

集合知予備実験結果報告

有馬 淑子

従来、群集は誤った判断に同調してバブルや群集雪崩を起こす愚かなものと考えられてきた。ところが近年、検索システムなどのインターネット技術の発達により、個々人の選択を機械的に集約すれば正確な判断が導かれることが知られるようになった。このような集約効果に対する厳密な学術用語はまだないが、ここではスロウィッキー(2006)の言葉を使い、集合知と呼ぶ。集合知は、他者の選択が関わる選挙や株価などの社会現象を予測する課題への適用が期待されている。実用的には、専門家の予測と、多くの人が予測を競う情報市場の予測とでは、どちらがより正確なのか問題となる。

本稿では、本学における心理学基礎実験Aにおいて実施した集合知実験結果の一部を報告する。全体のデータ分析の結果は、別途報告する予定である。この集合知実験は、正解は存在するが判断の困難な課題を用いて、4名集団による集団決定と集合知(クラス平均値)の、いずれが正解に近いかを探索的に検討したものである。過去7年にわたり14クラス(それぞれ20~30名程度)の実験結果が蓄積されているが、このテーマの性質上、1名を1ケースとはカウントできず、1クラスが1ケースとなる。本報告では、過去に得られたデータから代表的なパターンと考えられる2クラスを例にとり、事例報告を行う。

事例報告を行う前に、まだ明確な定義のない集合知という概念に関する概説と、この問題を社会心理学的に検討する意義について述べる。

1 集合知とは何か

1.1 集合知とは

近年、集合知という言葉が人口に膾炙されるようになってきた。集合知とは、生物の群れなどの集合体が、各個体が保持する以上の知性を見せる現象である。一般的にはインターネットのシステムによって実現される機能を指して使われる。その事例となるシステムは多様である。たとえば、リンク数を指標とする検索システム、知識を協同編集するシステムであるウィキペディア、ECサイトにおける商品レビュー、そのレビューワーカーのスター獲得システム、などである。

このようなシステムでは個人に分散された知識を集める機能が期待されているため、マーケティング、予測市場、民主的意志決定、コミュニティにおける課題解決など多彩な分野において、集合知の応用が試みられている。特に近年は、刻々と流入する人々の動きや購買情報などを分析するビッグデータを機械学習を用いて分析するシステムが構築されており、今後、様々な予測に力を発揮すると考えられている。さらに、クラウドソーシングやコモンズなどの、人々の働き方を変えつつあるシステムも集合知に絡む問題として研究されており、将来的には組織や社会のあり方にも影響する問題でもある。

1.2 社会心理学と集合知

このように集合知という言葉は重要かつ、なじみあるものになりつつあるが、一方でその概念は曖昧でとらえどころがない。さまざまな分野の研究者が応用領域として研究しているものの、理論としては多様な分野の複合体であり統一した研究分野ではない。しかし、徐々にその基盤となる科学は姿を現しつつある。中心となる分野は数理生物学および情報工学であるが、筆者自身が専攻する集団心理学も重要な一角を占める。たとえば、

2014年度に MIT で開催された集合知学会に散見された研究としては、集団の意志決定、小集団のタスクパフォーマンスや協同、ネットワーク、リーダーシップなど、過去の集団研究で行われてきたトピックが扱われている。

次に、集団意思決定研究の観点から集合知について検討する。

2 集団の意志決定と集合知

2.1 社会心理学における集団意思決定研究

意志決定研究には、大きく分けて二つの古い研究文脈がある。一つはグループダイナミクス研究からの流れ、もう一つは経済学的モデルを用いる数理社会学の流れである。前者は、集団が個人に及ぼす効果を検討する物である。1950年代のグループダイナミクス研究創設時から、集団意思決定により食事習慣を変える確率が高くなる、など、個人の態度変容に及ぼす集団の効果が検討されてきた。この流れは、多数者 vs. 少数者の社会的影響過程研究として展開した後、個人と集団のパフォーマンス(正確さ、遂行時間などの効率性)を比較する研究が数多く行われてきた。

