HOKKAIDO UNIVERSITY

Title	情報化社会における「集合知」発生メカニズムの社会生態学的検討
Author(s)	豊川, 航
Degree Grantor	北海道大学
Degree Name	博士(文学)
Dissertation Number	甲第11635号
Issue Date	2015-03-25
DOI	https://doi.org/10.14943/doctoral.k11635
Doc URL	https://hdl.handle.net/2115/58772
Type	doctoral thesis
File Information	Wataru_Toyokawa.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

博　士　論　文

情報化社会における「集合知」発生メカニズムの社会生態学的検討

北海道大学大学院 文学研究科
人間システム科学専攻
豊川 航

はじめに

1951 年夏のミュンヘン市街地を，マルティン・リンダウアーはミツバチの大群を見上げながら走っていた。リンダウアーの追いかけるミツバチの群れは古巣を飛び立ち，今まさに新しい住処を目指して移動中である。彼はその引越しの一部始終を観察していたのである。リンダウアーは負傷のためスターリングラード戦を前に兵役を退くと，動物行動学の祖の 1 人であるカール・フォン・フリッシュに弟子入りしていた。以来リンダウアーはミツバチのグループ・ダイナミクスに魅せられ，どうしてもそれを，自らの手で解明しようと思っていた。特に，ミツバチの群れの引越しは謎に満ちていた。さまざまな候補地がある中からどのような基準で 1 つの巣へと引越し先を決めるのか，その決定は一体誰が行うのか，また，その決定はいつ下されるのか。彼はその年，合計 3 回のミツバチ追跡に成功した。そして彼が辿り着いた場所はどれも，飛び立つ前に多数の働きバチの尻振りダンスが指し示していた，まさにその場所であった（Seeley, 2010）。このことは，ミツバチが多数決を実現するような集団意思決定メカニズムを備えていることを示唆していた。尻振りダンスを介したどのような情報伝達メカニズムによって意見を集約し，ミツバチは多数決を実現するのだろうか。このミツバチの引越しメカニズムの全貌が解明されるまでには，リンダウアーの観察から半世紀ほども待たなければならない。しかしリンダウアーはこのとき，ミツバチの引越しが集団意思決定プロセスであることを認識するという，重要な一歩目を踏み出したことになる。

彼によるこの発見は，単なる博物学的な喜びを超えた，より深い科学的価値を持っていた。1930 年代から 1950 年代にかけては，人間の集団意思決定に関する研究が社会心理学などで数多く行われていた時期である（e.g., Lorge & Solomon, 1955; Shaw, 1932）。しかしこの時代に，ヒト以外の動物における集団意思決定の存在をデータによって示した例は他に無かった。つまり彼の観察は，集団意思決定という現象が人間だけに特有でない，もしかしたら生物界に広く見られる現象である可能性を示したのである。そこから半世紀以上経った現在では，ミツバチにとどまらず，他の社会性昆虫や，魚類や鳥類，霊長類など多数の動物種において集団意思決定の存在が実証されている（Sumpter, 2010）。もはや，集団意思決定は人間だけに見られる例外的な現象ではなくなった。集団意思決定とは紛れもなく，自然科学者が興味を持つべき，自然界に繰り返し現れるパターンの 1 種だったのである。

系統的に離れた複数の生物種において類似の行動が観察されるとき，その行動の背後には共通する何らかの適応的意義があると考えるのは，ダーウィン以来の定石である。では，生物界に広く見られる集団意思決定の適応的意義とは何だろうか。その 1 つとして，集合知（collective intelligence）の効果が考えられる。集合知の効果とは，単独で意思決定を行う場合よりも集団意思決定を採用する場合の方が，不確実性下の意思決定における成績が上昇することである。これまでに，様々な動物種において集団意思決定の集合知効果が観察されてきた。次章ではそれら実証研究を鳥瞰しながら，種を越えて見られる集団意思決定の共通性を説明する。そこで挙げる「様々な動物種」の中には，もちろん人間も含まれている。様々な動物との比較を通じて，人間社会にみられる集団意思決定への理解は深まるはずである。またそれは同時に，人間という動物の特殊さを，他の動物との相対的な

位置関係から理解することへとつながるだろう。今日われわれの社会でめざましい発展を遂げている情報伝達技術は，明らかに我々人間の集団意思決定プロセスへ大きな影響を及ぼしている。人間はいまや，集団意思決定の側面からみて非常に特殊な状況の中で生きる動物なのである。World Wide Web が世界中へ公開されインターネットが普及してから，まだわずか 1 世代時間ほどしか経っていない。この急速で，大きな環境変化に対して人間はどのように対応すべきなのだろうか。その特徴の多くが進化的時間スケールの中で形成されてきた他の動物の集団意思決定とは対照的に，この情報伝達プロセスの急速な変化は人間特有の問題だといえるだろう。

情報化がもたらした集団意思決定環境の変化は，これまでの動物研究では見過ごされてきた問いを提起させる。動物の集団意思決定に関するこれまでの多くの研究では，デモンストレーターの行動をオブザーバーが観察するという過程での情報伝達に焦点を当ててきた。つまり,「情報の受信」側の戦略を主に扱ってきたのである（e.g., Mesoudi, 2009; Laland, 2004）。つまり，オブザーバー個体がいつ・誰から・どのような情報を取得するかに分析の焦点が当てられ，デモンストレーター個体による「情報の発信」にはあまり焦点が当てられてこなかった（しかし，真社会性昆虫のコロニーにおける情報伝達は例外である: e.g., Camazine et al., 2001; Seeley, 2010）。しかし，現代の人間の集団意思決定においては，情報を発信する側の振る舞いも情報伝達に大きな影響を与えているといえるだろう。インターネット上には多くの口コミサイトや推薦システムがあり，多くの情報が能動的に発信されている。このような情報発信の多様化は，集団意思決定のダイナミクス，特にその集合知効果へ，果たしてどのような影響を及ぼすだろうか。これが，本論文全体を通じたテーマである。

興味深いのは，人間の集団意思決定へ大きな影響を及ぼしている情報伝達プロセスの変化は，それ自体われわれ人間自身の手によるものだという点である。すなわち，「情報化」という環境の変化も，結局は人間の意思決定が産み出したものであるから，環境変化と意思決定とは互いに相手の変化へと影響を及ぼし合う関係なのである。これは，「ニッチ構築」の過程だということができる（Odling-Smee, Laland & Feldman, 2003 佐倉・山下・徳永訳 2007）。したがって，本稿のテーマである「情報の共有プロセスが集団意思決定へどのような影響を与えるのか」の理解を深めることは，裏を返せば，「集団意思決定から集合知を生み出すために，われわれはどのような情報共有の仕組みを持つべきなのか」を考察するための足がかりを与えるはずである。その足がかりを築くための一歩目として，本稿にまとめられた一連の研究が位置づけられるならば，これ以上の喜びはない。

本稿は以下のように展開される。まず第一章では，集団意思決定と集合知との関係を，人間を含むさまざまな動物のシステムを例に採りながら概観し，その普遍性を探る。第二章および第三章では，第一章において他の動物との比較から浮上した人間社会に特有の問に対して，実験室での行動実験から得られたデータを示しながら議論していく。そして最後の第四章において，個々の実験結果を総合的に考察しながら，将来の研究へ展望を述べる。

第一章

集団意思決定と集合知[1]

1．集団意思決定という普遍的な現象

多かれ少なかれ我々人間の意思決定は，周囲の他者の振る舞いから影響を受けるものである。反対に，自分の行動は他の誰かの意思決定へ影響を及ぼすことがあるだろう。本章では，このような意思決定の相互作用を「マイクロ・レベルでの個体のインプットが，何らかのメカニズムによってマクロ・レベルとしての集団のアウトプットに変換され，集団全体としての秩序やパターンを生み出す現象」（亀田，1997）と捉え，集団意思決定（collective decision-making）と定義する。この定義によれば，様々な規模・種類の「意思決定のために組織された集団」（e.g.，漁場の保全・管理を担う漁業協同組合，科学における査読システム，陪審員制度，国家規模のサミット）だけが「集団意思決定」の対象となるのではない。それに加えて，「他者との間で情報のやりとりがある個人」もまた，集団意思決定を形成する主体となる。すなわち，従来の社会心理学において社会的影響（social influence）や同調（conformity），あるいは社会的学習（social learning）として扱われてきた現象も，ここでは集団意思決定を生み出すメカニズムとして捉えることになる。しかしこれは，決して強引な統合ではない。後に見ていくように，動物行動学や進化生態学における動物の集合行動（collective animal behavior）の議論と非常に整合的であり，見晴

[1] 本章は，『豊川航・亀田達也（2013）ヒトと動物の「集団意思決定」をつなぐ．動物心理学研究 第 63 巻 第 2 号 107-122. 』として公刊された論文から一部抜粋し改変したものである。

らしがよくなる（Kameda, Wisdom, Toyokawa, & Inukai, 2012)。

人間の場合，他者との相互作用の多くの場面において，言語を介した情報伝達が普遍的である。したがって人間における多くの集団意思決定場面には，少なからず人間特有の言語能力や認知能力が関与している。この事実は，集団意思決定があたかも人間だけに実現可能な，特異的な現象であるかのように感じさせるかもしれない。しかし，動物行動学の研究者らがこれまで明らかにしてきたように，動物の集合行動もまた，まぎれも無く集団意思決定と呼ぶべき現象なのである。集合行動のメカニズムは，人間の持つ高度な言語コミュニケーションや認知能力に限定されるものではない。自然界を見渡せば，アリやミツバチ，ゴキブリなどの社会性昆虫，魚類や鳥類，霊長類の群れに至るまで，動物界の様々な種が，それぞれユニークな集団意思決定メカニズムを有していることがわかる。

動物行動学者らは近年，種ごとにユニークな集団意思決定システムを統合的に捉え，その根底に共通する原理を理解しようと試みている(Conradt & Roper, 2005; Sumpter, 2010)。種ごとに大きく異なっているように見える集団意思決定システムには，どのような共通点があり，そしてまたどのようなバリエーションが存在するのか。またそれらはなぜ生じたのか。研究者らは自身が対象としている種にとらわれず，広い視野からの眺めを得つつある。

本章では，これまで様々な動物で積み重ねられてきた動物の集団意思決定の知見から，特に「集約型意思決定（combined decision-making)」と呼ばれる現象に焦点を当てる。集約型意思決定とは，個体レベルの個々の意思決定が互いに影響を及ぼし合いながら，集団レベルでの行動を創発する現象のことである（Conradt & Roper, 2005)。つまり集団意思決定といっても，集団のメンバーに共有された目的は必要なく，集団の向かう方向へ指

揮を執るような中枢的個体が存在しない状況である[1]。リーダー不在の動物の群れは，どのようにして秩序を生み出すのだろうか。動物の集約型意思決定を支配するメカニズムは，どのようなかたちで，人間の「群れ」の中でも働いているだろうか。

2．秩序を創る正のフィードバック

動物の群れはしばしば，あたかも群れ全体が 1 つの意思を持つかのように秩序立った振舞いを見せることがある。その多くを理解する上で 1 つの鍵となるのは，個体同士の情報伝達の結果生じる正のフィードバック過程（positive feedback loop）である。例えばいくつかの真社会性昆虫のワーカーは，好適な餌場を発見すると巣へ戻り，尻振りダンスやフェロモン等のシグナルを用いて，他のワーカーをその餌場へと動員する（Goss, Aron, Deneubourg, & Pasteels, 1989; Seeley, Camazine, & Sneyd, 1991）。すると今度は，動員されたワーカー自身が動員する側（リクルーター）に加わり，さらなる未稼働個体を動員する。これが動員の正のフィードバックを生み出し，結果としてコロニーの採餌努力は好適な餌場へと集中分布することになる。

このような正のフィードバックの形成は，個体間の情報伝達が作り出している。正のフィードバックは真社会性昆虫の採餌システムに限らず，多くの動物の群れの振舞いにとって重要な役割を果たしている。例えば，伝書バトは帰巣の経路を選択する際に他の個体の飛行する方向へ誘引されることが観察されている（Biro, Sumpter, Meade, & Guilford, 2006）また，いくつかの魚類や霊長類の群れの移動パターンも，群れの個体間に働く正の

[1] 社会心理学において「集団」を定義する際，個人間に共有された社会的同一性を基準とする場合が多い（e.g., Brown, 1988 黒川・橋口・坂田訳 1993）。しかしここでは動物行動学との連続性を重視し，意思決定における情報伝達の及ぶ範囲をもって集団と呼ぶことにする。

フィードバックで説明する事ができる（Ward et al., 2008; King & Sueur, 2011）。

正のフィードバックによって作られる集約型の集団意思決定は，動員の性質によって特徴づけることができる。例えば，いま環境中に A，B の２つの餌場があるとする。そして餌場 A と餌場 B それぞれに対して，他個体からの情報（i.e., 社会的情報）があり，その社会的情報の存在比が６：４だと仮定する。このとき，これから餌場を選ぼうとしている個体が選択肢 A・B を選択する確率が，それぞれちょうど 60％と 40%のときと（i.e., 選択確率の分布が社会的情報の存在比に対して線形な場合），選択肢 A を選択する確率が 60％よりも大きいとき（i.e., 選択確率分布が，社会的情報の存在比に対して非線形な場合）とで，集団レベルの振舞いに質的な違いが生じることがある。以下の節からは，個体レベルの反応の線形・非線形の違いが集団レベルの振舞いに及ぼす影響を，いくつかの動物の例をもとに概観し，ヒト社会において見られる現象との繋がりについて考察する。

２.１ 社会的情報に対して線形に反応する場合

ミツバチ（*Apis mellifera ligustica*）のコロニーは食料を集めるために数万匹のワーカーを周囲の環境へ送り出す。Seeley, Camazine and Sneyd (1991) は野外実験を通じて，ミツバチのコロニーが質の高い蜜源へいつも採餌努力を集中させることを示した。彼らの実験は，巣箱から 400m ほど離れた南北の２地点に餌場を設置し（ただし，一方の餌場は他方よりも高い濃度のショ糖を含んでいる），好適な餌場の位置関係を実験開始から４時間後に南北逆転させるという手順で行われた。結果，コロニーは採餌努力をこの環境変化に対応させ，常に蜜濃度の高い方の餌場へ多くのワーカーを分布させた。ほとんどの採餌個体は南北いずれか片方の餌場にしか訪れないので，環境に関する全般的な知識を持つ個体はほ

とんどいない。にもかかわらず，コロニー全体としては環境中の利用可能な餌場の中から相対的に質の高いものをいつも選ぶことができたのである。

この実験が示すミツバチの集合知現象のもつ驚きは，主に２つの要素からなる。１点目は，ミツバチのコロニーが，個体レベルでは部分的な情報しか持ち合わせない（i.e., 各ワーカーは複数の餌場へは訪問せず, 大抵の場合１つの餌場にしか訪れない）にも関わらず，コロニーレベルでは好適な餌場を見つけ出せることである。そして２点目は，ひとたび特定の餌場へ採餌努力を集中させた後でも，より好適な餌場を見つけたならばそちらへ採餌努力をスイッチすることができたことである。前者は，資源の質に応じて調節される尻振りダンスを用いた動員によって説明することができる。餌場を経験し巣へ戻ったワーカーは, 経験した餌場の質が高い（i.e., 糖の濃度が高い）ほど尻振りダンスを行う確率が高く，またダンスの継続時間が長くなる（Seeley & Towne, 1992)。このような質依存的な動員によって，結果的に好適な餌場に対してより強い正のフィードバックをかけることになるのである。

２つ目のコロニーレベルでの選択の可塑性は，尻振りダンスというシグナルに対する未稼働個体の反応の線形性が鍵を握っている。尻振りダンスは巣の入り口付近のダンス場で密集して行われる。しかしその場に居合わせた各未稼働個体は，複数の尻振りダンスを比較せず，個々が偶然出会った１つの尻振りダンスだけの情報を頼りに，それが指し示す餌場の方角へと巣を出発する（Seeley & Towne, 1992)。したがって，未稼働バチの餌場選択確率は，ダンス場でダンサーに出会う確率に比例すると考えられる。ここで，いまコロニーの周囲には２つの餌場 A，B が存在する状況を考える。未稼働個体が餌場 A， 餌場 B へそれぞれ動員される確率 P_A および P_B は，以下の数式で表現することができる

（Camazine & Sneyd, 1991）：

$$P_A = \frac{d_A D_A}{d_A D_A + d_B D_B} \ ,$$

$$P_B = \frac{d_B D_B}{d_A D_A + d_B D_B} \ .$$

ただし D_A と D_B は，それぞれ餌場 A, B に対するある時刻でのリクルーターの数を表し，d_A と d_B はそれぞれの尻振りダンス発生確率およびダンス継続時間を表現するための重みである。したがって両者を掛け合わせた量が，その時刻における尻振りダンスの数を表している。このモデルから言えることは，もし仮に餌場 B が餌場 A よりも質の劣る蜜源だったとしても，餌場 B を宣伝するダンサーの頻度（$d_B D_B$）が 0 でなければ，餌場 B へも採餌バチが分布するということである。この性質によって，コロニーはほとんどの採餌バチがたった 1 カ所の餌場だけに集中してしまうのを防ぐことができる。実際に Seeley, Camazine and Sneyd（1991）の実験では，貧栄養の餌場へも常に少数のワーカーが訪問していた。線形動員のもつ「1 つの餌場へ集中しすぎない」という性質が，環境変化後にコロニーの採餌努力を再分布させることを可能としたと考えられる。ミツバチの利用する花蜜や花粉の分布は時間的に変化しやすく，このように採餌努力をほどほど分散させて環境の変動へ対応するシステムは機能的意義をもつだろう（Camazine, Deneubourg, Franks, et al., 2001）。

線形な動員システムは，ミツバチだけではなく，いくつかのアリにおいても見られる。Beckers, Deneubourg, Goss & Pasteels, 1990）はトビイロシワアリ（*Tetramorium caespitum*）のコロニーに 2 種類の餌場を与える実験を行った。実験では，まず始めに 0.1 M (0.1 mol/L）の濃度のショ糖を含んだ餌場をアリーナに設置し，そして実験開始から 60 分

後に，別の場所に 1 M を含んだ質の高い餌場を設置するという操作を行った。結果，トビイロシワアリのコロニーは，はじめ 0.1 M の餌場へ採餌努力を集中させた後でも，1 M の餌場が導入されれば，そちらへ動員を切り替えることが観察された。トビイロシワアリの用いる動員システムでは，スカウト個体が直接他のワーカーを導く（group recruitment）ため，システムの機能としてはミツバチの場合と同様に，リクルーターの頻度分布がそのまま動員確率に写像されるのである。また同様に，リクルーター 1 個体あたり未稼働個体を 1 個体動員する “tandem running” という動員方法を用いる ムネボソアリ（*Temnothorax albipennis*）なども，線形な動員反応を示す（Pratt, Mallon, Sumpter, & Franks, 2002)。

２.２ 社会的情報に対して非線形に反応する場合

また Beckers, Deneubourg, Goss, et al. (1990) は，上述したトビイロシワアリの採餌実験と同様の実験を，フェロモンによる動員を行うケアリの一種（*Lasius niger*）に対しても行った。するとケアリの場合には，始めに 0.1 M の餌場へ蟻道が形成されてしまうと，その後より好適な 1 M の餌場を導入しても，採餌努力をスイッチさせることができなかった。また別の実験では，ケアリのコロニーは餌場までの最短ルートを発見できるかテストされた（Beckers, Deneubourg & Goss, 1992; Goss, Deneubourg & Pasteels, 1989)。その実験でも，餌場まで続く蟻道が一度できてしまうと，後から餌場までより短い距離で移動できる橋が導入されてもそちらに切り替える事ができなかった。このように，ケアリはひとたび蟻道を作ってしまうと，たとえ環境が変化しても，なかなかその環境変化へ応答することができない。

さらにケアリでは，２カ所ある餌場にそれぞれ等しい濃度のショ糖液を配置した場合にも，採餌努力はどちらか一方の餌場に極端に偏り，集中的な採餌が生じる（Beckers, Deneubourg, Goss, et al., 1990)。どちらの餌場に対して蟻道が作られるかは全くの偶然レベルである。このような「対称性の破れ」が生じることも，ミツバチやトビイロシワアリなどのような線形システムには観察されない性質である。

これら２つの性質(環境変化へ可塑的に対応できないこと, および対称性が破れること）をもたらす至近的メカニズムは，どちらも個体の反応の非線形性によって説明することができる。いまコロニーの周囲に餌場 A と餌場 B が存在する場合を考える。Beckers, Deneubourg, Goss, et al. （1990） は *L. niger* の未稼働個体が各餌場を選択する確率を以下の方程式でモデル化した：

$$P_A = \frac{(k + C_A)^n}{(k + C_A)^n + (k + C_B)^n} \quad ,$$

$$P_B = \frac{(k + C_B)^n}{(k + C_A)^n + (k + C_B)^n} \quad .$$

ただし，餌場 A および餌場 B へ敷かれた蟻道フェロモン濃度をそれぞれC_A, C_Bとする。また k とは, 個体が餌場探索へ向かう際にフェロモンがどの程度影響力を持つかに関するパラメータである（ただし$k \geq 0$)。k が大きくなるほどフェロモンの効果が小さくなり，選択は偶然レベルに近づくと仮定されている。そしてパラメータ n がフェロモン濃度の比率に対する反応の非線形性を表している。パラメータ n が 1 より大きければ，反応は非線形である。Beckers らの用いたケアリの場合，n = 2，n = 6 のときに実験データをよく説明できた（Beckers, Deneubourg, & Goss, 1993)。この非線形性によって，餌場のクオリティの僅かな違いは増幅され，動員される個体数はフェロモン濃度の高い方へ大幅に偏

ることになる。こうして一度安定した蟻道ができてしまうと，その後少数の個体がクオリティの高い餌場や近道を発見したとしても，既存の蟻道に敷かれた濃度の高いフェロモンを前に，せっかくの新しい宣伝はかき消されてしまうのである（フェロモンはいずれ揮発してしまう)。また，たとえ両方の餌場が等しいクオリティであっても，初期に偶然フェロモン濃度に違いが生じれば, それがいずれ個体分布の不均一さを生むことになるのである。

ただし，注意しなければならないのは，ここでの議論は動員システムにおける反応の線形・非線形性の違いを強調するあまりに，厳密さには欠ける部分があることである。例えば，Nicolis and Deneubourg (1999) は，反応の非線形性は対称性の破れを生み出す必要条件の 1 つであるが，十分条件ではないことを示している。彼らはフェロモンが揮発してしまうことを考慮した数理モデルを解析し，揮発のスピードよりも十分速く巣から未稼働個体が動員されるのであれば対称性は破れるが，個体の動員速度が遅いと対称性は破れないことを示した（Nicolis & Deneubourg, 1999; Sumpter, 2010)。つまりフェロモンは次第に揮発していくため，それを十分な速度で補わなければ頑健な蟻道は築かれないのである。その他，動物の群れにおける自己組織的な秩序形成メカニズムの詳細な議論は，Camazine, Deneubourg, Franks, et al. (2001); Sumpter (2010); Sumpter and Pratt (2003) などによくまとまっている。

