

<下條研究の概要>

- 下條の研究では、比較対象がない場合の  $C_i$  の満足度と、ひとつの比較対象  $C_j$  がある場合の  $C_i$  の満足度はどう異なるかを検討している
  - これは、(確率ではなく満足度だが)「比較対象がない場合の  $C_i$  の満足度」を  $P(C_i|E)$ 、「ひとつの比較対象  $C_j$  がある場合の  $C_i$  の満足度」を  $P(C_i|E,C_j)$  と書ける
  - 仮説として、以下の2つが挙げられた
    - ①  $C_j$  が  $C_i$  より良いかどうかによって、 $C_i$  に対する満足度が変わる
      - $C_j$  が  $C_i$  より良いとき:  $P(C_i|E) > P(C_i|E,C_j)$
      - $C_j$  が  $C_i$  より悪いとき:  $P(C_i|E) < P(C_i|E,C_j)$
    - ② 比較対象があることによって、 $C_i$  の満足度が無条件で低下する
      - $P(C_i|E) > P(C_i|E,C_j)$
  - 実験の結果、②が支持された (比較対象があれば、無条件で満足度が低下した)

<下條研究のイントロと輪講論文との関連性>

- 人間は、観測された事象に対して「なぜその事象が起こったのか」という説明を求める
  - ある事象に対して、説明が1つしか存在しない場合もあれば、複数の説明の候補がある場合もある
  - しかし、ある事象に対する説明が複数存在し、どちらも同等に妥当である場合、どちらの説明を受容するかを判断するのが難しい (Douven & Mirabile, 2018)
- 説明に関する従来の研究では、説明の因果構造 (explanatory virtues) によって心理的指標 (e.g., 満足度) がどのように変化するかを検討する立場である Explanationism と、事前確率や尤度などの確率的情報によって確率的指標 (e.g., 主観的事後確率) がどのように変化するかを検討する立場である Bayesianism の2つに大別されてきた
  - ただし、従来の研究は、代替説明と比較した際の満足度のみを測定している
  - 比較しない場合の満足度と比べて、比較した場合の満足度がどのように変化しているかは未検討
    - Bayes の考え方に基づく Explaining Away (輪講論文のメインテーマ) が、この未検討部分に対応しているのではないか?
    - Explaining Away は、2つの原因候補があるとき、両原因の生起確率が低く見積もられるということの意味している? → 違った

## Explaining Away: significance of priors, diagnostic reasoning, and structural complexity

Alice Liefgreen<sup>1</sup>, Marko Tesic, & David Lagnado

In *Proceedings of the Annual Meeting of the Cognitive Science Society* (2018)

### <Introduction>

- Explaining Away とは、観察された事象（結果）を引き起こした可能性がある事象が、独立して複数存在し、競合している場合に行われる推論である
  - まず、結果が観察され（例：納屋が焼けた）、考えられる原因（例：不注意な喫煙、電気配線の不良）の事後確率を推定する
  - そして、一方の原因（電気配線の不良）の存在が確認されると、もう一方の原因（不注意な喫煙）の確率が更新（減少）され、「電気配線の不良が納屋の焼失を説明している」と判断する
    - 逆に、一方の原因（電気配線の不良）が起こっていなかったことがわかると、もう一方の原因（不注意な喫煙）の確率を増加させる
    - このような推論パターンは、社会的帰属、医療診断、法的領域など幅広い文脈で見られる (Kelley, 1973; Rottman & Hastie, 2016)
- この状況は、CBN (Causal Bayesian Network) でモデル化できる (Pearl, 2009)
  - Fig. 1-a は、2つの原因 ( $C_1$ ,  $C_2$ ) のいずれかが結果  $E$  を引き起こすモデル
  - CBN を完全にモデル化するためには、以下の2点が必要
    - ① 2つの原因の事前確率  $P(C_1)$ ,  $P(C_2)$
    - ② 各原因の存在/不在で条件付けたときの結果  $E$  の生起確率  
 $P(E | C_1, C_2)$ ,  $P(E | \neg C_1, C_2)$ ,  $P(E | C_1, \neg C_2)$ 、 $P(E | \neg C_1, \neg C_2)$
  - $P(C_1 | E, C_2) < P(C_1 | E) < P(C_1 | E, \neg C_2)$  となる (Morris & Larrick, 1995)