経済学的モデルによる意志決定研究としては、たとえば、民主的決定の不可能性を示すアローの定理などの投票制度の研究などが代表となる。経済学的モデルは社会心理学における集団意思決定研究に影響を与え、現実の集団の意思決定ルールを数学的に記述する試みが行われてきた。この流れは進化心理学的シミュレーション研究としても発展している。現在の社会心理学における集団研究は、小集団を用いて実際に実験する研究と、シミュレーション研究の二つの流れが分離した状態で行われている。集合知研究はこのどちらにも属しないが、両者の橋渡しになる分野として発展することを期待している。

2.2 集団の意思決定

社会心理学における集団意思決定研究を概観すると、集団決定は、個人の決定の平均よりはよいパフォーマンスを示すが、集団の中のベストメンバーのパフォーマンスには及ばない、という結果が一貫して得られてきた (Brown, 1988)。すなわち、集団はもっともよいパフォーマンスを発揮するメンバー (課題によってそのメンバーは異なる) の力を抑制するのである。この抑制の原因を探る研究が行われており、たとえば発言機会を失う、同調などさまざまな要因が検討されている。

同じことは集団記憶についても言える。集団記憶研究とは、集団で記銘または再生再認作業を行う場合と、個人で記銘または再生再認作業を行う場合で、どちらのパフォーマンスが優れているかを調べる実験パラダイムである。集団記憶においてもやはり、集団は個人のパフォーマンスをすべてあわせた記憶力を示さないことが示されている (Basden, Basden, Bryner, & Thomas, 1997)。この現象は協同抑制と呼ばれ、その原因を探る研究が行われている (Weldon, Blair, & Huesch, 2000)。

集団は記憶の総量においては個人の合計に劣るものの、個人より勝るパフォーマンスを示す側面もある。集団は多数決ルールに従って決定を行う確率が高いため、間違いが少数者であれば場合、多数のメンバーがそれを指摘して間違いを正す可能性が高い (Davis, 1973)。よって、正確性については集団の方が個人平均よりも高くなると予測され、この予測はおおむね支持されている (Hinsz, 1990)。

2.3 集合知と集団

「みんなの意見は案外正しい」という日本語タイトルで知られるスロウイッキー (2006) の著書では、集合知に関わる問題を、多様性、独立性、分散性、集約性の4つの要因にまとめている。そして、個人の多様性が高く独立性が保たれている場合に、分散された知識を集約することにより、集合知が集団決定よりもよりよい決定を生み出すことを示唆している。

この過程を明確にしたものが、スコット・ペイジ(2007)によるシミュレーション研究である。彼は、個人の回答の誤差を足しあわせた数値は、個人の回答の分散と集合知(全体平均)の誤差を足したものに等しいとする定理を示して、これを多様性予測定理と名付けた。すなわち、集合知は、全体の誤差から多様性を引いた数値となるため、多様性が高いほど集合知は正確になることが、必然となる。彼はさらに、数学モデルとシミュレーションを用いて、認知的多様性の高い集団の方がよい決定を行いうることを示している。

これらの研究は、正解のある課題についてのものである。本稿で報告する実験課題も正解のある課題とした。課題遂行時に正解が確定している課題、未来に正解が確定する課題(クラスメンバーのパフォーマンスの推測)、推定範囲の異なる課題(最小値予測と最大値予測：前者は0から自己の推定値までと範囲が定まるが、後者は自己の推定値から上限がないため範囲が定まらない)における集団のパフォーマンスと、集合知のパフォーマンスを比較することを目的とした。さらに、クラスの中でもっとも正解に近い推定を行った個人を専門家とみなし、集合知と専門家のパフォーマンスの比較を行った。

今回検討する集合知(平均のパフォーマンス)、個人のパフォーマンスの平均、そして集団決定の比較については、次のように予測される。

各個人が予測を独立に行えば、誤差が互いに相殺される効果により集合知が現れると期待される。しかし、予測が同じ方向への偏りを持つものであれば、むしろ集約されることによって誤りが大きくなる。従来の社会心理学の知見によれば、小集団相互作用は認知的スキーマ(偏り)を共有させるため、平均値からの偏りを極端化させる影響がある。よって、集団決定の正確さは集合知よりも劣ると予想される。もっともパフォーマンスのよい個人(専門家)と集合知の比較については仮説を立てずに検討する。

