

# 2 択のクイズを用いた情報カスケード実験

## データ解析

守 真太郎<sup>†</sup> 入江 洋介<sup>†</sup> 久門 正人<sup>††</sup> 高橋 泰城<sup>†††</sup>

<sup>†</sup> 北里大学理学部物理学科, 〒 252-0373 神奈川県相模原市北里 1-15-1

<sup>††</sup> 金融庁, 〒 100-8967 東京都千代田区霞ヶ関 3-2-1

<sup>†††</sup> 北海道大学大学院文学研究科, 〒 060-0810 北海道札幌市北区北 10 条西 7 丁目

E-mail: <sup>†</sup>mori@sci.kitasato-u.ac.jp, <sup>††</sup>ms12804z@st.kitasato-u.ac.jp,

<sup>†††</sup>masatohisakado@standardandpoors.com, <sup>††††</sup>taikitakahashi@gmail.com

あらまし 2010 年から 2013 年まで 5 回に渡り実施した 2 択のクイズを用いた情報カスケード実験のデータ解析とその結果について説明する。

キーワード 情報カスケード, 2 択のクイズ, データ解析

## Information cascade experiment with two-choice quiz Data Analysis

Shintaro MORI<sup>†</sup>, Yosuke IRIE<sup>†</sup>, Masato HISAKADO<sup>††</sup>, and Taiki TAKAHASHI<sup>†††</sup>

<sup>†</sup> Department of Physics, Kitasato University, Kitasato 1-15-1, Sagami-hara-shi, Kanagawa, 252-0373 Japan

<sup>††</sup> Financial Services Agency, 3-2-1 Kasumigaseki, Chiyoda-ku, Tokyo, 100-8967 Japan

<sup>†††</sup> Department of Behavioral Science, Hokkaido University, N10W7 Kitaku-ku, Sapporo-shi, Hokkaido, 060-0810 Japan

E-mail: <sup>†</sup>mori@sci.kitasato-u.ac.jp, <sup>††</sup>ms12804z@st.kitasato-u.ac.jp,

<sup>†††</sup>masatohisakado@standardandpoors.com, <sup>††††</sup>taikitakahashi@gmail.com

**Abstract** We explain in detail about the procedures and the results of data analysis of the voting experiments from 2010 to 2013.

**Key words** information cascade, two-choice quiz, data analysis

### 1. 情報カスケード転移とは

本実験は論文 [1] で予言された情報カスケード転移の可能性の検証を行うために実施したものである。ここで情報カスケード転移とは、一般化されたポリアの壺でのパラメータ変化により安定状態の個数が増える非平衡相転移の一種である [2], [3]。ここでポリアの壺とは、赤玉・青玉が合計  $t (\geq 1)$  個入った壺があり、赤玉の比率を  $z(t)$ 、青玉の比率を  $1 - z(t)$  とすると、次の  $t + 1$  個目の玉の色を確率  $z(t)$  で赤、確率  $1 - z(t)$  で青として壺に加えていくものである [4]。この操作を繰り返すことにより、赤玉の比率の時系列  $z(t), z(t+1), \dots$  が得られる。この「比率  $z(t)$  に等しい確率で赤玉を追加する」を任意の関数  $q: [0, 1] \rightarrow [0, 1]$  を用いて「確率  $q(z(t))$  で赤玉を追加する」としたものが一般化されたポリア壺である [2]。オリジナルのポリア壺と異なり、 $q(z)$  の選択によっては孤立した値に  $z(t)$  が収束する。こうした  $z(t)$  が収束する確率が正の状態を安定状態とよぶ。そして  $q(z)$  の安

定不動点 (stable fixed point) :  $q(z) = z$  かつ曲線  $(z, q(z))$  が対角線  $(z, z)$  を上から下に横切るもの、は安定状態である。一方、 $q(z) = z$  かつ曲線  $(z, q(z))$  が対角線を下から上に横切る不動点は不安定不動点と呼ばれ、 $z(t)$  が収束する確率はゼロである。こうした  $q(z)$  が対角線と交叉する不動点以外に、 $q(z)$  が対角線に接するものを touchpoint と呼ぶ。touchpoint が安定状態かどうかは、 $q(z)$  と対角線の接点での左微分の差の絶対値に依存し、 $1/2$  未満なら安定、 $1/2$  より大きいなら不安定である [5]。

関数  $q(z)$  の持つパラメータを変化させることにより安定不動点の数が 1 つの状態と複数の状態の間の非平衡相転移が起きる。例えば、2 つのパラメータ  $q_*, p$  を  $q_* \in [1/2, 1], p \in [0, 1]$  の範囲で選び、

$$q(z) = (1 - p) \cdot q_* + p \cdot \theta(z - 1/2) \quad (1)$$

の場合を考える。このモデルをデジタル模型と呼ぶ [1]。  $p$  は玉の色の間の相関をコントロールするパラメータ、 $q_*$  は  $p = 0$  の

相関がない場合の赤玉を追加する確率を表すパラメータである。確率  $1-p$  で独立に、確率  $p$  で  $z(t)$  に依存して玉を追加する。独立に赤玉を追加する確率は  $q_*$ 。  $z(t)$  に依存して追加する場合は多数決ルール (majority rule) に従い、  $z(t) > 1/2$  なら赤玉、  $z(t) < 1/2$  なら青玉、  $z(t) = 1/2$  なら赤・青ランダムに玉を追加する。  $q(z)$  は  $p = p_c(q) \equiv 1 - 1/2q_*$  のとき  $z_* = 1/2$  で対角線に接し、  $p_c(q)$  が相の境界を決める閾値となる。  $p < p_c(q)$  では安定不動点は  $z = z_+ = (1-p) \cdot q_* + p$  に1個で、  $z(t)$  は必ず  $z_+$  に収束する。一方、  $p > p_c(q)$  では、  $z = z_{\pm} = (1-p) \cdot q_* \pm p$  に2個の安定不動点を持つ。  $z(t)$  が  $z_{\pm}$  のどちらに収束するかはランダムである。一方、  $p = p_c(q)$  では、  $z_* = 1/2$  が touchpoint となるが、  $q_* = 1/2$  の場合と  $q_* > 1/2$  で状況が異なる。  $q_* = 1/2$  の場合  $p_c = 0$  であり、  $q(z) = 1/2$  より単純対称ランダムウォーク (RW) に他ならない。 touchpoint は  $z_+$  と同じで、  $z(t)$  は  $z_* = z_+$  に収束する。  $q_* > 1/2$  の場合、 touchpoint での対角線との左微分の差が1なので  $z_*$  は不安定で、  $z(t)$  は  $z_+ = (1-p) \cdot q_* + p$  に収束する。

### 1.1 秩序変数

情報カスケード転移の秩序変数は吸収状態相転移との関係から相関関数の極限值として定義するのが自然である [6]。玉の色の時系列を 0,1 の2値の時系列  $X(1), X(2), \dots, X(t) \in \{0, 1\}$  として表すことにする。ここで、  $t$  番目の玉の色が赤 (青) なら  $X(t) = 1(0)$  とする。  $t$  番目までの変数の1の比率を  $z(t) = \sum_{s=1}^t X(s)/t$  とすると、  $X(t+1)$  は確率  $q(z(t))$  で1、確率  $1-q(z(t))$  で0をとることになる。  $X(1)$  は 0,1 を等確率でとるものとする、  $X(1) = 1(0)$  の場合  $z(1) = 1(0)$  となり、一般化ポリア壺は  $z$  軸上の右端 (左端) を初期状態とした時系列  $z(1), z(2), \dots$  を生成する確率過程となる。相関関数  $C(t)$  を、  $X(1)$  と  $X(t+1)$  の共分散  $\text{Cov}(X(1), X(t+1))$  を  $X(1)$  の分散  $\text{Var}(X(1))$  で割ったものとして定義する。

$$C(t) \equiv \text{Cov}(X(1), X(t+1))/\text{Var}(X(1)). \quad (2)$$

規格化により  $C(0) = 1$  となる。  $C(t)$  は、最初の玉が赤 (青) での  $t+1$  番目が赤玉の確率の差と書くことができる。

$$C(t) = \Pr(X(t+1) = 1|X(1) = 1) - \Pr(X(t+1) = 1|X(1) = 0) \quad (3)$$

つまり、  $C(t)$  は  $X(1)$  の影響が  $t$  だけ離れた後の  $X(t+1)$  にどれだけ伝わるかを表すものであり、1種のドミノ効果の強さの指標となる。

$C(t)$  の極限値を  $c \equiv \lim_{t \rightarrow \infty} C(t)$  と書くと、  $c$  が一般化ポリア壺の示す相転移の秩序変数になる。安定な状態が  $z_+$  に1個しかない場合、  $t \rightarrow \infty$  で  $z(t)$  は  $z_+$  に収束し、  $X(t)$  は極限で  $z_+$  の確率で1をとり、  $X(1)$  の影響はない。つまり、  $c = 0$  となる。一方、安定状態が複数存在する場合、  $z(t)$  が  $z_+$  に収束する確率は  $X(1)$  に依存する。すると、  $X(t+1)$  が1をとる確率も  $X(1)$  に依存し、一般に  $c \neq 0$  となる。例えば安定状態が  $z_{\pm} (z_- < z_+)$  の2個存在する場合を考え、  $X(1) = 1(0)$  で  $z(t)$  が  $z_+$  に収束する確率を  $p_1(p_0)$  と書くものとする。  $X(1) = x \in \{0, 1\}$  のとき、  $X(t+1)$  が1をとる確率の  $t \rightarrow \infty$  での極限値は  $p_x \cdot z_+ + (1-p_x) \cdot z_-$

となる。よって  $c$  の値は次のように書くことができる。

$$c = (p_1 z_+ + (1-p_1) z_-) - (p_0 z_+ + (1-p_0) z_-) \\ = (p_1 - p_0)(z_+ - z_-). \quad (4)$$

変数間に正の相関がある場合、  $p_1 > p_0$  となり、  $c > 0$  となる。このように、  $c$  は安定状態が1個の相と複数の相を区別する秩序変数となる。秩序変数  $c$  のパラメータ依存性から臨界指数  $\beta$  が定義される。例えばデジタル模型の場合、パラメータとして  $p$  を用いて  $c \propto (p - p_c(q))^\beta$  で定義すると  $\beta = 1$  となる<sup>(注1)</sup>。