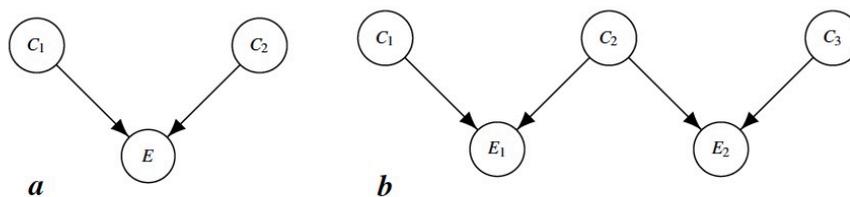


Figure 1: (a) 3-node CBN and (b) 5-node CBN

- 先行研究の欠点
- 1. 事前確率が明示されておらず、参加者が推定した主観的事後確率と、規範モデルが推定した事後確率を比較していない
  - Rottman & Hastie (2016) は、原因の事前確率を「中程度」or「低い」で操作
  - また、医療診断の文脈では、結果 E が観測されたときの原因の確率  $P(C_i|E)$  を参加者に求めさせることが多い
    - しかし、原因の事前確率  $P(C_i)$  と比較して、その確率の変化の方向性を考慮していない ( $P(C_i|E) < P(C_i)$ ?  $P(C_i|E) = P(C_i)$ ?  $P(C_i|E) > P(C_i)$ ?)
  - そのため、本研究では、以下の対処法によって、上述の欠点をクリアした
    - 事前確率を明示する
    - $P(C_i|E)$  が  $P(C_i)$  より小さいか、大きいか、または等しいかといった定性的な質問をした後、 $P(C_i|E)$  の確率の推定値を求める定量的な質問をする
- 2. Explaining Away に関する知見が、3-node 構造に限定されたものなのか、それとも、より複雑な構造 (5-node) を持つ因果構造にも及ぶものなのかを検討していない
  - 5-node の場合、3-node の場合よりも Explaining Away が強く起こると予想
  - また、5-node を導入することで、これまで説明してきた直接的 (direct) な Explaining Away と、連鎖的 (chained) な Explaining Away との比較が可能に
    - 【direct】ひとつの生起が、もうひとつの生起に全面的に影響する状況 (e.g.,  $C1 \rightarrow E \leftarrow C2$ 。C1 が生起したら、C2 の生起は不要)
    - 【chained】ひとつの生起が、もうひとつの生起に部分的に影響する状況 (e.g.,  $C1 \rightarrow E1 \leftarrow C2 \rightarrow E2 \leftarrow C3$ 。C1 が生起したら、C2 より C3)

#### <Experiment Overview>

- 【between】2(node 数: 3 or 5) × 2(コインの事前確率: 中程度 or 低)
  - Model 1 【3-node】:  $P(\text{Coin1}) = 0.5, P(\text{Coin2}) = 0.5$
  - Model 2 【3-node】:  $P(\text{Coin1}) = 0.2, P(\text{Coin2}) = 0.1$
  - Model 3 【5-node】:  $P(\text{Coin1}) = 0.5, P(\text{Coin2}) = 0.5, P(\text{Coin3}) = 0.5$
  - Model 4 【5-node】:  $P(\text{Coin1}) = 0.1, P(\text{Coin2}) = 0.2, P(\text{Coin3}) = 0.1$

<Methods>

- Participants:
  - 197 人 (母語が英語、平均 16.7 分の参加、謝金 1.25 ドル)
  - $N_{group1} = 51$  人、 $N_{group2} = 47$  人、 $N_{group3} = 49$  人、 $N_{group4} = 50$  人
  
- Design and Materials:
  - Group<sub>i</sub> には Model<sub>i</sub> (i = 1~4) が割り当てられた
  - 参加者は、各 Group のカバーストーリーを読み、アンケートに取り組んだ
    - Group1, 2 の質問数は 10 個、Group3, 4 の質問数は 17 個
      - Priors (片方のコインが表である事前確率。なお、もう片方のコインが表であるかどうか、ライトが点灯しているかどうかは不明)
      - Independence (片方のコインが表であることを知った上で、もう片方のコインが表である確率。なお、ライトが点灯しているかどうかは不明)
      - Diagnostic Reasoning (ライトが点灯したことを知った上で、片方のコインが表である確率。なお、もう片方のコインが表であるかどうかは不明)
      - Explaining Away (ライトが点灯していること and 片方のコインが表であることを知った上で、もう片方のコインが表である確率)
        - Direct (Model1, 2 の Explaining Away と、3, 4 の一部の質問)
        - Chained (Model3, 4 の一部の質問)

Table 1: Key inferences per group.