3 方 法

3.1 実験参加者

集合知実験は、20～30名前後の実験クラスごとに1ケースとなる。各実験の実施時期、実験参加者の詳細は、結果にて記述する。

3.2 課題

曖昧だが正解のある課題としてランダムドット図の数を推定する課題を用いた。

それぞれのドット図に対して、4問設定される。

① 推定値

10秒で正確にドット数をカウントできる上限は約26個と見込み、26個を下限としてランダムにドット数を設定した。100個以上のドット数になると、偶然に正解する場合を除いて、ほぼ正解は不可能となる。この数の差によって、曖昧性の高さを操作した。

② 確信度

各推定値に対して、正解である確信度を0から100で評定する。教示の際には、正解に近いと考えられる場合は100に近い評定値を、正解から遠いと考えられる場合は0に近い数値を記入するように指示した。

③ 最小値の推定

実験は30名前後のクラス単位で実施されている。このクラス内の推定値の最大値と最小値をあらかじめ予測させた。その最小値の推定を先に推測させる。

④ 最大値の推定

クラス内の推定値の最大値を推測させる。

この③と④は未来に正解が確定する予測となる。本稿では③と④の結果については報告しない。

3.3 手続き

実験者が簡単に実験内容を説明した後、解答用紙を配付した。この実験に先立って、授業として心理学実験倫理規定の説明の時間が取られており、解答用紙には、実験結果の分析を許可する場合には、性別年齢を記入するようにとの教示文が示されている。回答者は、自分のデータを提供するかどうかについて、他者に知られることなく個別に判断することが可能な状況で回答した。

解答用紙には推定値、確信度、最小値推定、最大値推定を行う4列10行印刷されている。実験者はスクリーン上にドット図をプロジェクタで映写する形で刺激呈示が行われた。

パワーポイントを用いて10枚のランダムドット図を提示した。提示時間は各図について10秒間であった。1枚提示するごとに解答用紙に記入を求め、全員が記入し終わった後に次の図を提示した。課題終了後、4名集団に分かれて、10枚のドット数の推定値について集団決定を行った。

3.4 従属変数

推定値と正解の差の二乗を指標とした。この値が小さいほど、正確な推定が行われていたことを示す。

個人誤差	各個人の推定値と正解の差(10項目合計、以下同様) → この値がもっとも小さかった人が専門家となる
個人多様性	誤差平均(時期別)と各個人の誤差との差
集合知	個人誤差—個人多様性(この変数の平均値が集合知となる)
最小値誤差	推定された最小値と実際の最小値の差
最小値多様性	最小値の誤差平均と各個人の最小値誤差の差
最小値集合知	この変数の平均値が最小値推定に関する集合知となる
最大値誤差	推定された最大値と実際の最大値の
最大値多様性	最大値の誤差平均と各個人の最大値誤差の差
最大値集合知	この変数の平均値が最大値推定に関する集合知となる

確信度 推定値確信度の10項目合計

事例報告については、各指標のグラフを持って報告する。左から順に、個人誤差、多様性、集合知、専門家、集団誤差、集団多様性、集団集合知、専門家集団となる。縦軸は誤差の大きさなので、小さいほど賢かったことになる。

I：個人の知	(各個人の推定値－正解) 2
C：集合知	(推定値の平均－正解) 2
G：集団決定	(集団決定－正解) 2
B：専門家	もっとも誤差(I)の小さかった個人の値
平均値ルール	討議集団メンバーの初期推定値の平均値
中央値ルール	討議集団メンバーの初期推定値の中央値

なお、多様性予測定理より、D：多様性を(各個人の推定値-推定値の平均)²とすると、 $I=C-D$ である。すなわち、Dが大きい(個人の成績のばらつきが大きい)ほど、集合知は賢くなる。

今回の課題は、ランダムドット数推定であり、集団極化現象実験パラダイムで用いられるような「どちらでもない」という態度の中間が存在しない。よって、最初の集団の偏りは、正解数からの食い違いによって判定する。最小値・最大値予測の正解は、クラスメンバーの予測値が出てから定められる。この課題については集団決定は行われていない。

