# Surprisingly Popular アルゴリズムによる集合知生成

小野 滋 (株式会社インサイト・ファクトリー)

INSIGHT FACTORY

# 本日のテーマ

- Surprisingly Popular アルゴリズムについて紹介します
- その拡張として提案された、McCoy & Prelec (2023) の生成的可能世界モデルについて紹介します

# Surprisingly Popular アルゴリズムとは

# Surprisingly Popular アルゴリズムとは

- 二値・多値選択型の質問に対する回答(投票)を集約する方法 (Prelec, Seung, & McCoy, 2017 Nature)
  - 回答とともに、他の人の各選択肢の選択率を予測するように求める
  - **選択肢のうち、Surprisingly popular な選択肢(人々の予測よりも実際の選択率が高かった選択肢)を選択する**

- SPアルゴリズムで選択された選択肢は、現実の正解と一致する (… と主張されている)

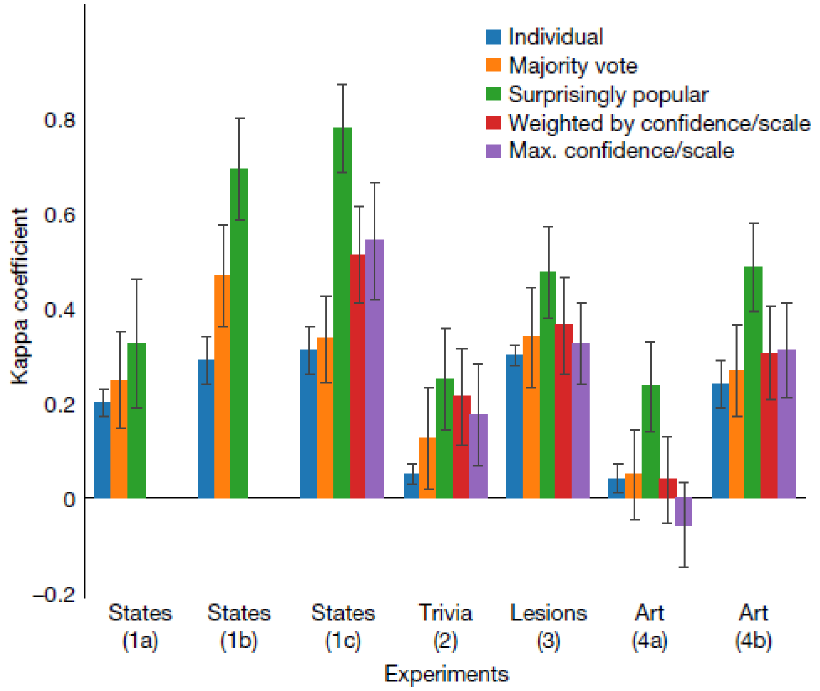



**Figure 4 | Results of aggregation algorithms on studies discussed in the text.** Study 1a, b, c: $n$ (items per study) = 50; Studies 2 and 3: $n$ = 80; Study 4a, b: $n$ = 90. Agreement with truth is measured by Cohen's kappa, with error bars showing standard errors. Kappa = $(A - B)/(1 - B)$, where $A$ is per cent correct decisions across items in a study, and $B$ is the probability of a chance correct decision, computed according to answer percentages generated by the algorithm. Confidence was not elicited in Studies 1a, b and 4a, b. However, in 4a, b we use scale values as a proxy for confidence[27], giving extreme categories (on a four-point scale) twice as much weight in scale-weighted voting, and 100% weight in maximum scale. The results for the method labelled 'Individual' are the average kappa across all individuals. SP is consistently the best performer across all studies. Results using Matthews correlation coefficient, F1 score, and per cent correct are similar (Extended Data Figs 1–3).

Prelec, Seung, & McCoy (2017)より

- 直観的説明
  - ここにコインがある。可能世界1(裏世界)でコインを投げると60%の確率で表が出る。可能世界2(表世界)では90%の確率で表が出る。つまり、表世界のほうが表が出やすい。しかし、我々は現実世界がどちらの可能世界なのかを知らない
    - 我々は、表が出る頻度は60%と90%の間だと予測できる
    - 仮に現実世界が可能世界1であるなら、コインを投げ続けると表が出る頻度は60%に収束する。このとき、裏が出る頻度は予測よりも大きい。すなわち裏がsurprisingly popularである
    - 仮に現実世界が可能世界2であるなら、コインを投げ続けると表が出る頻度は90%に収束する。このとき、表が出る頻度は予測よりも大きい。すなわち表がsurprisingly popularである
    - このように、裏世界では裏がsurprisingly popular, 表世界では表がsurprisingly popularである

  - 「PhiladelphiaはPennsylvania州の州都ですか？」という質問に対する回答の集約方法について考えよう
  - 可能世界1(実はPhiladelphiaが州都でない世界)と、可能世界2(実はPhiladelphiaが州都である世界)を想像しよう。可能世界1より可能世界2のほうがYes回答率が高いだろう
    - 仮に現実世界が可能世界1であるなら、Yes回答率は予測より低くなる。つまり回答「州都でない」がsurprisingly popularである
    - 仮に現実世界が可能世界2であるなら、Yes回答率は予測より高くなる。つまり回答「州都である」がsurprisingly popularである
    - よって、surprisingly popularな回答が正解である

- ベイジアン自白剤 (Prelec, 2004 Science)との関係
  - SPアルゴリズムは、当時"surprisingly common criterion" と呼ばれていたものと、たぶん同じ
    - 各選択肢の情報スコア $\log \frac{\bar{x}_k}{\bar{y}_k}$が、その選択肢がsurprisingly popularである程度を表現している
  - ベイジアン自白剤はゲーム理論的モデルに基づくが、SPアルゴリズムはそうでない
  - ベイジアン自白剤は個人のスコアリングに、SPアルゴリズムは回答の集約に焦点を当てている

■ BTSスコア

対象者 i のカテゴリ k に対するQ1での選択有無を $\boldsymbol{x_{ik}}$ , Q2での回答を $\boldsymbol{y_{ik}}$ とする。

下式で定義するBTSスコアの期待値は、真実回答において最大化される。
BTSスコアをインセンティブとすると、真実回答がベイジアン・ナッシュ均衡となる。

$$\bar{x}_k = \frac{1}{n}\sum_i^n x_{ik}$$

$$\log \bar{y}_k = \frac{1}{n}\sum_i^n \log y_{ik}$$

$$BTS\ Score_i = \underbrace{\sum_k^K x_{ik} \log \frac{\bar{x}_k}{\bar{y}_k}}_{\text{情報スコア}} + \alpha \underbrace{\sum_k^K \bar{x}_k \log \frac{y_{ik}}{\bar{x}_k}}_{\text{予測スコア}} \qquad (0 < \alpha < 1)$$

# SPアルゴリズムの正当化

- Prelec, Seung, & McCoy (2017) のSupplementary Informationで示されている
    - 正直言って、ほぼ理解できない

- にもかかわらず、以下では無謀にも紹介を試みます
    - Supplementary Information の1.1-1.2
    - ふたつの定理が証明されている
    - 二値設問についてのSPアルゴリズムに対応するのは、定理2と思われる