Inference Type	Key Inferences	
	Group 1 and 2	Group 3 and 4
<b>Priors</b>	$P_{1,2}(C_1 = H), P_{1,2}(C_2 = T)$	$P_{3,4}(C_1 = H), P_{3,4}(C_2 = T), P_{3,4}(C_3 = H)$
<b>Independence</b>	$P_{1,2}(C_2 = H   C_1 = H), P_{1,2}(C_1 = H   C_2 = T)$	$P_{3,4}(C_2 = H   C_1 = H), P_{3,4}(C_3 = H   C_2 = T), P_{3,4}(C_3 = H   C_2 = H)$
<b>Diagnostic Reasoning</b>	$P_{1,2}(C_1 = H   LB = on), P_{1,2}(C_2 = H   LB = on)$	$P_{3,4}(C_1 = H   LB_1 = on), P_{3,4}(C_2 = H   LB_1 = on)$
<b>Explaining Away</b>	$P_{1,2}(C_1 = H   LB = on, C_2 = H)$ $P_{1,2}(C_2 = H   LB = on, C_1 = H)$ $P_{1,2}(C_2 = H   LB = on, C_1 = T)$	<b>Direct:</b> $P_{3,4}(C_2 = H   LB_1 = on, C_1 = H)$ $P_{3,4}(C_2 = H   LB_1 = on, LB_2 = on, C_1 = H)$ $P_{3,4}(C_2 = H   LB_1 = on, LB_2 = on, C_1 = T)$ <b>Chained:</b> $P_{3,4}(C_1 = H   LB_1 = on, LB_2 = on)$ $P_{3,4}(C_3 = H   LB_1 = on, LB_2 = on, C_1 = H)$

あなたは、2つの部屋 (1, 2号室) と倉庫 (ユニット 1) がある建物にいます。各部屋には、自動でコインを投げる機械があり、ユニット 1 には電球が取り付けられています。各部屋の機械は、それぞれ別のケーブルで、ユニット 1 に取り付けられている電球に接続されています。つまり、次のような内容です。

- 部屋 1 の機械は、ユニット 1 に取り付けられている電球につながっている。
- 部屋 2 の機械は、ユニット 1 に取り付けられている電球につながっている。

部屋 1 の機械は、コイン 1 をトスします。コイン 1 は、50[20]%の確率で表になり、50[80]%の確率で裏になります。同様に、部屋 2 の機械は、コイン 2 をトスします。コイン 2 は、50[10]%の確率で表になり、50[90]%の確率で裏になります。

コイン 1 とコイン 2 は同時に投げられます。投げられた後、それぞれの機械は、ユニット 1 の電球に信号を送ります。投げられた 2 枚のコインのうち、少なくとも 1 枚が表ならば、電球が点灯します。つまり、次の場合にのみ、電球が点灯します。

- コイン 1 とコイン 2 の両方が「表」のとき
- コイン 1 が「表」、コイン 2 が「裏」のとき
- コイン 1 が「裏」、コイン 2 が「表」のとき

部屋 1 と 2、ユニット 1 は常に閉じられています。コイン 1 (2) が表か裏かを知るためには、部屋 1 (2) に入る必要があります。同様に、電球が点灯しているかどうかを調べるには、ユニット 1 に入る必要があります。ただし、あなたはどの部屋にもユニットにも入っていません。つまり、コイントスの結果や、電球のオン・オフは見えません。