1番目に呈示される刺激は、10秒以内にはカウントできない曖昧な刺激である。このような刺激は実験参加者にはなじみのない課題であるため基準点がない。よって誤差の外れ値が多く出ると予想される。次の2番目の刺激に10秒以内に数えることで正解が導出できる刺激が用意されたため、ここで基準点ができると考えられる。そこで、3番目の刺激以降には各個人内時系列での基準点が設定されていたと期待できる。よって、パフォーマンスの指標を算出する際には、10項目合計点の他、最初の2刺激を削除した8項目合計点も計算した。

二乗値を計算するために、外れ値が全体平均に大きく影響する。よって、外れ値の除去が必要となるが、個人の多様性が集合知の要因となるため、その基準は慎重でなければならない。今回の事例研究では外れ値の除去を行った場合と行わなかった場合を比較して、予備実験として外れ値の除去の是非について検討した。さらに、10項目合計(より外れ値がしやすい)と8項目合計の結果の違いからも、外れ値の影響の程度を推測する。

4 結 果

4.1 事例1：2012年4月実施

実験参加者：大学生32名(男性19名、女性13名、平均年齢19.41歳)が本実験に参加した。集団決定時は4名の集団を8集団構成して行った。

誤差の大きさを図1に、このデータの推定値の分布図を図2に示す。

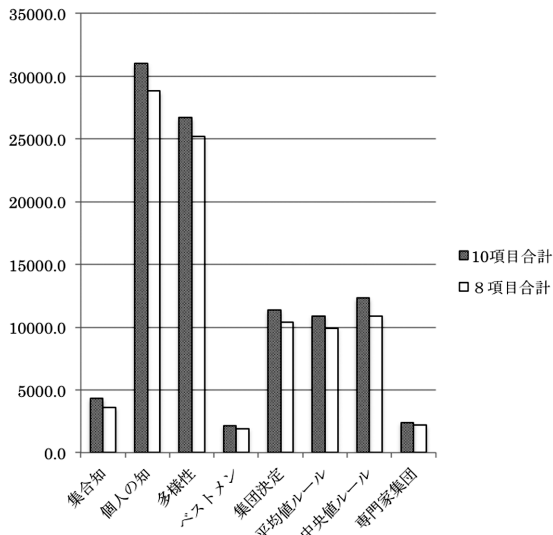


図1 事例1の結果

事例1では、10項目合計点をみた場合、もっともよい成績(誤差が少ない)となる指標は、クラス平均値を算出してから正解との差をとる集合知であった。数値は2204.9である。この数値の値は、多様性予測定理より、個人の知から多様性をひいた値に等しい。すなわち、この事例では多様性が高かったため、集合知の値が低くなっていることがわかる。

集合知の成績は、ベストメンバー(もっとも誤差の少なかった個人)3082や、専門家集団(もっとも誤差の少ない決定を行った討議集団)3137の成績よりもよい。集団の成績は、集合知に劣るだけでなく、集団が仮に平均値ルールに沿って決定を行った場合や、中央値ルールに沿って決定を行った場合に比べても成績が劣る物となっている。

一方、最初の2刺激を外した8項目合計点を見た場合は、集団の専門家(もっとも誤差が少なかった集団決定)1912の成績がもっともよい。集合知がそれに継ぐ成績となり2078.25、ベストメンバーの成績3001は集合知や専門家集団よりも劣る結果となっている。

事例1の分布図と外れ値検定の結果を図2に示す。正規性の検定結果は有意でこの分布は正規分布ではない。この事例では、推定値10項目合計点が107230を超える3名は外れ値と見なすことができる。外れ値となった実験参加者3名のデータを除去した後の10項目を指標とした集合知は1016.2、ベストメンバーは3082、8項目合計を指標とした場合は、集合知は973.592、ベストメンバーは3001であった。いずれも、ベストメンバーよりも集合知の成績の方がよい結果となる。よって、この事例からは、外れ値データを除去しても集合知とベストメンバーの優劣関係は変わらないことが示された。事例1は、外れ値の存在とは関係なく、集合知がベストメンバーよりも勝るケースが存在することを示すものである。