- おわび
    - 適宜、記号を付け加えています。原文ではもっと簡略化されています
    - すみません

## モデル

- $m$個の可能世界を考える。ここでは$m = 2$
- 現実世界を確率変数$A \in \{a_1, \dots, a_m\}$ で表す
    - 事前分布$\Pr(A = a_i)$は対象者の共有知識と整合していると考える
    - $Pr(A = a_i) > 0$とする
    - 正解を$a_{i^*}$とする
- 対象者$r$は私秘的シグナル $S^r \in \{s_1, .., s_n\}$を持つ
    - 対象者のシグナルは、同一の確率分布$\Pr(S^r = s_k | A = a_i)$に独立に従う
    - 対象者間の知識の差異はすべてシグナルで表現されていると仮定する
    - $Pr(S^r = s_k) > 0$とする
- すべての対象者は、同時分布 $Pr(S^r = s_k, A = a_i)$を知っていると仮定する
- 対象者は以下の2つの信念を持つ
    - 正解についての事後分布 $Pr(A = a_i | S^r = s_k)$
        - 同時分布 $Pr(S^r = s_k, A = a_i)$ から得られる
    - 他の対象者$q$が受け取っているシグナル$S^q$についての信念

$$Pr\left(S^q = s_j \middle| S^r = s_k\right) = \sum_i Pr\left(S^q = s_j \middle| A = a_i\right) P\,r(A = a_i | S^r = s_k) \Pr(A = a_i)$$

**定理1.**

現実のシグナルの確率 $\Pr(S^r = s_k|a_{i^*})$ についての知識と、シグナルによって含意される事後確率$\Pr(A = a_i|S^r = s_k)$についての知識のみから、正解を演繹できるアルゴリズムは存在しない。

証明: [略]

**モデル (つづき)**

- 対象者$r$は、投票$V^r \in \{v_1, \dots, v_m\}$を行う
  - $v_1, \dots, v_m$は$a_i, \dots, a_m$に対応している
- シグナル$s_k$を持っている人がどの投票をするかを返す関数を$V(s_k)$とし、以下のように仮定する。ただし$c_2 = 1 - c_1$とする。

$$V(s_k) = argmax_i \frac{\Pr(A = a_i|S = s_k)}{c_i}$$

- 対象者は、他者$q$の投票について条件付き確率を求めることができる

$$\Pr(V^q = v_i|S^r = s_k) = \sum_{j:V(s_j)=v_i} \Pr(S^q = s_j|S^r = s_k)$$

**定理2.**

すべての人が正解に投票するわけではないならば、正解に対する投票の平均的推定値は過小推定である。

証明: まず、正解に対する現実の投票は、正解に対する反事実的投票を上回ることを示す。

$$\frac{\Pr(V^q = v_{i^*}|A = a_{i^*})}{\Pr(V^q = v_{i^*}|A = a_k)} = \frac{\Pr(A = a_{i^*}\,|V^q = v_{i^*})\Pr(A = a_k)}{\Pr(A = a_k\,|V^q = v_{i^*})\Pr(A = a_{i^*})} = \frac{\Pr(A = a_{i^*}\,|V^q = v_{i^*})}{1 - \Pr(A = a_{i^*}|V^q = v_{i^*})}\frac{1 - \Pr(A = a_{i^*})}{\Pr(A = a_{i^*})}$$

[小野注: $\frac{A}{(1-A)}\frac{(1-B)}{B} > 1 \Leftrightarrow A > B$ だから] 上の式が1を上回る必要十分条件は

$$\begin{aligned} P(A = a_{i^*}|V^r = v_{i^*}) &> \Pr(A = a_{i^*}) \\ &= \Pr(A = a_{i^*}|V^r = v_{i^*})\Pr(V^r = v_{i^*}) + \Pr(A = a_{i^*}|V^r = v_k)\Pr(V^r = v_k) \end{aligned}$$

投票関数の定義より$\Pr(A = a_{i^*}|V^r = v_k) < c_{i^*} < \Pr(A = a_{i^*}|V^r = v_{i^*})$だから、この不等式は常に成り立つ。よって、 $\Pr(V^q = v_{i^*}|A = a_{i^*}) > \Pr(V^q = v_{i^*}|A = a_k)$ である。

いっぽう

$$\Pr\left(V^q = v_{i^*}\middle|S^r = s_j\right) = \Pr(V^q = v_{i^*}|A = a_i^*)\Pr\left(A = a_{i^*}\middle|S^r = s_j\right) + \Pr(V^q = v_{i^*}|A = a_k)\Pr\left(A = a_k\middle|S^r = s_j\right)$$

$\Pr(V^q = v_{i^*}|A = a_{i^*}) > \Pr(V^q = v_{i^*}|A = a_k)$より

$$\Pr\left(V^q = v_{i^*}\middle|S^r = s_j\right) \le \Pr\left(V^q = v_{i^*}\middle|A = a_{j^*}\right)$$

等号成立は$\Pr\left(A = a_{i^*}\middle|S^r = s_j\right) = 1$のときである。

すべてのシグナルについて弱い不等性が成り立ち、そのうちいくつかでは強い不等性が成り立つ。よって、平均的な推定値では強い不等性が成り立つ。(証明終)

# 感想

- 「人々の回答をうまく集約し、もっとも真実に近いと思われる選択肢を選びたい」という場面は少なくない
  - 例) マーケティング・リサーチ

- SPアルゴリズムの魅力
  - 正解の情報にも、多数派であることや合意にも依存しない
  - 実装がきわめてシンプル

- SPアルゴリズムに対する疑問
  - モデルの想定は現実場面に即しているか？
  - モデルは実際の回答者の行動と整合しているか？
  - もっとたくさんの実証研究がほしい

- SPアルゴリズムの限界
  - 正解と思しき選択肢をひとつ選ぶだけ。事後確率は推定されない

# SPアルゴリズムの後続研究

Wikipedia
The Free Encyclopedia

Search Wikipedia | Search

Create account Log in

# Surprisingly popular

2 languages

Article Talk — Read Edit View history Tools

Contents hide



From Wikipedia, the free encyclopedia

The **surprisingly popular** answer is a wisdom of the crowd technique that taps into the expert minority opinion within a crowd.[1] For a given question, a group is asked both "What do you think the right answer is?" and "What do you think the popular answer will be?" The answer that maximizes the average difference between the "right" answer and the "popular" answer is the "surprisingly popular" answer.[2] The term "surprisingly popular" was coined in a 2017 paper published in *Nature* entitled "A solution to the single-question crowd wisdom problem", which outlined the technique.[2][3]

## Example [ edit ]

Suppose the question to be determined is: Is Philadelphia the capital of Pennsylvania? The two questions asked of the group, and the numbers of responses, are:

Is Philadelphia the capital of Pennsylvania?

- Yes: 65%
- No: 35%

What do you think most people will respond to that question?

