサイズを大きくしてもその確率ほとんど変化がない。一方、 $r = O$  の場合、 $p_c(O) = 94.8\%$  なので、 $p = 90\%$  の系は

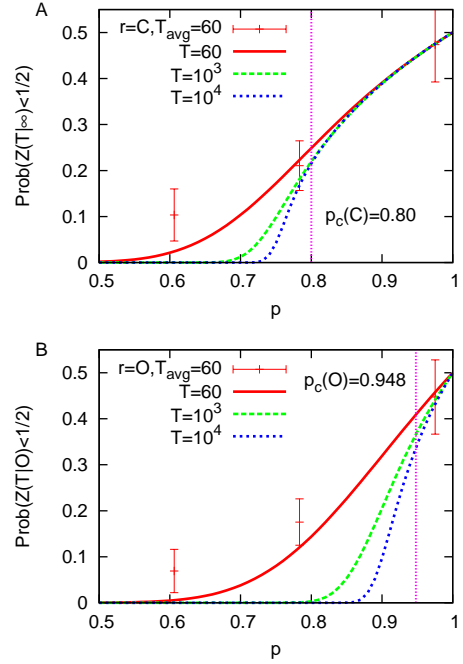


図 6  $p$  に対し、多数派が間違える確率 = 正答率が 50% 未満の確率  $\text{Prob}(Z(T|r) < 1/2)$  をプロットした。(A) $r = C$ , (B) $r = O$ 。

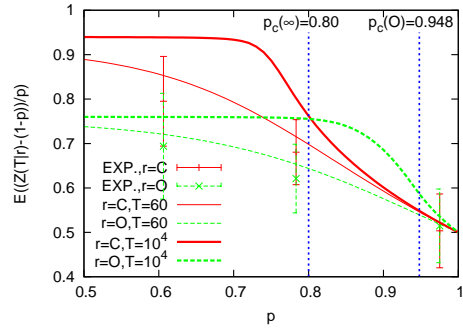


図 7  $p$  に対し、ハーダーの正答率  $E((Z(T|r) - (1-p))/p)$  をプロット。細い (太い) 線は  $T = 60 (10^4)$  の投票モデルの結果。

One Peak 相にある。 $T = 60$  では、多数派が間違える確率は約 30% だが、 $T$  を大きくしていくと  $T = 10^3$  で、約 20%、 $T = 10^4$  だと約 10% となる。一般に  $p < p_c$  では、 $T \rightarrow \infty$  の極限で多数派が間違える確率はゼロに収束する。 $r = C$ 、 $p > p_c(C)$  の場合、多数派の選択が間違っても独立投票者が修正できず、多数派が誤る確率が正となる。 $r = O$  の場合、 $p > p_c(C)$  でも、 $p < p_c(O)$  なら、独立投票者が多数派の間違った選択を修正し、多数派が間違える確率は  $T$  とともに急速に減少し「正しい集合知」[Surowiecki, Lorenz2011] が形成される。

ハーダーのパフォーマンスを計測してみる。最終正答率  $Z(i, T_i|r)$  のうち、独立投票者の寄与は  $(1-p(i))$  なのでそれを引いた残りをハーダーの比率  $p(i)$  で割ったものがハーダーの正答率である。ハーダーの正答率と  $p(i)$  の関係を見るために、表 1 の  $r = 0$  の各瓶に対

し、ハーダーの正答率を計算する。 $N(0)$  個のサンプルを  $k \in \{1, 2, \dots, N(0)\}$  でラベルし、次の平均値を計算する。

$$\sum_{k=1}^{N(0)} \frac{\{(Z(k, T_k|r) - (1-p(k)))/p(k)\}}{N(0)} \quad (7)$$

投票モデルでは、各  $p$  に対し、 $E((Z(T|r) - (1-p))/p)$  を計算する。その結果を図 7 に示した。 $T = 60$  程度のシステムサイズが小さい場合、 $r = C$  でのパフォーマンスが  $r = O$  より上回る。けれど、 $T = 10^4$  の場合、 $p$  が  $p_c(C)$  に近づくと、 $r = C$  でのパフォーマンスは急激に悪化し、 $p > p_c(C)$  でパフォーマンスは逆転する。 $r = O$  のパフォーマンスも  $p$  が  $p_c(O)$  に近づくと低下するが再度逆転されることはない。この結果は、 $p < p_c(C)$  だと多数派のコピーが最適戦略であるが、 $p > p_c(C)$  では系の非自己修正性から多数派が間違える確率が非常に高いため、 $p_c$  がより高く自己修正的な  $r = O$  のシステムのパフォーマンスが勝ったと考えられる。

## 5. ま と め

この論文では、二択のクイズを用いた情報カスケード実験を行い、ハーダーによる多数派のコピーに対するオッズの影響を調べたものである。被験者は各自の知識で回答したあと、他の被験者の回答情報に関する次の 2 つの状況で回答する。(1) 各選択肢を選んだ回答者数、(2) 各選択肢のオッズ。(1) のコピーコストゼロの状況での最適戦略が「デジタル=必ず多数派を選ぶ」だったのが、(2) のコピーコストありの状況では「アナログ=得票数に比例した確率で投票」に変化する。研究の目的は、ハーダーがこの最適な戦略を行うのか、また、(1) と (2) の異なる状況での集団および個体のパフォーマンスを比較することである。