3．人間の集約型意思決定

以上では，真社会性昆虫であるミツバチやアリの採餌システムに焦点を絞って述べてきた。しかし，それは何も正のフィードバックによる集約型意思決定の状況がそれ以外の動物に存在しないことを意味するものではない。たしかに動員システムというコロニーレベ

ルの形質が進化したのは，真社会性が大きく関与していただろう。しかし，採餌分布パターンなどの集団レベルの現象を形作る個体間の情報伝達それ自体は，フェロモンや尻振りダンスなどのシグナルを媒介せずとも，他個体を模倣するなどの「社会的学習」によっても生じる。この節では，ヒトにおける集約型意思決定の研究文脈を概観する。

社会心理学では，人間の示す他者への模倣や同調傾向は繰り返し報告されてきた（Asch, 1956; Davis, 1969; 1973; Festinger, 1954; Latane & Wolf, 1981; Sherif, 1936）。また，近年では行動生態学的な視点からの実験室実験が盛んに行われるようになり，より定量的に，人間の同調傾向へのアプローチがなされている。McElreath らは，実験環境下における人間の社会的学習プロセスをモデル化し，実際の参加者の社会的学習戦略を推定した（McElreath et al., 2005; 2008）。McElreath らの実験では，２本腕バンディット問題（2-armed bandit problem）という強化学習理論で古典的に扱われる不確実性下の意思決定課題が用いられた。参加者らはそれぞれがバーチャルな農場経営者であり，各農期において「小麦」と「じゃがいも」という２種類の作物のうちからいずれを植えるか選択しなければならなかった。初めのうちは，どちらの作物がより収益を上げるのかに関する知識を参加者は持っていない。しかし参加者は農期を繰り返しながら植えた作物の収益を経験し，どちらの作物が平均的により利益を上げるのかを学習することができた。加えて各参加者は，同時に参加した他の参加者が前の農期でどちらの作物を植えたかの情報を参考にすることができた。したがって参加者は，個人の経験を通じた個人学習と，周囲の参加者の行動を観察することによる社会的学習の両方を用いることができたのである。考えられる様々な学習モデルをデータへ当てはめ，その予測力を比較した結果，多くの場合で同調バイアス（conformity-biased）を示すパラメータが支持された（McElreath et a., 2005;

2008)。

McElreath ら（2005; 2008）の実験で示された同調バイアスは，社会的情報に対して非線形に反応するという意味で，2.2 節で挙げたケアリ（*L. niger*）などの示す性質と同型であるといえる。ということは，集団レベルにおいてもまた，ケアリの集団に見られたような挙動（e.g., 正のフィードバックによる選択の集中，および対称性の破れ）を生むだろうか。残念ながら，McElreath らは直接集団レベルの動態を分析してはいない。しかし，Salganik, Dodds and Watts (2006)によるインターネットを駆使した大規模な実験によって，人間社会にもケアリ同様の集団ダイナミクスが生じるということが示されている。Salganik らは，インターネット上に，あまり知られていない楽曲からなる音楽ダウンロードサイトを構築した。Salganik らは実験参加者をランダムに８つの独立した市場のいずれかに割り振り，それぞれの市場においてどの楽曲がどのように人気を得るか観察した。参加者の見る画面上には４８曲のタイトルが並べられており，参加者はそれらを自由に視聴し，自由にダウンロードすることができた。また各曲のタイトルの隣には，その曲が当該市場にて既に何回ダウンロードされているかという社会的情報が数字で示されていた。Salganik らは社会的情報を一切示さない対照群も用意し，その対照群での各曲のダウンロード回数の順位が潜在的な人々の「真の選好の分布」であると操作的に定義した。結果は，全ての市場が同じ４８曲で構成されていたにもかかわらず，８つの市場それぞれにおいて曲の流行のパターンは大きく異なっていた。特に，「真の選好の分布」の中間順位に入る曲では，各市場における順位の間には大きな違いがあった（人気の予測不可能性）。つまり，同程度の質を持つ選択肢であっても，その人気の行方は偶然の効果に支配されていていた。これは，ケアリで見られた「対称性の破れ」と類似の現象だと解釈できる。また，８つ用

意したどの実験マーケットにおいても，人気を得た曲とそうでなかった曲とのダウンロード率の格差は，社会的情報の無い対照群よりも大きくなった（i.e., スーパースター現象）。この結果は，同調による正のフィードバックが働いたことを示唆している。

以上の知見を総合すれば，人間の集約型意思決定の根底には，個人レベルにおける同調バイアスが作り出す正のフィードバックが働いていると考えられそうである。

４．集約型意思決定と集合知

社会科学において同調バイアスといえば，それが及ぼす負の側面に焦点を当てた研究の中で扱われることが多かった。負の側面とは，例えば Salganik らが実験で示したようなスーパースター現象やマーケットの予測不可能正などである。これらは，一般的にいえば「群衆行動（herd behavior）」の 1 つである（Raafat et al. 2009）。他者の行動をコピーするという行為は，たとえそうすることが個人にとって合理的な判断であるとしても，しばしば愚かな行為を集団全体に拡散しかねない（Bikhchandani, Hirshleifer & Welch, 1992）。

しかし，このような負の側面が研究者によって多く採りあげられるからといって，そのことは集約型意思決定が愚かな現象であることを意味しないだろう。第一に，人間社会で生じる群衆行動と類似の現象は，上で挙げたケアリなどの社会性昆虫においても同様に観察されるものである。動物たちは，群衆行動に陥るコストを支払ってでも余りある利益によって，集約型意思決定を進化させてきたに違いない。また進化生物学における理論研究からも，幅広い環境範囲において同調バイアスが自然選択によって進化可能でることが示されている（Boyd & Richerson, 1985; Nakahashi, 2007）。したがって，同調バイアスな

どの社会的学習が作り出す集約型意思決定は適応的なのであり，群衆行動はその副産物に過ぎないという見方ができそうである。

では，集約型意思決定が生み出す，群衆行動の危険を打ち消すほどの利益とは何だろうか。その候補の 1 つが，集合知の効果である（Krause et al., 2010）。多くの個体が情報を伝達することには，環境のもつ不確実性を低減させる効果が期待できそうである。例えばあるレストランの質について，自分 1 人の経験だけで判断するのは難しいかもしれない。1 人の経験は, その日またまシェフの調子が良かったり, あるいは悪かったりするという，偶然の効果に支配されやすいからだ。しかし，100 人が美味しいと評価しているのであれば，そのレストランが真に質の良いレストランである可能性は高いといえるだろう。このように, レストランの質に関する不確実性が社会的な情報共有によって減少するのである。

集団の利益を最大化する方向へと自然選択が働きやすい真社会性の動物では，群れとして利益が上がる集合知発生のメカニズムが進化するのはさほど不思議なことではない（Hamilton, 1964）。実際にミツバチやアリなどの真社会性昆虫たちは，情報伝達のプロセスによってこの集合知効果の恩恵を受けている（e.g., Seeley, 2010; Sasaki et al., 2013; Shaffer et al. 2013）。しかし，人間を含むその他の動物ではどうだろうか。この場合は，集約型意思決定から集合知が生まれるということは自明ではない。真社会性ではない動物の集団において集合知が発生するには，以下に挙げるいくつかの疑問へ答えなければならない。

5．浮上する問い

5.1 フリーライダーの問題

社会的学習によって集合知が生まれうる 1 つの理由は，他者がそれぞれの私的経験に基づいた最良の選択を行っている限りにおいて，それら他者の示す行動は，数ある選択肢中から質の高いものだけをふるい分けるフィルターとして働くからである。すなわち，可能な選択肢の中からデタラメに 1 つ選ぶよりも，他者の経験によってフィルタリングされた選択肢セットから選ぶ方が期待利得を高めることができるというのが，社会的学習の集合知効果である。当然，それがうまく機能するためには，集団中には自ら学習を行って選択肢をフィルタリングする個体, すなわち個人学習者がいなければならない。しかし問題は，だれがその役を担うのかである。概して，個人学習には社会的学習に比べて多くのコストがかかるだろう。メンバーのほとんどが非血縁個体からなるヒトなどの集団では，自らは個人学習をせずに社会的情報に頼るという，情報フリーライダー（i.e., information scrounger）の存在が予想される（Kameda et al., 2011）。情報伝達のなされる集団中では，情報は公共財とみなすことができる。したがってこの公共財供給をめぐるゲームが，人間の集約型意思決定の根底に存在しているのである（Bolton & Harris, 1999）。他者の生産する情報を利用するばかりで自身は情報を生産しないフリーライダーが発生することは，集合知の効果を減少させてしまうことが予想できる。Rogers （1988）は進化ゲームモデルを用いて，社会的学習の機会があっても集団の平均適応度の上昇（i.e., 集合知効果）は必ずしも導けないことを示した。しかし，仮に集団内での情報伝達が集合知効果を生み出さないのだとすると，社会的学習による文化伝達が人口増加や分布域拡大に大きく貢献してきたという素朴な観察事実の説明を困難にする。この Rogers のパラドクスと呼ばれる

問題は，理論・実証の両側から多くの関心を集めてきた（e.g., Kameda & Nakanishi, 2003; Enquist, 2007; Rendell et al., 2009）。第二章では，これまで行われてきた Rogers のパラドクスをめぐる研究をレビューしつつ，従来とは異なるモデルにおける情報フリーライダー問題を提起する。そして，実験室実験の結果を元に，フリーライダーの存在下での集合知について考察する。

５.２ 独立性と相互依存性の問題

人間の集団において集合知が発生するために乗り越えなければならないもう 1 つの問題は，独立性と相互依存性の両立である。各個体の行動がフィルターとして適切に機能し，集合知が生じるには，各個体が独立に，それぞれがベストだと思う選択を行うことが重要である。フィルターとはつまり，大数の法則によるノイズ低減の効果にほかならないからである（Condorcet, 1785/1994, Simons, 2004）。しかし，集団内に情報が伝達するということはすなわち，各個体の意思決定が相互依存的であることを意味する。個体同士が相互作用をしながら，同時に互いが独立であることなど可能なのであろうか。真社会性昆虫のミツバチを例にすると，ミツバチのワーカーは互いの尻振りダンスに動員されるという相互依存的な意思決定を行っている。しかし，ミツバチの各ワーカーは，動員されて行き着いた餌場や巣の候補地に対してそれぞれ独立した評価を行うと考えられている。個体ベース・モデルを用いたシミュレーションから，尻振りダンスを介した相互依存性と，各個体の評価の独立性とがうまく組み合わさって初めて，野外のミツバチで観察されるような集合知が生じることが示された（List et al., 2009）。

古典的な社会心理学の研究では，ヒトが意思決定において高い相互依存性を示すことが

発見されてきた（Asch, 1951; Sherif, 1936）。集団で同じ課題を解くといった場面において，メンバーの知覚的・認知的エラーが相関してしまうのである。さらにヒトの場合は，社会的・文化的な文脈が各個人への同調圧力を強めることがある（Janis, 1972; 1982）。また規範的な信念の共有も，個人の判断に相互依存性をもたらすだろう（e.g., Bateson, Nettle & Roberts, 2006）。

相互依存的な性質は，情報化と言われて久しい人間社会が集合知を応用していく際に 1 つの障壁となるだろう。しかも，コミュニケーション技術の発展は情報伝達を促進し，意思決定の相互依存性を高めるかもしれない。特にインターネット上では沢山の「口コミ」（i.e., 選択肢の評価情報）が共有され，人々の意思決定に影響を及ぼしている。しかし，人間がミツバチのように評価の独立性を保てるかどうかは自明でない。では，情報化に伴う口コミの共有は集合知へどのように寄与するのだろうか。第三章では，口コミを共有することが集合知効果へどのように影響するかを調べた実験結果を報告し，情報化が集合知へもたらす効果について考察を行う。

第二章

多腕バンディットにおける情報共有と集合知効果

1．背景

人間の集約型意思決定や文化進化の基礎は社会的学習による情報伝達である。これまでに，社会的学習の進化メカニズムや適応的意義については数多くの研究がなされてきた（e.g., Feldman, Aoki, & Kumm, 1996; Boyd & Richerson, 1985; Kameda & Nakanishi, 2002; 2003; McElreath, Bell, Efferson, Lubell, Richerson, & Waring, 2008; Morgan, Rendell, Ehn, Hoppitt, & Laland, 2012)。そこで問われてきた重要なテーマの 1 つが, Rogers のパラドクス（Rogers, 1988）に関するものである。 Rogers のパラドクスとは，以下に述べるように，「社会的・文化的な学習は個体の適応度を増す」という素朴な直感が, ゲーム理論による解析からは支持されないというパラドクスである。

Rogers（1988）は簡単なゲーム理論モデルによって, 個体学習戦略と社会的学習戦略という２戦略の進化ゲームを解析した。ここで２つの母集団を考えよう。１つは個体学習戦略のみからなる母集団であり（母集団１），もう１つは個体学習戦略と社会的学習戦略の両方が存在しうる母集団である（母集団２）。Figure 2-1 に，それぞれの学習戦略の平均適応度が示されている。X軸はその母集団における社会的学習戦略の比率である。個体学習戦略は,社会的学習を一切行わないと仮定されており，その適応度は社会的学習戦略の頻度に依存せず一定である（母集団 1)。一方, 母集団２においては，社会的学習戦略が存在しうる。社会的学習戦略の適応度は, 社会的学習戦略それ自身の頻度に負の依存性をもつ（i.e.,

社会的学習戦略が増えすぎるとうまくいかない)。比喩的には，「試験でのカンニング（社会的学習）が成功するためにはまじめに勉強してくる者（個体学習戦略）の存在が必要だ」と考えても良い。こうしたトレード・オフの結果, 2 つの学習戦略を考慮する母集団 2 では，個体学習戦略と社会的学習戦略の共存する進化的安定状態（Evolutionarily Stable State）が生まれる。興味深いことに，そこでの平均適応度は，個体学習のみからなる母集団 1 の平均適応度と等しくなる。すなわち，社会的学習の能力があるからといって，個人学習を上回る適応度は実現されないという予測が導かれてしまう。これは，人間が社会的学習によって未知なる不確実な環境へ適応してきたという従来の素朴な認識に反する予測であるため，パラドクスとして捉えられてきた。

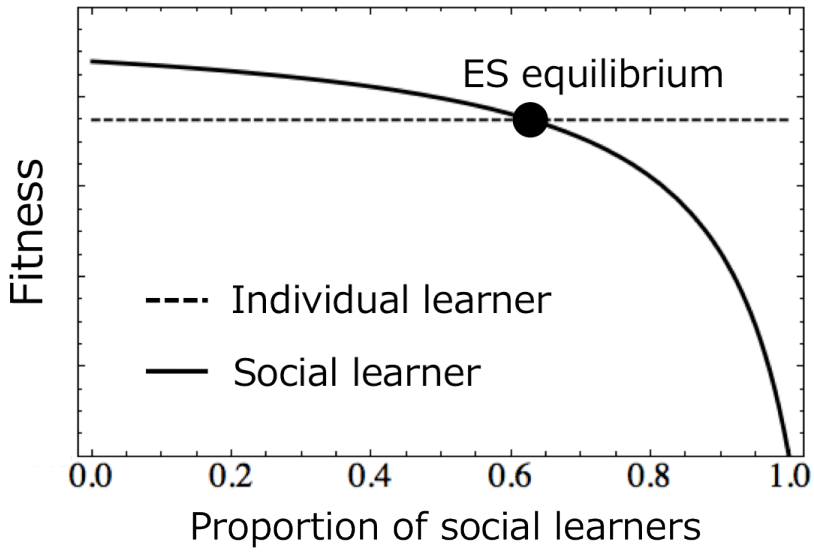


Figure 2-1. Illustration of Rogers' (1998) model. Fitness of individual learners and social learners are shown as a function of the proportion of social learners in the population. Social learning provides no fitness advantage to the population. When individual learners and social learners are at the evolutionary stable (ES) equilibrium, their average fitness is the same as that of individual learners in a population without social learners.

モデルの予測が現実の現象を上手く説明できない場合，そのモデルにはなにか重要な要素が欠けているか，あるいはモデルの前提が間違っていると考えられる。よって，ヒトの社会的学習に関する進化人類学や行動生態学における研究では，Rogers モデルに欠けていた重要な要因は何なのかを解明することが 1 つの主要なテーマとなってきた（e.g., Enquist, Eriksson, & Ghirlanda, 2007; Kameda & Nakanishi, 2003; Rendell, Fogarty, & Laland, 2010）。Rogers のモデルでは，環境へ適応するための情報は全て個体学習者が生産し，他者を模倣するだけの社会的学習者は情報生産について何の貢献もしない。すなわち，社会的学習者がいるからといって集団内にプールされる情報の質が上昇するなどといったことは，何も起こらないモデルなのである。Rogers のパラドクスが消えるためには，社会的学習による情報伝達が意思決定の不確実性を低減させるような，何らかのメカニズムを考慮する必要があるだろう。そうしたメカニズムの 1 つの例として，「個体学習で得られる情報が不確実な場合に他者を模倣する」という条件付き学習戦略を考えることができる（Kameda & Nakanishi, 2003）。この条件付き学習戦略を採用すると，比較的確かな情報を得られる個体が個体学習によって情報を生産し，情報精度の低い個体は社会的学習をすることになるので，全員が個体学習を行う場合よりも質の良い情報だけが集団中に広まることになる。

このように，Rogers のパラドクスを解消させるメカニズムの多くは，社会的学習が不確実性を低減させるという集合知効果を利用するものである。しかしこれまで，Rogers のパラドクスに関する研究は，社会性動物の群れにおいて見られる集合知の研究とはあまり結びついてこなかった（Grüter & Leadbeater, 2014）。動物の集合知研究からの視点を取り入れることによって，従来の社会的学習研究からは見えてこなかった重要な要素が浮かび

上がってくる。

それは，動物が学習し意思決定するプロセスにおいて普遍的に存在する 「探索と収穫のジレンマ」である。動物の意思決定のパフォーマンスは，多く場合，探索と収穫の配分の取り方に大きく依存する。動物が試行錯誤による学習によって環境に適合した行動を獲得するためには，まず様々な選択肢を試すこと，すなわち探索することが重要である。しかし，動物の個体がその活動に費やせる時間やエネルギーには限りがあるため，いつまでも探索し続けるのは得策ではない。探索で得た知識を活用する前に寿命が来てしまえば，探索した甲斐がなくなるからである。かといって，全く探索しなければ良い選択肢を利用できる見込みは低いだろう。よって，限られた時間やエネルギーを，探索に費やす部分と収穫に費やす部分とに適当に配分することが，パフォーマンスを最適化するための重要な課題となる。例えばミツバチによる集団採餌では，いつまでも全てのワーカーが探索するのではなく，徐々に好適な餌場へと集中（収穫）が生じることで，コロニーレベルで効率の良い採餌が実現される（Camazine & Sneyd, 1991）。真社会性昆虫であるミツバチは，探索と収穫への時間・エネルギーを適切に配分することによって集合知を実現していると解釈できる。

しかし，真社会性でない人間の場合，探索と収穫のバランスを集団レベルで最適化できるかは自明でない。なぜならば，集団中には探索を怠るフリーライダーが生じる可能性があるからである（Bolton & Harris, 1999）。果たして，探索と収穫のジレンマを含む不確実性下の意思決定場面において人間はどのような社会的学習戦略をとるのだろうか。また社会的学習によって，集合知は生まれるのであろうか。

本研究では，意思決定科学において 「多腕バンディット問題（multi-armed bandit problem; 以下，MAB 問題）」と呼ばれる状況に焦点を当てる。MAB 問題とは，探索と収

穫のジレンマを含む不確実性下の意思決定問題を定式化したものであり，具体的には，スロットマシンを有限回だけプレイ（i.e, 選択）して利得の最大化を目指すギャンブラーをイメージすればよい。スロットマシンは複数台あり，利得の得やすい台から得にくい台まで，それぞれの設定が異なっている。しかしギャンブラーはあらかじめ台の設定を知らず，設定の良し悪しは実際にプレイすることで得られる報酬量から学習するしかない。利得を最大化するためには高設定の台をなるべく多くプレイしたいが，高設定の台を知るためには様々な台を探索的にプレイしてみなければならない。したがって，有限の選択機会を探索（i.e., 新しい知識の獲得）と収穫（i.e., 現在の知識の利用）とにうまく選択を配分することが重要な課題となる。MAB 問題は，動物の採餌場面をはじめ，新規技術の採用や，病気に対する治療法の選択，投資先の決定など，多くの意思決定場面を抽象化した問題として，行動科学や機械学習の分野においてよく用いられるパラダイムである（e.g., Daw et al., 2006; McElreath et al., 2008; Sutton, Andrew, & Barto, 1998 三上・皆川訳 2000）。本研究では，古典的な MAB 問題を社会的学習が可能な集団場面へと拡張する。すなわち，同一の MAB 問題に直面する複数の個人がおり，各人が互いの選択内容とそこから生じる利得の情報を共有できる場合を考えるのである。各個人は自分が選択したマシンから得られる利得の情報に加え，他人の選択行動とそこで得られた利得情報をも観察することができる。

これまでの多くの社会的学習研究で「個人学習のコスト」として考慮されていたものは，MAB 問題の状況では探索に伴う機会コストとして組込まれている。探索とは，将来の収穫の質を高めるために不確実な選択肢を試してみる投資であり，短期的には損になる行動である。したがって，社会的学習が可能ならば，他者が探索を行ってくれる限り，自分は探索

せずに収穫に専念していた方が目先の利得は高まる。その結果, 利己的な個体ならば探索はほとんど行わなくなると予測される（Bolton & Harris, 1999）。

本研究では, MAB 問題において社会的情報を共有することの影響を実験室実験によって検討した。社会的情報の存在下で個人の「探索-収穫の配分」はどのように変化し，それが課題のパフォーマンスへどのように影響するのか，という２点を経験的に知ることで，今後の研究の基礎となる知見を得ることを目的とする。

２．方法

２.１ 実験参加者

北海道大学の学生 94 名（女性：25 名, 平均年齢（± 1 *s.d.*）：19.8 (±1.54) 歳 ）が実験に参加した。実験参加者は，北海道大学文学研究科行動システム科学講座が毎年春に行っているアンケートで実験参加の意思を表明した学生の中から無作為に抽出され，任意で参加した。実験を実施する前に，すべての参加者から実験参加への同意を書面上で得た。参加者へは，実験終了後に報酬が支払われた。報酬額は実験中の課題の成績に依存して参加者ごとに決まることが，実験開始前に参加者へ伝えられた。

２.２ 実験手続き

実験は，北海道大学社会科学実験研究センターで実施された。６～８名の参加者が１つのセッションへ参加し，そのうちランダムに割り当てられた５名が集団条件へ，残りが個人条件へと参加した。各参加者はそれぞれ１台のコンピュータの前に着席した。座席間に

は仕切りが設けられ，参加者が互いのコンピュータの画面を覗けないようになっていた。課題の開始前に，各参加者は課題の説明のためのスライドショーをそれぞれのコンピュータ画面上で視聴した。実験には z-tree（Fischbacher, 2007）ソフトウェアが用いられ，参加者はすべての課題を各モニター上で行った。