- Yes: 75%
- No: 25%

The difference between the answers to the *right* question and the *popular* question:

- Yes: 65% − 75% = −10%
- No: 35% − 25% = 10%

Thus, the *No* answer is surprisingly popular (10% > −10%). Because of the relatively high margin of 10%, there can be high confidence that the correct answer is *No*. (The capital is indeed not Philadelphia, but Harrisburg.)

An illustrative breakdown of this follows. There are four groups of people.

- A – "Philadelphia is the capital, and others will agree." (This group answers yes/yes.)
- B – "Philadelphia is the capital, but most others won't know that". (This group answers yes/no.)
- C – "Philadelphia is not the capital, and others will agree." (This group answers no/no.)

もはやWikipediaに項目ができている…

＜Prelecらの研究＞

- Cvitanic, Prelec, Radas & Sikic (2020 Theory of Probability and its Application)
    - 理論研究。おそらくSPアルゴリズムというよりベイジアン自白剤(真実申告メカニズム)の研究
- Radas & Prelec (2019 PLOS ONE)
    - 他者予測質問を用いた対象者の絞り込み手法を提案
- Radas & Prelec (2021 PLOS ONE)
    - SPアルゴリズムを利用したコンジョイント分析を提案
- Prelec & McCoy (2022 psyarxiv)
    - SPアルゴリズムの拡張について証明を与えている
- **McCoy & Prelec (2023 Management Science)**
    - SPアルゴリズムの拡張である生成的可能世界モデルを提案

＜SPアルゴリズムの実証研究＞

- Lee, Danileiko, & Vi (2018 Judgment & Decision Making)
  - スポーツ予測に適用。他の手法より優れていた
- Hosseini, Mandal, Shah, & Shi (2001 IJCAI)
  - 選択肢の真の順位の推定に適用。古典的手法より優れていた
- Rutchick, Ross, Calvillo, & Mesik (2020 Cognitive Research)
  - クラウドソーシングにおける予測課題に適用。成功

＜SPアルゴリズムと類似した提案＞

- Kong & Schoenebeck (2018 ACM Conf. Economic & Computation)
    - 集合知抽出の手法を提案
- Palley & Sholl (2019 Management Science)
    - 意見集約の手法 → McCoy & Prelec(2023)で言及
- Martinie, Wilkening, Howe (2020 PLOS ONE)
    - 他者予測を使った集約手法
- Wilkening, Martinie, & Howe (2022 Management Science)
    - SPアルゴリズムと類似しているが、確信度を用いる模様 → McCoy & Prelec(2023)で言及
- Pally & Satopaa (2023 Management Science)
    - 意見集約の手法。→ McCoy & Prelec(2023)で言及

＜機械学習・最適化＞

- Cui, et al. (2019 IEEE Congress on Evolutionary Computation)
  - SPアルゴリズムを粒子群最適化アルゴリズムの改善に利用
- Luo & Liu (2022 Machine Learning)
  - SPアルゴリズムを機械学習におけるアンサンブル手法の改善に利用
- Wu, et al. (2023 Swarm & Evolutionary Computation)
  - SPアルゴリズムを粒子群最適化アルゴリズムの適合度評価のメトリックとして適用
- Li, et al. (2023 IEEE Transactions on Cybernetics)
  - SPアルゴリズムをスケジューリング・アルゴリズムに適用

# 生成的可能世界モデル (McCoy & Prelec, 2023)

# 論文の概要

次の論文について紹介します

- McCoy, J., & Prelec, D. (2023) A Bayesian Hierarchical Model of Crowd Wisdom Based on Predicting Opinions of Others. *Management Science*, Article in Advance.
    - 2023年10月公開
    - 二値質問への回答を、ベイジアン階層モデルによって集約する方法を提案
    - SPアルゴリズムの拡張と主張されている

John McCoyさん
(U. Penn)

Drazen Prelecさん
(MIT)

**目次**

# 1. イントロダクション

- 集合知を抽出する方法
  - 多くの領域で開発と応用がなされている
    - 予測市場、売上予測のパリミュチュエル的メカニズム、需要予測の算術的平均…
  - 課題: いかにして個人の判断を集約(aggregate)するか、いかにして個人の専門性を識別するか

**1.1 標準的な集合知手法とその弱点**

- 民主的な平均
  - 回答の平均や最頻値を選ぶ
  - → 情報や洞察の個人差を無視している；情報の共有を考慮していない
- 個人ごとに異なる重みづけをする
  - 集団の正確さへの貢献で重みづける → 多くの設問への回答が必要
  - 決定の類似性で重みづける
  - ピアからの評価で重みづける
  - …

## 1.2 他者についての予測と＜意外に人気がある答え＞

- Prelec, Seung, & McCoy (2017 Nature)
  - 問題の質問に対して集団の他のメンバーのどう回答するかについての、個々の対象者の予測を利用する
    - なぜ他者予測が役に立つのか？
      - 直観的説明: 合理的な回答者が、自分の答えは少数派だろうと予測しながらある判断を行った場合、そのことは、彼の知識が広く共有されていないと彼が信じていることを示している
  - 複数の可能世界について推論するベイジアン・エージェントのモデルを提案
    - 集団における回答の分布を正しく解釈するには「もし世界がこれこれなら分布はこれこれになるはずだ」ということを知っている必要がある
      - いま70%の人がYesと回答したとしよう。そのことだけから、良い答えはYesだと結論することはできない。もし「正しい答えがYesだったら95%の人々がYesと答えるはずだ」と知っているなら、Yes率が70%しかないので、正解は”No”だと考えることになるだろう
    - 真の答えと回答分布との関係についての情報を、対象者の他者予測から得る試み
  - ＜意外に人気がある答え＞アルゴリズム surprisingly popular algorithm (SPアルゴリズム) を提案
    - 二値質問の場合、予測された投票数よりもより多くの投票数を得た回答を採用する
    - 多値質問へと一般化可能
    - 重みなし平均や重み付き平均よりも優れている

- Palley & Soll (2019 Mgmt.Sci.)
    - 私秘的情報と共有情報を分離する方法を提案
    - 多くの情報構造の下で判断の誤りを最小化できる
- Palley & Satopaa(2023 Mgmt.Sci)
    - 個々の対象者の点推定を重みづけて結合する。重みは、すべての対象者の推定の平均についてのその対象者の予測によって決める
    - 共有情報を過剰に重みづけるのを避けることができる
- Wilkening , Martinie, & Howe (2002 Mgmt.Sci.)
    - SPアルゴリズムの確信度ベース版。豊かな情報を持つ対象者に重みをつけることができる
    - 確率的な回答と、その他者を通じた平均の予測を求める

## 1.3 生成的可能世界モデル (GPWM)

- 本論文の提案モデル
- 他者予測とベイジアン階層モデルを結合
- 利点
  - 単一の答えを選ぶだけでなく、事後確率分布を提供
  - 回答・他者予測におけるノイズを明示的にモデル化
  - 対象者の専門性の個人差を識別
  - 回答の事前分布を明示的にモデル化。他の種類の情報を容易に統合できる (例, 経験的なベースレート頻度)
  - 多値質問の場合、SPアルゴリズムよりも弱い想定で成立する
- Prelec, et al. (2017)との関係
  - Prelec, et al. (2017)で定式化した可能世界モデルの生成バージョン。潜在変数の同時分布とそこからのデータの生成をモデル化
- GBWMについての直観的説明
  - 対象者の回答と他者予測を回答の事前確率に依存した形でモデル化。個々の対象者が持っている情報と、それらの情報と正解との関係について推論し、集合知を抽出できる
- 複数質問でも使える。その場合、対象者を専門性(expertise)で順序付けることができる
- 質問への正解の情報は不要