情報カスケード転移はある場合に吸収状態相転移となる [6]。吸収状態相転移とは確率過程において、系がその状態になると変化しない吸収状態という状態を持ち、  $t$  において吸収状態に変化していない確率を  $D(t)$  と書くと、  $t \rightarrow \infty$  での  $D(t)$  の極限値がゼロの状態と有限の状態の間の非平衡相転移である。  $D(t)$  は、相関時間  $\xi$  によってスケールされることが知られている。

$$D(t) \sim t^{-\alpha} \cdot f(t/\xi). \quad (5)$$

ここで  $f(x)$  は  $x = 0$  で解析的な関数である。臨界点では  $\xi$  が発散し、  $D(t) \propto t^{-\alpha}$  というべき関数として振る舞う。一方、  $t \rightarrow \infty$  において  $c \equiv \lim_{t \rightarrow \infty} D(t)$  が有限であるためには、  $f(x) \propto x^\alpha$  として振る舞う必要がある。よって  $c = \xi^{-\alpha}$  となる。  $\xi$  の臨界点での発散の臨界指数を  $\nu_{||}$  と定義すると、  $\alpha, \beta, \nu_{||}$  の間に次のスケーリング関係式が成立する。

$$\beta = \alpha \cdot \nu_{||} \quad (6)$$

$D(t)$  と  $C(t)$  の対応関係を示す。デジタル模型で  $q = 1$  の場合を考える。このとき、  $z(t)$  は  $z(t) > z_* = 1/2$  となると  $z < 1/2$  の領域に戻ってくることはできず、  $z > 1/2$  の状態が吸収状態となる。情報カスケード転移は  $z$  軸上に安定な状態が  $z_+ = 1$  にひとつの状態と  $z_{\pm} = (1-p) \pm p$  に2個の状態間の相転移であった。ひとつの場合、  $z(t)$  は確率1で  $z_+$  に収束し、吸収状態に変化していない確率  $D(t)$  もゼロに収束する。2個の場合、  $z_- = (1-p)$  に収束する確率は正となり、  $D(t)$  は正の値に収束する。  $X(1) = 1$  はすでに吸収状態なので、  $\Pr(X(t+1)|X(1) = 1) = 1$  となる。よって

$$C(t) = 1 - \Pr(X(t+1) = 1|X(1) = 0) \\ = \Pr(X(t+1) = 0|X(1) = 0) \quad (7)$$

となる。  $\Pr(X(t+1) = 0)$  は、  $z(t)$  が  $< z_*, = z_*, > z_*$  でそれぞれ  $p, p/2, 0$  より、

$$C(t) = p \cdot \Pr(z(t) < z_* | X(1) = 0) \\ + p/2 \cdot \Pr(z(t) = z_* | X(1) = 0) \quad (8)$$

となる。  $t$  が十分大きいと、  $\Pr(z(t) = z_* | X(1) = 0)$  の項は無視できるのと、  $D(t)$  は  $\Pr(z(t) \leq z_* | X(1) = 0)$  に他ならないことに注意すると

$$C(t) \simeq p \cdot D(t) \quad (9)$$

(注1):  $q_* > 1/2$  の場合。

となる。これが  $q = 1$  のデジタル模型におけるカスケード転移での  $C(t)$  と吸収状態相転移での  $D(t)$  の間の対応関係である。一般の  $q$  のデジタル模型、また、一般化ポリア壺の相転移では、完全な吸収状態が存在せず、情報カスケード転移は厳密な意味では吸収状態相転移ではない。しかしデジタル模型では一般の  $q > 1/2$  の場合でスケーリング関係  $\beta = \alpha \cdot \nu_{||}$  を満たし、同じ普遍類に属することが示されている。このことは一般化ポリア壺の相転移も吸収状態相転移と共通の構造を持つことを示唆する。

## 1.2 相関関数の漸近形

相転移の検証には秩序変数  $c$  の評価が必要になる。 $t$  が十分に大きい場合、 $C(t)$  の値から  $c$  の値を推定することは難しい。しかし、情報カスケードの実験では  $t$  の値に限界があり、かつサンプル数が少ないため  $C(t)$  の誤差も大きい。そこで、 $C(t)$  の漸近的な振る舞いをもとに、相転移の検証において重要な物理量を導入する。

$C(t)$  の漸近的な振る舞いは、 $q(z)$  が touchpoint を持たない場合、次のように書くことができる [3]。

$$C(t) \simeq c + a \cdot t^{l-1} + a' \cdot t^{l-1-b}. \quad (10)$$

ここで、 $c$  は秩序変数で安定状態が複数の場合の  $C(t)$  のリーディング項となる。その場合のサブリーディング項  $a \cdot t^{l-1}$  の次数  $l-1$  の  $l$  は、安定状態が 2 個の場合、 $q(z)$  の微分の安定状態での値を  $l_{\pm} \equiv q'(z_{\pm})$  として、そのなかの大きいほうの値に等しい。安定状態がひとつの場合、 $c = 0$  で  $a \cdot t^{l-1}$  がリーディング項となる。 $l$  は、 $q(z)$  の微分の  $z_+$  での値  $q'(z_+)$  となる。デジタル模型の場合、 $q(z)$  の微分は touchpoint 以外ではゼロで、 $C(t)$  は  $c$  と指数関数的に減衰する項の 2 つからなる。また、第 3 項は  $o(t^{l-1})$  の項を表しているので、 $b > 0$  とする。

一方、 $q(z)$  が touchpoint を持つとき、 $C(t)$  の振る舞いは系の対称性、touchpoint 近傍での  $q(z)$  の様子に依存する。例えばデジタル模型の場合、touchpoint は不安定なので  $c = 0$  である。 $q = 1/2$  では  $p_c = 0$  なので対称単純 RW に他ならず、 $C(t) = \delta_{t,0}$  となる。一方、 $q > 1/2$  の場合、 $p_c = 1 - 1/2q$  で、 $p = p_c$  では  $C(t) \propto t^{-1/2}$  となり  $\alpha = 1/2$  となる。デジタル模型の  $q > 1/2$  でのカスケード転移の普遍類は  $\alpha = 1/2, \beta = 1$  で特徴づけられる。

$q(z)$  として滑らかな関数を用いたとき、 $q(z)$  は touchpoint で接するため左微分の差がゼロとなり touchpoint は安定状態である。とくに  $q > 1/2$  の場合 touchpoint が  $z_+$  と異なるため  $c > 0, q = 1/2$  の場合は touchpoint が  $z_+$  と等しいので  $c = 0$  となる。またサブリーディング項は  $\log t$  のべき関数として振る舞う。 $C(t)$  の振る舞いは、

$$C(t) \simeq c + a \cdot (\log t)^{-\alpha'} + o((\log t)^{-\alpha'}) \quad (11)$$

と書くことができる。普遍類は  $C(t)$  が  $\log t$  のべき関数であることから  $\alpha = 0$ 。また、秩序変数  $c$  は  $q > 1/2$  の場合、touchpoint をもったときに有限の値をとるためゼロと有限の値のあいだの不連続変化を行い  $\beta = 0$  である。 $q = 1/2$  の場合、 $c$  はゼロから連続的に有限の値に変化するが対数関数的な特異性を示し、 $\beta = 0$  である。

## 1.3 相関時間の漸近形

$c$  の評価に便利かつ物理的にも重要な相関時間を導入する [3], [6]。まず、 $t$  に対し、 $n$  次モーメント  $M_n(t)$  を

$$M_n(t) \equiv \sum_{s=0}^{t-1} C(s)s^n. \quad (12)$$

とする。そして、積分相関時間  $\tau(t)$  を 0 次モーメント  $M_0(t)$  として、また、二次モーメント相関時間  $\xi(t)$  を 2 次モーメントと 0 次モーメントの比の平方根として定義する。

$$\tau(t) \equiv M_0(t), \quad (13)$$

$$\xi(t) \equiv \sqrt{\frac{M_2(t)}{M_0(t)}}. \quad (14)$$

以下、積分相関時間を緩和時間、二次モーメント相関時間を相関時間と呼ぶことにする。この 2 つのタイムスケールのうち、相転移の特徴的な長さスケールを決めるのが相関時間  $\xi(t)$  である。

$C(t)$  の漸近的な振る舞い (10) を用いて  $\tau(t)/t$  を評価すると、

$$\tau(t)/t \simeq \begin{cases} \frac{a}{t} \cdot t^{l-1} (1 + \frac{a'}{a(l-b)} \cdot t^{-b}), & c = 0, \\ c \cdot (1 + \frac{a}{cl} \cdot t^{l-1}), & c > 0. \end{cases} \quad (15)$$

となる。 $\tau(t)/t$  は、 $t \rightarrow \infty$  の極限で秩序変数  $c$  に一致する。

$$\lim_{t \rightarrow \infty} \tau(t)/t = c. \quad (16)$$

また、 $\xi(t)/t$  は

$$\xi(t)/t \simeq \begin{cases} \sqrt{\frac{l}{l+2}} \cdot (1 + \frac{a'}{2a} (\frac{l+2}{l+2-b} - \frac{l}{l-b}) t^{-b}), & c = 0, \\ \sqrt{\frac{1}{3}} \cdot (1 + \frac{a}{2c} (\frac{3}{l+2} - \frac{1}{l}) t^{l-1}), & c > 0. \end{cases} \quad (17)$$

となる。 $\xi(t)/t$  の  $t \rightarrow \infty$  の極限値は  $c > 0$  と  $c = 0$  で異なる。

$$\xi(t)/t \rightarrow \begin{cases} \sqrt{l/l+2}, & c = 0, \\ \sqrt{1/3}, & c > 0. \end{cases} \quad (18)$$

これらの  $\tau(t)/t, \xi(t)/t$  の漸近形を用いて実験データで評価した  $\tau(t)/t, \xi(t)/t$  をフィットすることにより、 $c, l$  を推定することができる。<sup>(注2)</sup>

## 2. 実験データについて

過去 5 回の実験状況をまとめたものが表 1 である<sup>(注3)</sup>。被験者集団を  $G$ 、問題の集合を  $I$ 、参照人数の集合を  $R$ <sup>(注4)</sup>、被験者の識別  $id$  の集合を  $ID$ 、要素数を  $|G|$  などで表しまとめたものである。最後の列には被験者が行った選択回数を示した。

(注2):  $l$  は  $z(t)$  の分散のべき指数  $\gamma(\text{Var}(z(t))) \propto t^{-\gamma}$  と次の関係にある。

$$\gamma = \begin{cases} 2 - 2l & \text{if } l > 1/2 \\ 1 & \text{if } l \leq 1/2 \end{cases}$$

とくに、 $l > 1/2$  のとき、 $\gamma < 1$  となるスーパー拡散相、 $l < 1/2$  のときノーマル拡散相と呼ばれる。

(注3): 実験の詳細については「2 択のクイズを用いた情報カスケード実験 実

表 1 実験データ：実験番号, グループ数  $|G|$ , 問題数  $|I|$ , 被験者数  $|ID|$ , 被験者の各  $r$  に対する選択回数. 実験 3 の括弧内の選択回数は  $r = M$  に対するものである. 実験 4, 5 では,  $r \in \{0, C\}$  と  $r \in \{0, M\}$  を異なる問題に対して行っている. そのため実際の問題数は  $2|I|$  となる.

No.	$ G $	$ I $	R	$ ID $	選択回数
1	2	100	$\{0, 1, 2, 3, 5, 7, 9, C\}$	62	6200
2	2	120	$\{0, 1, 5, 11, 21, CM\}$	104	12195
3	2	120	$\{0, C, M\}$	120	14400(14350)
4	1	10	$\{0, C\} (\{0, M\})$	218	1909(1941)
5	1	120	$\{0, C\} (\{0, M\})$	125(85)	10103(5696)

グループ  $g \in G$ , 問題  $i \in I$ , 参照人数  $r \in R$  の  $t \in \{1, \dots, T(g, i, r)\}$  番目に回答した被験者  $id$  を  $id(g, i, r, t) \in ID$ , 回答を  $X(g, i, r, t) \in \{0, 1\}$  で表す.  $X = 1$  は正解,  $X = 0$  は不正解の選択肢を選んだことを意味する.

$$X(g, i, r, t) \in \{0, 1\}, g \in G, i \in \{1, \dots, |I|\}$$

$$t \in \{1, \dots, T(g, i, r)\}, id(g, i, r, t) \in ID$$

グループ  $g$ , 問題  $i$ , 参照人数  $r$  に対し  $T(g, i, r)$  回の回答が行われ,  $g, i$  に関する和が表 1 での選択回数である. 実験 1 の場合,  $T(g, i, r) = 31$  で  $g, i, r$  に依存しないので,  $31 \times 100 \times 2 = 6200$  (注5) である (注6).