本実験で参加者が行った課題は，利益率の異なる 30 の選択肢からの選択を繰り返しながら利得を稼ぐという，30 本腕バンディット問題（30-armed bandit problem）であった（Figure 2-2）。各参加者には合計 100 ラウンド（i.e., 100 試行）の箱を選択する機会が与えられ，すべての選択を通じて得たポイントの総額が実験終了後に日本円に換算され報酬額として参加者へと支払われた。1 回のラウンドは 2 つのステージ：「選択ステージ」と「確認ステージ」で構成されていた。まず選択ステージにおいて，参加者は画面上の 30 個の箱から 1 つをクリックし，任意の箱を選択した。次の確認ステージにおいて，先の選択において獲得したポイントの額が提示された。参加者が獲得するポイントは，選択した箱の質に依存して異なっていた。ポイントは，区間［*Min*, *Min* + 150］の定常な一様確率分布から整数値として出された。自然界に多く見られる，「質の高い資源ほど稀である」という資源分布の様子を再現するため，箱の質の分布は以下のように設定された：11 個の箱が最も質の低いカテゴリー1 であり，Min = 0; また 8 個の箱はカテゴリー2 で Min = 15; 5 個の箱はカテゴリー3 で Min = 30; 3 つの箱はカテゴリー4 で Min = 45; 2 つの箱はカテゴリー5 で Min = 60; そして 1 つの箱が最も質の高いカテゴリー6 で Min = 75 であった。これら選択肢の空間配置は同一のセッションでは等しく，また試行中は固定されていた。ただし，異なるセッション間においては，その配置は異なっていた。

実験の開始段階において参加者は，（a）多数の選択機会があること，（b）選択から得ら

れるポイントは各選択肢に固有の確率分布に従ってランダムに得られること，そして（c）実験で獲得した累積のポイントは実験の最後に１０分の１倍され，その値が１０円単位に繰り上げられて現金で支払われることを，それぞれ知らされていた。

したがって実験の開始段階では，ポイントの出される確率分布や，箱の配置，そして具体的な試行数に関する知識を，各参加者は持っていなかった。しかし，ラウンドを重ねるに連れて，確認ステージで表示されるポイント額を通じた学習が可能であった。

実験の所要時間は，説明のためのスライドショーと謝金の支払いを含めて合計で約 75 分であった。

２.３ 集団条件

集団条件は，1 グループあたり 5 名が参加する形式で実施された。集団条件には計 70 名の参加者（ = 14 グループ）が参加した。

集団条件では，確認ステージにおける各自の獲得ポイントの教示に加えて，グループに所属する他のメンバーの前試行における行動からも情報を得ることができた（Figure 2-2B）。集団条件の選択ステージにおいて，モニター上の各箱には「前試行において何名から選択されたか」と「前試行において平均何ポイントが出されたか」についての情報が表示されていた。ただし共有されたのは情報のみであり，それぞれの獲得したポイントは参加者ごとに個別に加算された。選択ステージと確認ステージは各参加者が任意のタイミングで進めることができたが，メンバー全員が確認ステージを終えるまでは次のラウンドへ進めなかった。

2.4 個人条件

対照実験として，24 名の参加者を個人条件に割り当てた。集団条件と異なり，個人条件では獲得ポイントに関する情報は完全に私的情報であり，他の参加者との情報共有はなかった。それ以外の手続きは集団条件と同一であった（Figure 2-2A）。

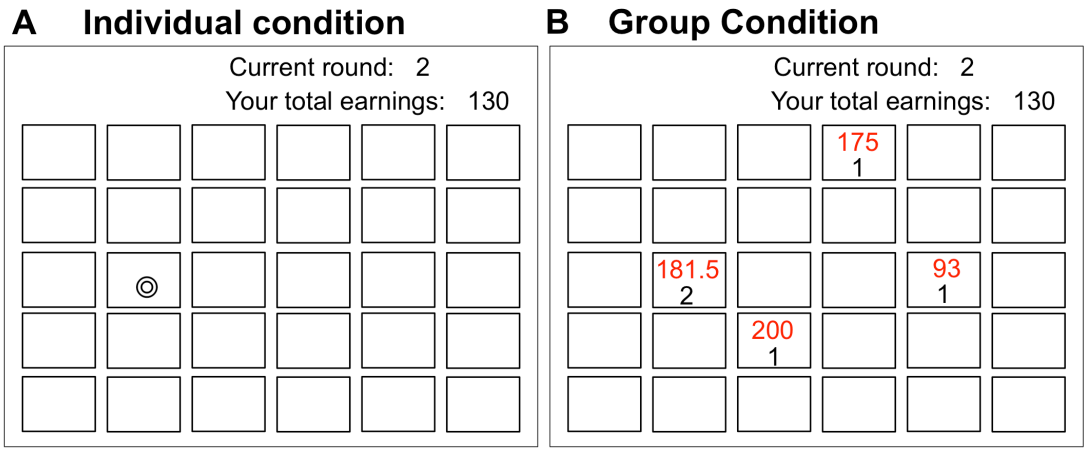


Figure 2-2. Example displays of choice stage. A: An example display of individual condition. The circle shown on a box (i.e., option) indicates the participant's choice in the preceding round (round 1 in this example). B: An example display of group condition. Social information about all 5 group members' choice behaviors in the preceding round was made available to each participant. The lower black number indicates how many members chose that option in the preceding round. In addition, participants were informed of average payoffs (upper red numbers) that members had earned from their chosen options in the preceding round.

2.5 学習モデル

社会的な情報共有がもたらす行動の変化を定量的に分析するため，以下のような学習モデルを考える。

個人学習モデル

個人学習のモデルとして，基礎的な強化学習モデルを用いる（Sutton & Barto, 1998 三上・皆川訳 2000）。各個人は，ラウンド t+1 における選択肢 i の価値 $A_{i,t+1}$ を以下のように更新する：

$$A_{i,t+1} = A_{i,t} + \left(\frac{1}{k_{i,t}}\right) * \left(r_{i,t} - A_{i,t}\right),$$

ただし，$r_{i,t}$ はラウンド t において選択肢 i から観察された利得ポイント，$k_{i,t}$ はラウンド t を終えた時点における選択肢 i の累積選択回数である。この更新式は結局，当該個人が選択肢 i から観察した過去全てのポイント値の平均を算出している（標本平均規則; sample-averaging rule）。本実験で用いた課題のように定常な（i.e., 時間的環境変化のない）学習課題においては，学習率（$1/k_{i,t}$）が学習の進行に反比例し低減していくプロセスは不自然な仮定ではない（Sutton & Barto, 1998）。なお，選択されなかった選択肢の価値は不変であると仮定した。また全ての選択肢に関して，価値の初期値 A_0 は 0 と置いた。

各選択肢の価値は，以下の関数（soft-max choice rule）に従って各選択肢の選択確率へと変換される。すなわち，ラウンド t+1 において箱 i を選択する確率は以下のように表される：

$$Pr(i|\overline{\boldsymbol{A}}_t, \tau)_{t+1} = \frac{\exp\left(\frac{1}{\tau} * A_{i,t}\right)}{\sum_m \exp\left(\frac{1}{\tau} * A_{m,t}\right)},$$

ただし，$\overline{\boldsymbol{A}}_t$ は全ての選択肢に対するラウンド t 時点でのすべての価値の集合である。また τ は選択がどのくらい探索的かを調整する「温度」と呼ばれるパラメータである。$\tau \to 0$ のときは価値が最大の選択肢（$max(\overline{\boldsymbol{A}}_t)$）のみを選択し，$\tau \to \infty$ のときは価値とは無関係に全ての選択肢から一様な確率でランダムに選択する。

社会的学習モデル

集団条件における社会的情報を学習に反映させるため，上述の個人学習モデルを以下のように改変した。

まず，個人条件では利得は私的情報であるため，$\overline{\boldsymbol{A}}_t$ の更新は個人ごとに異なったが，集団条件では 5 人のメンバー全員の得た利得が公共情報となって共有されるため，$\overline{\boldsymbol{A}}_t$ の更新は集団内で共通であると仮定した。すなわち，自分が選択した箱以外についても，他のメンバーが選択した箱に関しては，その利得情報に基づいた標本平均規則によって評価値が更新された。

加えて，集団条件においては，前試行における各選択肢の選択頻度（i.e., 何名から選択されたか）を知ることができた。つまり，頻度依存型社会的学習が可能な状況である。これを反映させるため，選択確率は以下のモデルを用いた：

$$Pr(i|\overline{\boldsymbol{A}}_t,\tau,\lambda,\theta)_{t+1} = (1-\lambda) * \frac{\exp\left(\frac{1}{\tau} * A_{i,t}\right)}{\sum_m \exp\left(\frac{1}{\tau} * A_{m,t}\right)} + \lambda * \frac{n_{i,t}^{\theta}}{\sum_m n_{m,t}^{\theta}},$$

ただし，$\lambda\,(0 \le \lambda \le 1)$ は社会的情報への重みであり，$n_{i,t}$ はラウンド t において箱 i を選択している人数を表し（i.e., 社会的頻度），また θ は社会的頻度への依存性を調節している。$\theta = 0$ のとき，頻度に依存しないランダムな選択となる。$0 < \theta < 1$ のときは，ゆるやかな正の頻度依存性が現れる（Claidière, Bowler & Whiten, 2012）。$\theta = 1$ のときは，バイアスのない模倣である（i.e., 社会的頻度に対する線形な反応）。$\theta > 1$ のときは，多数のメンバーから選択される箱はその人気の比率以上にさらに選択されやすくなるという同調バイアスが出現する。また θ が負であれば，負の頻度依存性を持つことになる。

MCMC によるパラメータ推定

上述の学習モデルのパラメータを，マルコフ連鎖モンテカルロ法（以下 MCMC）によって

推定した。推定したパラメータは，個人学習モデルでは温度パラメータ τのみ； 社会的学習モデルでは温度 τ, 社会的情報への重み λ，そして頻度依存性 θ である。個人学習モデルの推定には 100 ラウンド×24 名分の計 2,400 個の行動データを用いた。社会的学習モデルの推定には 100 ラウンド×70 名分の計 7,000 個の行動データを用いた。MCMC の詳細な手続きは付録 2-1 を参照のこと。

2.6 探索量の比較

探索量の定義

探索におけるフリーライダーの存在を定量的に調べるため，個人がどの程度の探索を行ったのかを，学習モデルに組み込まれた選択肢の価値 $\overline{\boldsymbol{A}}_t$ を用いて定義する。価値が最大である箱 $max(\overline{\boldsymbol{A}}_t)$ を選択することを「収穫」と定義し，それ以外の箱を選択することを「探索」と定義する。100 ラウンドを通じて何回探索を行ったかを各個人について算出し，それぞれの探索量とした。

また，上述の計算論モデルを用いた探索量に加えて，シャノンのエントロピー（H）を用いて「選択のゆらぎ」の大きさを算出する。エントロピーは以下の式で表される：

$$H_{t,t+9} = -\sum_{i=1}^{M_t} \frac{count_i}{10} \ln\left(\frac{count_i}{10}\right),$$

ただし，M_t はラウンド t からラウンド t+9 までの 10 ラウンド中に選択した箱の種類であり， $count_i$ はその 10 ラウンド中に箱 i を選択した回数である。仮に 10 ラウンドの中で一種類の箱だけを選択した場合には $H = 0$ という最低値をとり，10 ラウンドを全て異なる箱の選択に使った場合には最大値 $H = 2.30$ をとる。H の値が大きいほど選択の種類が多様であり，選択は大きくゆらいでいたと解釈できる。

階層ベイズ法による分析

個人条件と集団条件とで個人あたりの探索確率（q）に差が見られたかどうかを，階層ベイズ法によって分析した。階層ベイズ法とは，個人ごとのランダム効果や集団ごとのランダム効果を考慮したモデルをベイズ推定する方法である。モデルは以下の混合ロジスティックモデルで表される：

$$q_{ij} = \frac{1}{1+\exp(-z_{ij})} \ ,$$

ただし，$z_{ij} = \beta 1 + \beta 2 * condition(i) + r(i) + rg(j)$,

$r(i) \sim Normal(0, s),\ rg(j) \sim Normal(0, sg)$.

パラメータは，データ全体に共通する大域的パラメータである切片（$\beta 1$）と条件（個人/集団）の効果を表す係数（$\beta 2$）の他に，個人ごとのランダム効果を表す $r(i)$，集団ごとのランダム効果を表す $rg(i)$ という２つの局所的パラメータがある。また$condition(i)$ はダミー変数（個人条件 = 0，集団条件 = 1）である。このモデルでは，２つのランダム効果がそれぞれの階層事前分布（平均ゼロで標準偏差がそれぞれ s および sg である正規分布）から生成されると仮定する。

また，個人の選択のゆらぎ（エントロピーH）についても，個人条件と集団条件とで差が見られたか比較する。分析には，以下の混合線形モデルを用いた：

$H_{i,block(j)} \sim Normal(\mu_{i,block\ j}, v)$,

$\mu_{i,block(j)} = intercept + \beta_1 * block(j) + \beta_2 * condition(i) + \beta_3 * block(j) *$

$condition(i) + r(i) + rg(group(i))$,

ただし，$block(j)\ (j = 1,2,3,\ldots,91)$ はラウンド j からラウンド $j+9$ までの 10 区間を一区切りにしたもの，$condition(i)$ は個人 i の所属する条件のダミー変数（$individual = 0$; $group$

= 1）である。また，$r_i(i)$ は個人ごとのランダム効果を，$r_g(group(i))$ は個人 i の所属するグループの効果を，それぞれ表している。興味があるパラメータは，条件の固定効果を示す β_2 である。ベイズ推定の詳細な手続きは付録 2-2 を参照のこと。

2.7 集合知効果の分析

各参加者の課題での成績として，選択肢の質（カテゴリー1~6）を用いた。隣り合うカテゴリー間での得られるポイントの期待値は 15 ずつと一定であるため，カテゴリーの表す箱の質は間隔尺度として扱うことができる。各参加者の 100 試行全ての選択を集計した合計のスコアを算出し，各個人の成績とした。スコアの最低値は 100 試行全てでカテゴリー1 の箱を選び続けた場合の 100 であり，最大値は全ての試行でカテゴリー6 を選び続けた場合の 600 である。

個人条件と集団条件とで，個人の成績（$Score_i$）に変化が見られたかどうかを分析するために，以下のモデルを階層ベイズ法によって分析した（分析の手続きについては付録 2-3 参照）：

$$Score_i \sim Normal(\mu_score_i, v_score),$$

$$\mu_score_i = intercept + \beta * condition(i) + rg(group(i)).$$

3．結果

3.1 学習モデルのパラメータ

Table 2-1 は，個人学習モデルと社会的学習モデルそれぞれのパラメータ事後分布を示している。Gelman-Rubin 統計量が 1.1 を下回っていることから，MCMC は収束したといえる。

個人学習モデルの温度パラメータ（τ）の 95%信用区間は社会的学習モデルの温度パラメータの 95%信用区間を上回っていた。このことから，社会的学習モデルの温度パラメータは個人学習よりも有意に低かったと解釈する（以下の分析においても同様に，95%信用区間に基づき結果の有意性を解釈する)。温度パラメータが低いということは，価値（A_t）が高く見積もられた選択肢をより選択しやすかったということである。すなわち，集団条件においては，個人条件と比較して個人の探索行動が少なかったことが示唆された。

社会的学習モデルにおいて，社会的情報への重みλは有意に 0.5 を上回っていた（95%信用区間の下限は 0.72)。すなわち，集団条件における個人の選択は，個人学習よりも社会的頻度情報からより大きな影響を受けたことが示された。

また，社会的頻度依存性については，θ の 95%信用区間が区間（0，1）の範囲に入っていることから，参加者全体の傾向としては，社会的学習は弱い正の頻度依存（Claidière, Bowler & Whiten, 2012）だったことが示唆された。

Table 2-1. MCMC result of the learning models

	Parameters	Mean	SD	Quantiles			Gelman-Rubin Statistics	
				2.5%	50.0%	97.5%	median (upper C.I.)	Effective sample size
Individual learning	τ	49.13	1.07	47.1	49.1	51.2	1.00 (1.01)	2621
Social learning	τ	13.14	0.63	12.0	13.1	14.4	1.00 (1.01)	1000
	λ	0.74	0.01	0.72	0.74	0.76	1.01 (1.02)	901
	θ	0.44	0.006	0.43	0.44	0.45	1.00 (1.01)	1257

３.２ 探索量の比較

Figure 2-3 は，各条件における探索量の分布を表している。階層ベイズモデル推定の結果，探索確率への条件の有意な影響はみられなかった（Table 2-2; β_2)。

Figure 2-4 は，探索量（選択のエントロピー）の時間変化を示している。階層ベイズ推定の結果，ブロックの効果（β_1）および条件の効果（β_2）はともに有意に負であり，課題が進むに連れてエントロピーが有意に減少したこと，および集団条件において個人条件よりも有意にエントロピーが低かったことがそれぞれ示された（Table 2-3)。また，ブロックと条件との交互作用が有意に 0 より大きかったことから，試行が進むことによるエントロピーの減少速度（i.e., Figure 2-4 の傾き）は，個人条件のほうが集団条件よりも大きかったことが示唆された（Table 2-3)。

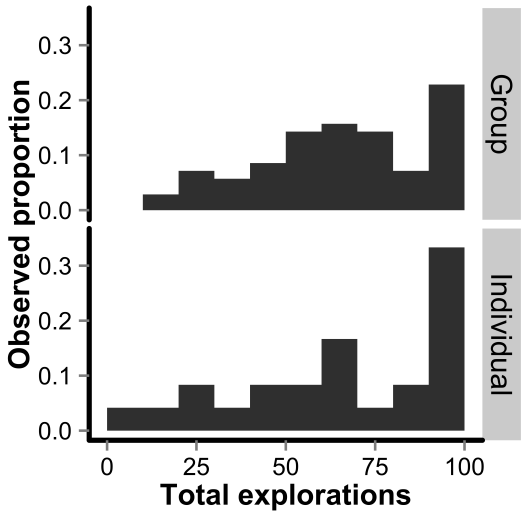


Figure 2-3. Histograms of participants' exploratory choices. The x-axis refers to the total number of "exploratory" choices (when a participant selected options other than the one that had yielded the largest cumulative payoff in her/his own experience so far) out of 100 rounds. The y-axis refers to observed proportions of participants with a given exploratory-choice frequency in each condition.

Table 2-2. MCMC result of the exploration model

Parameters	Mean	SD	Quantiles			Gelman-Rubin Statistics	
			2.5%	50.0%	97.5%	median (upper C.I.)	Effective sample size
β_1	1.01	0.57	-0.079	1.00	2.11	1.03 (1.07)	735
β_2	0.0046	0.66	-1.35	0.014	1.28	1.03 (1.07)	666
s	1.29	0.12	1.07	1.29	1.56	1.00 (1.00)	4365
sg	1.14	0.28	0.68	1.11	1.30	1.00 (1.01)	3034

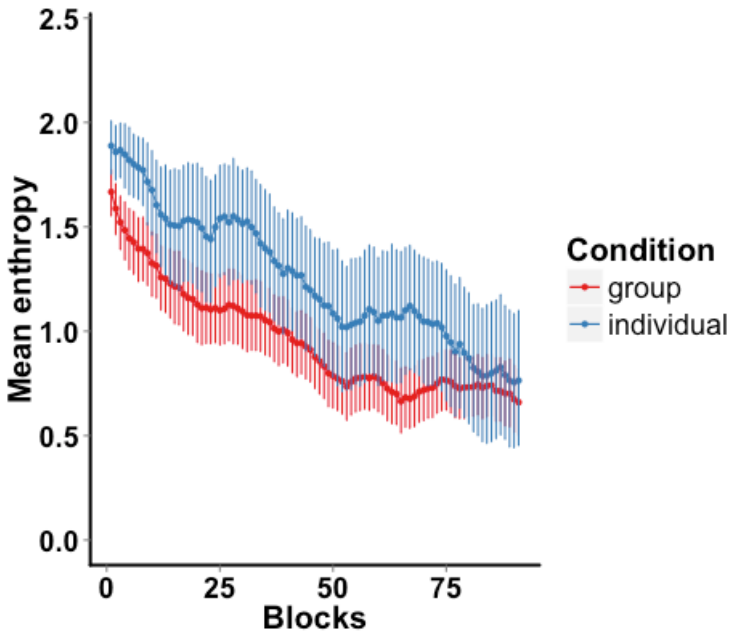


Figure 2-4. Trajectories of participants' average enthropy. Block t indicates the section of rounds [t, t+9], in which each enthropy was caluculated. The red line shows group condition and the blue line shows individual condition. The error bars indicate 95% confidence intervals.

Table 2-3. MCMC result of the enthropy model

Parameters	Mean	SD	Quantiles			Gelman-Rubin Statistics	
			2.5%	50.0%	97.5%	median (upper C.I.)	Effective sample size
intercept	1.78	0.12	1.56	1.77	2.01	1.01 (1.01)	422
β_1	$-1.15*10^{-2}$	$3.5*10^{-4}$	$-1.22*10^{-2}$	$-1.15*10^{-2}$	$-1.1*10^{-2}$	1.00 (1.00)	2000
β_2	-0.406	0.15	-0.706	-0.402	-0.143	1.00 (1.01)	370
β_3	$2.5*10^{-3}$	$4.1*10^{-4}$	$1.7*10^{-3}$	$2.5*10^{-3}$	$3.3*10^{-3}$	1.00 (1.00)	1946
variance	0.44	$3.3*10^{-3}$	0.43	0.44	0.45	0.99 (1.00)	1896
individuals' effect	0.51	0.046	0.43	0.51	0.61	1.00 (1.00)	1401
groups' effect	0.19	0.10	0.015	0.19	0.40	1.01 (1.02)	739

3.3 条件間での成績の比較

Figure 2-5 は，各条件における個人の成績分布を示している。MCMC の結果，成績に対する条件の効果（β）の 95%信頼区間は 0 を上回っており，集団条件における個人の成績は個人条件に比べて有意に高かったことが示された（Table 2-4）。

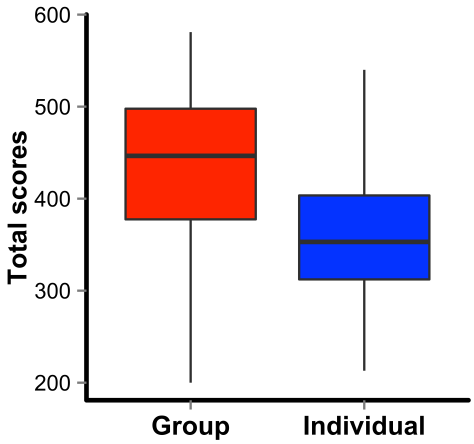


Figure 2-5. Participants' scores in the individual and the group conditions (N = 24 and N = 70, respectively). The y-axis refers to total scores of participants summed over 100 rounds, ranging from 100 (always choosing one of the 11 lowest-quality options) to 600 (always choosing the highest-quality option). The horizontal bars represent the median of each condition, while the boxes represent 50% ranges, and the upper and lower whiskers represent the highest and lowest values respectively that are within 1.5 * IQR (the inter-quartile range).