## 1.4 GPWM vs. 既存のベイジアン階層モデルによる集約

- ベイジアン階層モデルによる集約

  A. 項目反応理論と関連した諸モデル。個々の対象者の能力・バイアスと、質問の困難度を推論

    - Raykar et al.(2010 J.MachineLearn.Res.), Merkle et al.(2016 Decision), Karabastos & Batchelder(2003 Psychometrika)

  B.「シグナル＋ノイズ＆バイアス」モデル。対象者は正解の周りに分布するシグナルを受け取り、それに基づいて回答する

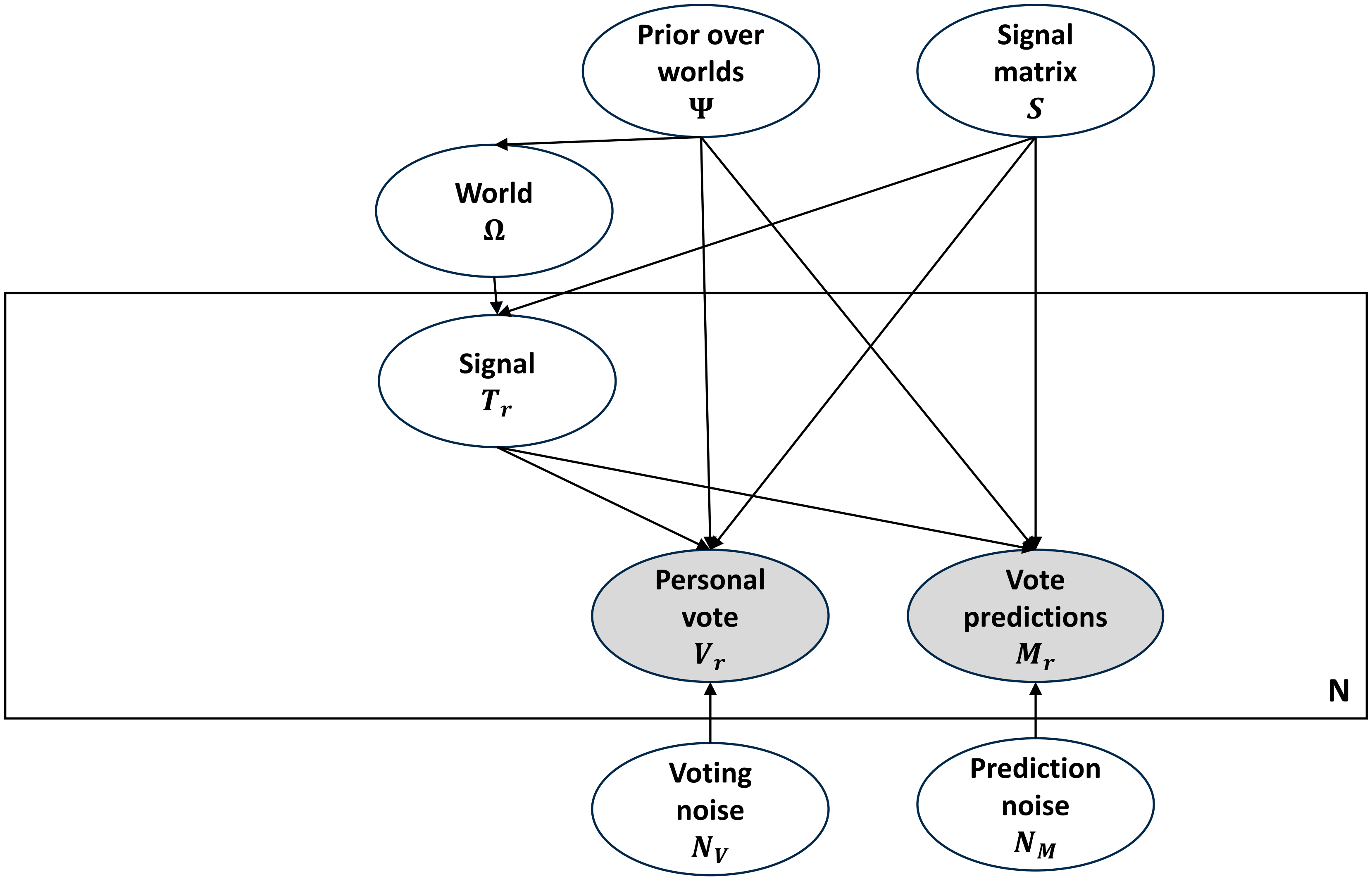
    - Budescu & Johnson(2011 JudgmentDecisionMaking), Turner et al.(2014 MachineLearn.)

- 本論文では以下の2つのモデルとの比較を行う
  - **Bayesian cultural consensus (BCC)モデル**
    - 上記Aに相当。Oravexz et al.(2013 Psychometrika), Oravexz et al.(2014 Field Methods)
  - **Bayesian cognitive hierarchy (CH) モデル**
    - 上記Bに相当。Lee & Danileiko (2014 JudgmentDecisionMaking)
- GPWMとベイジアン階層モデルとのちがい
  - ベイジアン階層モデルは、正解と合意のあいだになんらかの関係があると考えている。GPWMはそうでない
  - ベイジアン階層モデルは個人レベルパラメータの学習に依存しており、複数質問を想定している。GPWMはそうでない

## 1.5 本論文のパースペクティブ [略]

# 2. 生成可能世界モデル

- 以下の順に説明する
    - 単一設問GPWM
    - 複数設問GPWM
- 二値質問のGPWMについて説明する
    - 多値設問への拡張について後述する
- 実データに適用できるStanコードをご用意している




[単一質問GPWMのグラフィカル・モデル。Figure 1 より抜粋。対象者数をNとする。観察されるノードをアミカケ]

## 2.1 世界状態と世界事前分布

- 世界状態を $\Omega \in \{a_0, a_1\}$とする
  - 個々の世界状態は、質問へのありうる回答と対応している
  - 質問への正解 (現実の世界状態と対応している回答) を $a_{i^*}$とする
- Ωは事前分布Ψのもとでベルヌーイ分布に従う
  - $\Psi_0 = p(\Omega = 0), \Psi_1 = p(\Omega = 1)$として、$\Omega | \Psi \sim Bernoulli(\Psi_1)$
  - 事前分布$\Psi = (\Psi_0, \Psi_1)$の超事前分布は一様分布。$\Psi_0 \sim Uniform([0,1])$