### 3. データ解析：マクロな法則

ここでは, 問題毎の正答率, 選択の時系列の相関関数, 緩和時間, 相関時間など系のマクロな性質の解析結果について述べる.

#### 3.1 正答率の $r$ 依存性

$g, i, r$  に対する正答率  $Z(g, i, r)$  は

$$Z(g, i, r) = \frac{1}{T(g, i, r)} \sum_{i=1}^{T(g, i, r)} X(g, i, r, t)$$

で定義される.  $r = 0$  の場合は自己情報のみで回答したときの正答率, それ以外は何らかの他の被験者の選択情報をもとでの正答率を表している. もし, 被験者が他の被験者の選択情報に影響されずに回答する場合,  $Z(g, i, r)$  は  $r$  に依存せず  $r = 0$  での正答率と等しくなるであろう. そこで, 図 1 に実験 1, 2 での  $r = 0$  と過去  $r$  人の選択を参照した場合での正答率の散布図をプロットした.

験の手順とデータアーカイブについて」(統計数理研究所共同研究リポート 311 経済物理とその周辺 (10), pp.1-12.) を参照のこと. 実験 1, 2 のデータを用いた論文は [7], 実験 2, 3 のデータを用いた論文は [8] である.

(注4):  $C$  は過去の被験者全員の選択者数 (Cumulative number),  $M$  は過去の被験者全員の選択者数をもとに, 各選択肢の倍率 (Multipliers) を表示する. 「2 択のクイズを用いた情報カスケード実験 実験の手順とデータアーカイブについて」では, 前者を  $r = \infty$ , 後者を  $1/\infty$  と記している.

(注5): 実験 2 の場合,  $T(g, i, r)$  は  $g, i$  毎に異なり約 50.8 であるので, 約  $50.8 \times 120 \times 2 = 12192$  である.

(注6):  $t$  番目の被験者の  $r = C$  での回答は  $t - 1 \in R$  なら,  $r = t - 1$  での回答と等しい.  $t \leq r$  のとき  $r$  での回答は行われていないので,  $r = C$  での回答を  $r$  での回答とする. 例えば,  $r = 5$  の場合,  $t = 6$  から  $r = 5$  での選択が始まる.  $t \leq 5$  に対しては,  $t - 1$  人参照の選択を  $r = 5$  での選択でもあるとする. こうして, 選択の時系列  $\{X(g, i, r, t)\}$  は  $t = 1$  から始まると考える.

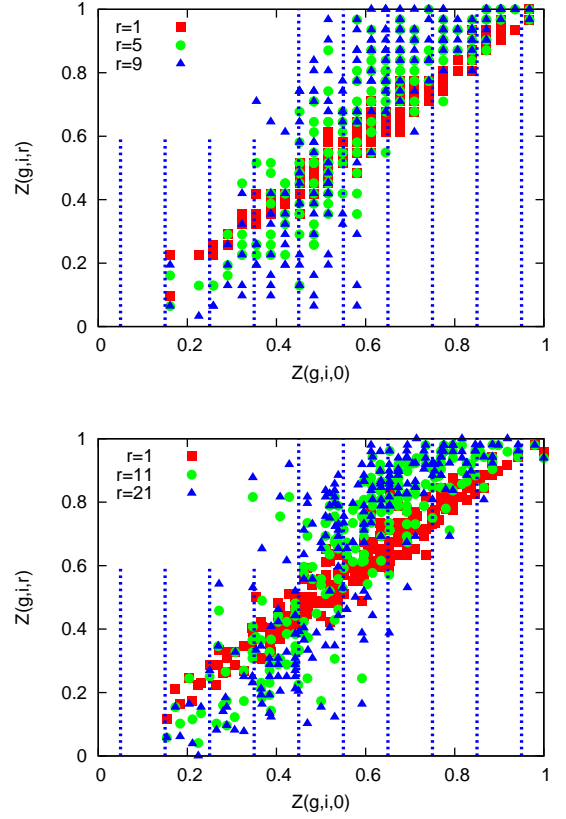


図 1 同じ問題  $i \in I$  に対する  $r = 0$  の正答率  $Z(g, i, 0)$  と有限の  $r$  での正答率  $Z(g, i, r)$  の散布図. 実験 1 では  $r \in \{1, 5, 9\}$ , 実験 2 では  $r \in \{1, 11, 21\}$  とした.

$r = 1$  ではほぼ対角線上に分布し, 他の被験者の影響が小さいことが分かる. 一方, 実験 1 の  $r = 9$ , 実験 2 の  $r = 21$  の場合, 対角線から大きく外れた位置に分布している. 一般に,  $r$  が 1 から増加するに従い, 正答率の分布は対角線から離れていくことが分かる. また,  $Z(g, i, 0)$  が 1 近傍や 0 近傍の正答率が高い (低い) 場合,  $r > 1$  での正答率はより高く (低く) なっている. 一方,  $r = 0$  での正答率が 50% 前後の場合,  $r > 1$  での正答率は高くなるか低くなるかのどちらかである.

図 2 に, 全 5 回 ( $r = M$  の場合は 4 回) の実験での  $r = 0$  と  $r = C$  および  $M$  の正答率の散布図を示した.  $r = C$  の場合, 実験番号が 1, 2, 3, 5 と下がるとともに,  $T(g, i, C)$  が大きくなる. すると, 一般に, 正答率の  $r = 0$  から  $r = C$  での変化も大きくなり, 図 1 で見た傾向がより鮮明になる.  $r = 0$  での正答率が高く 1 に (低く 0 に) 近づけば,  $r = C$  での正答率はより高く 1 (より低く 0) に近づき,  $r = 0$  での正答率が 50% 前後では,  $r = C$  の正答率は極端に高くなるか低くなるかである.  $r = M$  の場合も同様に正答率の分布が対角線からずれることが分かる.  $r = C$  の場合と比較して正答率の変化は小さく 1 や 0 までの極端な変化は見られない. 一方, 実験 4 に関しては他の実験室実験と比較して変化は小さいことが分かる. 特に,  $r = M$  の倍率の場合, 正答率の変化は小さく,  $r = 0$  での正答率とほぼ無相関であることも分かる. 実験室実験と異なり他の被験者の選択の情報とし

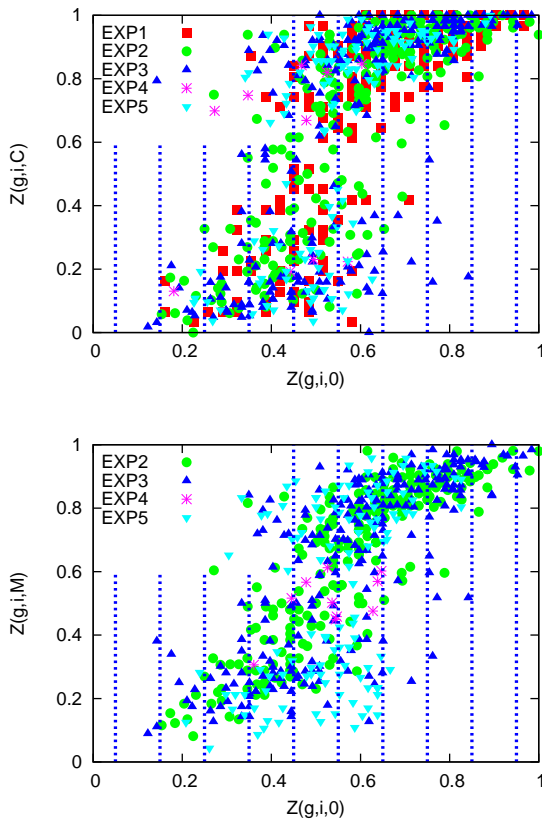


図2  $r = 0$  と  $r = C$  (上図) および  $r = M$  (下図) での正答率の散布図を示した。

て、過去のある大学での選択情報に基づくとしたため、被験者の信用度が低かったからと考えられる。後にハーダーの選択法則について述べるが、実験4の  $r = M$  のみ少数派を選択する傾向があった。被験者は信用できない多数派の選択より倍率の大きな選択肢を選んだと考えられる。

### 3.2 ハーダー比率とデータの分類

情報カスケード転移が起きるには  $q(z)$  がパラメータの変化により、安定状態が1個の状態と複数の状態を持つようになる必要がある。正答率の  $r = 0$  と  $r = C, M$  での変化から、 $r = 0$  での正答率が  $q(z)$  のパラメータの役割を果たし、それが高い、または低い場合は安定状態が1つ、50%前後の場合は2個存在すると考えられる。 $r = 0$  において被験者は自分の知識のみを用いて回答している。一方、 $r \neq 0$  においては被験者は自分より前の  $r$  人、または全員 ( $r = C, M$ ) の選択情報のもとに回答している<sup>(注7)</sup>。自分の知識に絶対の自信がない場合、多数派の選択肢を選択することが合理的である、なぜなら他人はより確かな情報にもとづいて選択していると考えられるからで、経済学ではこの選択の傾向のことを合理的ハーディング (rational herding) と呼ぶ。そして、 $r = 0$  と  $r \neq 0$  での正答率の変化は合理的ハーディングの影響と考えられる。情報カスケードとは、合理的ハーディングにより、自己情報の示す選択肢が多数派の選択肢と異なった場合、多数派の選択肢を選ぶことをいう。その仕組みから情報カスケードは連鎖する傾向がある。情報カスケード転移は、

(注7): 前者を  $r$  有限と呼び、 $r = C, M$  と区別する。

合理的ハーディングにより誤った連鎖が無限に続く状態<sup>(注8)</sup>と十分多数の被験者が選択するといずれ正しい選択肢を選ぶようになる状態<sup>(注9)</sup>のどちらになるか分からない相と後者のみの相の間の相転移ということも出来る。

情報カスケード転移には多数派にハードする被験者の存在が必要である。こうした被験者を一般にハーダー (herder) と呼ぶことにする。実験におけるハーダーの比率  $p$  の推定方法を述べる。ハーダーの比率の推定には、被験者の分類を行う必要がある。以前述べたように、2 択のクイズを用いた情報カスケード実験における最大の困難は、データ解析における被験者の分類である。再度、被験者の分類を述べておく。

- (1) 正解を知る被験者。正答率  $q_*$  は 1。
- (2) 「正解を知っている」と考える被験者。  $q_* < 1$ 。
- (3) 「正解を何となく知っている」と考える被験者。
- (4) 正解がまったく分からない被験者。