Table 2-4. MCMC result of the performance model

			Quantiles			Gelman-Rubin Statistics	
Parameters	Mean	SD	2.5%	50.0%	97.5%	median (upper C.I.)	Effective sample size
intercept	349.6	17.6	314.2	349.3	383.6	0.99 (1.00)	2642
β	83.6	22.1	40.9	83.3	126.8	1.00 (1.00)	2688
variance	83.3	17.6	70	82.8	98.2	1.00 (1.00)	2545
groups' effect	34.9	15.2	5.1	34.6	64.6	1.00 (1.01)	2199

４．考察

本研究の結果から，探索と収穫のジレンマが存在する不確実性下の意思決定場面において，社会的学習は個体あたりの意思決定パフォーマンスを高め，集合知効果を生じさせることが示された（Figure 2-5）。以下では，本研究で経験的に得られた結果から何が分かり，どのような問いが新たに生じたのかを論じる。

４.１ 社会的学習への重みと頻度依存性

Rendell らは，Robert Axelrod がかつて実施したように世界中から様々な学習戦略を募集してそれらの進化シミュレーションを行い，いかなる学習戦略が適応進化するかを調べた（Rendell, Boyd, Cownden et al., 2010）。その結果，広い環境パラメータ範囲において成績が上位となった戦略は，いずれも社会的学習には強く依存するが個人学習はあまり行わないという共通の特徴を持っていた。これは本研究の社会的学習モデルの推定結果と整合している。社会的学習モデルの推定からは，個人の選択確率が個人学習よりも社会的頻度情報により大きく依存していたことが示された（Table 2-1, $\lambda \approx 0.74$）。

しかしながら，本研究で推定した学習モデルは強化学習（i.e., 個人学習）の項と社会的頻度依存の項とをλによって線形に結合しただけの単純なものであり，実際の参加者はより複雑で洗練された戦略を使っていた可能性はある。例えば，選択肢間の価値の差が小さい時に社会的学習へより強く依存する（i.e., copy-when-uncertain）ような，条件付き戦略などが考えられる。Rendell らのトーナメントで上位となった戦略においても，個人学習と社会的学習とが様々な条件のもとでバランス調整されていた（Rendell, Boyd, Cownden et al., 2010）。今後，より複雑な混合戦略をも含めたモデル比較を行うことによ

って，人々が実際にはどのような社会的学習戦略を用いているかが分かってくるだろう。

また，本研究の結果からは人々があまり強い頻度依存性を示さなかったことが示唆された（Table 2-1, $\theta \approx 0.44$）。人間がどの程度の「同調バイアス」を示すのかという問題は，ここ数年盛んに議論されている（e.g., Bond, 2005; Claidiere et al., 2012）。しかし，本研究のようなモデルベースの推定は，少数の例外を除きほとんどなされてこなかった。例外として，McElreath らは２本腕バンディット問題における人々の行動を社会的学習モデルにあてはめている（McElreath et al., 2008）。McElreath らの結果では，本研究とは反対に，$\theta = 3.39$ という強い正の頻度依存性が示された。しかし興味深いことに，彼らの結果では社会的学習への重みが $\lambda = 0.13$ と非常に低くなっており，本研究の結果である $\lambda = 0.74$ とは対照的である。l つの可能性として，McElreath らの実験課題では２つの選択肢しか与えられていなかった（= two-armed bandit）ことによるアーチファクトが働いたことが考えられる。２つの選択肢しかない状況では，多数派に強く同調すること（i.e., θ を大きくすること）と個人学習を無視して社会情報への依存を高めること（i.e., λを大きくすること）は両方とも同じ働きを持ちうる。例えば，個人学習によって支持される選択肢は箱 A で，一方集団の多数派が選択するのは箱 B だったとする。このとき，社会的学習モデル上の θ とλのどちらを大きくしても，多数派に支持される箱 B の選択確率を上昇させることができる。このことから，本研究結果と McElreath らの結果との差異が本質的に重要な意味を持つのかどうかは判断し難い。今後，より緻密な実験によって，モデルベースに，人間の社会的学習戦略への理解が進むだろう。

４.２ 探索のフリーライダーと集合知

個人あたりの探索確率（q）については，個人条件と集団条件とで差が見られなかった（Figure 2-3, Table 2-2）。Table 2-2 が示すように，探索確率は主に個人のランダム効果（s）と集団のランダム効果（sg）に依存していたようである。すなわち，個人差・集団差の効果が大きかったといえる。しかしながら，この結果だけからは，集団条件における探索のフリーライダーが生じなかったと結論付けることはできない。なぜなら，学習モデルにおける温度パラメータτは，個人条件に比べて集団条件でより低かった（Table 2-1）。また，選択の多様さ（エントロピーH）も集団条件の方が個人条件より低かった（Figure 2-4, Table 2-3）。これらの結果はいずれも，社会的情報の共有によって個人あたりの探索行動が少なくなるという仮説を支持している。

ではなぜ，探索確率そのものについての差は出なかったのか。１つには，本研究における探索の定義が厳しすぎた可能性が考えられる。本研究では，強化学習の理論から，各時点で価値が最大の選択肢（i.e., $\max(\overline{\boldsymbol{A}}_t)$）を選択することを「収穫」，それ以外の選択を「探索」と定義した。しかしその定義によると，２番目以降に価値が高い選択肢をとり続けるような行動をも全て探索とみなしてしまうのである。例えば，課題の初期に大きな値が出て，価値が最大となった箱 P があったとする。しかし，社会的な選択頻度が大きいのは他の箱 Q だったとしよう。このとき，多数派同調を行って，学習モデル上は価値最大ではない箱 Q を繰り返し選択したとする。定義によれば，この行動は探索行動になってしまう。しかしこの場合，箱 Q を「選び続ける」行動は「収穫」とみなすのが現象の捉え方としては自然だろう。この問題は，箱 P についての価値が，しばらく選ばれていないにも関わらずいつまでも低減しないことに由来する。したがって学習モデル中に何らかの忘却メカニ

ズムを入れれば回避できる問題だと思われる。しかし，パラメータ推定の困難さから，モデルをこれ以上複雑にすることはできなかった。よって本研究では，エントロピーによる「選択のばらつき」の大きさの方が探索行動の様子をより的確に捉えていると考える。

個人あたりのエントロピー（i.e., 選択のばらつき度合い）低下にも関わらず，集団条件における平均成績は個人条件よりも有意に高かった（Figure 2-5, Table 2-4）。つまり，集合知効果は生じたと言える。ではなぜ，フリーライダーの存在にもかかわらず集合知が生じたのか。そのメカニズムとして以下の２つが考えられる。

一つ目は，ゲーム理論的な均衡よりも高い割合での協力（ここでは，探索行動を行うこと）が実現していた可能性である。これまでに行われてきた様々な実験研究において，ゲーム理論的な均衡よりも高頻度での協力行動が実現される事例が報告されてきた（e.g., Fehr & Fischbacher, 2004; Kameda et al., 2011）。したがって本研究で見られた集合知効果も，従来の実験研究で報告されてきた結果と同様に，理論的に予測されるよりも高頻度での協力行動（i.e., 未知の選択肢への探索）が維持された結果として生じたのかもしれない。しかし，本実験状況での進化的に安定な探索確率（i.e., 協力率）を解析的に求めることは数学的に難しい（Rosenberg, Salomon, & Vieille, 2013）。今後の研究では，よりシンプルな MAB 問題の課題を用いて，実際に参加者が理論的な進化的安定均衡を上回るほどに探索を行うかどうかを検証する必要があるだろう。

二つ目は, MAB 問題における探索と収穫の関係に由来する性質である。本実験での確率的な設定では，探索しないメンバーによる収穫からも，「本当にその選択肢の価値が高いか」を確認することができ，情報を得ることができた。したがって，たとえ探索量が減ったとしても，個人が単独でいる場合よりは豊富な情報が集団中に共有される。すなわち，仮

に利己的な個体だけからなる集団であっても，情報伝達と社会的学習は各個人の意思決定精度を上昇させうるのである。この効果は，学習と意思決定に関する至近メカニズムを考慮していなかった従来の社会学習モデルでは説明できなかったものである。

4.3 今後の展望

本研究で用いた MAB 問題のパラダイムでは，個人間での情報共有は完全であった。すなわち，各人が獲得した利得ポイントの情報が全て公共情報となっていた。しかし現実には，それが不可能な場合がほとんどである。行動は観察できるが他者の得た利得は分からない場合や，曖昧な手掛かりしか伝達されない場合などの方が，より現実的だろう。そのような仮定の変更は，学習と意思決定のプロセスや集合知に対してどのような影響与えるだろうか。

続く第 3 章では，この点を考慮して本章での実験パラダイムに変更を加えた実験の結果を報告する。特に，近年インターネット上の「口コミサイト」で見られるような，各個人による選択肢への評価を社会的情報として共有することが，集合知効果へどのような影響をおよぼすかに焦点を当てる。情報が不完全にしか共有されない場合，社会的学習による集合知はどの程度担保されるのであろうか。

第三章

口コミの共有は社会的学習の集合知効果を高めるか[1]

1．Introduction

In the modern human environment, it is often argued that the “multi-armed bandit” (MAB) problem may be solved collectively by mass information-sharing systems (Krause et al., 2009). For instance, with the growth of online shopping sites like Amazon or review sites like Yelp, consumers can learn how other consumers have evaluated products (e.g., Amazon’s 5-star rating system), as well as how many others have purchased those products. Do such information-sharing systems in fact help us solve the MAB problem? And, if so, how do the two types of social information (frequency information in conjunction with evaluation information) affect our decisions?

Although several previous studies have demonstrated that consumer choices are actually influenced by social information on the Internet (Salganik et al., 2006; Muchnik et al., 2013) and in the MAB task (McElreath et al., 2005; 2008; Toelch et al., 2010), they did not directly assess how social information could improve the objective qualities of people’s choice behaviors above and beyond individual choices. We thus

[1] 本章は，『Toyokawa, W., Kim, H. & Kameda, T. 2014. Collective human intelligence under dual exploration-exploitation dilemmas. *PLoS ONE*, 9 (4): e95789.』 としてすでに公刊された論文を一部改変したものである。

implemented a test-bed for the MAB problem in the laboratory.

In the current experiment, a group of 5 participants worked on a 30-armed bandit problem (see Figure 3-1; also see the methods section, Supporting figure S3-1 and S3-2 for details). Participants worked on the MAB task for a total of 100 rounds (the number of rounds was not specified in advance). In each round, participants chose one of the 30 alternatives individually, receiving associated payoffs as personal rewards. For each chosen option, payoffs were randomly generated from a stationary uniform probability distribution. To simulate some foraging environments in which qualities of patches are negatively correlated with their commonality (i.e., high quality patches are rare), one of the 30 options was made to have the highest mean ("category 1" option), followed by two "category 2", three "category 3", five "category 4", eight "category 5", and eleven "category 6" options (see Methods section for details about these categories). In addition to their own private payoff-feedback, participants could observe information about other members' choice behaviors from the preceding round. Within the group condition, we created two sub-conditions in which the richness of the social information was varied: (A) the "frequency only" sub-condition where individuals were informed of how many members had chosen each of the 30 alternatives in the preceding round; and (B) the "frequency plus evaluation" sub-condition where participants were provided opportunities to rate their chosen options on a 5-point scale (see Figure 3-1B) and could learn averages of those ratings for each option in addition to the social-frequency information. Rating the chosen option was entirely optional, and participants could

skip the evaluation and proceed to the next round (see Methods section for details). As a baseline, we also ran an individual condition where participants worked on the MAB task alone without any social information. Our focus is thus on how the availability of social information may affect human collective performance on the MAB.

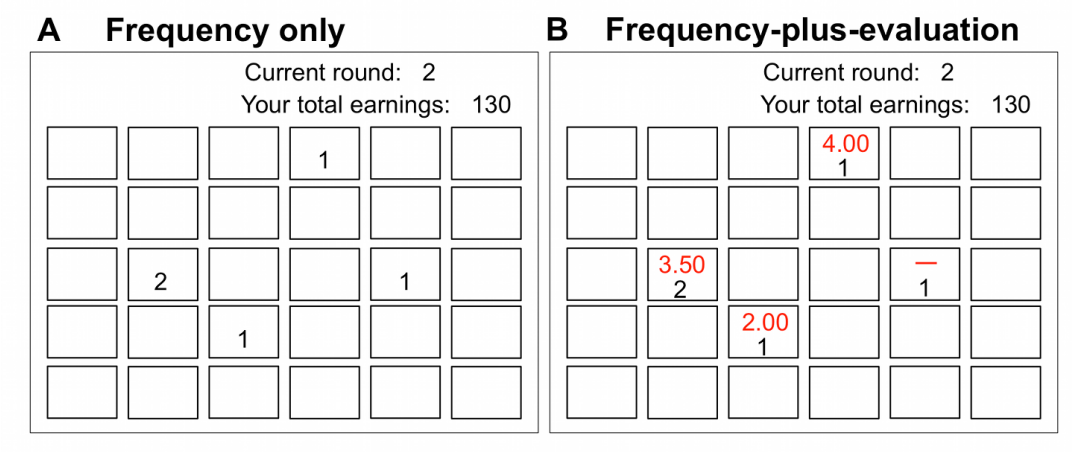


Figure 3-1: Example displays of choice stage. Except for round 1, social information about all 5 group members' choice behaviors in the preceding round was made available to each participant. A: An example display in the frequency-only sub-condition. The number displayed in each box (i.e., option) indicates how many members chose the option in the preceding round (round 1 in this example). B: An example display in the frequency-plus-evaluation sub-condition. In addition to the social-frequency information (black colored numbers), participants were informed of average evaluations (red colored numbers) that members had provided about chosen options in the preceding round, on a 5-point scale that ranged from 1 (very bad) to 5 (very good). Evaluating options was not mandatory, and a horizontal red bar indicated that no evaluation was contributed about the option.

2. Methods

2.1 Ethics statements

This study was approved by the Institutional Review Board of the Center for Experimental Research in Social Sciences at Hokkaido University (No. H24-3). Written informed consent was obtained from all participants before beginning the task.

2.2 Participants

One hundred and ninety-four undergraduates (42 females; age: mean ± S.D. = 19.8 ± 1.44) were randomly selected from a subject pool at Hokkaido University in Japan to participate in the experiment. After the experiment, participants received monetary rewards based on their performance in the MAB task as compensation for their participation (mean ± S.D. = 1175.7 ± 139.6 JPY).

2.3 Experimental procedure

For each experimental session, six to eight participants were brought into the laboratory. Five were randomly assigned to the group condition and the remaining participants were assigned to the individual condition. Upon arrival, each participant was seated in a separate semi-soundproof cubicle equipped with a computer terminal, and received further instructions individually on a computer screen. Participants remained strictly anonymous to each other throughout the experiment, and the monetary reward was also paid individually after the experiment. The experimental

task software was constructed using z-Tree (Fischbacher, 2007). The entire session lasted about 75 minutes.

The MAB task

The experimental task assigned to participants was a 30-armed bandit problem in which they had to choose between 30 options that could yield different payoffs. Participants worked on the MAB task for a total of 100 rounds, and each round consisted of two stages: a choice stage followed by a feedback stage. First, each participant was asked to choose from an array of 30 options (see Figure 3-1) by clicking their preferred option. Next, participants were told how much their payoffs were. Payoff amounts depended on quality of the option each participant chose. For each chosen option, payoffs were randomly generated from a stationary uniform probability distribution with an interval (min, min + 150) (where min is the option's minimum payoff), rounded up to the next integer. To simulate some foraging environments in which qualities of patches are negatively correlated with their commonality (i.e., high quality patches are rare), we varied the profitability of the 30 options as follows. Eleven of the 30 options were in the lowest-quality category (category 1) with min = 0, eight options in category 2 with min = 15, five options in category 3 with min = 30, three options in category 4 with min = 45, two options in category 5 with min = 60, and only one option was in the highest-quality category (category 6) with min = 75. The locations of these options on the computer screen were fixed across 100 rounds and were common for all participants in the same experimental session, but were randomly

re-shuffled across different experimental sessions. At the outset of each session, participants were told that (a) they would have multiple (unspecified) choice opportunities in the experiment, (b) payoffs from their choices would be randomly determined by some unknown probability distribution unique to each option, and (c) those individual payoffs would be summed, multiplied by 0.1, and rounded up to the nearest ten in Yen to determine their individual rewards after the experiment (see Supplementary Figure S3-1 and Figure S3-2 for more details about the program).

Thus, at the beginning, participants had no knowledge about the properties of the options, the exact shapes of the payoff distributions, and the total number of rounds. As time passed, however, participants could learn about the task structure via the feedback from their choice outcomes in each round.

Conditions

In addition to the private payoff-feedback, participants in the group condition could learn other members' choice behaviors in the preceding round. Two sub-conditions were nested in the group condition. Half of the 28 groups were randomly assigned to the "frequency only" sub-condition in which, at the choice stage, each participant could learn how many group members had chosen each option in the preceding round (see Figure 3-1A). The remaining 14 groups were assigned to the "frequency plus evaluation" sub-condition where, in addition to the social-frequency information, participants could learn the average ratings that participants had given to their chosen options in the feedback stage (see Figure 3-1B). Rating the chosen option at the feedback stage was

entirely optional, and participants could skip the evaluation and proceed to the next round. Contributing or skipping evaluation had no monetary impact. When all 5 members finished the feedback stage, they proceeded to the next choice stage. As a control, we also had an individual condition where 54 participants worked on the 30-armed bandit problem alone, without any social information.

2.4 Statistical analysis

Comparing participants' performance

We used the qualities (category 1–6) of each option to score participants' choice performance in the MAB task. Because the difference in expected payoffs between any two adjacent categories were held constant (15 points), the qualities of options can be treated as an interval scale. We summed participant's choices over the 100 rounds, which yielded scores that could range from 100 (always choosing one of the 11 lowest-quality options) to 600 (always choosing the highest-quality option). Because participants in the group condition were nested in the same 5-person groups, we used a hierarchical linear model with the individual vs. group condition as a fixed effect and group-specific effect as a random effect to analyze their performance.

Comparing frequencies of participant's exploratory choices

We used hierarchical Bayesian logit models with the Markov Chain Monte Carlo (MCMC) method to estimate per-individual exploration probability in the two conditions (see Supplementary SI 3-2) and in the two sub-conditions (SI 3-3) (51). The

individual vs. group condition, and the frequency-only vs. frequency-plus-evaluation sub-condition were treated as fixed effects, and individual- and group-specific effects were treated as random effects. We then examined the 95% Bayesian credible interval of a parameter ($\lambda 1$) for the condition (or sub-condition) effect to see if the interval contained zero (if not, we interpret the effect as significant). The MCMC simulation was conducted using the package rjags under R 3.0.3 and JAGS v 3.3.0 (further technical details are provided in Supporting Information section SI 3).

Comparing the two sub-conditions

We divided the 100 rounds into 5 blocks, and calculated each participant's average score in each block. Because the activities of participants in the same group were correlated, we averaged all 5 members' scores in each group to derive its aggregate scores to assure independence of samples. Therefore, the unit of analysis was the group, with N = 28. We applied a 2 (Sub-conditions) × 5 (Blocks) repeated-measures ANOVA (available in the package "anovakun v.4.3.3" under R 3.0.1) to this data set.

The relation between the volume of evaluation information in a group and participants' exploration behaviors

We constructed two mixed logit models treating both individual- or group-specific effects as random effects. One model posits that an individual's exploration probability at round t+1 is influenced by the total number of ratings contributed in the group at round t (evaluation effect model; SI3-3). The other, converse model assumes that probability of each member's contributing rating in the feedback stage at round t is

influenced by whether or not she/he engaged in exploration in the choice stage at round t (exploration effect model; SI3-4). We used the hierarchical Bayesian method to infer the parameters using the package rjags under R 3.0.3 and JAGS v 3.3.0 (further technical details of this analysis are provided in Supplementary SI 3). We used 95% credible intervals to see the significance of each parameter.

Analysis of individual differences in scale-use

We first examined the modal point that each participant used in evaluation (i.e., which point of the 5-point scale, from 1 to 5, was used most frequently by each participant). We then applied a variance-ratio test to see whether the modes were more variable between participants belonging to the same group (as an index of individual idiosyncratic biases in scale-use) than between group-means.

3. Results

In the following, we first examine whether participants benefited from social information to yield "collective intelligence" at the group level. We then examine how the richness of social information affected collective performance by comparing the frequency-only and the frequency-plus-evaluation sub-conditions.

3.1 Group versus individual performance

Figure 3-2A shows box-plots of participants' cumulative scores over 100 rounds in the individual and group conditions. Participants' choices in each round were assigned

scores ranging from 1 (choosing one of the 11 lowest-quality options) to 6 (choosing the single highest-quality option). Because participants in the group condition were nested in the same 5-person groups, we used a hierarchical linear model to analyze their individual performances (see Methods section for details). On average, participants achieved higher scores in the group condition (M = 397.3) than in the individual condition (M = 321.7), ΔD (difference between model deviances) = 19.50, p = 1.0 * 10-5.

Figure 3-2B shows frequency distributions of how often participants engaged in exploratory choices out of the 100 rounds in the individual and group conditions. Here, exploration was defined as a choice in which a participant selected options other than the "greedy" option that had yielded the largest cumulative payoff so far in her/his own experience (see Methods section for details). We estimated per-individual exploration probability by a hierarchical Bayesian logit model with the Markov Chain Monte Carlo (MCMC) method (see SI 3-1). The 95% Bayesian credible interval of the fixed effect for the conditions (individual vs. group) was (-1.63, -0.27), which indicates that participants engaged in significantly fewer exploratory choices in the group condition than in the individual condition (the full results are shown in Supplementary Table S3-1). Together with the results from Figure 3-2A, this implies that social information about other members' choices in the preceding round reduced each participant's total exploration time, helping her/him to exploit better options for a longer time in the group condition than in the individual condition. In other words, the within-individual exploration-exploitation dilemma was more efficiently resolved in

the group condition than in the individual condition.

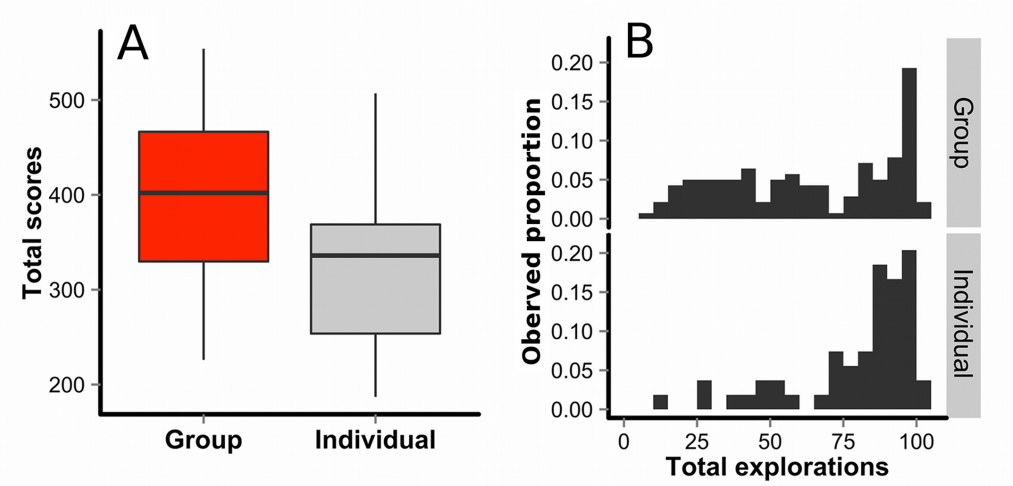


Figure 3-2: Comparison between individual and group condition. A: Participants' scores in the individual and group conditions (N = 54 and N = 140, respectively). The y-axis refers to total scores of participants summed over 100 rounds, ranging from 100 (always choosing one of the 11 lowest-quality options) to 600 (always choosing the highest-quality option). The horizontal bars represent the median of each condition, while the boxes represent 50% ranges, and the upper and lower whiskers represent the highest and lowest values respectively that are within 1.5 * IQR (the inter-quartile range). B: Histograms of participants' exploratory choices among the 100 rounds in the individual and group conditions (N = 54 and N = 140, respectively). The x-axis refers to the total number of "exploratory" choices (when a participant selected options other than the one that had yielded the largest cumulative payoff in her/his own experience so far) out of 100 rounds. The y-axis refers to observed proportions of participants with a given exploratory-choice frequency in each condition.