## 2.2 シグナルとシグナル行列

- 対象者 $r$ は私秘的シグナル $T_r \in \{s_0, s_1\}$ を受け取る
  - 個々の世界状態は、質問へのありうる回答と対応している
  - 正しいシグナルを$s_{i^*}$とする
  - シグナルの分布は世界状態 Ω とシグナル行列 $S$ (後述)によって決まる
- 対象者はシグナルを除き交換可能であり、独立に同一の確率分布に従ってシグナルを受け取ると仮定する
  - 以下、$p(\cdot | T_r = s_i)$を$p(\cdot | s_i)$ と略記する

- シグナル行列 $S$
  - 個々の世界状態のもとでのシグナルの条件付き確率を表す

$$S = \begin{bmatrix} p(T_r = s_0|\Omega = a_0) & p(T_r = s_0|\Omega = a_1) \\ p(T_r = s_1|\Omega = a_0) & p(T_r = s_1|\Omega = a_1) \end{bmatrix}$$

  - $a_{i^*}$の下での条件付き分布を「実際のシグナル分布」と呼ぶ
    - $T_r|S, \Omega \sim Bernoulli(S_{1,a_i^*})$ となる
  - シグナル行列は、ある集合から一様にドローされる
    - その集合とは、列和が1であり、かつ$S_{ii} > S_{ij}$である行列の集合
    - こう書くこともできる: $S' \sim Uniform([0,1]), S'' \sim Uniform([0,1]), S_{0,0} = \max(\{S', S''\}), S_{0,1} = \min(\{S', S''\})$
  - つまり、$p(s_i|a_i) > p(s_i|a_j)$ という制約をかけている
    - この点が、ほかの多くのモデルと異なる
    - 単に、世界状態が$a_j$である場合よりも$a_i$である場合のほうがシグナル$s_i$を受け取りやすい、と述べているだけ
    - モデルの識別のための制約。情報の分布についての実質的な仮定ではない
    - 正しいシグナルを受け取る人が多数派だ ( $p(s_i|a_i) > p(s_j|a_i)$ ) という仮定ではないことに注意

## 2.3 世界事前分布とシグナル行列の知識の下でベイジアン対象者が計算できる量

- すべての対象者は、世界事前分布$\Psi$とシグナル行列$S$についての知識を持っていると仮定する
- 対象者はシグナルなしで、以下を決定論的に算出できる：
    - シグナルの周辺分布 $F = S\Psi$
        - $F_0 = p(s_0|\Psi, S) = p(s_0|a_0, S)p(a_0|\Psi) + p(s_0|a_1, S)p(a_1|\Psi)$
    - 受け取ったシグナルの下での、世界状態の事後分布を表す行列$W$
        - $W_{i,j} = p(a_i|s_j) = \frac{S_{j,i}\Psi_i}{p(s_j|\Psi, S)}$
    - 対象者$r$が受け取ったシグナルの下での、対象者$s$が受け取るシグナルの事後分布を表す行列$O = SW$
        - $O_{i,j} = p(T_s = s_i|T_r = s_j) = \sum_{a_i} p(s_i|a_i, S)p(a_i|s_j, S, \Psi)$
    - 受け取ったシグナルの下での、他者の投票の予測の行列$P$
        - $P_{ij}$は、シグナル$s_j$を受け取った時の、「世界$a_i$の事後確率のほうが大きい人」の割合の予測
        - シグナル$s_0$が世界$a_0$を支持し、シグナル$s_1$が世界$a_1$を支持している場合 ($W_{0,0} > 0.5, W_{1,1} > 0.5$ ) → $P = O$
        - どちらのシグナルも世界$a_0$を支持している場合($W_{0,0} > 0.5, W_{0,1} > 0.5$) → $P_{00} = 1, P_{01} = 1$
        - どちらのシグナルも世界$a_1$を支持している場合($W_{1,0} > 0.5, W_{1,1} > 0.5$) → $P_{10} = 1, P_{11} = 1$

- 数値例
  - 世界事前分布が$\Psi = (0.45, 0.55)$, シグナル行列が$S = \begin{bmatrix} 0.4 & 0.1 \\ 0.6 & 0.9 \end{bmatrix}$ だとする
  - シグナルの周辺分布 $F = S\Psi$

$$F_0 = p(s_0|\Psi, S) = p(s_0|a_0, S)p(a_0|\Psi) + p(s_0|a_1, S)p(a_1|\Psi) = 0.4 \times 0.45 + 0.1 \times 0.55 = 0.24$$
$$F_1 = p(s_1|\Psi, S) = p(s_1|a_0, S)p(a_0|\Psi) + p(s_1|a_1, S)p(a_1|\Psi) = 0.6 \times 0.45 + 0.9 \times 0.55 = 0.76$$

  - 受け取ったシグナルの下での、世界状態の事後分布を表す行列$W$

$$W = \begin{bmatrix} \frac{S_{0,0}\Psi_0}{p(s_0|\Psi, S)} & \frac{S_{1,0}\Psi_0}{p(s_1|\Psi, S)} \\ \frac{S_{0,1}\Psi_1}{p(s_0|\Psi, S)} & \frac{S_{1,1}\Psi_1}{p(s_1|\Psi, S)} \end{bmatrix} = \begin{bmatrix} \frac{0.4 \times 0.45}{0.24} & \frac{0.6 \times 0.45}{0.76} \\ \frac{0.1 \times 0.55}{0.24} & \frac{0.9 \times 0.55}{0.76} \end{bmatrix} = \begin{bmatrix} 0.76 & 0.35 \\ 0.23 & 0.65 \end{bmatrix}$$

  - 対象者$r$が受け取ったシグナルの下での、対象者$s$が受け取るシグナルの事後分布を表す行列$O = SW$

$$O = \begin{bmatrix} \sum_{a_i} p(s_0|a_i, S)p(a_i|s_0, S, \Psi) & \sum_{a_i} p(s_0|a_i, S)p(a_i|s_1, S, \Psi) \\ \sum_{a_i} p(s_1|a_i, S)p(a_i|s_0, S, \Psi) & \sum_{a_i} p(s_1|a_i, S)p(a_i|s_1, S, \Psi) \end{bmatrix} = \begin{bmatrix} 0.4 \times 0.76 + 0.1 \times 0.23 & 0.4 \times 0.35 + 0.1 \times 0.65 \\ 0.6 \times 0.76 + 0.9 \times 0.23 & 0.6 \times 0.35 + 0.9 \times 0.65 \end{bmatrix} = \begin{bmatrix} 0.33 & 0.21 \\ 0.67 & 0.79 \end{bmatrix}$$

  - 受け取ったシグナルの下での、行列$P$ (他者の投票を予測する行列)
    - $W_{0,0} > 0.5, W_{1,1} > 0.5$ なので、$P = O$

## 2.4 対象者の投票と投票予測

- 対象者 $r$ の投票 $V_r \in \{v_0, v_1\}$
  - 世界状態の事後分布の対数のsoftmax関数としてモデル化する
  - ノイズパラメータ$N_V$を組み込む。全対象者で共通と仮定する
  - 対象者$r$がシグナル$s_k$を受け取ったとき

$$p\left(V_r = v_j\right) = \frac{\exp(N_V \log p(a_j|s_k))}{\sum_i \exp(N_V \log p(a_i|s_k))}$$