以下の問いに答えてください。

Q1. コイン 1 が表になる確率は何%か? (0~100%で回答)  $P(C1)$  prior

Q2. コイン 2 が表になる確率は何%か? (0~100%で回答)  $P(C2)$  prior

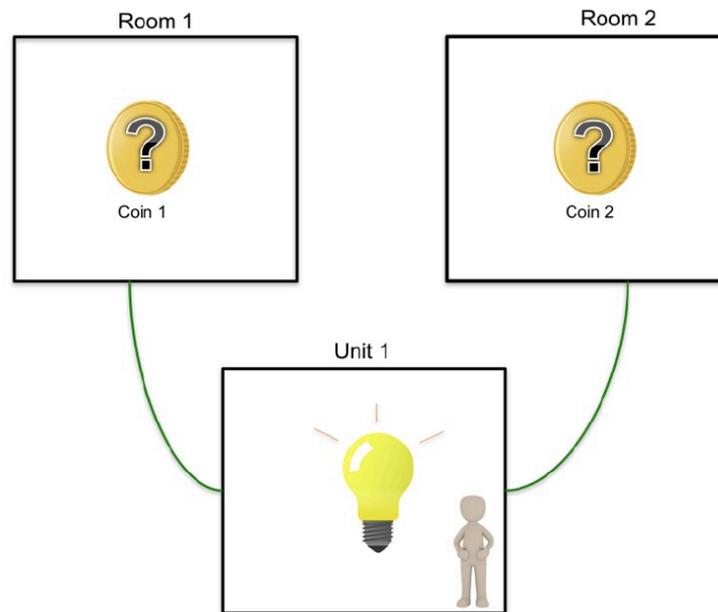
Q3. コイン 1 が表になった場合、コイン 2 も表になる確率は変わりますか?

(「はい (上がる)」 「はい (下がる)」 「いいえ」)  $P(C2 | C1)$  independence

Q4. コイン 2 が裏になった場合、コイン 1 が表になる確率は変わりますか?

(「はい (上がる)」 「はい (下がる)」 「いいえ」)  $P(C1 | \neg C2)$  independence

あなたは、ユニット1に入り、電球が点灯しているのを知りました。



以下の問いに答えてください。

Q5-1. 電球が点灯したことがわかった後、コイン1が表になる確率は変わりますか？

(「はい (上がる)」 「はい (下がる)」 「いいえ」) ☆ Diagnostic reasoning (定性)

Q5-2. コイン1が表になった確率はどの程度だと思いますか？

(0~100%) 【A'】  $P(C1 | LB)$  Diagnostic reasoning (定量)

Q6-1. 電球が点灯したことがわかった後、コイン2が表になる確率は変わりますか？

(「はい (上がる)」 「はい (下がる)」 「いいえ」) ☆ Diagnostic reasoning (定性)

Q6-2. コイン2が表になった確率はどの程度だと思いますか？

(0~100%) 【A】  $P(C2 | LB)$  Diagnostic reasoning (定量)

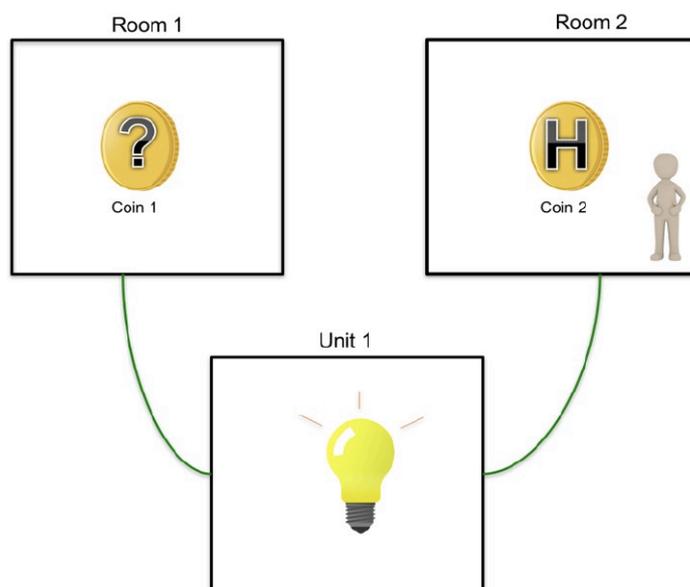
Q7. Q5とQ6の回答を考えると、コイン1が表になる確率は、コイン2が表になる確率とは異なりますか？

(はい： コイン1が表である確率は、コイン2が表である確率よりも大きい)

(はい： コイン1が表である確率は、コイン2が表である確率よりも小さい)

(いいえ： コイン1が表である確率は、コイン2が表である確率と等しい)

あなたはユニット 1 から出て、部屋 2 に入り、コイン 2 が表になっているのを知りました。



以下の問いに教えてください。

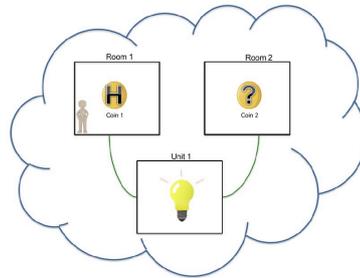
Q8-1. コイン 2 が表になったことがわかったとき、コイン 1 が表になる確率は (Q5 の確率と比べて) 変わりますか？

(「はい (上がる)」 「はい (下がる)」 「いいえ」) 【定性】 Explaining Away (direct)