2 択のクイズではこうした4つの分類が考えられる<sup>(注10)</sup>。最初の分類に属する正解を知っている被験者は他の被験者の選択に影響されず正解を選ぶと考えられるので、正答率  $q_* = 1$  の独立投票者と考えてよいであろう。問題は残りの分類に属する被験者である。実験2と実験5では自信度の回答を求めたので、それが100%以外の被験者を分類3,4とすることは可能であろう<sup>(注11)</sup>。一方、自信度100%の被験者で他の被験者の選択を見ても選択を変えず自信度100%のままの被験者も分類1,2に属すると考えてもよさそうだが、適当に100%の自信があると回答し、たまたま自分の選択肢が多数派だったのでそのまま100%とただけで、少数派だったら変えていた可能性もある。一方、実験1,3,4では自信度の回答はなく、情報は  $X(g, i, r, t) \in \{0, 1\}$  しかない。 $q_* = 1$  の独立投票者ならすべての  $r \in R$  に対して  $X(g, i, r, t) = 1$  と考えられるが、逆にすべての  $r \in R$  に対して  $X(g, i, r, t) = 1$  だからと言って  $q_* = 1$  の独立投票者とは限らない。また、すべての  $r \in R$  に対して  $X(g, i, r, t) = 0$  だとしても、2番目の分類に含まれるとは限らない。正解がまったく分からなかったけれど、 $r = 0$  で偶然選んだ選択肢が多数派の選択肢だったのでそれを再度選んだ可能性もある。

このように、自信度の回答があった場合はその正確さや主観的な自信度というものの解釈の難しさがある。自信度がない場合、異なる  $r$  に対して選択を変えた被験者は分類3,4に属すると考えられても、選択を変えなかった被験者を分類1,2に属するとは言いきれない。また、2 択のバイアス<sup>(注12)</sup>から、分類4の被験者の正答率が50%とも言いきれない。ここでは、もっとも単純な分類に基づいた解析を行うことにする。それは、分類1の被験者を独立投票者、それ以外の被験者をハーダーと考える。前者の正答率は任意の  $r$  で  $q_* = 1$ 、後者の  $r = 0$  での正答率は

(注8): non-self-correcting

(注9): self-correcting, fully revealing

(注10): もちろん、これ以外に、問題が何であれ適当に選択するノイズな被験者の存在も考えられる。

(注11): もちろん被験者が嘘をついている可能性は否定できない

(注12): 正解がまったく予想がつかない場合でも二つの選択肢の選択率が50%になるとは限らない。この50%からのゆがみを「2 択のバイアス」と呼んだ。



1/2 とする。2 択のバイアスを考えるとこれは大胆な仮定である。そこで、この仮定に基づいて解析を行い、2 択のバイアスが原因と考えられるデータを削除することとした。これは都合の悪いデータを恣意的に削除する操作ではない。実験の前提条件を、正解を知っている被験者 ( $q_* = 1$ ) と正解を知らず  $r = 0$  での正答率が 1/2 のハーダーの集団と設定することに対応する。この前提条件に反するデータを削除するわけである<sup>(注13)</sup>。

$g, i$  でのハーダーの比率を  $p(g, i)$ 、独立投票者の比率を  $1 - p(g, i)$  と書くことにする。 $g, i$  に対する  $r = 0$  の正答率の期待値は

$$(1 - p(g, i)) \cdot 1 + p(g, i) \cdot \frac{1}{2}$$

となる。実際の正答率  $Z(g, i, r)$  はこの期待値のまわりで標準偏差  $\sqrt{p(g, i)/4T(g, i, 0)}$  でゆらく。正答率の期待値の最小値は  $p(g, i) = 1$  のとき 1/2 なので、実際の正答率がゆらぎを考量してもなお 1/2 からかけ離れて小さい場合、実験の前提条件に反していると考えられる。つまり、不正解の選択肢を選びやすい 2 択のバイアスが存在したため、正答率が 1/2 以下になったと考えられる。

$Z(g, i, 0)$  の期待値に対して

$$E(Z(g, i, t)) = (1 - p(g, i)) \cdot 1 + p(g, i) \cdot \frac{1}{2}$$

が成立することから、この式を解いて  $p(g, i)$  を求めると、

$$p(g, i) = 2 \cdot (1 - E(Z(g, i, 0)))$$

となる。 $Z(g, i, 0)$  のゆらぎを考慮し、 $p(g, i)$  の推定値として

$$p(g, i) = 2 \cdot (1 - Z(g, i, 0))$$

を用いることにする。 $Z(g, i, 0)$  のゆらぎから  $p(g, i)$  の誤差は  $\sqrt{p(g, i)/T(g, i, 0)} < 1/\sqrt{T(g, i, 0)}$  と見積もることができる。 $T(g, i, 0) = 10^2$  で 10% になることが分かる。

$p(g, i)$  の評価式を用いて実験の前提条件に反するデータを削除する。まず、 $r = 0$  での正答率に対する条件  $Z(g, i, 0) \geq 0.5$  から、 $Z(g, i, 0) < 0.45$  のデータを削除することとした。 $Z(g, i, 0) < 0.5$  としないのはゆらぎを考慮してのことである。任意の  $r \in R$  での正答率  $Z(g, i, r)$  の最小値は、比率  $p(g, i, r)$  のハーダーすべてが間違える場合を考慮して  $1 - p(g, i, r)$  である。もし、ある  $r \in R$  に対して  $Z(g, i, r)$  が  $1 - p(g, i, r)$  を下回っている場合、 $p(g, i)$  を過小評価していることになる。これは正しい選択肢を選びやすい 2 択のバイアスが原因であると推測されるので、こうしたサンプルも除外する。こうした除外後の、任意の  $r \in R$  に対して  $Z(g, i, r) \geq 1 - p(g, i)$ 、かつ  $Z(g, i, 0) \geq 0.45$  のサンプル  $(g, i)$  を  $I$  と書くことにする。図 3 で赤線を境界とする領域に任意の  $r \in R$  で含まれている  $(g, i)$  の組が  $I$  の要素である。

$$I = \{(g, i) | Z(g, i, 0) \geq 0.45, Z(g, i, r) \geq 1 - p(g, i), \forall r \in R\}$$

(注13) :  $p$  の推定が  $q(z)$  の推定に必要な不可欠とも言い切れない。 $p$  の推定を諦めて、 $r = 0$  での正答率のみで被験者集団を分類し、 $q(z)$  を推定する方法もある。そのほうが独立投票者、ハーダーというモデルを導入しなくてすむため客観的な解析とも言える。

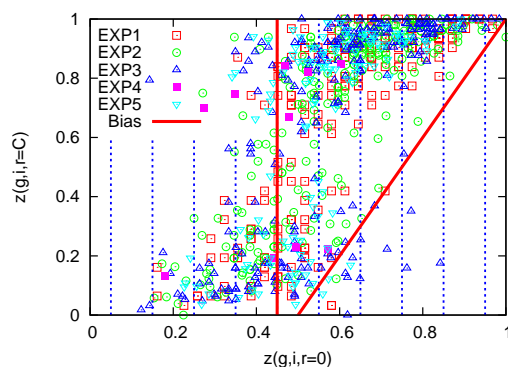


図 3  $r = 0$  と  $r = \infty$  での正答率  $Z(g, i, r)$  の散布図。 $Z(g, i, 0)$  の値  $z$  が  $45\% \leq z < 55\%$ ,  $55\% \leq z < 65\%$ ,  $65\% \leq z < 75\%$ ,  $75\% \leq z < 85\%$  の 10% 刻みで分類し、それぞれに属するサンプル  $(g, i)$  の集合を  $I(1), I(2), I(3), I(4)$  と呼ぶことにする。 $I(1)$  の分類に含まれるサンプルは正答率がほぼ 50% なので、 $p$  の値はほぼ 100% となる。 $Z(g, i, 0) = 1/2$  と  $(1/2, 0)$  と  $(1, 1)$  を結ぶ赤い直線は、2 択のバイアスを除去するための境界線である。

そして図 3 に示したようにデータ  $\{X(g, i, r)\}$  を  $Z(g, i, 0)$  の値で分類する。 $Z(g, i, 0)$  の値  $z$  が  $45\% \leq z < 55\%$ ,  $55\% \leq z < 65\%$ ,  $65\% \leq z < 75\%$ ,  $75\% \leq z < 85\%$ ,  $z \geq 85\%$  に含まれる  $g, i$  の組の集合をそれぞれ  $I(1), I(2), I(3), I(4), I(5)$  と呼ぶことにする。

$$I = I(1) + I(2) + I(3) + I(4) + I(5)$$

表 2 各実験での  $I(n)$  の要素数  $|I(n)|$  一覧。実験 4, 5 の数字は  $R = \{0, C\}$ 、括弧内は  $r = \{0, M\}$  のものである。

$ I(n) $	EXP1	EXP2	EXP3	EXP4	EXP5
$ I(1) $	52	43	38	4(5)	35(24)
$ I(2) $	39	45	52	2(3)	27(36)
$ I(3) $	28	45	38	0(0)	19(18)
$ I(4) $	28	33	38	0(0)	7(5)
$ I(5) $	16	11	14	0(0)	2(0)
$ I $	163	177	168	6(8)	90(83)

$45\% \leq z < 55\%$  の  $I(1)$  に含まれるサンプルは正答率が約 50% であり、 $p = 2(1 - z)$  から、 $p$  はほぼ 100% であることが分かる。また  $Z(g, i, 0)$  の値を 10% 刻みで分類した結果、 $I(n)$  に含まれるサンプルの  $p(g, i)$  は  $n \in \{1, 2, 3, 4\}$  に対しそれぞれ約 100%, 80%, 60%, 40% となる。正答率  $Z(g, i, 0)$  の 10% 刻みでサンプルを分類した理由は、各  $I(n)$  中のサンプル数  $|I(n)|$  を確保するために細かく分類することが出来なかったからである<sup>(注14)</sup>。また  $p$  の評価には  $\pm 10\%$  のゆらぎがあるため、あまり細かく設定する意味もないと考えたこともある。もし、十分なサンプル数が確保でき  $Z(g, i, 0)$  を 1% 刻みで分類したとしても、 $T(g, i, r) = 10^2$  なら  $p(g, i)$  の値に  $\pm 10\%$  程度のバラツキはあり、 $p$  に対する物理量の値の評価の精度もこのバラツキで制限される。 $I(n)$  のサンプルでの  $p(g, i)$  の平均値を  $p(n)$  と書くこ

(注14) : ここで、 $I(n)$  の要素数を  $|I(n)|$  と書いた。

とにする.

$$p(n) \equiv \frac{1}{|I(n)|} \sum_{(g,i) \in I(n)} p(g,i)$$

$I(n)$  中のサンプル  $(g,i)$  に対するハーダーの比率  $p(g,i)$  は, この平均値のまわりに  $\pm 14\%$  程度のばらつきをもつ<sup>(注15)</sup>.

### 3.3 相関関数 $C(t)$ のハーダー比率依存性

時系列データ  $\{X(g,i,r,t)\}$  に対して相関関数  $C(t)$  を計算する. モデルの理論的解析では  $X(1)$  と  $X(t+1)$  の共分散  $\text{Cov}(X(1), X(t+1))$  を  $X(1)$  の分散  $\text{Var}(X(1))$  で割ったものとして  $C(t)$  を定義した(式(2)). 一方,  $C(t)$  は  $X(1) \in \{0,1\}$  での  $X(t+1)$  の期待値の差でもあった(式(3)). ここでは後者の計算式を用いる. 理由は期待値の誤差評価が容易なことにある.  $C(t)$  の標準誤差は期待値推定での標準誤差を  $X(1) \in \{0,1\}$  に対してもとめ, それらの自乗和の平方根として評価する.