  - すなわち

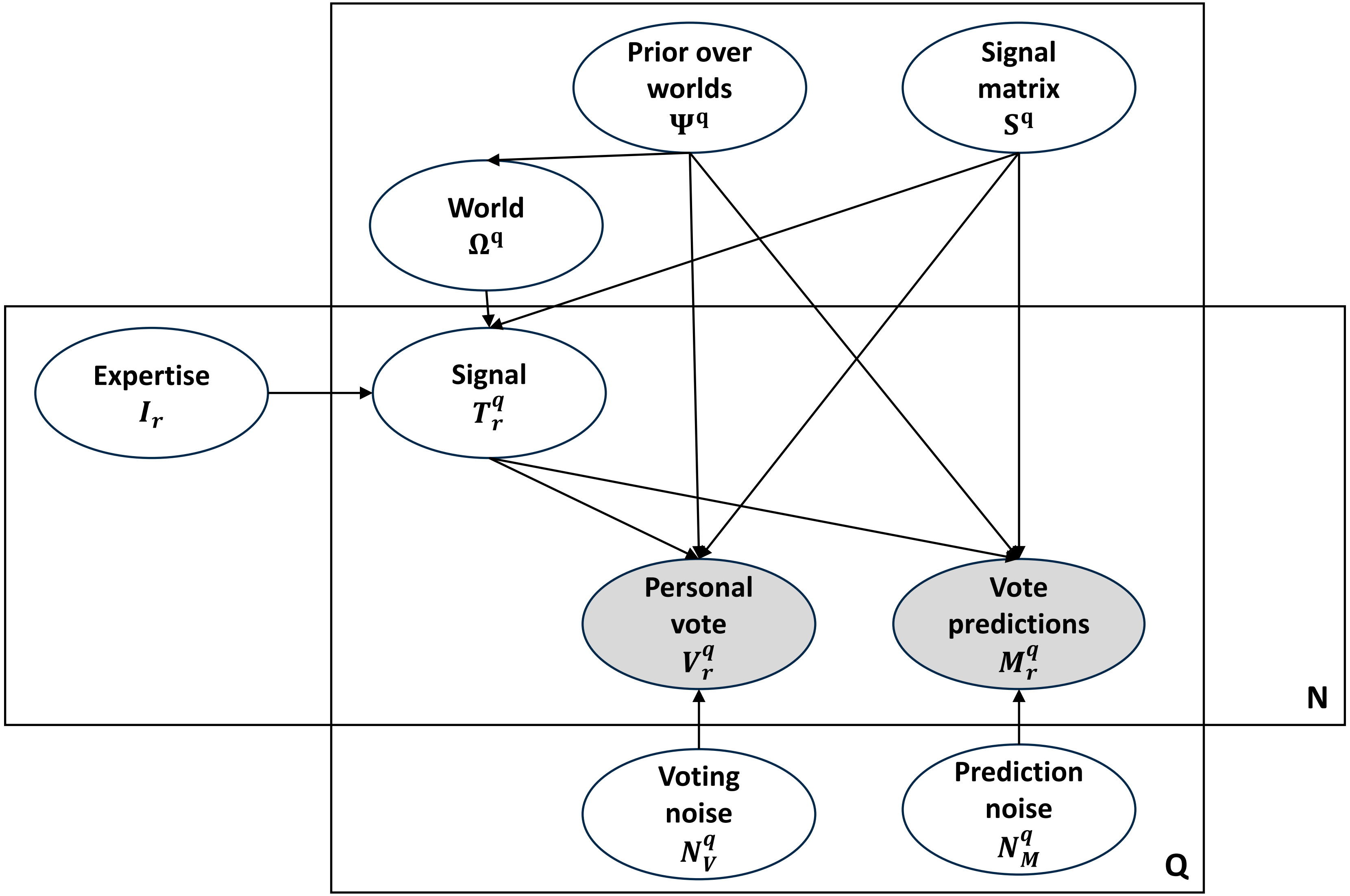
$$V_r|N_V, W, T_r \sim Bernoulli\left(\frac{\exp(N_V \log W_{1,T_r})}{\exp\left(N_V \log W_{0,T_r}\right) + \exp\left(N_V \log W_{1,T_r}\right)}\right)$$

- 対象者 $r$ による、他者の投票の予測 (投票が$v_0$の人の割合の予測) $M_r$
  - ノイズパラメータ$N_M$を組み込む。全対象者で共通と仮定する
  - 「世界$a_i$の事後確率のほうが大きい人」の割合の予測$P_{0,T_r}$を平均とする正規分布を[0,1]で打ち切った分布に従うと仮定する

$$M_r|N_M, P, T_r \sim Normal_{[0,1]}(P_{0,T_r}, N_M)$$

## 2.5 ノイズの事前分布

- $N_V \sim Gamma(3,3), N_M \sim Uniform([0,0.5])$ とする




複数質問GPWMのグラフィカル・モデル。Figure 1より(一部改変)。対象者数をN, 質問数をQとする。観察されるノードをアミカケ

## 2.6 専門性と複数質問GPWM

- $Q$個の質問がある場合について考える
  - 質問$q$についての変数を$\Omega^q, \Psi^q, S^q, N_V^q, N_M^q, T_r^q, V_r^q, M_r^q$とする。対象者が決定論的に計算できる量を$F^q, W^q, O^q, P^q$とする
- 対象者$r$ は専門性パラメータ$I_r \in [0,1]$を持つと仮定する
  - 正しいシグナルを受け取る確率が、$I_r$の実現値$i_r$ とともに線形に増大すると仮定する
  - $a_{i^*}^q = a_j^q$のとき

$$p\left(T_r^q = s_j \middle| a_{i^*}^q, i_r, S^q\right) = S_{jj}^q + i_r\left(1 - S_{jj}^q\right)$$
$$p\left(T_r^q = s_i \middle| a_{i^*}^q, i_r, S^q\right) = S_{ij}^q - i_r(1 - S_{ij}^q)$$

- 対象者は専門性を考慮せず、すべての対象者において$i_r = 0$であるとみなすと仮定する
  - 2.7.2 で拡張する
- $I_r \sim Uniform([0,1])$ と仮定する

## 2.7 GPWMのふたつの拡張: 確信度と専門性の自己知識

### 2.7.1 対象者の確信度

- 単一質問GPWMの場合について説明する
- 投票$V_r$とともに、確信度$C_r$を回答してもらう
- シグナルが$s_k$, 投票が$v_j$であるとして、$C_r$は$W_{jk} = p(a_j|s_k)$のノイズ付き報告であると仮定する

$$C\_r|N_C, T_r, V_r, W \sim Normal_{[0,1]}(W_{V_r,T_r}, N_C)$$
$$N_C \sim Uniform([0,0.5])$$

### 2.7.2 専門性の自己知識

- 複数質問GPWMで、対象者$r$は自分の専門性$i_r$を知っており、かつほかの対象者の専門性を0とみなすと仮定する
- 対象者$r$の世界状態の事後分布は以下となる

$$W_{jk}^r = p\left(a_j \middle| s_k, i_r, S, \Psi\right) = p\left(s_k \middle| a_j, i_r, S\right) p\left(a_j \middle| \Psi\right) / p(s_k|i_r, S)$$