Interestingly, regarding the between-individual dilemma, the exploratory costs essential for emergence of collective intelligence (i.e., efficient collective exploitations of better options in the group condition) were not borne equally by all group members. Figure 3-3 displays participants' exploratory choices over the course of the experiment separately for each group sub-condition. We classified the experimental rounds (except for the 1st round where any choices were counted as "exploration") into three blocks of 33 rounds each, and examined how often participants in the group condition engaged in exploratory choices in each block. As the figure shows, the frequency distributions changed over time, approaching a more U-shaped pattern in the later blocks. It can also be seen that such U-shaped pattern as well as the prevalence of explorations in groups were more evident in the frequency-plus-evaluation sub-condition than in the frequency-only sub-condition (the 95% Bayesian credible interval of the fixed effect for the sub-condition in individual exploration probability was (0.07, 1.78); see SI3-2 and Table S3-2 for details). The observed U-shaped pattern implies that participants were divided largely between "information producers" who engaged in exploratory choices most of the time and information scroungers who free-rode on those efforts, exploiting others' findings through social learning (Galef & Giraldeau, 2001; Kameda & Nakanishi, 2002; 2003). Indeed, consistent with results from a recent social-learning-strategies tournament (Rendell et al., 2010), participants who engaged in more exploratory individual learning attained smaller net profits within each 5-person group ($r(140) = -0.52$, $p = 0.45 * 10\text{-}11$; see Figure S3-4). Yet, it is noteworthy

that, despite the emergence of information scroungers, groups still benefited overall from exchanging social information about each other's behavioral choices in the MAB task (Kameda & Nakanishi, 2003; Rendell et al., 2010).

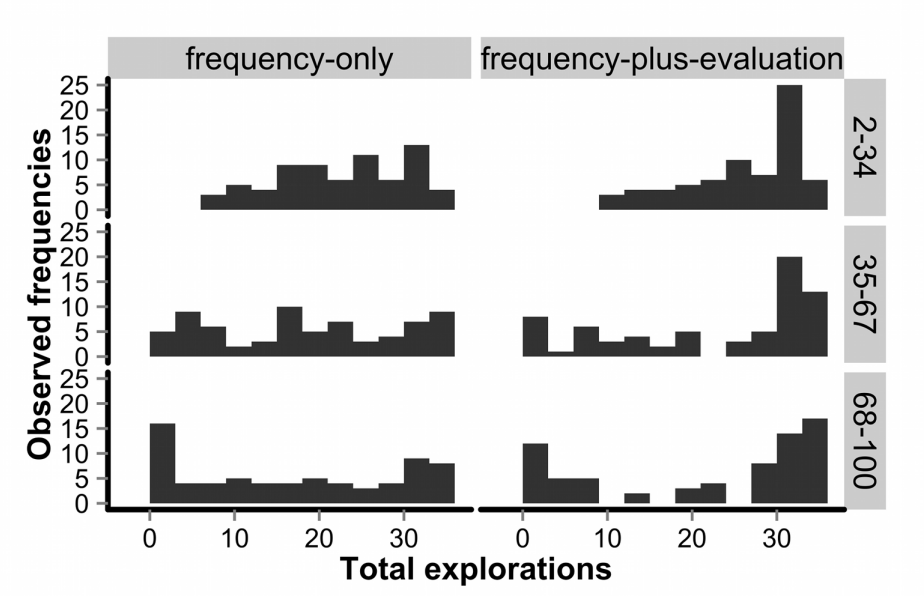


Figure 3-3: Temporal changes in frequency distributions of exploratory choices in the group condition. Shown from top to bottom, the first block (2nd through 34th round), the second block (35th through 67th round) and the last block (68th through 100th round). The y-axis refers to observed frequencies of participants with a given exploratory-choice frequency in each sub-condition (N = 140).

3.2 The frequency-only versus frequency-plus-evaluation sub-conditions

Next we examine how enriched social information (i.e., evaluation information alongside frequency information) affected the collective performance of participants in the MAB task. A simple statistical intuition suggests that richer information (i.e., more predictors) should lead to better collective performance, as instantiated in many websites posting updated customer reviews of products in addition to sales volume or rank (e.g., Amazon or Yelp). However, this intuition may not necessarily hold for the group MAB situation. First, given the exploration-exploitation dilemma operating at the between-individual level (see Figure 3-3), some people may intentionally abuse the evaluation opportunity. For instance, they may rate options randomly, or even evaluate good options as "bad" in order to manipulate others into exploring different options while they take advantage of their own new findings. Second, if participants use the rating scale differently from each other, the evaluation information could be a statistically less reliable signal as compared to the frequency information, which is free from such interpersonal scaling differences. In short, the evaluation information may be susceptible to strategic manipulations as well as individual biases in scale-use.

Figure 3-4A shows the trajectories of participants' average scores in the frequency-only and frequency-plus-evaluation sub-conditions (as a comparison, the trajectory in the individual condition was also displayed). For this analysis, we divided the 100 rounds into 5 equal blocks. A 2 (Sub-conditions) x 5 (Blocks) repeated-measures Analysis of Variance (MANOVA) yielded a main effect of blocks ($F(4, 26) = 114$, $p = 3.0$

* 10-6) and an interaction effect ($F(4, 26) = 2.53$, $p = 0.0448$), while a main effect of sub-conditions was not significant ($F(1, 26) = 2.67$, $p = 0.11$). As can be seen in Figure 3-4A, performance generally improved over time, but the improvement was smaller in the frequency-plus-evaluation sub-condition than in the frequency-only sub-condition (see also Figure S3-3 for trajectories of all participants' performances). A post hoc Shaffer's multiple comparison test revealed that, in block 5, the average score was significantly lower in the frequency-plus-evaluation sub-condition ($M = 4.49$) than in the frequency-only sub-condition ($M = 5.10$; $F(1, 26) = 4.27$, $p = 0.0488$), which suggests that, contrary to the aforementioned intuition, the additional evaluation information may have had a detrimental effect on participants' performance in the group MAB task.

To shed some light on the potentially detrimental effects of the evaluation information, we examined the relationship between the total number of times 5 members chose to rate their payoffs over the 100 rounds and the group's overall score where 5 members' total scores were averaged (see Figure 3-4B). Figure 3-4B also displays each group's total number of exploratory choices over the 100 rounds. As can be seen, more evaluation was associated with lower group score ($r(14) = -0.613$, $p = 0.0198$), indicating that evaluation information may indeed have had detrimental effects. Group score was also negatively correlated with the total number of exploratory choices ($r(14) = -0.854$, $p = 9.95 * 10\text{-}5$). A causal analysis using a hierarchical Bayesian logit model with the MCMC method (see Methods section) revealed that evaluations

and exploratory choices had a circular relation, whereby more evaluations led participants to engage in more exploratory behaviors, which facilitated further evaluations. This behavioral loop led participants in the frequency-plus-evaluation sub-condition to engage in over-exploration and fail to exploit better options for enough time (see Table S3-2 for details of the MCMC results).

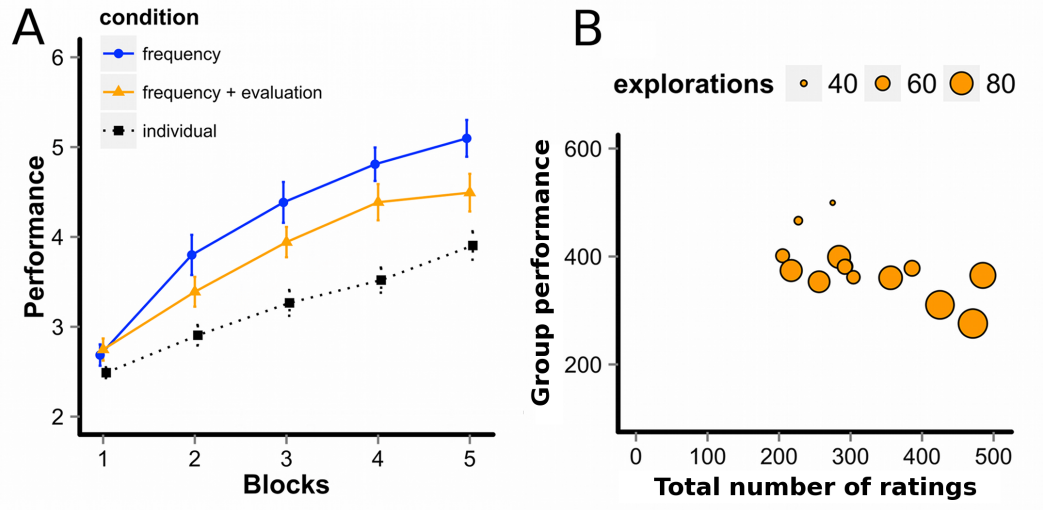


Figure 3-4: Effects of additional social information. A: Trajectories of participants' average scores. The scores of all 5 members of each group were averaged to yield aggregated group performance scores in each block that could range from 1 (choosing the lowest-quality options) to 6 (choosing the highest-quality option). Thus, the units of analysis was the group, with N = 14 in each sub-condition. Means of these group scores were plotted for the 2 sub-conditions and 5 blocks in the figure. To show the baseline, the performance of the individual condition is also plotted (N = 54). The error bars indicate ± 1 SEM. B: Scatter plots of the collective scores against the total number of ratings contributed and the total number of explorations in groups of the frequency-plus-evaluation sub-condition (N = 14). The x-axis refers to the total number of times group members chose to rate their payoffs over the 100 rounds (minimum: 0, maximum: 500), and the y-axis refers to the group's overall score (minimum: 100, maximum: 600). The size of circle represents the total number of exploratory choices in the group (minimum: 0, maximum: 500 — see Fig. 2 for the definition of exploration).

3.3 Fragility of the evaluation information

As argued earlier, one possible reason for the detrimental effects of the evaluation information was that it could be faked much more easily than the frequency information. Figure 3-5 displays the validities of each participant's evaluation information in terms of correlation coefficients between the participant's ratings of options on the 5-point scale and actual experienced outcomes. As can be seen, 20 out of 70 participants in the frequency-plus-evaluation sub-condition provided invalid evaluations of options (i.e., correlations between their ratings of options and actual experienced payoffs were either non-significant or significantly negative). Three of them even rated bad options as "good" and/or good options as "bad", suggesting strategic manipulations of evaluation.

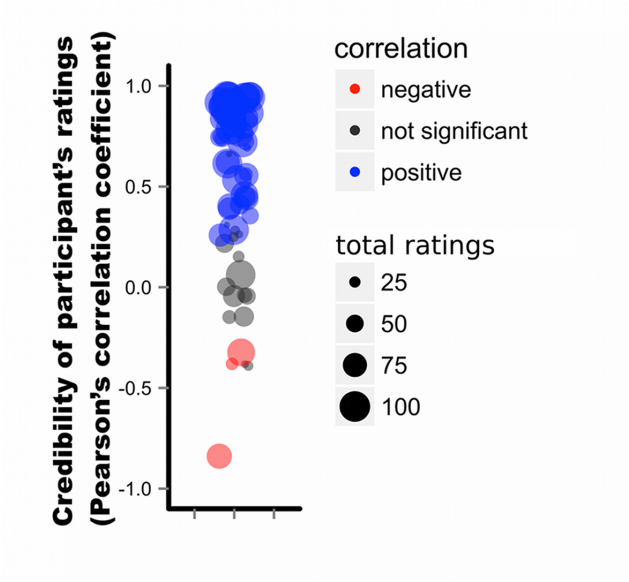


Figure 3-5: Credibility of evaluation information. The plot shows validities of each participant's evaluation information in the frequency-plus-evaluation sub-condition (N = 70). Each participant is represented by a circle. The y-axis refers to Pearson's correlation coefficient between a participant's ratings of options on the 5-point scale and actual earned payoffs. The size of each circle represents the total number of ratings that the participant provided out of the 100 rounds. To avoid overlap, the circles were jittered horizontally.

Finally, Figure 3-6A and 3-6B show overall validities of the two types of social information. Although both types were positively correlated with the objective option-qualities (frequency: $r = 0.45$, $p = 2.2 * 10\text{-}16$; evaluation: $r = 0.22$. $p = 2.2 * 10\text{-}16$), the correlation coefficient was significantly lower for the evaluation than the frequency information ($z = 13.39$, $p = 3.52 * 10\text{-}41$). The lower validity of the evaluation information was caused by individual differences in use of the rating scale (see Methods section for details). Modal points that participants used most frequently on the 5-point scale were different from each other ($F(69, 13) = 2.88$, $p = 0.0378$), as reflected in the broad 50% ranges in Figure 3-6B.

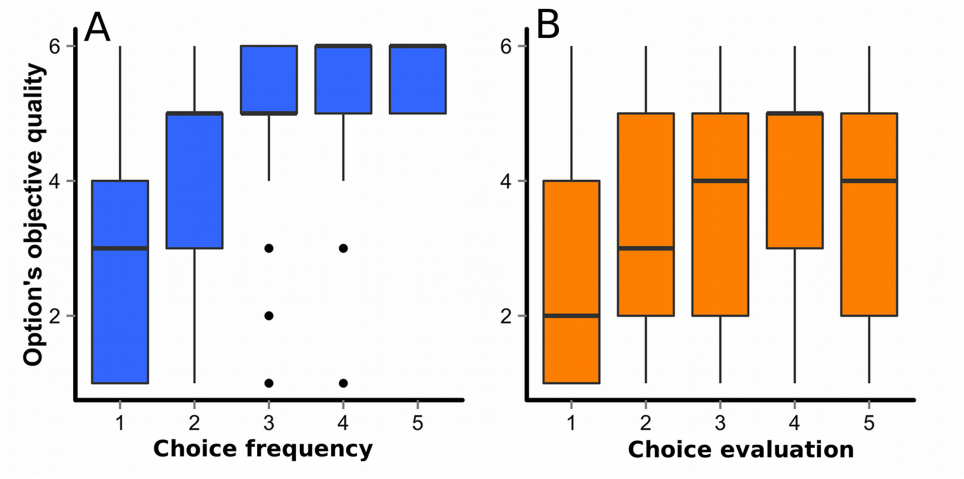


Figure 3-6: validity of frequency and evaluation information in the frequency-plus-evaluation sub-condition. (A) Relations between the choice frequencies and the chosen options' average qualities. The y-axis refers to mean objective qualities of options chosen by 1, 2, 3, 4, or 5 participants in each round. (B) Relations between evaluations of options and the options' average qualities. The y-axis refers to mean objective qualities of options rated on the 5-point scale ranging from 1 (very bad) to 5 (very good). We merged all rounds of all 14 groups into one dataset, thus both graphs represent 5000 observations each. The horizontal bars represent medians, while the boxes represent 50% ranges, and the vertical bars represent 1.5 * IQR (the inter-quartile range).

4. Discussion

The "multi-armed bandit" (MAB) problem is a pervasive adaptive task for humans as well as non-human animals. Eusocial insects often show an impressive array of performances on the MAB task through evolved coordinated mechanisms that function at the group level. Different from their colonies, however, the exploration-exploitation dilemma exists both within- and between-individuals for human groups composed of non-kin. This study thus focused on how a group of human agents may solve such dual dilemmas collectively with the aid of social information about each other's behavior.

Results from the experiment showed that average choice performance was substantially better in the group condition than in the individual condition (see Figure 3-2A). Analysis of choice behaviors suggest that on average, participants in the group condition were able to shorten total exploration time and exploit better options for a longer time by referring to each other's past behaviors (see Figure 3-2B). Interestingly, such exploratory costs were not borne evenly by all members — some members behaved as information scroungers who free-rode on other members' exploratory efforts via social learning, harvested the "greedy" option (Sutton & Barto, 1998) most of the time (see Figure 3-3) and earned more profits than information producers within each group (see Figure S3-4). However, opportunities to share social information were still beneficial for producing collective intelligence, which suggests that a group of human agents working on the MAB task can deal with the within- and between-individual dilemmas successfully, if not perfectly. In other words, social learning opportunities can provide greater net benefits to individuals above and beyond asocial learning, which is consistent with recent arguments from evolutionary theories emphasizing individual-fitness advantages of cultural learning (Rogers, 1988; Kameda & Nakanishi, 2003; Rendell et al., 2009; Enquist et al., 2007).

For better performance in the MAB problem, it is important to switch from the exploration phase to the exploitation phase at some point and harvest from a reasonably good option, rather than keep exploring for the very best option all the time. In this sense, our study introduces an important dimension for the rapidly growing

literature on human collective intelligence. It has been argued that social influence sometimes undermines collective performance on simple estimation tasks (Asch, 1956). For example, in a recent study (Lorenz et al., 2011), each participant was first asked to make numeric estimates about some factual questions (e.g., population density of Switzerland, crime statistics), and then provided social information about other participants' estimates to reconsider her/his initial responses. Compared to the initial estimates, estimates after the social information were found to be less accurate, by which the researchers concluded that even mild social influence could undermine the wisdom of crowds effect. Although such demonstrations are useful in illuminating the functions of social influence, we conjecture that the experimental paradigm that focuses only on accuracy may miss a critical aspect of natural decision-making — cost-benefit trade-offs under a fixed time/energy budget. Given that most natural decisions are made under time/energy constraints, striking an optimal balance between the benefits of accuracy and necessary cognitive/physical costs is essential to maximizing the overall profitability of choices. By extending the MAB problem to a group context, this study has demonstrated that the benefit of social information arises not only from improving individual decision accuracies per se, but also from enabling agents to save exploration costs collectively, even in the face of the between-individual exploration-exploitation dilemma.

This study also examined how the richness of social information may affect collective performance. Contrary to the statistical intuition that more predictors should lead to

better choices, sharing evaluations of options in addition to their frequency of choice had a detrimental effect on choice performance (see Figure 3-4A and 3-4B). Even though frequency information and evaluation information were both expected to be useful cues to predicting objective option-qualities, participants performed worse when both cues were combined than when only the frequency information was available, similar to the "less-is-more" effect in the judgment and decision making literature (Gigerenzer et al., 1999). Our results suggest that at least two factors were responsible for this decline. First, nearly 30% of the participants seemed to try to manipulate others toward exploration of new options by providing unreliable evaluations (see Figure 3-5). Notice that for such manipulative purposes, faking evaluation is much cheaper and easier than faking the frequency information, which would require sacrificing the benefits of one's own behavioral choices. Second, participants used the rating scale differently from each other, especially in which of the 5 rating levels to use most frequently in evaluation. With such idiosyncratic individual biases, the average of those ratings across individuals could be misleading (see Figure 3-6A and 3-6B). In short, the evaluation information seems to be susceptible to manipulations and individual biases in scale-use, and thus more fragile than the social frequency information.

Given these results, it is tempting to argue that websites such as Amazon or Yelp that implement similar rating systems may also suffer from these problems. However, users of these websites may not have strong incentives to deceive others into exploring.

Furthermore, rather than the 5-star rating system implemented in our study, many recent websites (the most famous being Facebook) have adopted a "Like" button where only positive evaluations are allowed. Interestingly, in the animal kingdom, most eusocial insects use only positive signals to recruit other workers to valuable resource patches (e.g., waggle dances by honeybees (Seeley et al., 1991), pheromone trails (Beckers et al., 1990) or tandem-running by ants (Shaffer et al., 2013), while negative or mixed signals that can inform others of the "badness" of patches are rather rare. Given their impressive choice performance as colonies, it seems important to examine how these positive-negative differences in rating systems may affect group dynamics under uncertainty. Comparing different species from insect colonies to human societies and examining similarities and differences in their computational algorithms will be helpful in illuminating such questions (Sumpter, 2010; Krause & Ruxton, 2002). In future research, collaborations between biologists and social scientists on these topics will be essential for a better understanding of the nature of collective intelligence, which is much desired in our rapidly connecting societies.

5. Supplementary materials

5.1 Supporting methods

Defining exploration and exploitation

We followed the traditional definitions of exploration and exploitation in the reinforcement learning literature [1]. Individual x facing a choice at round t has a unique vector $Q(x,t) = \{Q^x_{1,t}, Q^x_{2,t}, Q^x_{3,t}, \ldots, Q^x_{30,t}\}$, whose elements represent expected payoffs of the 30 options based on x's cumulative knowledge up to round t. As x's experience accumulates, *Q (x, t)* is updated by an averaging rule, in which $Q^x_{i,t}$ is equal to the mean payoff that x has earned from option *i* until round *t*. The averaging updating rule is a suitable assumption of the action-value function in a stationary multi-armed bandit task [1]. This model assumes that players can hold up to 30 values for the different options simultaneously in working memory and calculate a new average for each option. Although such assumptions are too strong for human cognitive capacity, we believe that this model can provide a first approximation to distinguish between exploration and exploitation in a stationary multi-armed bandit task. We categorized choice behaviour as "exploitation" when the individual x chose option *i* with max $Q^x_{i,t}$ in round $t + 1$ (i.e., the greedy choice). Otherwise, we categorized the choice as "exploration." The initial *Q* values were set to zero, i.e., $Q(x,0) = \{0,0,0,\ldots,0\}$, because x had no prior knowledge about the expected payoff from each option. For the first round ($t = 1$) only, we categorized all choices as "exploration."

Comparing exploration frequencies between the individual and group conditions

To compare the frequencies of exploratory choices over all 100 rounds between the individual and group conditions, we used a hierarchical Bayesian method. We modeled the probability of individual x in group g engaging in exploration (q_(x,g)) as a mixed logit model:

$$\ln\left(\frac{q_{x,g}}{1-q_{x,g}}\right) = \lambda_{1,0} + \lambda_{1,1} * (condition) + \varepsilon_1(x) + \mathrm{E}_1(g), \quad \text{(S1)}$$

where $\lambda_{1,0}$ indicates an intercept, $\lambda_{1,1}$ indicates a fixed effect for the condition (dummy coded: *individual condition* = 0; *group condition* = 1), and local parameters $\varepsilon_1(x)$ and $\mathrm{E}_1(g)$ indicate random effects specific to the individual and to the group respectively ($\varepsilon_1(x) \sim Normal(0, \sigma_1)$ and $\mathrm{E}_1(g) \sim Normal(0, \Sigma_1)$).

To estimate the four parameters ($\lambda_{1,0}$, $\lambda_{1,1}$, σ_1, and Σ_1), we set the uninformed priors of $\lambda_{1,0}$ and $\lambda_{1,1}$ to a normal distribution with mean = 0 and variance = 10^4 (i.e., a very stretched distribution), and those of the hyper parameters (σ_1 and Σ_1) to a uniform distribution [0, 10^4]. We conducted Markov Chain Monte Carlo (MCMC) simulations for 5 independent sequences. The number of total iterations per chain was 240,000 (thinning rate was 250), and first 40,000 steps were discarded from analysis because of initial-value dependencies. We used 95% Bayesian credible intervals to determine the significance of each parameter.