$p$  の違いによる相関関数  $C(t)$  の振る舞いの差異が本質なので, 個々の  $I(n)$  のサンプルに対して評価する.  $I(n)$  サンプルを  $X(1) \in \{0,1\}$  で分類し, サンプル数を  $I_0(n), I_1(n)$  と書くものとする. まず,  $X(1) = x \in \{0,1\}$  に対して  $X(t+1)$  の期待値  $E_1(n,r,t+1|x)$ , その標準誤差  $\Delta E_1(n,r,t+1|x)$  を評価する.

$$E_1(n,r,t+1|x) \equiv \frac{1}{|I_x(n)|} \sum_{(g,i) \in I_x(n)} X(g,i,r,t+1)$$

$$\Delta E_1(n,r,t+1|x) \equiv \sqrt{\frac{E_1(n,r,t+1|x)(1-E_1(n,r,t+1|x))}{|I_x(n)|}}. \quad (19)$$

$C(n,r,t)$  とその標準誤差  $\Delta C(n,r,t)$  の計算式は次のものである.

$$C(n,r,t) \equiv E_1(n,r,t+1|1) - E_1(n,r,t+1|0)$$

$$\Delta C(n,r,t) \equiv \sqrt{\sum_x \Delta E_1^2(n,r,t+1|x)}. \quad (20)$$

図4は,  $r=C$  の場合の相関関数  $C(t)$  を  $t$  に対してプロットしたものである<sup>(注16)</sup>.  $n=3,4$  に対して  $C(t)$  はすぐにゼロに減衰し, ゼロのまわりでゆらぐ. これは  $c=0$  を示唆している. 一方,  $n=1,2$  の場合, 減衰後ある正の定数のまわりでゆらぐようになる. 実験1,2では  $n=1$  の場合その定数は0.1から0.2, 実験3,5では約0.3となり, これは  $c>0$  を示唆する. 実験1,2と実験3,5の違いは実験での被験者の情報の与え方の差にあると考えられる. 実験3,5では  $r=C$  での回答は  $r=0$  の直後に行われるか, 別の問題を回答したあとに行われる. 一方, 実験1,2において  $t$  番目に回答する被験者は  $r \leq t-1$  を満たす  $r \in R$  での回答後に  $r=C$  で回答する. そのため, 自分の直近の被験者の回答の影響が強くなり, 結果  $t=1$  の回答の影響が弱くなっていると考えられる.

(注15):  $T(g,i,r) = T = 10^2$  に対し  $\sqrt{(0.1)^2 + \frac{1}{T}} \simeq \sqrt{2} \cdot 10\%$  と評価した. 1%刻みでサンプルを分類した場合でも,  $\sqrt{(0.01)^2 + \frac{1}{T}} \simeq 10\%$  となり, 大幅な改善はない.  $T=400, 5\%$  刻みでサンプル  $g,i$  を分類できれば  $p(g,i)$  のゆらぎを10%未満に抑えることが可能である.

(注16): 実験5ではサンプル数が少ないため  $n=4$  を除外した.

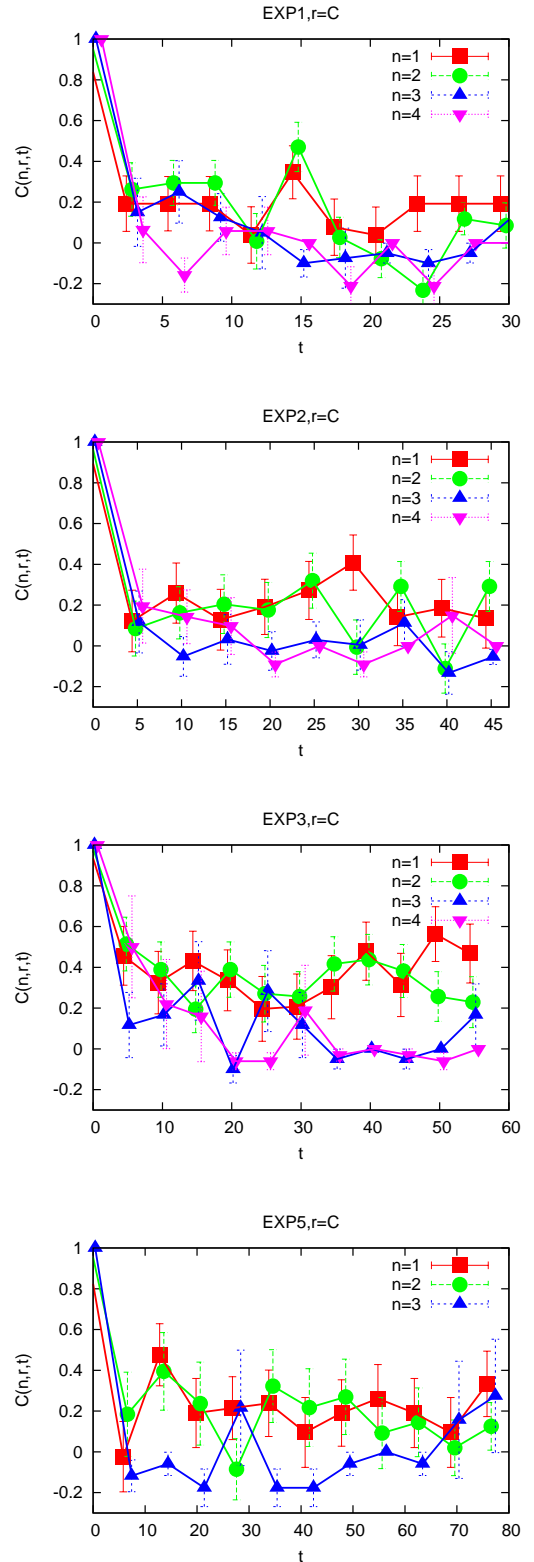


図4  $r=C$  の場合,  $I(n)$  の時系列データ  $\{X(g,i,r,t)\}$  に対する相関関数  $C(n,r,t)$  を計算し  $t$  に対してプロットした. 上から順に実験1,2,3,5の結果を示している.  $x$  座標の値は見易いように定数ずらしている. 以下のグラフも同様.

図5は実験2において  $r$  が有限のときの  $C(t)$  を  $t$  に対してプロットしたものである. 上図は実験2で  $r \in \{1,5,11,21\}, n=1$ , 下図は実験2で  $r=21, n \in \{1,2,3,4\}$  の場合である.  $C(t)$  は

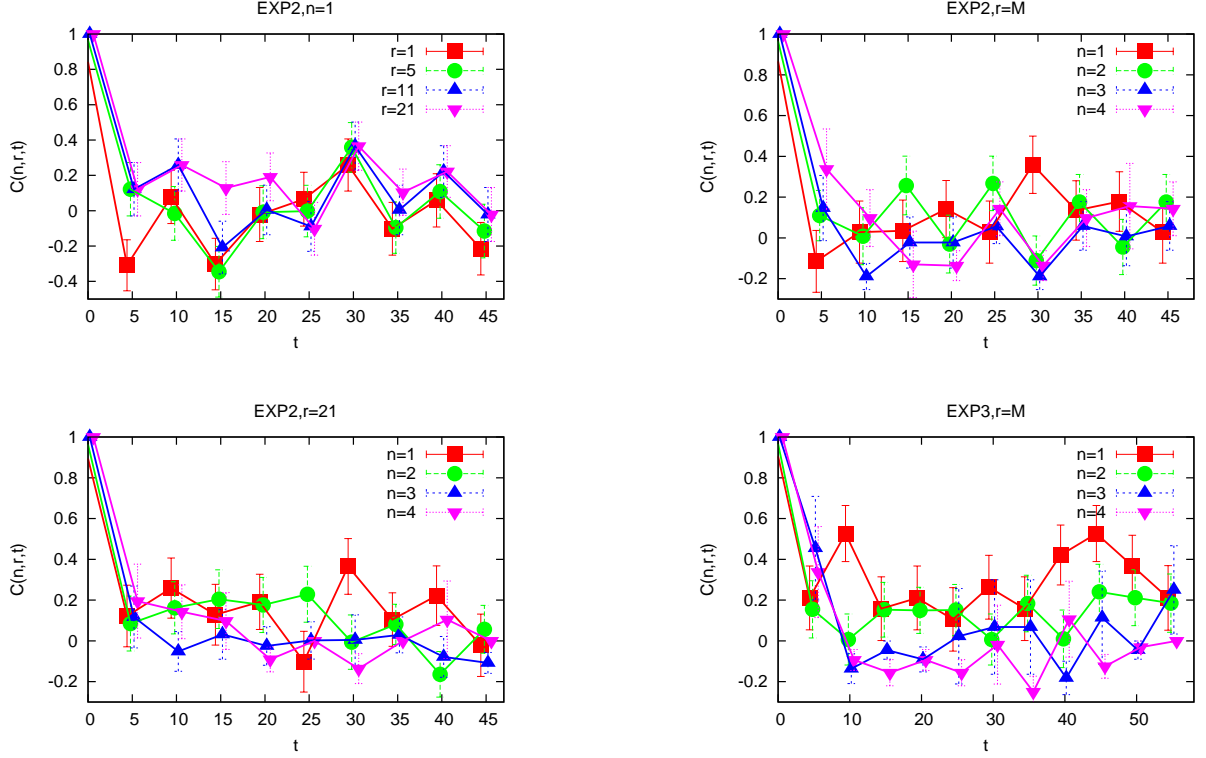


図5 実験2で  $r$  が有限の場合,  $I(n)$  の時系列データに対する  $C(t)$  を  $t$  に対してプロットした。上:  $n = 1, r \in \{1, 5, 11, 21\}$ , 下:  $r = 21$  で  $n \in \{1, 2, 3, 4\}$ 。

$t$  とともに減少するが,  $n = 1$  の場合,  $r$  の増加とともに減少の割合が減っていることが分かる。また,  $r = 21$  に固定した下図の結果から,  $n$  の増加とともに減少の割合が増えている。ただ, どの場合も十分大きな  $t$  に対して  $C(t)$  が正の定数のまわりでゆらぎ, ゼロまで減衰しているように見える。このことは,  $r$  が有限の場合,  $c = 0$  を示唆する。  $r$  が有限のモデルの理論解析の結果,  $r = 1$  の相関のある RW と同じ普遍類に属し,  $c = 0$  である [6].  $C(t)$  は  $t$  のべき関数ではなく指数関数的な減衰をする。

図6は  $r = M$  に対して  $C(t)$  をプロットしたものである。  $r = C$  と同様に実験3の  $n = 1, 2$  の場合に  $C(t)$  は減衰後定数に収束するようである。実験2の  $n = 1$  の場合,  $C(t)$  は他の  $n$  に比べて大きい, 定数に収束するとは言いえない。実験5の場合は  $C(t)$  のゆらぎが大きく結論不可能。よって,  $c > 0$  可能性が高いのは実験3の  $n = 1, 2$  の場合のみである。  $r = C$  と比較して  $c$  の値は小さく, ドミノ効果は弱いと考えられる。この差の原因は,  $r = C$  と  $r = M$  での  $q(z)$  の差異に起因する。