- 計算が大変になる

**参考: GPWMモデルに基づく推論について** (Online Appendix 2.1)

- 単一質問GPWMについて説明する。$V_r, M_r$のベクトルを$V, M$とする
- 世界状態$a_i$の事後確率は

$$p(a_i|V, M) = \prod_r p(a_i|V_r, M_r)$$
$$\propto \int_\Psi p(\Psi)p(a_i|\Psi) \int_S p(S) \int_{N_V} p(N_V) \int_{N_M} p(N_M) \prod_r \sum_{s_k} p(s_k|a_i, S)p(V_r|\Psi, S, s_k)p(M_r|\Psi, S, s_k)$$

  - $\Psi, S, N_V, N_M$を固定し$a_i$を$a_{i^*}$に固定すると、混合ウェイト$p(s_k|a_{i^*}, S)$の混合モデルになっている。$V_r$も$M_r$も同じ混合要素で決まっている

- GPWMに基づきもっとももありそうな答えを選ぶ際には、単に $argmax_i\ p(a_i|V, M)$ を選ぶ
  - SPアルゴリズムとは異なる。SPアルゴリズムでは$argmax_j\ p(a_{i^*}|s_j)$

# 3. 比較モデル

[略。主な内容は1.4で既述]

# 4. データ

- Prelec et al.(2017) のデータを再分析。すべての質問で正解が既知
- states
  - US 50州それぞれについて、 “Is city X the capital of state Y?” と質問。Xはその州でもっとも人口の多い都市。(1) 真偽判断 (2)真と判断する人の割合を予測
  - **MIT class states:** MITのMBA学生, n=51。
  - **Princeton states:** プリンストン大の実験室, n=32
  - **MIT laboratory states**: MITの実験室, n=22。あわせて、(3)真偽判断の確信度評定、(4)他の人の確信度の平均を予測
- art
  - 20世紀アート90点の複製をみせる。それぞれについて(1)市場価格を4段階で判断。(2) 他者の回答を予測。
  - 2段階(High,Low)に併合して分析する
  - **Art professionals**: ギャラリーのオーナーなど, n=20
  - **Art laypeople**: 素人, n=20
- **Lesion**
  - 皮膚科医, n=25。皮膚の病変の写真をみせる。(1)両性か悪性かを判断。(2)他者の回答を予測。(3)自分の判断の確信度を評定。
- **Trivia**
  - Amazon MT, n=39。歴史・科学・地理・言語の80個のトリビアを提示し、(1)真偽判断。(2)他者の回答を予測。(3)自分の判断の確信度を評定。

# 5. 結果

### 5.1 データセットの比較

[データセットの記述統計の紹介。略]

### 5.2 方法

- 回答を以下の手法で集約する
    - 個々の質問について、{単一質問GPWM、多数決、SP回答(実際の投票頻度が予測された投票頻度を超えている回答)}
        - 確信度を聴取しているデータについては、投票と確信度から個人レベル確率を求め、{平均(線形意見プール), 基準化した幾何平均(対数意見プール), extremized mean, logit aggregator}
    - 各データセットの全質問を用いて、{複数質問GPWM, BCCモデル}
        - 確信度を聴取しているデータについては、{CHモデル}

- [「回答をGPWMで集約する」とは、回答データにGPWMモデルをあてはめ、モデルから世界状態の事後確率を推論すること、ないしもっとも事後確率が高い世界状態を選択することを指しているのだと思います]

## 5.3 回答の正確さ

- カテゴリカルな正確性 (各手法で支持された回答と正解が一致するか)
  - 各手法について、得られた答えと正解との一致係数 (Cohenの $\kappa$) を算出
  - 単一質問GPWMは、すべてのデータセットで多数決、線形意見プール、対数意見プール、logit aggregatorを上回った。またSP回答と類似した結果となった
  - 複数質問GPWMは、BCCモデル、CHモデルと同程度。データセットにより優劣が異なった
- 確率的な正確性 (各手法で正解に与えられた確率) [略]

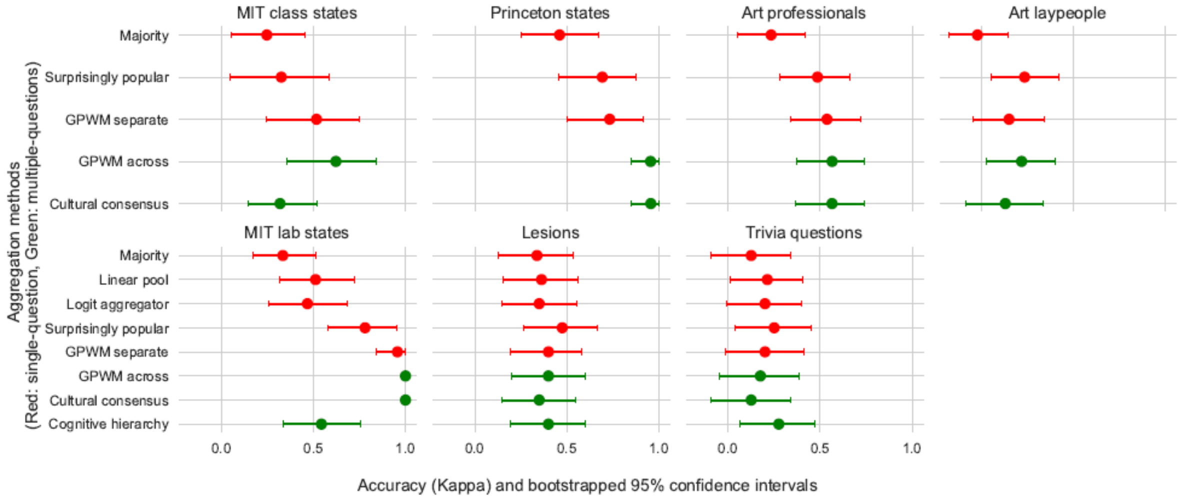



Figure A.5: Cohen's kappa coefficients for each aggregation method, including 95% confidence intervals. Methods with higher Cohen's kappa are more accurate. Methods shown in red operate on single questions independently, and methods shown in green require multiple questions. Confidences were only elicited in the studies shown in the bottom row, and so some methods are only applied to these studies.