Comparing exploration frequencies between the frequency-only and frequency-plus-evaluation sub-conditions

To compare the frequencies of exploratory choices over all 100 rounds between the

two sub-conditions, we used a hierarchical Bayesian method. We modeled the probability of individual x in group g engaging in exploration $q_{x,g}$ as a mixed logit model:

$$\ln\left(\frac{q_{x,g}}{1-q_{x,g}}\right) = \lambda_{2,0} + \lambda_{2,1} * (subcondition) + \varepsilon_2(x) + \mathrm{E}_2(g), \quad \text{(S2)}$$

where $\lambda_{2,0}$ indicates an intercept, $\lambda_{2,1}$ indicates a fixed effect for the sub-condition (dummy coded: *frequency-only sub-condition* = 0; *frequency-plus-evaluation sub-condition* = 1), and local parameters $\varepsilon_2(x)$ and $\mathrm{E}_2(g)$ indicate random effects specific to the individual and to the group respectively ($\varepsilon_2(x) \sim Normal(0, \sigma_2)$ and $\mathrm{E}_2(g) \sim Normal(0, \Sigma_2)$).

Because the structure of the model was identical to the equation S1, the MCMC protocol was the same as described above.

Analysis of the causality between exploration and evaluation

To analyze factors affecting individual exploratory choices, we constructed two models. One model posits that an individual's exploration probability at round $t + 1$ is influenced by the total number of ratings contributed in the group at round t (evaluation effect model). The converse model posits that the individual probability of contributing rating information at the feedback stage of round t is influenced by whether or not she/he has engaged in exploration at the choice stage of round t (exploration effect model).

For the evaluation effect model, the probability of individual x in group g engaging in exploration at round $t + 1$ ($p_{x,g,t+1}$) was modelled:

$$ln\left(\frac{p_{x,g,t+1}}{1-p_{x,g,t+1}}\right) = \alpha_1 + \beta_1 * W_{g,t} + \varepsilon_3(x) + \mathrm{E}_3(g), \quad \text{(S3)}$$

where $z_{ij} = \beta 1 + \beta 2*\text{condition(i)} + r(i) + rg(j)$, α_1 indicates an intercept, β_1 indicates a fixed effect for the total number of ratings contributed in the group at the preceding round ($W_{g,t}$), and local parameters $\varepsilon_3(x)$ and $\mathrm{E}_3(g)$ indicate random effects specific to the individual and to the group respectively ($\varepsilon_3(x) \sim Normal(0, \sigma_3)$ and $\mathrm{E}_3(g) \sim Normal(0, \Sigma_3)$).

To estimate the four parameters (α_1, β_1, σ_3, and Σ_3), we set the uninformed priors of α_1 and β_1 to a normal distribution with mean = 0 and variance = 10^4 (i.e., a very stretched uniform distribution), and those of the hyper parameters (σ_3 and Σ_3) to a uniform distribution [0, 10^4]. We conducted MCMC simulations for 5 independent sequences. The number of total iterations per chain was 63,000 (thinning rate was 300), and first 3,000 steps were discarded from analysis because of initial-value dependencies. We used 95% credible intervals to determine the significance of each parameter.

For the exploration effect model, the probability of contributing evaluation at the feedback stage of round t was modeled:

$$\ln\left(\frac{\omega_{x,g,t}}{1-\omega_{x,g,t}}\right) = \alpha_2 + \beta_2 * (exploration)_{x,t} + \varepsilon_4(x) + \mathrm{E}_4(g), \quad \text{(S4)}$$

where $\omega_{x,g,t}$ indicates the probability of individual x contributing her/his evaluation at round t, α_2 for an intercept, β_2 for a fixed effect of exploration in the current period t (dummy coded: exploration = 1, exploitation = 0). Local parameters $\varepsilon_4(x)$ and $\mathrm{E}_4(g)$ indicate random effects specific to the individual and to the group respectively

($\varepsilon_4(x) \sim Normal(0, \sigma_4)$ and $\mathrm{E}_4(g) \sim Normal(0, \Sigma_4)$). Because the structure of the model was identical to the evaluation effect model, the MCMC protocol was the same as described above.

5.2 References

[1] Sutton, R. S. & Barto, A. G. Reinforcement Learning: An Introduction. (MIT press, Cambridge, 1998).

第四章

まとめと展望

1．集約型意思決定と集合知

人間社会には，投票による意見集約や陪審制度など，集団中に分布する意見を 1 つにまとめる手続きが多く見られる。従来の集合知研究では，主にそのような合意型意思決定（consensus decision making）へ焦点が当てられてきた。そこでは，各個人の意思決定と集団の意思決定とではどちらが優れているか，あるいは優れた集団意思決定をするためには意見をどのような手続きで集約させるべきか，といった問いが扱われてきた（e.g., Hastie & Kameda, 2005）。このような「合議の知は生まれるか」という問題設定は，民主主義の理論的基礎となるものであり，社会科学にとって古くから重要なテーマである。

１８世紀フランスの数学者であり哲学者でもあったコンドルセ侯爵は，シンプルな数理モデルを使って，多数決がいかに集合知効果を発揮する意見集約の仕組みであるかを示した（コンドルセの陪審定理，Condorcet, 1785/1994）。いま，各個人が２つの選択肢から 1 つを選択するという場面について考える（e.g., 有罪か無罪かを決める陪審場面）。選択肢のうちどちらか一方が，誰にとっても正解（期待効用を高めるもの）である。コンドルセの陪審定理は，各個人の意見が互いに独立であり，かつ個人あたりの正解確率が偶然よりも高いという 2 つの必要条件が満たされれば，多数決の決定はかならず平均個人よりも正解率が高まる（i.e., 集合知効果が生まれる）ことを示す。これまでの多くの集合知研究は，

コンドルセ陪審定理の必要条件がいつ満たされ，いつ満たされなくなるのかを実証的に探ってきた（e.g., Krause et al., 2011; Faria et al., 2009）。

一方，本稿の第二章・第三章では，合意型意思決定ではなく集約型意思決定（combined decision making）に焦点を当て，そこでの集合知効果を探った。集約型意思決定とは，集団として意見を 1 つにまとめて合意に至ることが目的の場面ではなく，各個人が相互作用を持ちながらもあくまで個人主体の意思決定を行う状況である。したがって，集約型意思決定では，各個人は集団レベルのパフォーマンスの最適化を目指すような目標を共有していない。究極的には，相互に情報のやりとりをする利己的エージェントの集合と捉えられる場面である。果たしてそのような「利己的な群れ」の中にも集合知が生まれるのか，というのが第二・三章での焦点であった。第二・第三章の結果を要約すれば，そのような人間の群れの中にも集合知効果は生まれたといえる。集団内での情報の共有は，ある程度の探索フリーライダーを生みながらも，平均的には不確実性下の意思決定パフォーマンスを向上させたのである。

このことは，行動生態学の大きなテーマである，動物の群居性生活の適応的意義を理解することへと繋がるだろう。これまで,「動物はなぜ群れるのか」を説明する究極要因として，捕食者の撹乱（Wood & Ackland, 2007）や捕食者発見確率の上昇（Powell, 1974）などの対捕食者防御機能，餌場発見確率の上昇（Ward & Zahavi, 1973），あるいは配偶効率の上昇など（Widemo & Owens, 1995），いくつもの説明が挙げられてきた。捕食者や餌場を発見する確率の上昇は，まさに集合知効果に着目した説明である。しかし，これまでの研究の多くは，個々の動物種それぞれに特有の現象へ特化した説明を求めてきた（e.g., Bijleveld et al., 2010）。一方，本研究で扱った多腕バンディット問題は，探索-収穫ジレン

マを含む普遍的な枠組みであり，様々な動物の意思決定場面に一般化可能である。したがって，本研究で示された集合知効果は，ヒトという種のみならず，集約型意思決定がみられる群居性動物一般に成り立ちうるものだと考えられるだろう。

2．集団サイズの増大と集合知

情報技術の進歩によって，人間の集約型意思決定の根幹をなす社会的情報の伝達プロセスは大きく変化した。その変化の 1 つとして，情報伝達範囲の拡大を挙げることができるだろう。インターネットは，地理的隔離のみならず，言語の壁による文化的隔離さえも乗り越えた情報伝達を可能にしつつある。一般的に，動物の集団内における社会的情報の伝達には物理的な限界がある。社会的ネットワークを介して個体から個体へと情報が伝わる場合，そのネットワークの外の個体（e.g., 他の群れの個体）には情報が伝播しない。あるいは，フェロモンなどの物理的刺激を介した情報伝達の場合にも，その及ぶ範囲には限界がある。したがって，情報を共有できる集団のサイズは，様々な物理的制約に支配されるのである。情報化の進展は，人間を唯一，この制約から解き放ちつつあるといえる。

コンドルセの陪審定理は，集団サイズの増大に伴う集合知効果の増大を予測する。すなわち，1 個体あたりの正解確率が変わらないならば，サイズが大きな集団ほど多数決の正解確率は最大値 1 へ近づくのである。コンドルセの陪審定理が説明するのは合意型意思決定における集合知であるが，本論文で焦点を当てた集約型意思決定においても集団サイズの増大は集合知へ正の効果をもたらすと予想できる。例えば多腕バンディット問題において，情報を共有するエージェントの数が増えるほど，選択肢の品質をフィルタリングしてくれる個体数が増加することになり，社会的情報（i.e., 選択頻度情報）の持つ情報量は増

加するだろう。集合知の基盤は統計的な大数の法則なのだから（Krause et al., 2009），集団サイズの増大が集合知効果を増すと考えるのは直観的にも自然な予測である。本論文で示した実験結果はいずれも，サイズが 5 人の集団に限定されていた。では，情報化に伴ってさらに大きな集団サイズが実現されていることを考えれば，実験室実験で示された集合知効果は，現実社会ではさらに「強化」されて見られるということだろうか。

この予測が成り立つためには，集団サイズが拡大しても個体レベルの行動規則は変化しないことが必要である。つまり，各個人の持つ社会的学習戦略が集団サイズによらず一定でなければ，集団サイズの大小によって集団レベルのダイナミクスが質的に変化してしまう可能性がある。Bond (2005) は，これまでに社会心理学分野で行われてきた同調行動の実験結果をメタ分析し，個人レベルの多数派同調傾向と集団サイズとの間に正の相関があることを示した。つまり，大きい集団に属する個人ほど強い同調バイアスを示す傾向である。Bond (2005) の結果を受け入れると，集団サイズの増加と集合知効果との間に興味深いトレード・オフ関係が浮かび上がってくる。それは，集合知と群衆行動との葛藤である。多数派への強い同調は，不適応な行動を集団中に伝播させるリスク，すなわち群衆行動のリスクを高めるだろう（Raafat et al., 2009）。これは，第一章で論じた，ケアリなどの非線形な強い動員を行う社会性昆虫にみられる「初期の人気への依存性」や「対称性の破れ」が生じるメカニズムと同様の理由による。強い同調行動のもたらすこのようなリスクは，社会的学習によって適応的な行動が伝播するという効果とは相反するものである。果たして，集団サイズの増加は，集合知効果と群衆行動のリスクとのどちらをより強く手繰り寄せるのだろうか。これは，今後実証的に確かめるべき課題だろう。インターネットを介した大規模な集団実験と，個体レベルの行動を詳しく測定する精緻な実験室実験を組み合わ

せることによって，それを検証していくべきである。

3．評価関数の違いにどう対処するか

第三章では，社会的情報の共有プロセスとして，人々による選択肢の評価，つまり「口コミ」の共有に着目した。人々の評価をまとめ，共有するサービスは，ネットショッピングを始めとし，今日のインターネット上には広く存在する。では口コミを共有することは，集合知効果を高めるのだろうか。第三章の実験結果は，口コミがむしろ集合知効果を弱める方向に働くことを示した。その原因は，各個人のもつ評価関数がそれぞれ異なっていたことに由来すると考える。

第二章で扱ったような，「他者の行動」と「それに続く結果」とを両方とも観察できる状況では，他者行動の観察を通じて行動価値（i.e., 強化学習モデル中の $\overline{A_t}$ の値）を直接更新することができる。つまり，行動と結果との関連を，擬似的に経験することができる状況である。一方，第三章の口コミ共有条件で参加者が観察できたのは，「他者の行動」と「その結果に対する当該他者からの評価」であり，その他者が実際に得た結果そのものは観察できなかった。つまり，他者行動の観察によって自身の行動価値（$\overline{A_t}$）を直接更新することはできなかった。計算論的に，他者が付けた評価値を使って自分自身の行動価値を更新しようとするならば，他者の付けた評価値を実際の結果（i.e., 実際には何ポイントだったか）へ復号化することが必要である。すなわち，観察された評価値 4 という数字は，果たして約何ポイントを示すのか，見当をつける必要がある。その復号化を経て初めて，他者が下した評価値から，選択肢の質を学習することができるからである。

集団中の全員が同一の評価関数（i.e., 「記号化関数」および「復号化関数」）を持って

いるのであれば，評価を介した情報伝達はうまくいくと思われる。しかし，実際の人間社会においてその条件を満たすことは容易ではないだろう。同じ物理的刺激を受けたとしても，個人によってそれをどう表現するか（i.e., 記号化関数）は異なるだろうし，伝えられた表現からどのような物理刺激を想起するか (i.e., 復号化関数) もまた人それぞれである。第三章の実験においても，評価情報の共有が集合知効果を下げた原因の 1 つは，記号化・復号化関数における人々の差異であった（Figure 3-6)。

評価関数が完全に一致する他者は存在しなくとも，なるべくそれが似ている他者を選んで社会的学習することは有効だろう。あるいは，いま観察している他者の状態や経験からその人の評価関数を推測する能力があれば，精度の高い社会的学習が可能かもしれない。例えば，すでに 100 軒のレストランを訪れたことがある人と，まだ数軒しか訪れたことのない人とでは，前者による高評価の方がより客観的に質の高い店を示す期待は高そうである。あるいは，同じく 100 軒のレストランを巡った経験のある二者でも，その経歴の違い（e.g., 一方は質の高いレストランを多く経験したが, 他方は質の低いレストランばかり経験してきた）は，２人の評価関数へ大きな違いを生み出しそうである。第三章で示した実験では，5 名の集団メンバーを個別に識別することはできなかった。つまり,「誰から社会的学習するか」型の戦略をとることはできない実験設定であった。将来の実験ではこの点を考慮し，各他者を区別可能な状況にすることで，果たして人々が特定の他者からの情報を多く利用するようになるか，という問いにアプローチが可能だろう。また，ここで述べた「誰から社会的学習するか」という戦略を使うためには, 他者の背景にある知識や経験, 評価関数などを推測する能力が鍵になると考えられる。他者の状態を推測する能力（e.g, 心の理論や共感性）と社会的学習戦略との間には，どのような関係があるだろうか。今後

は，行動実験と脳神経科学のアプローチをつなぎながら，個人内の社会的学習メカニズムへ迫っていくべきだろう。

４．これからの集団意思決定研究

集団意思決定を理解するということはすなわち，マイクロ・レベルでの個人の行動とマクロ・レベルで創発する集団の振る舞いとの相互作用を理解するということである。複数の個人が相互に影響しあうシステムは，多くの場合，非線形なダイナミクスである。したがって, 個体レベルにおける行動の理解から加法的に集団動態を説明することはできない。つまり，自然言語による論述モデルでは理解が追いつかないほど，マイクロ–マクロ相互作用は複雑であることが普通なのである。複雑なものを理解するためには，数理モデルの力が必要である。

特に計算論的モデルが，これからの集団行動研究に重要な役割を果たすと考える。計算論的モデルとは，第二章で扱った強化学習モデルのように個体が情報処理をする過程をモデル化したものである。これは神経行動科学では広く用いられており，個人レベルの行動理解にとってはすでに有効なアプローチとなっている（e.g., Suzuki et al., 2012; Daw et al., 2006; Jepma et al., 2011）。例えば，Daw ら（2006）は，強化学習モデルにおける学習率や探索率などのパラメータ値と相関する脳活動を特定した。また Jepma ら（2011）は，ノルアドレナリン系を調節する青斑核の活性パターンが瞳孔サイズと相関することを利用し、瞳孔サイズから強化学習における人々の探索行動を予測できることを示した。このように，計算論的モデルというレンズを通して脳活動を調べることによって，人間行動の至近的基盤が解明されている。

計算論的モデルは情報処理を記述するモデルであるから，個体間の相互作用もモデル中に容易に組み込むことがきでる（Suzuki et al., 2012）。つまり，個人内過程を扱うモデルから，個人間過程のモデルへと射程を拡大することができるのである。ということは，個体間相互作用を組込んだ計算論的モデルを個体ベース・シミュレーションの基礎として用いることができるということである。そうすれば，個人の行動アルゴリズムと集団レベルの動態とのつながりが見えてくるだろう。

神経メカニズムと行動レベル，そして行動レベルと集団レベル，そのどちらの階層を説明する場合であっても，計算論的なアプローチは有効である。計算論的モデルに基づいた行動実験が，これからの集団意思決定研究ではさらに重要となるだろう。

謝辞

本研究を遂行し学位論文をまとめるにあたり，多大なるご支援とご指導を賜りました，指導教員である現東京大学教授の亀田達也先生に深く感謝しております。未熟者の私に多くの挑戦とチャンスを与えて下さりました。先生と重ねた議論はいつも楽しく，そこからサイエンスの素晴らしさを，実感を持って学ばせていただきました。先生の研究への姿勢や情熱を模範としながら，自立した一人前の研究者となれるよう，これから益々精進しようと思います。

また，すばらしい研究環境を管理・運営してくださり，同時にまた研究生活には欠かすことのできない沢山の知的刺激と議論の場を惜しみなく与えてくださった，北海道大学大学院文学研究科・行動システム科学講座の先生方，秘書様方，そして大学院の学友の皆様には，深く感謝しております。

日常生活においては，私に大いなる活力と安らぎを共に与えてくれた北海道大学恵迪寮の寮生諸君に深く感謝しております。恵迪寮に漂う，「誰しもがイノベーターである」という文化をこれからも大切にしていただきたいです。

最後に，これまで私の思う道に対して，常に暖かく見守り応援し続けてくれた父と母，そして妹に深い感謝の意を示し，謝辞とさせていただきます。

引用文献

Aoki, K. & Feldman, M. W. 2014. Evolution of learning strategies in temporally and spatially variable environments: A review of theory. *Theoretical Population Biology* 91: 3-19.

Akerlof, G. A., & Shiller, R. J. 2009. *Animal spirits: How human psychology drives the economy, and why it matters for global capitalism*. Princeton, NJ: Princeton University Press.

Allen, J. L. 2002. *Conclave: The politics, personalities, and process of the next Papal election*. New York, NY: Doubleday.

Anderson, C. M. 2012. Ambiguity aversion in multi-armed bandit problems. Theor Decis 72: 15-33.

Asch, S. 1951. Effects of group pressure upon the modification and distortion of judgment. In H. Guetzkow (Ed.), *Groups, leadership and men. Pittsburgh*, PA: Carnegie Press.

Asch, S. 1956. Studies of independencies and conformity. A minority of one against a unanimous majority. *Psychological Monographs*, 70, 1-70.

Auer, P., Cesa-Bianchi, N. & Fischer, P. 2002. Finite-time analysis of the multi-armed bandit problem. Mach Learn 47: 235-256.

Bateson, M., Nettle, N. & Roberts, G. 2006. Cues of being watched enhance cooperation in a real-world setting. *Biology Letters*, 2, 412–414.

Beckers, R., Deneubourg, J. L., & Goss, S. 1992. Trails and U-turns in the Selection of a Path by the Ant *Lasius niger*. *Journal of theoretical biology*, 159, 379–415.

Beckers, R., Deneubourg, J. L., & Goss, S. 1993. Modulation of trail laying in

the ant Lasius niger (Hymenoptera: Formicidae) and its role in the collective selection of a food source. *Journal of Insect Behavior*, 6, 751–759.

Beckers, R., Deneubourg, J. L., Goss, S., & Pasteels, J. M. 1990. Collective decision making through food recruitment. *Insectes Sociaux*, 37, 258–267.

Berry, D. A. & Fristedt, B. 1985. Bandit Problems: Sequential Allocation of Experiments. Netherlands: Springer. DOI: 10.1007/978-94-015-3711-7

Bijleveld, A. I., Egas, M., van Gils, J. A., & Piersma, T. 2010. Beyond the information centre hypothesis: communal roosting for information on food, predators, travel companions and mates? Oikos, 119, 277-285.

Bikhchandani, S., Hirshleifer, D. & Welch, I. 1992. A theory of fad, fashion, custom, and cultural change as information cascades. *J Polit Econ*. 100 (5): 992-1026.

Biro, D., Sumpter, D. J. T., Meade, J., & Guilford, T. 2006. From compromise to leadership in pigeon homing. *Current Biology*, 16, 2123-2128.

Boehm, C. 1996. Emergency decisions, cultural- selection mechanics, and group selection. *Current Anthropology*, 37, 763–793.

Bolton, P. & Harris, C. 1999. Strategic experimentarion. *Econometorica*, 67, 349-374.

Boyd, R., & Richerson, P. J. 1985. *Culture and the evolutionary process*. Chicago, IL: University of Chicago Press.

Brezzi, M. & Lai, T. L. 2000. Incomplete learning endogenous data in dynamic allocation. Econometrica 68: 1511-1516.

Brooks, S. P. & Gelman, A. 1997. General methods for monitoring

convergence of iterative simulations. Journal of Computational and Graphical Statistics, 7, 434-455.

Brown, R. 1988. *Group Processes. Dynamics within and between Groups*. Oxford, UK: Blackwell. (黒川正流・橋口捷久・坂田桐子（訳）(1993). グループ・プロセス　（株）北大路書房)

Camazine, S., Deneubourg, J. L., Franks, N. R., Sneyd, J., Theraulaz, G., & Bonabeau, E. 2001. *Self-organization in biological systems*. Princeton studies in complexity. Princeton, New Jersey: Princeton University Press.

Camazine, S., & Sneyd , J. 1991. A model of collective nectar source selection by honey bees: Self-organization through simple rules. *Journal of Theoretical Biology*, 149, 547–571.

Christakis, N. A., & Fowler, J. H. 2007. The spread of obesity in a large social network over 32 years. *New England Journal of Medicine*, 357, 370–379.

Chudek, M., Heller, S., Birch, S., & Henrich, J. 2012. Prestige-biased cultural learning: bystanderd's differential attention to potential models influences children's learning. *Evolution and Human Behavior*, 33, 46-56.

Claidiére, N., Bowler, M., & Whiten, A. 2012. Evidence for weak or linear conformity but not for hyper-conformity in an everyday social learning context. *PLoS one*, 7, e30970.

Cohen, J. D., McClure, S. M. & Yu, A. J. 2007. Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philos T Roy Soc B* 362: 933-942.

Condorcet, M. 1785/1994 *Éssai su l'application de l'analyse à la probabilité des décisions rendues à la plu- ralité des voix* (trans. McLean & Hewitt). Aldershot: Edward Elgar.