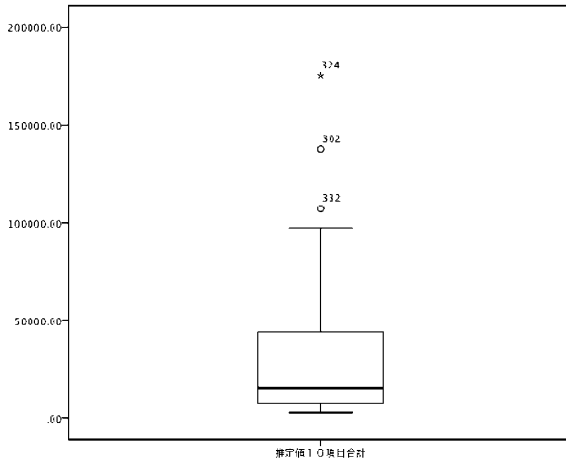
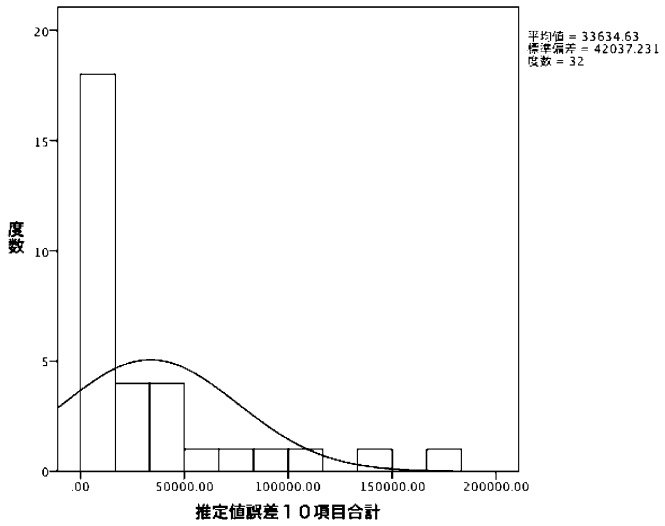


図2 事例1の推定値分布と外れ値検定結果

次に、集合知の方がベストメンバーよりも悪い例を見る。

4.2 事例2：2014年9月実施

実験参加者：大学生31名(男性18名, 女性13名, 平均年齢19.74歳)が本実験に参加した。集団決定時は4名の集団を7集団, 3名の集団を1集団構成して行った。

誤差の大きさを図2に, このデータの推定値の分布図を図3に示す。

事例2では, 10項目合計点をみた場合, もっともよい成績(誤差が少ない)となる指標は, ベストメンバー2155である。専門家集団がそれに継ぐ成績をおさめて(2414)おり, 集合知の成績はこれらよりも悪い(4315)成績であった。この値は個人の知から多様性を引いた値であることは事例1と同じである。個人の知と多様性の比は, 事例1では93.44%であったのに対して, 事例2では86.08%である。この多様性の低さが, 集合知がベストメンバーより低くなった原因と考えられる。

集団決定の成績は, 集合知に劣るだけでなく, 中央値ルールよりは悪いが, 平均値ルールに勝るものとなっている。以上の傾向は, 最初の2刺激をはずした8項目合計点でも同じ順位となった。

事例2の分布図と外れ値の結果を図4に示す。正規性の検定結果は有意でこの分布は正規分布ではない。この事例では, 推定値10項目合計点が82818を超える2名は外れ値と見なすことができる。外れ値を除去後の10項目を指標とした集合知は6609.0, ベストメンバーは2155, 8項目合計を指標とした場合は, 集合知は6023, ベストメンバーは1921であった。いずれも, 集合知よりもベストメンバーの成績の方がよい結果となる。よって, この事例からも, 外れ値を除去しても集合知とベストメンバーの優劣関係は変わらないことが示されている。