### 3.4 $\xi(t)/t$ の振る舞い

$C(n, r, t)$  を用いて  $\xi(t), \tau(t)$  を評価し  $t$  依存性を調べる。  $C(n, r, t)$  の  $k$  次モーメント  $M_k(n, r, t)$  を

$$M_k(n, r, t) = \sum_{s=0}^{t-1} s^k C(n, r, s) \quad (21)$$

で計算する。緩和時間  $\tau(n, r, t)$  および相関時間  $\xi(n, r, t)$  を

$$\tau(n, r, t) \equiv M_0(n, r, t) \quad (22)$$

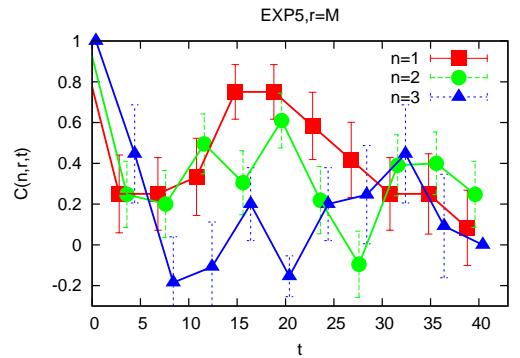


図6  $r = M$  の場合,  $I(n)$  の時系列データに対し  $C(t)$  を計算し  $t$  に対してプロットした。上から順に実験2, 3, 5の結果を示している。

$$\xi(n, r, t) \equiv \sqrt{M_2(n, r, t)/M_0(n, r, t)} \quad (23)$$

と評価する。それらの標準誤差は  $\Delta C(n, r, t)$  を用いて評価する。  $C(n, r, t)$  の標準誤差が大きくても  $M_k(n, r, t)$  は  $C(n, r, t)$  の和で評価するためそれらの相対誤差は  $t$  の平方根程度で小さくなる。ただし, 高次モーメントを用いる場合,  $t$  の大きなものの寄与が大きくなるため標準誤差はそれほど小さくならない。そのため, 緩和時間  $\tau(t)$  に比べて相関時間  $\xi(t)$  の誤差は相対的に大きい。注意として,  $\xi(n, r, t)$  の評価式で平方根の中が負になった場合,  $\xi(n, r, t)$  およびその標準誤差をゼロと評価している。

相関時間  $\xi(t)$  の  $t$  依存性を見ても,  $\xi(t)/t$  の極限值 (式 (18)) から,  $1/\sqrt{3}$  に収束すれば  $c > 0$ ,  $1/\sqrt{3}$  以下の値に収束すれば  $c = 0$  である。図7は実験3の  $r = C, M$  ので  $\xi(t)/t$  を  $t$  に対してプロットしたものである。  $n = 1, 2$  の場合,  $r = C, M$  のどちらのケースでのエラーバーの中に  $1/\sqrt{3}$  が入っている。これは  $c > 0$  を示唆する。一方,  $r = C, n = 3, 4$  の場合,  $\xi(t)/t$  は  $1/\sqrt{3}$



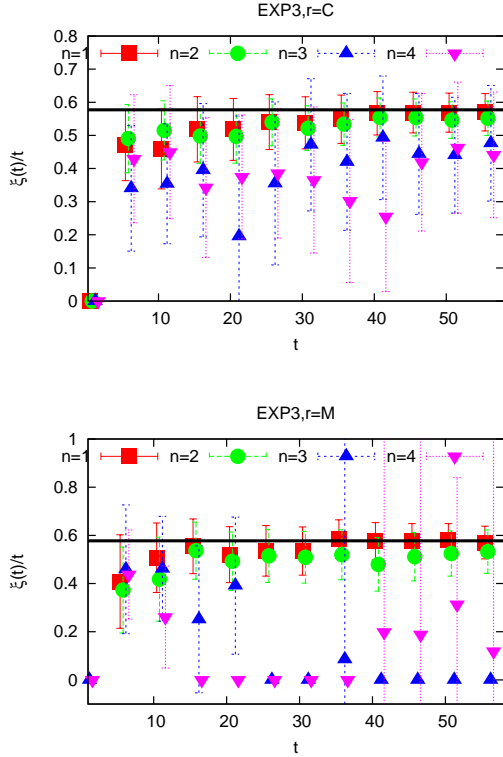


図7 実験3,  $r = C, M$  に対する  $\xi(t)/t$  を  $t$  に対してプロット。一点鎖線は  $1/\sqrt{3}$  を示している。

以下の値に収束しているように見え、 $c = 0$  を示唆する。しかし、 $c > 0$  という帰無仮説を棄却出来ない。 $r = M$  の場合もほぼ同様であるが、 $n = 3, 4$  のケースでは  $\xi(t)/t$  はほぼゼロに収束している。

### 3.5 $\tau(t)/t$ の振る舞いと $c$ の評価

$r = C, M$  での  $\tau(t)/t$  の  $t$  依存性を調べる。実験3では、式(15)より、パラメータ  $c, a, l$  と  $\tau(t)/t = c + a \cdot t^{l-1}$  の関数形を仮定して  $\tau(t)/t$  を最小二乗法でフィットし、 $c, l$  を評価する。ただし、 $c < 0$  となった場合  $c = 0$  として  $\tau(t)/t = a \cdot t^{l-1}$  で再度フィットし  $l$  を評価する。フィットする領域は  $t \in [10, 60]$  とした<sup>(注17)</sup>。図8はその結果である。 $r = C$  の場合  $n = 1, 2$  で  $r = M$  の場合  $n = 1$  で  $c > 0$  となっている。 $r = M$  で  $n = 2$  の場合、図7から  $c > 0$  の可能性を示唆した。図8のフィットの結果から  $c = 0, l = 0.78 \pm 0.01$  と推定される。このとき  $\xi(t)/t$  は  $\sqrt{l/(l+2)} \approx 0.53$  に収束し、図7の結果とも矛盾しない。これら以外の場合、 $c = 0$  である<sup>(注18)</sup>。

$r$  が有限の場合  $C(t)$  は指数関数的に減衰し、 $\tau(t)$  も  $t$  の増加とともに正の値に収束する。そこで、実験2の  $r$  有限の場合に  $\tau(t)$  をプロットしたものが図9である。 $n = 1, r \in \{1, 5, 11, 21\}$  (上図) および  $r = 21, n \in \{1, 2, 3, 4\}$  (下図) の場合の結果である。 $n = 1$  の場合、 $r$  の増加とともに  $\tau(t)$  の収束値も増加する。これは緩和時間  $\tau(t)$  が  $r$  の増加関数であることを意味する。ま

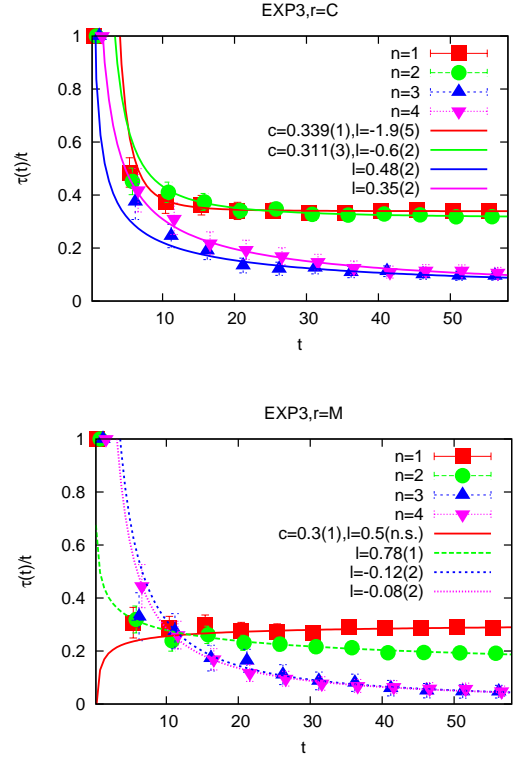


図8 実験3,  $r = C, M$  のケースでの  $\tau(t)/t$  を  $t$  に対してプロットした。各種ラインは式(15)でフィットした結果である。

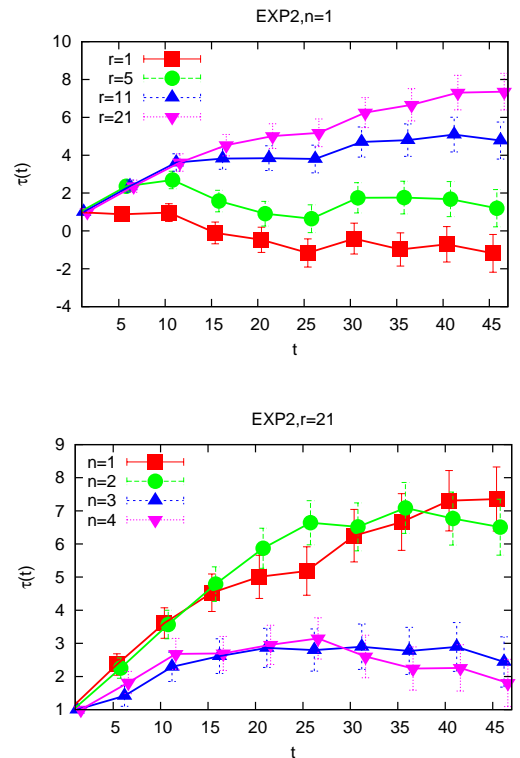


図9 実験2,  $r$  有限のケースでの  $\tau(t)$  を  $t$  に対してプロットした。

た、 $r = 21$  の場合、 $n$  とともに緩和時間が短くなることを示唆する。

(注17): 実験3,  $r = M, n = 1$  のみ有意な結果が得られなかったので  $t \in [5, 60]$  でフィットした。

(注18):  $l < 0.5$  よりノーマル拡散相

#### 4. データ解析：ミクロな法則

情報カスケード転移が起きるかどうかをミクロに決めるのは被験者の応答関数  $q(z)$  である。二択のクイズの情報カスケード実験において  $q(z)$  を変化させるのはハーダーの寄与の変化である。問題が簡単な場合、独立投票者 ( $q_* = 1$ ) が多く、 $q(z)$  は 1 に近い定数となる。問題が難しくなるとハーダーの比率が増える。ハーダーのレスポンス関数を  $q_h(z)$  と書くと、ハーダーは多数派を選択する傾向が強いが、2 つの選択肢のどちらが正解か分からないため、 $q_h(z)$  は  $z$  の増加関数で  $z = 1/2$  を中心とした点対称となる ( $q_h(z - 1/2) = -q_h(1/2 - z)$ )。これを  $Z_2$  対称性と呼ぶ。ハーダーの比率を  $p$  と書くと、正答率  $z(t) = z$  の状況で  $t + 1$  番目の被験者が正解を選ぶ確率  $q(z)$  は次のように書くことができる。

$$q(z) = (1 - p) \cdot 1 + p \cdot q_h(z). \quad (24)$$

ここでは実験データをもとに  $q_h(z)$  を推定する。ただし、個々のハーダーを個別に扱うのではなく、平均的なハーダーの  $q_h(z)$  を評価する。