Conradt, L., & List, C. (Eds.) 2009. Theme issue: group decision making in humans and animals. *Philosophical Transactions of the Royal Society B*, 364, 719–852.

Conradt, L., & Roper, T. J. 2005. Consensus decision making in animals. *Trends in Ecology and Evolution*, 20, 449–456.

Couzin, I. D., Loannou, C. C., Demirel, G., Gross, T., Torney, C. J., Hartnett, A., Conradt, L., Levin, S. A., & Leonard, N. E. 2011. Uninformed individuals promote democratic consensus in animal groups. *Science*, 334, 1578-1580.

Czaczkes, T. J., Grüter, C., Jones, S. M. & Francis, L. W. R. 2011. Synergy between social and private information increases foraging efficiency in ants. *Biol Letters* 7: 521-524.

Davis, J. H. 1969. *Group performance*. Reading, MA: Addison-Wesley.

Davis, J. H. 1973. Group decision and social interaction: A theory of social decision schemes. *Psychological Review*, 80, 97–125.

Davis, J. H., Kerr, N. L., Atkin, R. S., Holt, R., & Meek, D. 1975. The decision processes of 6- and 12- person mock juries assigned unanimous and two- thirds majority rules. *Journal of Personality and Social Psychology*, 32, 1–14.

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B. & Dolan, R. J. 2006. Cortical substrates for exploratory decisions in humans. *Nature*, 441, 876-879.

Detrain, C. & Deneubourg, J. L. 2013. Collective decision-making and foraging patterns in ants and honeybees. *Adv Insect Physiol* 35: 123-173.

Dubois, F., Morand-Ferron, J. & Giraldeau, L. A. 2010. Learning in a game context: strategy choice by some keeps learning from evolving in others. *Proc R Soc B* 277: 3609-3616.

Enquist, M., Eriksson, K. & Ghirland, S. 2007. Critical social learning: A solution to Rogers's paradox of nonadaptive culture. *Am Anthropol* 109: 727-734.

Fehr, E., & Fischbacher, U. 2004a. Third party sanctions and social norms. *Evolution and Human Behavior*, 25, 63–87.

Fehr, E., & Fischbacher, U. 2004b. Social norms and human cooperation. *Trends in Cognitive Sciences*, 8, 185–190.

Feldman, M. W., Aoki, K., & Kumm, J., 1996. Individual versus social learning: evolutionary analysis in a fluctuating environment. *Anthropological Science*, 104, 209-213.

Festinger, L. 1954. A theory of social comparison processes. *Human Relations*, 7, 117–140.

Fischbacher, U. 2007. z-Tree: Zurich toolbox for ready-made economic experiments. *Exp Econ* 10: 171-178.

Fowler, J. H., & Christakis, N. A. 2008. Dynamic spread of happiness in a large social network: Longitudinal analysis over 20 years in the Framingham Heart Study. *British Medical Journal*, 337, a2338.

Franks, N. R., Mallon, E. B., Bray, H., Hamilton, M. J., & Mischler, T. C. 2003. Strategies for choosing between alternatives with different

attributes: Exemplified by house-hunting ants. *Animal Behaviour*, 65, 215–223.

Galef, B. G. J. & Giraldeau, L. 2001. Social influences on foraging in vertebrates: causal mechanisms and adaptive functions. *Anim Behav* 61: 3-15.

Galton, F. 1907. Vox Populi. *Nature*, 75, 450–451.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. et al. 2013. *Bayesian Data Analysis, Third Edition*. New York: CRC press. 661 p.46.

Gigerenzer, G., Todd, P. M. & ABC Research Group. 1999. *Simple Heuristics That Make Us Smart.* Oxford: Oxford University Press. 416 p.

Gelman, A. & Rubin, D.B. 1992. Inference from iterative simulation using multiple sequences, *Statistical Science*, 7, 457-511.

Gilardeau, L-A., & Caraco, T. 2000. *Social foraging theory*. Princeton, NJ: Princeton University Press.

Gittins, J. C. 1979. Bandit processes and dynamic allocation indices. *J Roy Stat Soc B* 41: 148-177.

Gittins, J., Glazebook, K., & Weber, R. 2011. *Multi-Armed Bandit Allocation Indices 2nd Edition*. UK: John Wiley & Sons Ltd. 293 p.

Glaeser, E. L., Sacerdote, B., & Scheinkman, J. A. 1996. Crime and social interactions. *Quarterly Journal of Economics*, 111, 507–548.

Goss, S., Aron, S., Deneubourg, J. L., & Pasteels, J. M. 1989. Self-organized shortcuts in the Argentineant. *Naturwissenschaften*, 76, 579–581.

Grüter, C. & Leadbeater, E. 2014. Insights from insects about adaptive social information use. *Trends in Ecology and Evolution*, 29: 177-184.

Hamilton, W. D. 1964. The genetical evolution of social behaviour I and II. *Journal of Theoretical Biology*. 7, 1-52.

Hardin, R. 1982. *Collective action*. Baltimore, MD: Johns Hopkins University Press.

Hastie, R., & Kameda, T. 2005. The robust beauty of majority rules in group decisions. *Psychological Review*, 112, 494–508.

Hastie, R., Penrod, S. D., & Pennington, N. 1983. *Inside the jury*. Cambridge, MA: Harvard University Press.

Haun, D. B. M., Rekers, Y., & Tomasello, M. 2012. Majority-biased transmission in chimpanzees and human children, but not orangutans. *Current Biology*, 22, 727-731.

Henrich, J., & Boyd, R. 1998. The evolution of conformist transmission and the emergence of between-group differences. *Evolution and Human Behavior*, 19, 215-241.

Janis, I. L. 1972. *Victims of groupthink*. New York, NY: Houghton Mifflin.

Janis, I. L. 1982. *Groupthink: Psychological studies of policy decisions and fiascoes* (2nd ed). New York, NY: Houghton Mifflin.

Jepma, M. & Nieuwinhuis, S. 2011. Pupil diameter predicts changes in the exploration-exploitation trade-off: evidence for the adaptive gain theory. *Journal of Cognitive Neuroscience* 23, 1587-1596.

Kameda, T. 1991. Procedural influence in small group decision making:

Deliberation style and assigned decision rule. *Journal of Personality and Social Psychology*, 61, 245–256.

亀田達也 1997. 認知科学モノグラフ③ 合議の知を求めて−グループの意思決定− 共立出版

Kameda, T. & Nakanishi, D. 2002. Cost-benefit analysis of social/cultural learning in a nonstationary uncertain environment: An evolutionary simulation and an experiment with human subjects. *Evol Hum Behav* 23: 373-393.

Kameda, T. & Nakanishi, D. 2003. Does social/cultural learning increase human adaptability? Rogers's question revisited. *Evol Hum Behav* 24: 242-260.

Kameda, T. & Tamura, R. 2007. "To eat or not to be eaten?" Collective risk-monitoring in groups. *J Exp Soc Psychol* 43: 168-179.

Kameda, T. & Tindale, R. S. 2006. Groups as adap- tive devices: Human docility and group aggre- gation mechanisms in evolutionary context. In M. Schaller, J. Simpson, & D. Kenrick (Eds.), *Evolution and social psychology*. (pp. 317–341). New York, NY: Psychology Press.

Kameda, T., Tsukasaki, T., Hastie, R., & Berg, N. 2011. Democracy under uncertainty: The wisdom of crowds and the free-rider problem in group decision making. *Psychological Review*, 118, 76–96.

Kameda, T., Wisdom, T., Toyokawa, W., & Inukai, K. 2012. Is consensus-seeking unique to humans? A selective review of animal group decision-making and its implications for (human) social psychology. *Group Processes & Intergroup Relations*, 15, 673-689.

Kathehakis, M. N. & Robbins, H. 1995. Sequential choice from several

populations. *Proc Natl Acad Sci USA* 92: 8584-8585.

Keasar, T., Rashkovich, E., Cohen, D. & Shmida, A. 2002. Bees in two-armed bandit situations: foraging choices and possible decision mechanisms. *Behav Ecol* 13: 757-765.

Kendal, R. L., Coolen, I., & Laland, K. N. 2009. Adaptive trade-offs in the use of social and personal information. In R. Dukas, & J. Ratcliffe (Eds.), *Cognitive ecology: the evolutionary ecology of learning memory and information use* (pp. 249-271). Chicago: University of Chicago Press.

Kerr, N. L. 1983. Motivation losses in small groups: A social dilemma analysis. *Journal of Personality and Social Psychology*, 45, 819–828.

Kerr, N. L., MacCoun, R., & Kramer, G. P. 1996. Bias in judgment: Comparing individuals and groups. *Psychological Review*, 103, 687–719.

Kerr, N. L., & Tindale, R. S. 2004. Group performance and decision making. *Annual Review of Psychology*, 55, 623–655.

King, A. J., & Sueur, C. 2011. Where next? Group coordination and collective decision making in Primates. *International Journal of Primatology*, 32, 1245-1267.

Kurause, J., & Ruxton, G. D. 2002. *Living in groups*. Oxford series in ecology and evolution. Oxford: Oxford University Press.

Krause, J., Ruxton, G. D. & Krause, S. 2009. Swarm intelligence in animals and humans. *Trends Ecol Evol* 25: 28-34.

Laland, K. N. 2004. Social learning strategies. *Learning Behav* 32: 4-14.

Larrick, R. P., & Soll, J. B. 2006. Intuitions about combining opinions:

Misappreciation of the averaging principle. *Management Science*, 112, 111–127.

Latane, B., & Wolf, S. 1981. The social impact of majorities and minorities. *Psychological Review,* 88, 438–453.

Ledyard, J. 1995. Public goods experiments. In J. Kagel, & A. Roth (Eds.), *Handbook of experimental economics* (pp.111–173). Princeton, NJ: Princeton University Press.

List, C., Elsholtz, C., & Seeley, T.D. 2009. Independence and interdependence in collective decision making: an agent-based model. *Philosophical Transactions of the Royal Society B*, 364, 755–762.

Lorenz, J., Rauhut, H., Schweitzer, F. & Helbing, D. 2011. How social influence can undermine the wisdom of crowds effect. *Proc Natl Acad Sci USA* 108: 9020-9025.

Lorge, I. & Solomon, H. 1955. Two models of group behavior in the solution of eureka-type problems. *Psychometrica*, 20, 139-148.

McElreath, R., Lubell, M., Richerson, P. J., Waring, T. M., Baum, W., et al. 2005. Applying evolutionary models to the laboratory study of social learning. *Evol Hum Behav* 26: 483-508.

McElreath, R., Bell, A. V., Efferson, C., Lubell, M., Richerson, P. J., et al. 2008. Beyond existence and aiming outside the laboratory: estimating frequency-dependent and pay-off-biased social learning strategies. *Philos T Roy Soc B* 363: 3515-3528.

Mesoudi, A. 2011. *Cultural evolution: how Darwinian theory can explain human culture and synthesize the social sciences.* Chicago and London: University of Chicago Press.

Milgram, S., Bickman, L., & Berkowitz, L. 1969. Note on the drawing power of crowds of different size. *Journal of Personality & Social Psychology*, 13, 79-82

Muchnik, L., Aral, S. & Taylor, S. J. 2013. Social influence bias: a randomized experiment. *Science* 341: 647-651.

Nakahashi, W. 2007. The evolution of conformist transmission in social learning when the environment changes periodically. *Theoretical population biology*, 72, 52-66.

Nicolis, S. C., & Deneubourg, J. 1999. Emerging patterns and food recruitment in ants: an analytical study. *Journal of theoretical biology*, 198, 575-592.

Nowak, M. A. & Sigmund, K. 2005. Evolution of indirect reciprocity. *Nature*, 437, 1291–1298.

Odling-Smee, F. J., Laland, K. N. & Feldman, M. W. 2003. *Niche Construction: The Neglected Process in Evolution. Monographs in Pupulation Biology 37.* Princeton: Princeton University Press.（佐倉統・山下篤子・徳永幸彦(訳) (2007). ニッチ構築-忘れられていた進化過程-　共立出版）

Oldroyd, B. P., Gloag, R. S., Even, N., Wattanachaiyingcharoen, W., & Beekman, M. 2008. Nest site selection in the open-nesting honeybee *Apis florea*. *Behavioral Ecology and Sociobiology*, 62, 1643–1653.

Ostrom, E. 1990. Governing the commons: The evolution of institutions for collective action. Cambridge: Cambridge University Press.

Powell, G. V. N. 1974. Experimental analysis of the social value of flocking by starlings (*Sturnus vulgaris*) in relation to predation and foraging. *Animal Behaviour* 22, 501-505.

Pratt, S. C. 2005. Quorum sensing by encounter rates in the ant *Temnothorax albipennis*. *Behavioral Ecology*,16, 488–496.

Pratt, S. C., Mallon, E., Sumpter, D. J. T., & Franks, R. N. 2002. Quorum sensing, recruitment, and collective decision-making during colony emigration by the ant *Leptothorax albipennis*. *Behavioral Ecology and Sociobiology*, 52, 117-127.

Pratt, S. C., Sumpter, D. J. T., Mallon, E. B., & Franks, N. R. 2005. An agent-based model of collective nest choice by the ant *Temnothorax albipennis*. *Animal Behaviour*, 70, 1023–1036.

Raafat, R. M., Chater, N., & Frith, C. 2009. Herding in humans. *Trends in Cognitive Sciences*,13, 420–428.

Ravakov, D. V. 1973. *Schooling in the ecology of fish*. New York: John Wiley & Sons.

Rendell, L., Boyd, R., Cownden, D., Enquist, M., Eriksson, K., Feldman, M. W., Forgarty, L., Ghirlanda, S., Lillicrap, T. & Laland, K. N. 2010. Why copy others? Insight from the social learning strategies tournament. *Science*, 328, 208-213.

Rendell, L., Fogarty, L. & Laland, K. N. 2009. Rogers' paradox recast and resolved: population structure and the evolution of social learning strategies. *Evolution* 64: 534-548.

Richer, H. & Heeb, P. 1995. Is the information centre hypothesis a flop? *Advances in the Study of Behavior*, **24**, 1-45.

Richerson, P. & Boyd, R. 2004. *Not Genes Alone: How Culture Transformed Human Evolution*. Chicago, IL: University of Chicago Press. 332 p.

Robins, H. 1952. Some aspects of the sequential design of experiments. *Bull Am Math Soc* 58: 527–535.

Rogers, A. R. 1988. Does Biology Constrain Culture? *Am Anthropol* 90: 819–831.

Rosen, S. 1981 The Economics of Superstars. *The American Economic Review*, 71, 845-858.

Rosenberg, D., Salomon, A. & Vieille, N. 2013. On games of strategic experimentation. *Games and Economic Behavior*, **82**, 31-51.

Salganik, M. J., Dodds, P. S., & Watts, E. 2006. Experimental study of inequality and cultural market. *Science*, 311, 854–856.

Sasaki, T., Granovskiy, B., Mann, R. P., Sumpter, D. J. T. & Pratt, S. C. 2013. Ant colonies outperform individuals when a sensory discrimination task is difficult but not when it is easy. *Proceedings of the National Academy of Sciences*. 110(34): 13769-13773.

Seeley, T.D. 2010. *Honeybee democracy*. Princeton, NJ: Princeton University Press.

Seeley, T. D., Camazine, S., & Sneyd J. 1991. Collective decision-making in honey bees: how colonies choose among nectar sources. *Behavioral Ecology and Sociobiology*, 28, 277–290.

Seeley, T. D., & Towne, W. 1992. Tactics of dance choice in honey bees: do foragers compare dances? *Behavioral Ecology and Sociobiology*, 30, 59-69.

Seeley, T. D., & Visscher, P. K. 2003. Choosing a home: How the scouts in a honey bee swarm perceive the completion of their decision making. *Behavioral Ecology and Sociobiology*, 54, 511–520.

Seeley, T. D., & Visscher, P. K. 2004. Quorum sensing during nest-site selection by honeybee swarms. *Behavioral Ecology and Sociobiology*, 56, 594–601.

Shaffer, Z., Sasaki, T. & Pratt, S. C. 2013. Linear recruitment leads to allocation and flexibility in collective foraging by ants. *Anim Behav* 86: 967-975.

Shaw, M. E. 1932. Comparison of individuals and small groups in the rational solution of complex problems. *American Journal of Psychology*, 44, 491-504.

Sherif, M. 1936. *The psychology of social norms*. New York, NY: Harper Collins.

Simons, A. M. 2004. Many wrongs: The advantage of group navigation. *Trends in Ecology and Evolution*, 19, 453–455.

Sueur, C., Deneubourg, J., & Petit, O. 2010. Sequence of quorums during collective decision making in macaques. *Behavioral Ecology and Sociobiology*, 64, 1875-1885.

Sueur, C., Deneubourg, J., & Petit, O. 2011. From the first intention movement to the last joiner: macaques combine mimetic rules to optimize their collective decisions, *Proceedings of the Royal Society B*, 278, 1697-1704.

Sumpter, D. J. T. 2010. *Collective animal behavior*. Princeton, NJ: Princeton University Press.

Sumpter, D. J. T., & Pratt, S. C., 2003. A modeling framework for understanding social insect foraging. *Behavioral Ecology and Socioniology*, 53, 131-144.

Sumpter, D. J. T., & Pratt, S. C., 2009. Quorum responses and consensus decision making. *Philosophical Transactions of Royal Society B*, 364, 743-753.

Surowiecki, J. 2004. *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations*. New York, NY: Doubleday.

Sutton, R. S. & Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. Massachusetts: The MIT Press. (サットン R.S.・バート A.G. 三上貞芳・皆川雅章（訳）(2000). 強化学習　森北出版株式会社)

Suzuki, S. Harasawa, N., Ueno, K., Gardner, J. L., Ichinohe, N., Haruno, M., Cheng, K. & Nakahara, H. 2012. Learning to simulate others' decisions. *Neuron* 74, 1125-1137.

Tewari, A. & Bartlett, P. L. 2008. Optimistic linear programming gives Logarithmic regret for irreducible MDPs. In: Platt JC, Koller D, Singer Y, Roweis S, editors. *Advances in Neural Information Processing Systems 20*. Cambridge: MIT Press.

Toelch, U., Bruce, M., Meeus, M. T. H. & Reader, S. M. 2010. Humans copy rapidly increasing choices in a multi-armed bandit problem. *Evol Hum Behav* 31: 326-333.

Tokic, M. & Palm, G. 2011. Value-difference based exploration: adaptive control between epsilon-greedy and softmax. In: KI 2011: *Advances in Artificial Intelligence: 34th Annual German Conference on AI, Berlin, Germany, October 4-7, 2011, Proceedings.* Heidelberg: Springer. pp. 335-346.

豊川航・亀田達也 2013. ヒトと動物の「集団意思決定」をつなぐ　動物心理学研究, 63 巻, pp. 107-122.

豊川航・亀田達也 2014. "探索と収穫のジレンマ"における社会的学習と集合知効果の実験的検討. 北海道心理学研究, 36 巻, pp. 1-14.

Toyokawa, W., Kim, H. & Kameda, T. 2014. Human collective intelligence under dual exploration-exploitation dilemmas. *PLoS ONE*, 9(4): e95789

van Leeuwen, E., J., C., & Haun, D., B., M. 2013. Conformity in nonhuman primates: fad or fact? *Evolution and Human Behavior*, 34, 1-7.

Visscher, P. K., & Seeley, T.D. 2007. Coordinating a group departure: who produces the piping sig- nals on honeybee swarms? *Behavioral Ecology and Sociobiology*, 61, 1615–1621.

Wakano, J., & Aoki, K. 2007. Do social learning and conformist bias coevolved? Henrich and Boyd revisited. *Theoretical population biology*, 72, 504-512.

Walker, M. B., & Andrade, M. G. 1996. Conformity in the Asch Task as a Function of Age. *The Journal of Social Psychology*, 136, 367-372.

Ward, A. J. W., Sumpter, D. J. T., Couzin, I. D., Hart, P. J. B., & Krause, J. 2008. Quorum decision-making facilitates information transfer in fish shoals. *Proceedings of the National Academy of Science of United States of America*, 105, 6948-53.

Ward, P. & Zahavi, A. 1973. The importance of certain assemblages of birds as “information-centres” for food finding. *Ibis*, **115**, 517-534.

Whiten, A., Horner, V., & de Waal, F. B. M. 2005. Conformity to cultural norms of tool use in chimpanzees, *Nature*, 437, 737-740.

Widemo, F. & Owens, I. P. F. 1995. Lek size, male mating skew and the evolution of lekking. *Nature* 373, 148-150.

Wilson, D. S. 1997. Incorporating group selection into the adaptationist program: A case study involving human decision making. In J. Simpson, & D. Kenrick (Eds.), *Evolutionary social psychology* (pp. 345–386). Hillsdale, NJ: Erlbaum.

Wolf, M., Kurvers, R. H. J., Ward, A. J. W., Krause, S. & Krause, J. 2013. Accurate decision in an uncertain world: collective cognition increases true positives while decreasing false positives. *Proc R Soc B* 280: 20122777.

Wood, A. J. & Ackland, G. J. 2007. Evolving the selfish herd: Emergence of distinct aggregating strategies in an agent-based model. *Proc R Soc B* 274: 1637-1642.