実験手続きとしては外れ値の除去は心理学での標準的手続きであるが, 多様性が鍵となる集合知の実験に関しては, 意図的に外れ値を出したと判

定できるケースを除き、除去しない方がよいことが示唆される。また、最初の2刺激をトライアルとみなして除去した結果も、すべての項目を使った結果と差は見られなかった。

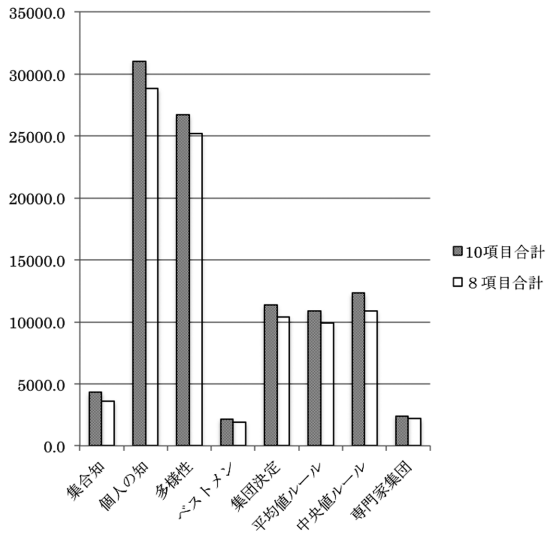


図3 事例2の結果

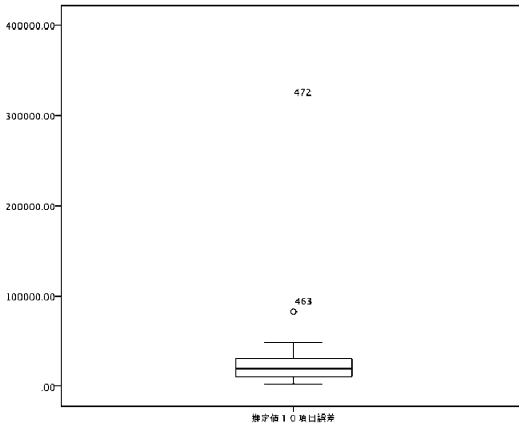
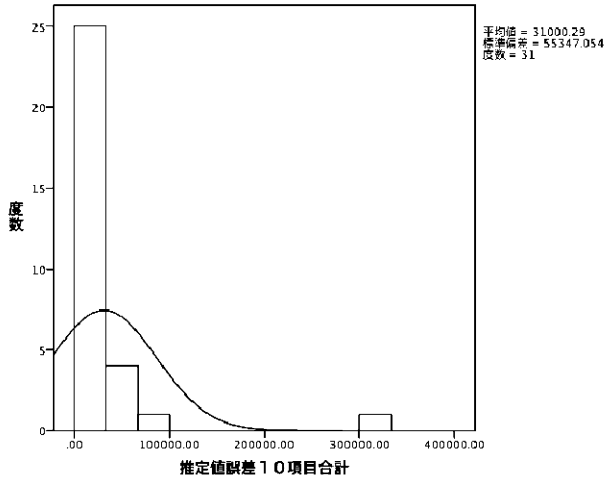


図4 事例2の推定値分布と外れ値検定結果

5 考 察

集合知のパフォーマンスが上がる仕組みについては、次の3つの可能性から論じられてきた。

- A) ローカルに分散していた知識を集合できる
- B) 正規分布をする事象の推定については、各個人の誤差を平準化できるために正しい推定が可能となる。
- C) 多様な観点により認知バイアスを低減できる

今回の事例報告は、B)の正しさを示している。すなわち、集合知は個人の成績の平均よりはよい成績を示すという結果である。ただし、集合知の効果が期待されている市場予測では、専門家を超えるパフォーマンスが期待されているが、今回の事例研究では、集合知が専門家(もっとも安定して成績のよい個人)を超えるかどうかについては、結果の安定性がないことが示された。

次に問題となるのは、集合知を形成する群集の多様性である。これを、外れ値の除去から見よう。

事例1では外れ値を除去することによって集合知の成績は良くなっているのに対して、事例2では外れ値を除去することによって集合知の成績は悪くなっていた。これは、集合知が個人の知が高いほど良くなる(外れ値を除去した方がよい)のに対して、多様性が高いほどよい(外れ値を除去しない方がよい)ためである。多様性が個人の知に対して十分に高い場合には、外れ値除去によって集合知のパフォーマンスはよくなるが、多様性が十分でない場合は、外れ値の除去によって集合知のパフォーマンスは低下する場合もありうる。