結果は、(1) コピーコストがない場合、ハーダーはデジタル的 (=必ず多数派を選ぶ傾向)、(2) コピーコストありの場合、ハーダーは倍率が 2 倍の近く (1.3 倍から 4 倍) でアナログ的 (=選択肢の選択者の数に比例する確率で選ぶ) であった。ハーダーのコピー確率のモデルをもとに、比率  $p$  でハーダー、比率  $1-p$  で独立投票者の混在する系をシミュレートする確率モデルを導入して解析した。結果はコピーコストの有無に関わらず情報カスケード転移が起きる。(1) のコピーコストがゼロの場合、ハーダーの比率  $p$  が  $p_c(C) = 80.0\%$  以下なら One Peak 相にあり、正答率は 100% 近くに収束する。一方、 $p$  が  $p_c(C)$  より大きい場合、系は Two Peaks 相にあり、正答率が 100% 近くに収束するか、10% から 20% に収束するかは確率的になる。(2) のコピーコストがゼロでない場合、カスケード転移の閾値  $p_c(O)$  は 94.8% になる。この転移点の差は、システムのパフォーマンスに重大な影響

を及ぼす。 $p < p_c(C)$  では (1) のコピーコストゼロの集団が優れているが、 $p = p_c(C)$  を境に逆転する。

今後の課題は、オッズでのハーダーの応答関数  $q_h(x|O)$  の詳細な評価である。本実験では、 $x = 1/2$  での傾きが 1.18 となり、最適なアナログハーダー  $q_h(x|O) = x$  での傾き 1 から多少ずれていた。しかし、システムサイズ、サンプル数の増加により、最終的に 1 になる可能性がある。その場合、 $r = O$  でカスケード転移は消滅し、系は自己修正的になる。また、我々の情報カスケード実験は、一次元ネットワークに対するものであったが、より複雑なネットワークでのカスケード転移もまた興味深い。

## 謝 辞

本実験の実施における入江洋介と Ruokang Han の協力を深く感謝する。

## ◇ 参 考 文 献 ◇

- [Anderson1997] L. Anderson and C. Holt, Information cascades in the laboratory, *Am. Econ. Rev.*, 87, 847.
- [Bikhchandani1992] S. Bikhchandani and D. Hirshleifer and I. Welch, A Theory of Fads, Fashion, Custom, and Cultural Changes as Information Cascades, *J. Polit. Econ.* 100, 992.
- [Kubler2004] D. Kubler and G. Weizacker, Limited depth of reasoning and failure of cascade formation in the laboratory, *Rev. Econ. Stud.* 71, 425.
- [Goeree2007] J. K. Goeree, T. R. Palfrey, B. W. Rogers and R. D. McKelvey, "Self-correcting information cascades", *Rev. Econ. Stud.*, 74, 733.
- [Hisakado2010] M. Hisakado and S. Mori, Phase transition and information cascade in voting model, *J. Phys. A* 43, 315207.
- [Hisakado2011] M. Hisakado and S. Mori, Digital herders and phase transition in a voting model, *J. Phys. A* 44, 275204.
- [Hisakado2012] M. Hisakado and S. Mori, Two kinds of Phase transitions in a Voting model, *J. Phys. A* 45, 345002.
- [Mori2010] S. Mori and M. Hisakado, Exact Scale Invariance in Mixing of Binary Candidates in Voting Model, *J. Phys. Soc. Jpn.* 79, 034001.
- [Mori2012] S. Mori, M. Hisakado and T. Takahashi, Phase transition to two-peaks phase in an information cascade voting experiment, *Phys. Rev. E* 86, 026109.
- [Rendell2010] L. Rendell and R. Boyd and D. Cownden and M. Enquist and K. Eriksson and M. W. Feldman and L. Fogarty and S. Ghirlanda and T. Lillicrap and K. N. Laland, Why copy others? Insights from the social learning strategies tournament, *Science* 328, 208.
- [Rendell2011] L. Rendell and L. Fogarty and W. J. E. Hoppitt and T. J. H. Morgan and M. M. Webster and K. N. Laland, Cognitive culture: theoretical and empirical insights into social learning strategies, *Trends Cogn. Sci.* 15, 68.
- [Surowiecki] J. Surowiecki, *The Wisdom of Crowds: Why the Many Are Smarter than the Few and How Collective Wisdom Shapes Business, Economics, Societies, and Nations*, Doubleday Books (NY), 2004.
- [Lorenz2011] J. Lorenz, H. Rauhut, F. Schweitzer and D. Helbing, How social influence can undermine the wisdom of crowd effect, *Proc. Natl. Acad. Sci. (USA)* 108, 9020.
- [Ziemba2008] D. B. Hausch, V. Lo and W. T. Ziemba Eds., *Efficiency of Racetrack Betting Markets*, Academic Press, World Scientific.