付録

付録 第二章

1．付録 2-1: 学習モデルパラメータのベイズ推定

実験において各参加者がどの選択肢を選んだかのデータは，本文中で記述した学習モデルで表現される選択確率分布 $\boldsymbol{Pr_t} = \{Pr_{1,t}, Pr_{2,t}, \dots, Pr_{30,t}\}$ からラウンドごとに確率的に得られると仮定する。ただし，選択確率分布は学習の進行にともなって変化することに注意されたい。

個人学習モデルにおいて推定するパラメータは温度パラメータτのみである。τの事前分布としては，定義域が 0 から 100 までの一様分布を用いた。JAGS v.3.3.0. によるマルコフ連鎖モンテカルロ法（MCMC 法）によってパラメータの事後分布を推定した。なお，推定は 12000 ステップの計算を 5 系列繰り返し，その平均を用いた。また初期値依存的である最初の 2000 ステップは推定結果に含めず，20 ステップごとの結果を推定に用いた。モデルの収束については，各パラメータの PSRF（Potential Scale Reduction Factor; Gelman-Rubin 統計; R-hut）が 1.1 を下回っているかで評価した（Brooks & Gelman, 1997; Gelman & Rubin, 1992）。MCMC の収束判定については，以下全て同様である。

社会的学習モデルにおける推定パラメータは，温度パラメータτ，社会的情報への重みλ，頻度依存性θの 3 つである。各パラメータの事前分布はすべて，平均 0，分散 10^4 である正規分布を用いた。ただし，τとθは 0 以上という制約を付け，λに関しては 0 以上 1 以下という制約を付けた。JAGS v.3.3.0. によるマルコフ連鎖モンテカルロ法（MCMC 法）によってパラメータの事後分布を推定した。なお，推定は 3000 ステップの計算を 5 系列繰り返し，その平均を用いた。また初期値依存的である最初の 1000 ステップは推定結果に含めず，20 ステップごとの結果を推定に用いた。

２．付録 2–2: 探索量の階層ベイズ分析

２.１ 探索確率のロジットモデル

探索確率の分析は，混合ロジットモデルを用いた。パラメータは，データ全体に共通する大域的パラメータである切片（β_1）と条件（個人/集団）の効果を表す係数（β_2）の他に，個人ごとのランダム効果を表す $r(i)$，集団ごとのランダム効果を表す $rg(j)$ という２つの局所的パラメータがある。また $condition(i)$ はダミー変数（個人条件 = 0，集団条件 = 1）である。このモデルでは，２つのランダム効果がそれぞれの階層事前分布（平均ゼロで標準偏差がそれぞれ s および sg である正規分布）から生成されると仮定する。

事後分布が推定されるパラメータは，切片 β_1，条件の効果を表す係数 β_2，２つのランダム効果の標準偏差 s および sg，の４つである。推定されるパラメータの事前分布として，無情報事前分布を用いた。それぞれ，β_1 および β_2 は平均ゼロで分散が 10^4 である正規分布，定義域が正である s および sg は 0 から 10^4 の範囲の一様分布を用いた。

JAGS v.3.3.0. によるマルコフ連鎖モンテカルロ法（MCMC 法）によってパラメータの事後分布を推定した。なお，推定は 240000 ステップの計算を 5 系列繰り返し，その平均を用いた。また初期値依存的である最初の 40000 ステップは推定結果に含めず，250 ステップごとの結果を推定に用いた。

２.２ 探索のゆらぎ（エントロピー）モデル

$block\ j$ における個人 i のエントロピー（$H_{i,j}$）は，期待値 $\mu_{i,j}$，分散 v の正規分布から出されると仮定する。$\mu_{i,j}$ は階層線形モデルによって記述され，そのパラメータは，切片 $intercept$, block の効果（β_1），条件の効果（β_2）, block と条件との交互作用（β_3）の３つである。他に，個人ごとのランダム効果を表す $r(i)$，集団ごとのランダム効果を表す $rg(j)$ という２つの項がある。推定パラメータは，v，$intercept$，β_1，β_2，β_3，s，sg の７つである。

JAGS v.3.3.0. によるマルコフ連鎖モンテカルロ法（MCMC 法）によってパラメータの事後分布を推定した。なお，推定は 110000 ステップの計算を 5 系列繰り返し，その平均を用いた。また初期値依存的である最初の 10000 ステップは推定結果に含めず，250 ステップごとの結果を推定に用いた。

3．付録 2-3: スコアの分析

課題の成績（スコア）についてのモデルは，上記のエントロピーモデルと同様，階層線形モデルである。ただしスコアは参加者 1 人あたり 1 つの値のみを持つので，個人のランダム効果は考慮しなくてもよくなる。したがってモデルでは集団のランダム効果だけが考慮されている。

MCMC によるパラメータの事後分布推定には 104000 ステップの計算を 5 系列繰り返し，その平均を用いた。また初期値依存的である最初の 4000 ステップは推定結果に含めず，200 ステップごとの結果を推定に用いた。

付録 第三章

1. SI 3-1: Defining exploration and exploitation

We followed the traditional definitions of exploration and exploitation in the reinforcement learning literature [1]. Individual x facing a choice at round t has a unique vector $Q(x,t) = \{Q^x_{1,t}, Q^x_{2,t}, Q^x_{3,t}, \dots, Q^x_{30,t}\}$, whose elements represent expected payoffs of the 30 options based on x's cumulative knowledge up to round t. As x's experience accumulates, Q (x, t) is updated by an averaging rule, in which $Q^x_{i,t}$ is equal to the mean payoff that x has earned from option i until round t. The averaging updating rule is a suitable assumption of the action-value function in a stationary multi-armed bandit task [1]. This model assumes that players can hold up to 30 values for the different options simultaneously in working memory and calculate a new average for each option. Although such assumptions are too strong for human cognitive capacity, we believe that this model can provide a first approximation to distinguish between exploration and exploitation in a stationary multi-armed bandit task. We categorized choice behaviour as "exploitation" when the individual x chose option i with max $Q^x_{i,t}$ in round t + 1 (i.e., the greedy choice). Otherwise, we categorized the choice as "exploration." The initial Q values were set to zero, i.e., $Q(x,0) = \{0,0,0,\dots,0\}$, because x had no prior knowledge about the expected payoff from each option. For the first round (t = 1) only, we categorized all choices as "exploration."

2. SI 3–2: Comparing exploration frequencies between the individual and group conditions

To compare the frequencies of exploratory choices over all 100 rounds between the

individual and group conditions, we used a hierarchical Bayesian method. We modelled the probability of individual x in group g engaging in exploration ($q_{x,g}$) as a mixed logit model:

$$\ln\left(\frac{q_{x,g}}{1-q_{x,g}}\right) = \lambda_{1,0} + \lambda_{1,1} * (condition) + \varepsilon_1(x) + \mathrm{E}_1(g),$$

where $\lambda_{1,0}$ indicates an intercept, $\lambda_{1,1}$ indicates a fixed effect for the condition (dummy coded: individual condition = 0; group condition = 1), and local parameters $\varepsilon_1(x)$ and $\mathrm{E}_1(g)$ indicate random effects specific to the individual and to the group respectively ($\varepsilon_1(x) \sim Normal(0, \sigma_1)$ and $\mathrm{E}_1(g) \sim Normal(0, \Sigma_1)$).

To estimate the four parameters (λ1, 0, λ1, 1, σ1, and Σ1), we set the uninformed priors of λ1, 0 and λ1, 1 to a normal distribution with mean = 0 and variance = 104 (i.e., a very stretched distribution), and those of the hyper parameters (σ1 and Σ1) to a uniform distribution [0, 104]. We conducted Markov Chain Monte Carlo (MCMC) simulations for 5 independent sequences. The number of total iterations per chain was 240,000 (thinning rate was 250), and first 40,000 steps were discarded from analysis because of initial-value dependencies. We used 95% Bayesian credible intervals to determine the significance of each parameter.

3. SI 3-3: Comparing exploration frequencies between the frequency-only and frequency-plus-evaluation sub-conditions

To compare the frequencies of exploratory choices over all 100 rounds between the two sub-conditions, we used a hierarchical Bayesian method. We modelled the probability of individual x in group g engaging in exploration (q_(x,g)) as a mixed logit model:

$$\ln\left(\frac{q_{x,g}}{1-q_{x,g}}\right) = \lambda_{2,0} + \lambda_{2,1} * (subcondition) + \varepsilon_2(x) + \mathrm{E}_2(g),$$

where λ2, 0 indicates an intercept, λ2, 1 indicates a fixed effect for the sub-condition (dummy coded: frequency-only sub-condition = 0; frequency-plus-evaluation sub-condition = 1), and local parameters ε2 (x) and E_2 (g) indicate random effects specific to the individual and to the group respectively (ε_2 (x) ~ Normal(0,σ_2) and E_2 (g) ~ Normal(0,Σ_2)).

Because the structure of the model was identical to the equation S1, the MCMC protocol was the same as described above.

4. SI 3-4 Analysis of the causality between exploration and evaluation

To analyze factors affecting individual exploratory choices, we constructed two models. One model posits that an individual's exploration probability at round t + 1 is influenced by the total numbervolume of ratingsevaluation information contributed in the group at round t (evaluation effect model). The converse model posits that the individual probability of contributing ratingevaluation information at the feedback stage of round t is influenced by whether or not she/he has engaged in exploration at the choice stage of round t (exploration effect model).

For the evaluation effect model, the probability of individual x in group g engaging in exploration at round t + 1 (p_(x,g,t+1)) was modelled:

$$ln\left(\frac{p_{x,g,t+1}}{1-p_{x,g,t+1}}\right) = \alpha_1 + \beta_1 * W_{g,t} + \varepsilon_3(x) + \mathrm{E}_3(g),$$

where z_{ij} = β1 + β2*condition(i) + r(i) + rg(j),α1 indicates an intercept, ß1 indicates a fixed effect for the total number of ratings contributed in the group at the preceding round ($W_{g,t}$), and local parameters ε3(x) and $\mathrm{E}_3(g)$ indicate random effects specific to the individual and to the group respectively ($\varepsilon_3(x) \sim Normal(0, \sigma_3)$ and $\mathrm{E}_3(g) \sim Normal(0, \Sigma_3)$).

To estimate the four parameters (α1, ß1, σ3, and Σ3), we set the uninformed priors of α1 and ß1 to a normal distribution with mean = 0 and variance = 104 (i.e., a very stretched uniform distribution), and those of the hyper parameters (σ3 and Σ3) to a uniform distribution [0, 104]. We conducted MCMC simulations for 5 independent sequences. The number of total iterations per chain was 63,000 (thinning rate was 300), and first 3,000 steps were discarded from analysis because of initial-value dependencies. We used 95% credible intervals to determine the significance of each parameter.

For the exploration effect model, the probability of contributing evaluation at the feedback stage of round t was modelled:

$$\ln\left(\frac{\omega_{x,g,t}}{1-\omega_{x,g,t}}\right) = \alpha_2 + \beta_2 * (exploration)_{x,t} + \varepsilon_4(x) + \mathrm{E}_4(g),$$

where ωx,g,t indicates the probability of individual x contributing her/his evaluation at round t, α2 for an intercept, ß2 for a fixed effect of exploration in the current period t (dummy coded: exploration = 1, exploitation = 0). Local parameters ε4(x) and E4(g) indicate random effects specific to the individual and to the group respectively ($\varepsilon_4(x) \sim Normal(0, \sigma_4)$ and $\mathrm{E}_4(g) \sim Normal(0, \Sigma_4)$). Because the structure of the model was identical to the evaluation effect model, the MCMC protocol was the same as described above.

5 . SI 3-5 Supporting figures

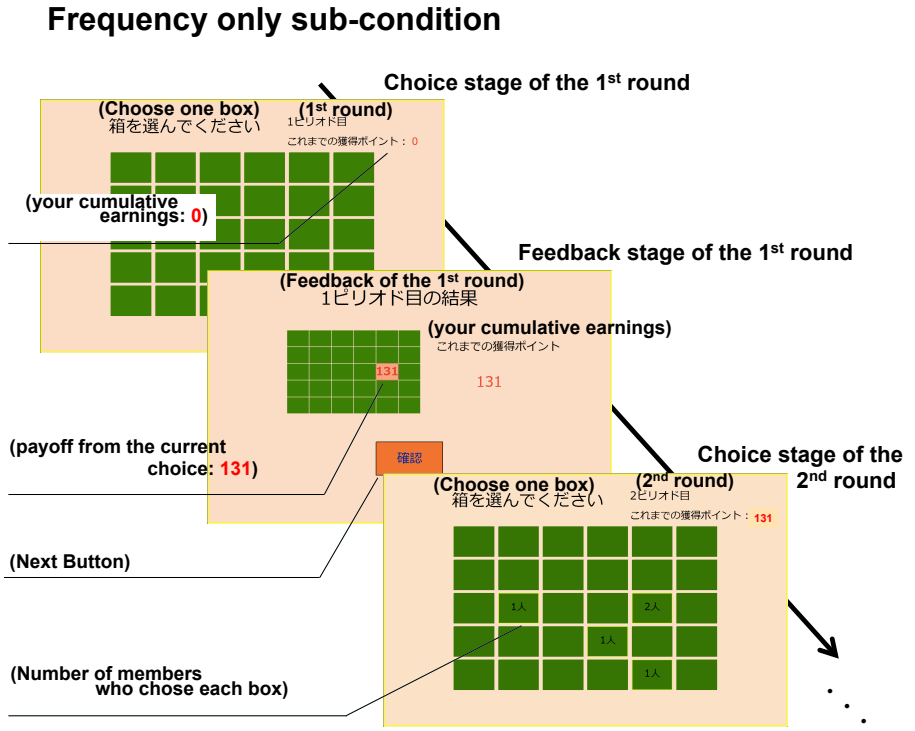


Figure S3-1: Time sequence in the frequency only sub-condition. At the choice stage, each participant chose one box from 30 green box icons, then proceeded to the feedback stage where she/he could learn how many points they earned in this round. By clicking the next button, they could proceed to the next round's choice stage. Participants were provided 100 rounds in total, although they were not informed of this ahead of time.

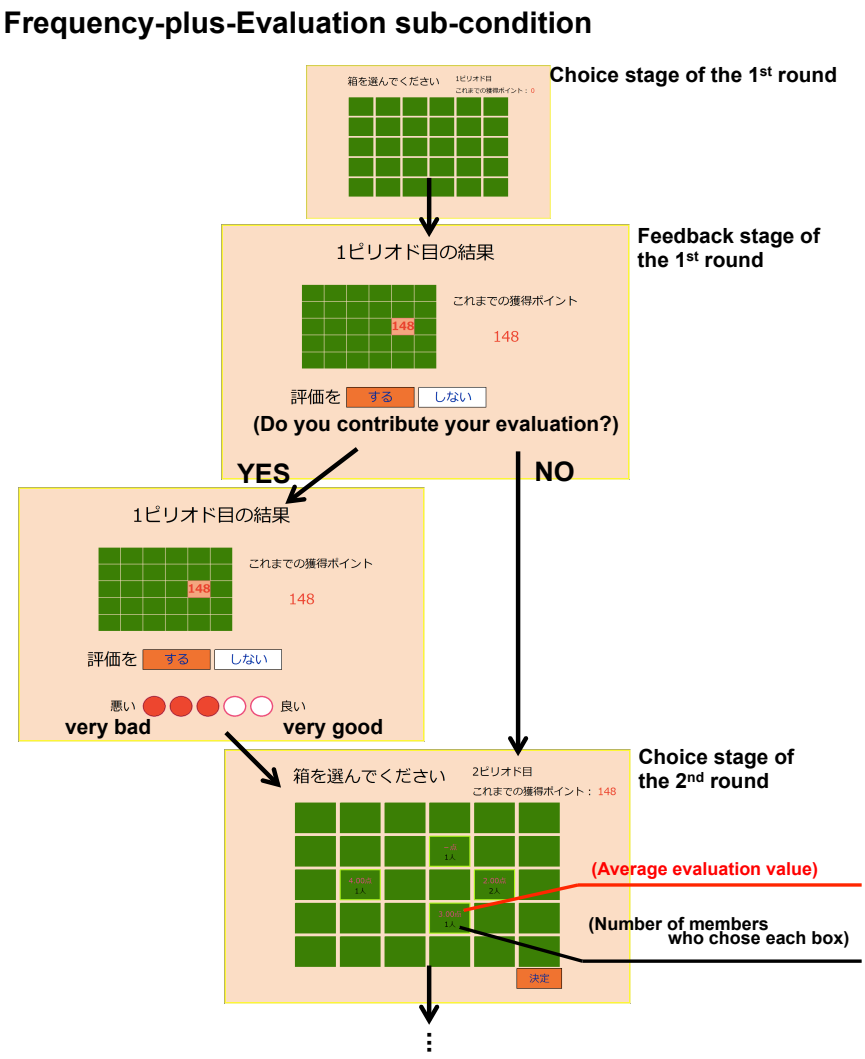


Figure S3-2: Time sequence in the frequency-plus-evaluation sub-condition. At the result stage of each period, participants were asked to decide whether or not to contribute an evaluation of the chosen option. If yes, they were asked to rate the option on a 5-point scale before proceeding to the next round's choice stage.

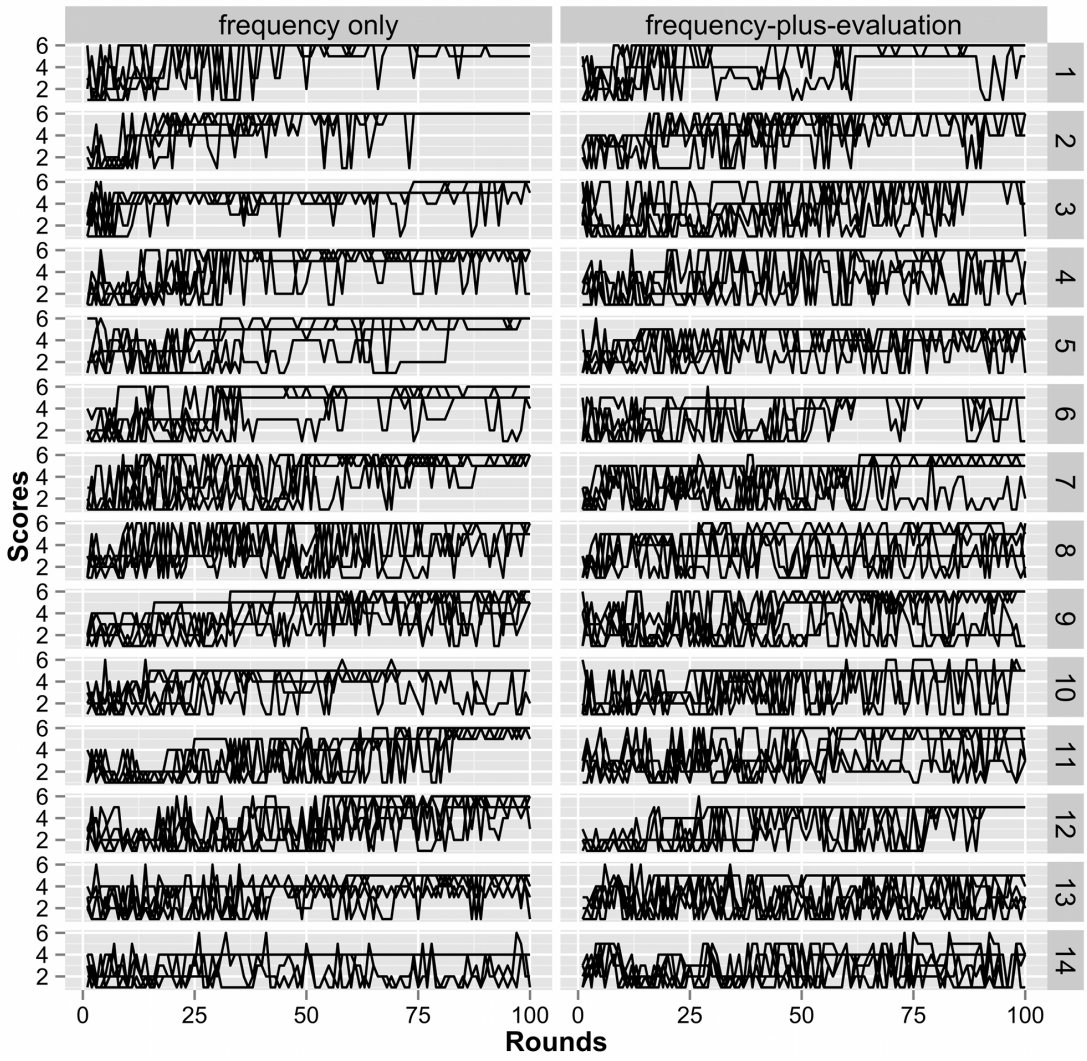


Figure S3-3: Trajectories of all participants' choices are shown separately for each group. Performances of 5 participants belonging to the same group are displayed by separate curves in each group. The y-axis refers to objective qualities of chosen boxes. The left column shows choice trajectories in the frequency-only sub-condition and the right columns shows the frequency-plus-evaluation sub-condition. The numbers on the right indicate the rank of the group in terms of total scores averaged over its 5 members, within the respective sub-condition. As can be seen, choices fluctuated more (i.e., were more exploratory) in the frequency-plus-evaluation sub-condition than in the frequency-only sub-condition, and more among lower ranked groups than among higher ranked groups.

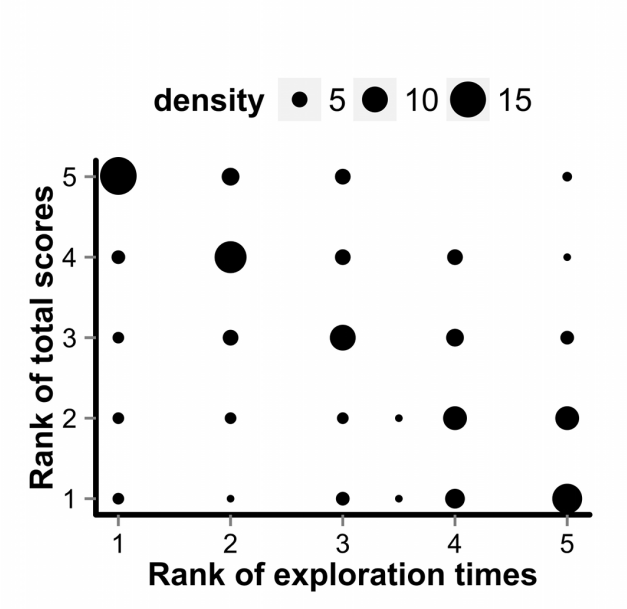


Figure S3-4: Relation between frequencies of explorations and participants' total scores. The x-axis refers to a participant's rank within a 5-person group in terms of exploration frequency. The y-axis refers to the participant's rank within the group in her/his total score. The size of circle represents the number of participants in the respective rank.

6. SI 3-6 Supporting tables

Table S3-1: MCMC results of the exploration probability model (equation S1)

Parameters	Mean	SD	Quantiles			Gelman-Rubin Statistics	
			2.5%	50.0%	97.5%	median (upper C.I.)	Effective sample size
$\lambda_{1,0}$	1.94	0.25	1.42	1.94	2.47	1.00 (1.01)	1527
$\lambda_{1,1}$	-0.94	0.32	-1.63	-0.94	-0.27	1.00 (1.01)	1528
σ_1	1.62	0.11	1.42	1.61	1.86	1.00 (1.00)	4051
Σ_1	0.83	0.19	0.44	0.83	1.20	1.00 (1.01)	2921

The Gelman-Rubin statistics of each parameter was lower than 1.1, which means the MCMC sampling converged.

Table S3-2: MCMC results of the exploration probability model (equation S2)

Parameters	Mean	SD	Quantiles			Gelman-Rubin Statistics	
			2.5%	50.0%	97.5%	median (upper C.I.)	Effective sample size
$\lambda_{2,0}$	0.54	0.30	-0.05	0.54	1.13	1.01 (1.02)	1675
$\lambda_{2,1}$	0.92	0.43	0.07	0.91	1.78	1.00 (1.01)	1959
σ_2	1.70	0.13	1.42	1.69	1.96	1.00 (1.00)	3611
Σ_2	0.80	0.25	0.28	0.80	1.31	1.00 (1.00)	2494

The Gelman-Rubin statistic for each parameter was lower than 1.1, which means the MCMC sampling converged.

Table S3-3: MCMC results of the causality between exploration and evaluation

	Parameter	Mean	SD	Quantiles			Gelman-Rubin Statistics
				2.5%	50.0%	97.5%	median (upper C.I.)
Evaluation effect model (q)	α_0	1.47	0.32	0.86	1.48	2.10	1.01 (1.03)
	α_1	0.42	0.040	0.34	0.42	0.50	1.00 (1.02)
	σ_q	1.93	0.20	1.57	1.91	2.38	1.01 (1.03)
	Σ_q	0.67	0.39	0.04	0.66	0.90	1.00 (1.01)
Exploration effect model (ω)	β_0	0.87	0.39	0.094	0.86	1.67	1.01 (1.02)
	β_1	0.34	0.089	0.17	0.34	0.52	1.00 (1.01)
	σ_ω	2.09	0.25	1.67	2.06	2.60	1.00 (1.01)
	Σ_ω	0.98	0.47	0.12	0.96	2.01	1.00 (1.02)

The Gelman-Rubin statistics of each parameter was lower than 1.1, which means the MCMC sampling converged.