時系列データ  $\{X(g, i, r, t)\}, t = 1, \dots, T(g, i, r)$  でのハーダーの比率  $p(g, i)$  は  $2(1 - Z(g, i, 0))$  とする<sup>(注19)</sup>。また、 $t + 1$  番目の被験者が参照するのは  $t$  番目の被験者までの直近  $r$  人または、過去全員の正当者数である。それを  $C_1(g, i, r, t)$  と書くことにする。

$$C_1(g, i, r, t) = \begin{cases} \sum_{s=1}^t X(g, i, r, s) & r \in \{C, M\} \\ \sum_{s=t-r+1}^t X(g, i, r, s) & r < \infty, t \geq r \\ \sum_{s=1}^t X(g, i, r, s) & r < \infty, t < r \end{cases}$$

問題は他の被験者の選択情報がハーダーの選択にどのような影響を与えたかである。ハーダーにとっては、2 択のバイアスがない場合、二つの選択肢に違いがなく、 $C_1(g, i, r, t)$  も単にある選択肢を選んだ被験者の数にすぎない。この情報がない場合、ハーダーが正解を選ぶ確率は 50%だが、 $C_1(g, i, r, t)$  という情報があった場合、それをもとに正解を選ぶ確率が 50%から変化するかどうかは過去  $t$  人の選択に依存する。ただ、ハーダーにとっては、 $C_1(g, i, r, t)$  が  $r/2$  より大きい ( $r \in \{C, M\}$  の場合、 $t/2$  より大きい) なら、それは正解である可能性が高いと推測し、その選択肢を選ぶ頻度が増えて、正解である比率も 50%を超える。逆に、 $C_1(g, i, r, t)$  が  $r/2$  ( $t/2$ ) 以下の場合、正解を選ぶ比率は 50%以下になる。ハーダーの正解確率  $q_h$  は  $C_1(g, i, r, t) = c_1$  のとき、 $r$  が有限の場合は  $\{r, c_1\}$  の関数  $q_h(r, c_1)$ 、 $r$  が  $C, M$  の場合は  $\{t, c_1\}$  の関数  $q_h(t, c_1)$  となる。また、 $q_h$  は次の  $Z_2$  対称性を持つ。

(注19): ある被験者が  $r = 0$  で回答し、 $r \neq 0$  で回答しなかった場合、 $r = 0$  の回答からその被験者の回答を削除して正答率を計算し、 $p(g, i, r)$  とすべきであるが、そうした操作は行っていない。実験 4 の解析では、 $r = 0$  と  $r \neq 0$  にも回答した被験者のデータのみで解析を行っている。他の実験と比較して実験 4 の場合は、 $r = 0$  で回答した被験者と  $r \neq 0$  で回答した被験者の人数差がかなり大きかったためである。

$$q_h(r, c_1) = 1 - q_h(r - c_1) \text{ for } r < \infty,$$

$$q_h(t, c_1) = 1 - q_h(t, t - c_1) \text{ for } r \in \{C, M\}.$$

##### 4.1 $r$ が有限の場合の $q_h(r, c_1)$

まず、 $r$  が有限の場合の  $q_h(r, c_1)$  を  $(g, i) \in I, \{X(g, i, r, t)\}, t = 1, \dots, T(g, i, r)$  を用いて推定する。 $X(g, i, r, t)$  の選択をハーダーが行った確率は  $p(g, i)$ 、独立投票者が行った確率は  $1 - p(g, i)$  である。前者の正答率は  $C_1(g, i, r, t) = c_1$  のとき、 $q_h(r, c_1)$ 、後者の正答率は  $q = 1$  より、 $X(g, i, r, t) = 1$  となる確率は

$$\Pr(X(g, i, r, t) = 1 | c_1) = (1 - p(g, i)) \cdot 1 + p(g, i) q_h(r, c_1) \quad (25)$$

となる。ここで、 $C_1(g, i, r, t) = c_1$  の条件を  $\Pr(\cdot)$  内の  $|$  の右側に明示した。 $c_1 \in \{0, \dots, r\}$  のうち、 $Z_2$  対称性から、 $c_1 \in \{(r + 1)/2, \dots, r\}$  に対して<sup>(注20)</sup> 値を推定すればよい。

$C_1(g, i, r, t) = c_1$  の条件での  $X(g, i, r, t) = 1$  の確率は期待値と等しく、期待値の推定には平均値を用いる。すると、左辺の評価には、 $t > r$  に対し  $C_1(g, i, r, t - 1) = c_1$  であったすべての  $g, i, t$  での  $X(g, i, r, t)$  のうちの 1 の比率を計算すればよい。右辺には独立投票者の寄与  $1 - p(g, i)$  も含まれるので、それを  $X(g, i, r, t)$  から引き、 $p(g, i)$  で割るとハーダーの寄与  $q_h(r, c_1)$  のみを抽出できる。そこで、 $q_h(r, c_1)$  の推定量として次の計算式を用いる。

$$\frac{\sum_{(g,i) \in I} \sum_{t=r+1}^{T(g,i,r)} \left\{ \frac{X(g,i,r,t) - (1-p(g,i))}{p(g,i)} \right\} \delta_{C_1(g,i,r,t-1), c_1}}{\sum_{(g,i) \in I} \sum_{t=r+1}^{T(g,i,r)} \delta_{C_1(g,i,r,t-1), c_1}}. \quad (26)$$

分母の  $\sum_{(g,i) \in I} \sum_{t=r+1}^{T(g,i,r)} \delta_{C_1(g,i,r,t-1), c_1}$  は、データのうち、ちょうど  $C_1(g, i, r, t - 1) = c_1$  が実現した回数を表す。 $C_1(g, i, r, t - 1) = r - c_1$  のとき、 $Z_2$  対称性から  $1 - q_h(r, c_1)$  となるので、分子に  $1 - \frac{X(g,i,r,t) - (1-p(g,i))}{p(g,i)}$ 、分母も 1 の寄与をするとする。つまり、分子は正確には

$$\frac{X(g, i, r, t) - (1 - p(g, i))}{p(g, i)} \delta_{C_1(g, i, r, t-1), c_1} + \left( 1 - \frac{X(g, i, r, t) - (1 - p(g, i))}{p(g, i)} \right) \delta_{C_1(g, i, r, t-1), r - c_1}$$

であり、分母は

$$\sum_{(g,i) \in I} \sum_{t=r+1}^{T(g,i,r)} (\delta_{C_1(g,i,r,t-1), c_1} + \delta_{C_1(g,i,r,t-1), r - c_1}) \quad (27)$$

である。分母は  $C_1(g, i, r, t - 1) = c_1$  または  $C_1(g, i, r, t - 1) = r - c_1$  が実現した回数である。 $N$  回の実現回数のとき、 $q_h(r, c_1)$  の推定値の標準誤差は、 $N$  回の試行での確率  $q_h$  に対する標準誤差と考えると、 $\sqrt{q_h(1 - q_h)/N}$  と見積もることができる。<sup>(注21)</sup>

(注20):  $r$  が奇数なら

(注21):  $q_h(r, c_1)$  の別の推定法として、 $X(g, i, r, t)$  にはハーダーの寄与が  $p(g, i)$  しかないと考え、推定量として

$$\frac{\sum_{(g,i) \in I} \sum_{t=r+1}^{T(g,i,r)} \{X(g, i, r, t) - (1 - p(g, i))\} \delta_{C_1(g, i, r, t-1), c_1}}{\sum_{(g,i) \in I} \sum_{t=r+1}^{T(g,i,r)} \delta_{C_1(g, i, r, t-1), c_1} p(g, i)}$$

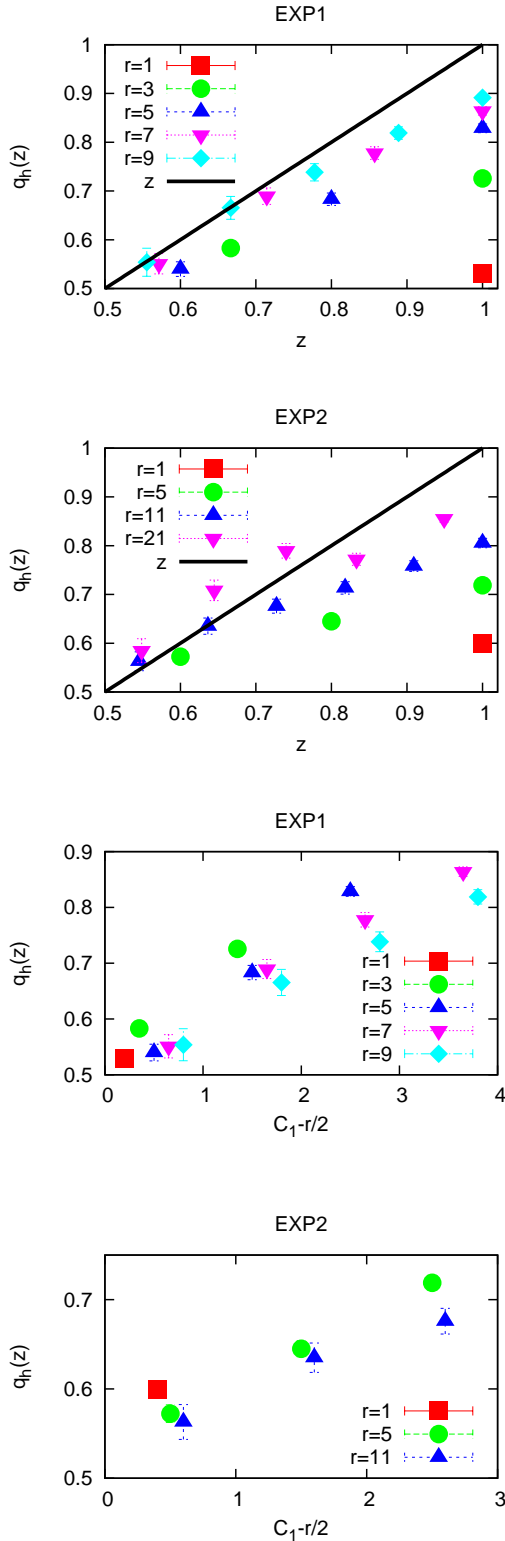


図 10  $q_h(r, c_1)$  を  $c_1/r$  (上 2 図) および  $c_1 - r/2$  (下 2 図) に対してプロットした。実験 1 では  $r \in \{1, 5, 9\}$ , 実験 2 では  $r \in \{1, 5, 11, 21\}$  とした。

図 10 は実験 1, 2 のデータを用いて  $q_h(r, c_1)$  を評価した結果を  $z = c_1/r$  および  $c_1 - r/2$  に対してプロットしたものである。 $Z_2$  対称性から  $c_1/r \geq 0.5$  の領域のプロットである。実験 2 の  $r = 21$  の場合、 $c_1/r \in \{11/21, \dots, 21/21\}$  と 11 個の点で  $q_h(r, c_1)$  を評価することになるが、データ数が少なくなりバラツキが大きくなるので、 $[0.5, 1.0]$  を 0.1 刻みの 5 個の領域に分割し、それぞれの領域での  $c_1/r$  の平均値に対して  $q_h(r, c_1)$  の平均値を計算している。