### 5.4 GPWMを損傷させたときの回答の正確性: 他者についての予測

- GPWMモデルを損傷させ、他者予測を無視するようにした
- 得られた回答は多数決とほぼ等しくなった

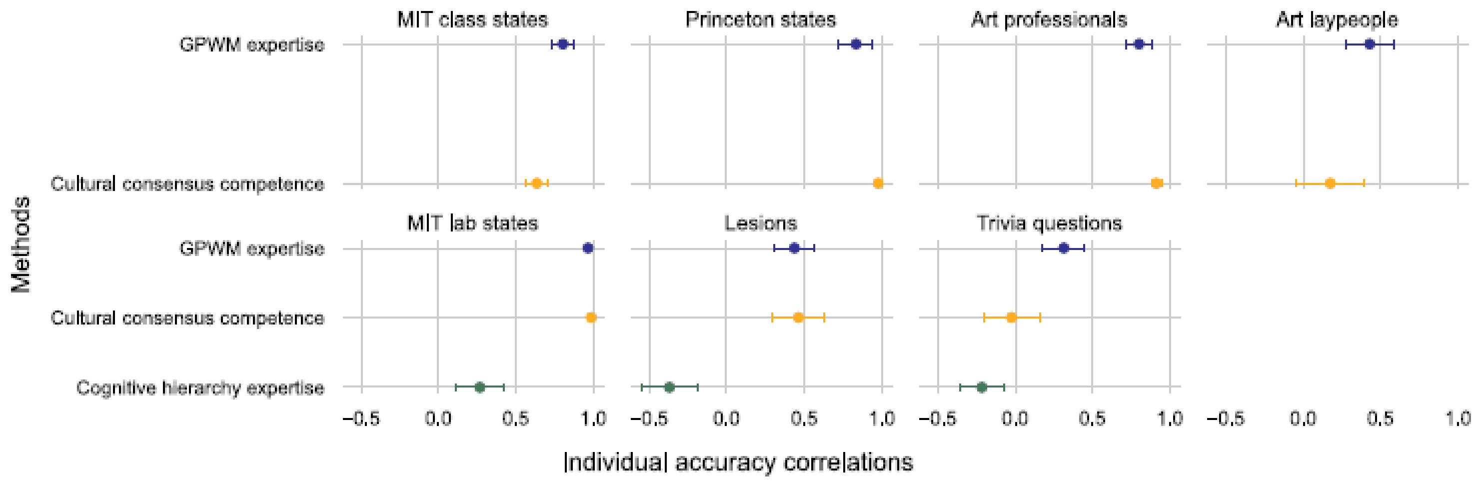
### 5.5 GPWMの拡張における回答の正確性

- 対象者の確信度を組み込んでも、結果はあまり変わらなかった
- 専門性の自己知識を仮定しても、結果はあまり変わらなかった

## 5.6 対象者レベル・パラメータ

- 複数質問GPWM, BCCモデル, CHモデルでは、個人レベルの専門性パラメータを推定できる
- 個人ごとの成績 (正解と回答との一致係数)と、専門性パラメータ推定値との相関は、複数質問GPWM・BCCモデルで高かった

Figure 2. (Color online) Pearson Correlations, Including Bootstrapped 95% Confidence Intervals, Between Inferred Respondent-Level Parameters and the Accuracy of Each Respondent




*Notes.* Respondent accuracy is evaluated with Cohen's κ. Models with a high correlation for a data set successfully identify actual respondent expertise.

## 5.7 他の潜在パラメータについての推論: 世界事前分布

- 世界事前分布 Ψ の推定値について、statesデータで評価
- Bingで”都市名, 州名” (例, “Birmingham, Alabama”)を検索したときの検索数(対数)と、ある程度の相関があった
- 都市名と州の関係の顕著性についての共有知識を反映しているものと思われる

## 5.8 回答コーディングの質問間での一致性

- 本研究のデータセットでは、データセット内で回答のコーディングが一貫しているが(回答カテゴリの一方は「真」で他方は「偽」)、現実の問題はそうでないことが多い
- 各データセットにおいて、質問の半分について回答カテゴリを入れ替えてみた(正解、回答、他者予測のすべてを逆にした)
- GPWM, CHモデルは影響を受けないが、BCCモデルの性能は下がった。「真」へのバイアスを推定しているから

# 6. 考察

- Palley & Soll (2019 Mgmt.Sci.): 共有情報へのアプローチには、情報焦点的アプローチと判断焦点的アプローチがある
    - GPWMは中間に位置する
    - SP回答を知識豊かな対象者の回答だとみなすという点では判断焦点的
    - すべての対象者のデータによって決まるという点では情報焦点的

- GPWMの認知科学的位置づけ
    - Marr (1982)のいう計算理論レベルのモデル。対象者が持っている情報と、対象者の判断・予測との関係をモデル化
    - 社会的投影 (自分の信念と他者予測が相関を持つ現象)を説明
    - 他の認知モデルとは異なり、対象者の答えと他者予測との間に直接的関係を仮定しない

- ベイジアン階層モデルに基づくほかの集合知モデル (BCCモデル, CHモデル)との違い
    - 合意と真実のあいだの関係を仮定しない
    - 共有事前分布を推定
    - 回答のコーディングに影響されない

## 6.1 モデルの想定と拡張

- 共有知識
    - GPWMは、世界事前分布とシグナル行列について対象者が正確な知識を共有していると仮定している
    - 現実には、人はこれらの量について正確な知識を持たないし、個人差もある
    - 今後の拡張案
        - 同じシグナルを受けた人だけが同じ世界事前分布を持つ
        - 各対象者は世界事前分布とシグナル行列にノイズが乗ったものを知る
        - 世界事前分布の超事前分布を非一様にし、領域知識を表現する

- 利用可能なシグナル
    - GPWMでは、世界状態の数とシグナルの数が同じで、対応がある
    - 現実には、人は多様なシグナルを受け取る
    - 今後の拡張案
        - 世界状態のどれが正しそうかとその確率のタプルをシグナルとして受け取る
        - シグナルにもっと豊かな内部構造を持たせる

- 対象者の計算
    - CHモデルでは、対象者が知覚している確率に応じて対象者別のカリブレーションを行う
        - GPWMにも同様のカリブレーションを組み込めるかもしれない
    - 偽の独自性効果
        - 社会心理学では、偽の合意効果と、偽の独自性効果(他者予測において自分の答えを過度に軽視する)の両方が指摘されている
        - GPWMは前者を説明できるが後者を説明できない
        - 他者予測をシグナルの事前分布と事後分布の混合としてモデル化し、偽の独自性効果パラメータで重みづければよいかも
    - ダニング・クルーガー効果 (低成績者は自分の能力についての知識を調整できない)
        - GPWMに、専門性についての自己知識を専門性の関数として組み込めばよいかも
        - GPWMに、他者の専門性についての知識を組み込めばよいかも

**6.2 多値質問**

- 多値設問に拡張するには、世界知識とシグナルを多項分布にする必要がある

以上

# 感想

- 魅力的な点
    - 認知的なモデルを目指している！
- よく理解できなかった点
    - なぜ他者のシグナルの事後分布 $O$ と他者予測 $P$ を分けているのか
    - SPアルゴリズムとの関係
        - 「SPアルゴリズムの拡張」と主張しているが、SPアルゴリズム自体を組み込んでいるわけではない模様
- 疑問点
    - 頑健性
        - モデルにおける対象者は、誠実に投票し他者予測している。そうでない場合に何が起きるか？
            - ノイズが大きいとき
            - カテゴリに対する系統的なバイアスがあるとき (例, 社会的望ましさバイアス)
    - 認知モデルとしての妥当性
        - そもそも、他者予測にはどのような手がかり・方略が用いられているのか
        - 知識とシグナルの下で投票と他者予測が独立だという仮定は、さすがにちょっと奇妙な感じが…