群集が多様であればあるほど、正解から外れた個人も存在する。著しく正解から外れた個人がいると全体のパフォーマンスを下げることになる。

この比率については個々の課題に応じて数学的に決定することが可能だろう。

社会心理が検討すべきは、C)の認知バイアスの共有性の問題である。この問題は、今回のような単純な課題では、推定する場合の個人の基準点(アンカリング)の収束として操作的に定義できる。これがどのように平準化されていくかは、過去の同調実験でも繰り返し扱われてきたテーマである。シェリフタイプの同調実験では正解のない課題を用いて基準点が収束していくことが示されてきた。事例1・2の集団決定の結果は、平均値ルールまたは中央値ルールとほぼ同じであり、今回のような価値判断を含まない課題では集団極化現象もおこらないことが示唆される。すべての集団が平均値ルールを採択すれば、集団決定は集合知に近づくはずであるが、小集団内のメンバー数が少なく(多様性が小さく)誤差のキャンセルができない分、集団決定は集合知よりも成績が劣る。価値判断を含まない課題についてはメンバー数の多い集合知の方が優秀であることが確実であるため、集団決定という手続きには意味がないことが示唆される。

今回の実験ではA)については検討していないが、集合知が有利になるアルゴリズムがわかれば、人工知能で代替可能である。集合知は集団のサイズが大きいほど、フィードバックがあるほど判断は正確になる。これは機械学習で実現可能であることを示している。知識は世界中の個人がブログやツイッターからテキスト情報として提供してくれているので、近い将来には機械的にそれらの分散された知性を集約することが可能となるだろう。

インターネットによる知識の提供と人工知能の発達により、集合知には民主主義に変わる決定規則として期待が寄せられている。しかし、よい面ばかりではない。集合知はそもそも認知主体ではないため、その結果を個人が運用することになる。しかし、集合知が予測した情報が広まると未来が変化するため、集合知は正解ではなくなる。そこで、情報機関が集合知を公開せず独占するようになれば、従来以上に大きな情報格差が生み出さ

れるだろう。

さらに、集合知は単純な軸でしか物事を判断できない「馬鹿」になる可能性がある。共有知識構造は情報を理解する枠組みであり、そこには多様な観点が存在する。多数者が共有する偏見は少数者の異なる視点を取り込めてこそ相対化しうる。しかし、認知バイアスがキャンセルされた結果の集合知を読もうとする各個人には、最大公約数としての意味しか読み取れない可能性が高い。このような情報を用いて決定が行われるとすれば、集合知は認知バイアスを薄めるというよりは、むしろ多数者の認知バイアスを増幅させる装置となる。集合知を全体の知識と見なせば、偏見を是正できない世界を生み出しかねない。

情報は伝播しやすいが、知識構造を複雑なまま伝播させることは難しい。今後の研究課題は、共有知識構造と集合知の関係を検討することである。

参考文献

- Basden, B. H., Basden, D. R., Bryner, S., & Thomas, R. L. I. (1997). A comparison of group and individual remembering: Does collaboration disrupt retrieval strategies? *Journal of Experimental Psychology: Learning, Memory, & Cognition.*, 23(5), 1176-1191.
- Brown, R. J. (1988). *Group processes Dynamics within and between groups.*: Basil Blackwell.
- Davis, J. H. (1973). Group decision and social interaction: a theory of social decision schemes. *Psychological Review*, 80, 97-125.
- Hinsz, V. B. (1990). Cognitive and consensus processes in group recognition memory performance. *Journal of Personality & Social Psychology.*, 59(4), 705-718.
- スコット・ペイジ, 2007, 「「多様な意見」はなぜ正しいのかー衆愚が集合知に変わる時」, 水谷淳訳, 日経 BP.
- ジェームズ・スロウィツキー, 2006, 「「みんなの意見」は案外正しい」, 小高尚子訳, 角川書店.
- Weldon, M. S., Blair, C., & Huebsch, D. (2000). Group remembering: Does social loafing underlie collaborative inhibition? *Journal of Experimental Psychology: Learning, Memory, & Cognition.*, 26(6), 1568-1577.
- Proceedings of Collective Intelligence in Boston, MIT 2014.