ここで、 $q_h(r, c_1)$  を  $c_1/r$  と  $c_1 - r/2$  に対してプロットした理由は、 $q_h(r, c_1)$  の関数形、特に  $r$  依存性を見るためである。つまり、ハーダーの選択は、選択肢の選択率  $c_1/r$  で決まるのか、それとも選択肢間の選択者数の差  $c_1 - r/2$  に依存するのか、という問題である。上二つの図は  $c_1/r$  という比率に対して  $q_h(r, c_1)$  をプロットしたものであり、 $r = 1, 5, 9$  (実験 1) または  $r = 1, 5, 11, 21$  (実験 2) と  $r$  を増やしていくと同じ  $c_1/r$  であっても  $r$  が大きいと  $q_h(r, c_1)$  の値が大きくなるのが分かる。これはハーダーの選択が  $c_1/r$  という比率だけではなく、 $r$  にも依存していることを意味する。同じ比率でも  $r$  が大きくなると選択者数の差が大きくなるので、この  $r$  依存性は  $q_h(r, c_1)$  が  $c_1 - r/2$  の関数の可能性を示唆する。下 2 図は実験 1 の場合に  $q_h(r, c_1)$  を  $c_1 - r/2$  に対してプロットしたものである。この図を見ると、同じ  $c_1 - r/2$  であっても  $r$  が大きいと  $q_h(r, c_1)$  は小さくなり、 $c_1 - r/2$  のみに依存するわけではないことが分かる。これらの事実から、 $q_h(r, c_1)$  の関数形は、 $r$  が小さければ  $c_1/r$  への依存性が増加するが、 $c_1 - r/2$  ほどではなく、 $r$  が十分大きくなると  $r$  依存性は消え、比率  $c_1/r$  のみで決まると考えられる。実際、 $r = C$  の場合の  $q_h(t, c_1)$  を  $c_1/t$  依存性と実験 2 の  $r = 21$  に対する  $q_h(r, c_1)$  の  $c_1/r$  依存性はほぼ同じであり、この主張を裏付ける。

#### 4.2 $r = C, M$ の場合の $q_h(r, c_1)$

$r = C, M$  の場合の  $q_h(t, c_1)$  の解析を行う際の最初の問題はその関数形の仮定である。 $r$  が有限な場合、 $q_h(r, c_1)$  は  $r, c_1$  の双方に依存し、 $c_1/r$  と  $c_1 - r/2$  の中間的な依存性であることを示した。一方、 $r$  が十分大きい場合、とくに  $r = C, M$  のように過去の  $t$  人の選択情報をすべて参照する場合、 $c_1 - t/2$  依存性は  $c_1/t - 1/2$  依存性であり、 $t \rightarrow \infty$  の極限では  $c_1/t$  が  $1/2$  を越えるか越えないかで正解か不正解というデジタル的:  $\theta(z - 1/2)$  な法則となる。ここでは、 $c_1/t$  依存性を仮定して解析を行う。もし、 $c_1/t = 1/2$  近傍で  $q_h(t, c_1)$  がデジタル的に振る舞うなら、 $c_1 - t/2$  依存性が十分大きな  $t$  に対しても残ることを意味し、そうでないなら  $c_1/t$  依存性で記述できることになる。以下の解析では時系列データ  $\{X(g, i, r, t)\}$  のうち、 $t \geq 2$  のデータを用いている。 $t$  が小さい場合、 $q_h(t, c_1)$  の  $t$  依存性が結果に反映される可能性もある。しかし、 $t \geq 11, 21$  などのデータに限って解析を行っても、時系列が短い実験 1 以外はほとんど影響を受けなかった。このことは、時系列の初期だけ  $t$  依存性が顕著で中盤以

と考えることもできる。 $Z_2$  対称性も適宜考慮する。このとき試行回数  $N$  を  $\sum_{(g,i) \in I} \sum_{t=r+1}^{T(g,i,r)} \delta_{C_1(g,i,r,t-1), c_1} p(g, i)$  と評価して標準誤差を見積もることになり、 $p(g, i)$  の分だけ試行回数が増えるためエラーバーが大きくなる。本実

験では、 $p(g, i)$  の平均は 80% から 90% だったので前者と比較してエラーバーは 10% 程度大きくなるのが分かる。以下の解析では前者の推定量と標準誤差を用いることにする。

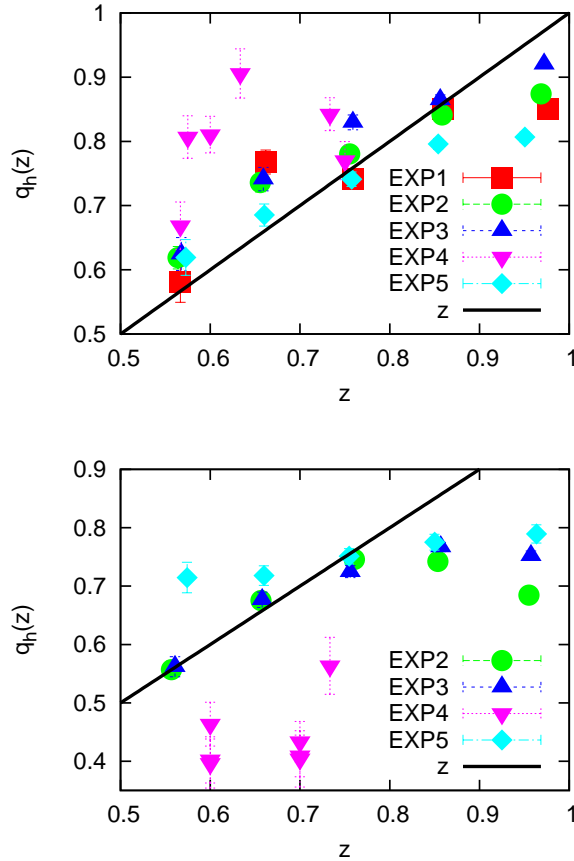


図 11 実験 1 から 5 に対し、 $q_h(t, c_1)$  を  $c_1/t$  に対してプロットした。上図が  $r = C$ 、下図が  $r = M$  の結果である。

降は  $q_h(t, c_1)$  が  $c_1/t$  に依存することを示唆している。

図 11 は、実験 1 から 5 までのデータを用いて  $r = C, M$  での  $q_h(t, c_1)$  を  $c_1/t$  に対してプロットしたものである。図 10 の  $r = 21$  と同様に、 $[0.5, 1]$  を 10% 刻みで 5 個の領域に分割し、 $c_1/t$  が入った領域のデータで  $q_h$  を評価している。x 軸の値は各領域に入った  $c_1/t$  の平均値である。まず、アンケート実験（実験 4）と実験室実験（実験 1,2,3,5）の結果に顕著な違いが見られる。アンケート実験では、過去の選択からヒントを計算する代わりに、ある大学で行った実験結果をもとにヒントを作成していると実験の最初に説明して、実験 3 のデータをもとに作成したヒントを与えている。そのためヒントに対する信用度が低く、 $r = C$  の場合は多数派の選択肢を選ぶが、 $r = M$  の場合、倍率に重きを置いて少数派の選択肢を選んだと考えられる。これはアンケート実験の問題点を示唆している。画面に表示された情報が操作されたものか、実際に他の被験者が選んだものかをどの程度信用するかで実験結果に有意な差異が見られる。一方、実験室実験 1,2,3,5 の結果は、 $r = C$  の場合、どの実験でもほぼ同じ振る舞いをしていることが分かる。特に  $c_1/t = 1/2$  の近くで傾きが 1.6 から 2 になっている。この傾きが 1 を越えるかどうか情報がカスケード転移の存在に関係しているため重要な点である [7]。

一方、 $r = M$  の場合、実験 2,3 と実験 5 の結果に違いが見られる。実験 2,3 では  $c_1/t = 1/2$  での傾きがほぼ 1 なのに対し、

実験 5 では  $c_1/t$  が約 0.57 で  $q_h$  が 0.7 を越え傾きが 2 を越える。実験 2,3 と振る舞いが異なることは実験の再現性に問題があることを意味する。実験 5 は実験 2,3 とほぼ同じ情報を被験者に与えているが、実験 2,3 での回答方法の問題を改善した実験<sup>(注22)</sup>である。一方、実験 5、 $r = M$  はデータ数をもっとも少なく、 $C(t)$  の振る舞いでも大きなバラツキが確認された（図 6）。論文 [8] では実験 2, 3 の結果： $q_h$  の傾きがほぼ 1 になることがハーダーによるゲーム理論での Max-Min 戦略の集団での採用と解釈できると結論している。しかし、実験 5 の結果はこの主張の前提が崩れていることを意味する。追試を行うことによりこの点を明らかにする必要がある。

## 5. 謝 辞

本実験の実施において、神田朋彦ヘンリー、石澤遼、熊谷直紀、辻崇史、南川晴紀（実験 1）、Ruokang Han 氏、中村文彦氏（実験 2,3）、NTT コムオンラインの藤森敬之氏（実験 4）、吉田俊介、桑波田康太（実験 5）に多大なご協力をいただきました。ここに深く感謝します。本研究は JPSJ KAKENHI Grant Number 21654054,25610109 の助成を受けています。

## 文 献

- [1] M. Hisakado and S. Mori, “Digital herders and phase transition in a voting model,” J. Phys. A, vol.44, pp.275204–275220, 2011.
- [2] B. Hill, D. Lane, and W. Sudderth, “A strong law for some generalized urn processes,” Ann. Prob., vol.8, pp.214–226, 1980.
- [3] S. Mori and M. Hisakado, “Correlation function for generalized plya urns: Finite-size scaling analysis,” arXiv:1501.0076, 2015.
- [4] G.Pólya, “Sur quelques points de la théorie des probabilités,” Ann. Inst. Henri Poincaré, vol.1, pp.117–, 1931.
- [5] R. Pemantle, “When are touchpoints limits for generalized plya urns?,” Proc. Amer. Math. Soc., vol.113, pp.235–243, 1991.
- [6] S. Mori and M. Hisakado, “Finite-size scaling analysis of binary stochastic processes and universality classes of information cascade phase transition,” arXiv:1404.4921, 2014.
- [7] S. Mori, M. Hisakado, and T. Takahashi, “Phase transition to two-peaks phase in an information cascade voting experiment,” Phys. Rev. E, vol.86, pp.026109–026118, 2012.
- [8] S. Mori, M. Hisakado, and T. Takahashi, “Collective adoption of maxmin strategy in an information cascade voting experiment,” J.Phys.Soc.Jpn., vol.82, pp.0840004–0840013, 2013.

(注22)：実験 2 では  $r = 0, 1, 5, 11, C, M$  の順に回答したため、 $r = C$  の選択と  $r = M$  の選択を組み合わせることが可能であった。実験 3 では、 $r = 0$  で 60 問回答、次に  $r = C$  で 60 問に回答し、最後に  $r = M$  で 60 問に回答としている。この場合、 $r = C$  と  $M$  の回答を組み合わせることは実験 2 より難しくなるが、それでも不可能ではない。実験 5 では、 $r = 0, C$  で回答する 80 問と  $r = 0, M$  で回答する 80 問は異なる問題であり、 $r = C$  の回答が  $r = M$  の回答に直接影響することはない。