Q8-2. コイン 1 が表になった確率はどの程度だと思いますか？

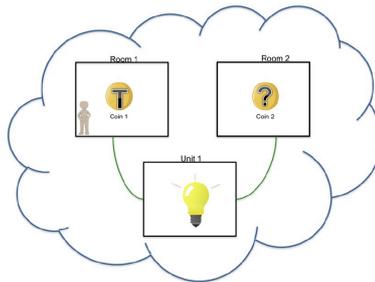
(0~100%) 【B'】  $P(C1 | LB, C2)$  【定量】 Explaining Away (direct)

ここで、電球が点灯したことはわかっていますが、部屋2に行ってコイン2が表になったことを知る代わりに、部屋1に行ってコイン1が表になったことを知るというシナリオを想像してみてください。



- Q9-1. このシナリオの場合、コイン1が表になったことがわかったとき、  
 コイン2が表になる確率は (Q6の確率と比べて) 変わりますか?  
 (「はい (上がる)」 「はい (下がる)」 「いいえ」) 【定性】 Explaining Away (direct)
- Q9-2. コイン2が表になった確率はどの程度だと思いますか?  
 (0~100%) 【B】  $P(C2 | LB, C1)$  【定量】 Explaining Away (direct)

ここで、また別のシナリオを想像してみてください。電球が点灯したことはわかっていますが、部屋2に行ってコイン2が表になったことを知る代わりに、部屋1に行ってコイン1が裏になったことを知るというシナリオです。



- Q10-1. このシナリオの場合、コイン1が裏になったことがわかったとき、  
 コイン2が表になる確率は (Q6の確率と比べて) 変わりますか?  
 (「はい (上がる)」 「はい (下がる)」 「いいえ」) 【定性】 Explaining Away (direct)
- Q10-2. コイン2が表になった確率はどの程度だと思いますか?  
 (0~100%) 【C】  $P(C2 | LB, -C1)$  【定量】 Explaining Away (direct)

<Results>

• Overall Performance

- 参加者 1 人あたりのスコアを算出し、それをパーセンテージに変換
  - Prior に関する質問では、提示された事前確率の  $\pm 5\%$  ( $\pm 10\%$ ) 以内であれば 0.5 (0.25) ポイント、そうでなければ 0 ポイントを与えた
  - Independence に関する質問では、正解すれば 1 点、そうでなければ 0 点
  - その他の定性的な質問では、正解すれば 2 点、不正解なら 0 点
  - その他の定量的な質問では、規範的な答えの  $\pm 5\%$  ( $\pm 10\%$ ) 以内であれば 1 (0.5) ポイント、そうでなければ 0 ポイント
- グループ間で、パーセンテージに有意差なし ( $M_{Group1} = 33.2\%$ ,  $M_{Group2} = 44.4\%$ ,  $M_{Group3} = 34.4\%$ ,  $M_{Group4} = 33.7\%$ ,  $\chi^2 = 3.37$ ,  $p = 0.34$ )
- モデルごとに 10,000 回のシミュレーションを行った結果、参加者の平均値以上のスコアが得られる可能性は極めて低かったため ( $p < .0001$ )、平均値がチャンスレベルでないと判断した

• Priors P(C1), P(C2)

- 【正確確率検定】事前確率に関するすべての質問に正しく答えた参加者の割合は、以下のモデル間で有意差あり ( $p_{all} < .0001$ )
  - グループ 1 (正答率 100%) vs. グループ 2、グループ 4
  - グループ 3 (正答率 94%) vs. グループ 2、グループ 4
  - これは、裏の確率と表の確率を単純に混同している (「10%の確率で表」を「10%の確率で裏」と勘違いしている) 参加者が多くいたからでは? (Fig. 2)

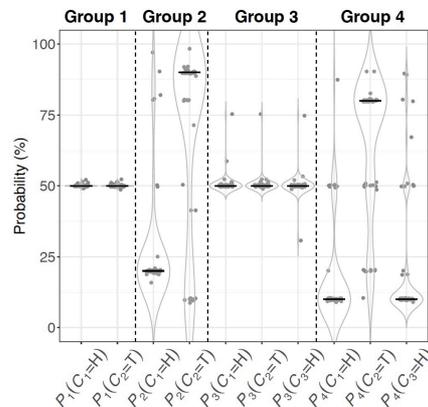


Figure 2: Distribution of prior probabilities. Black lines are stated priors.

- **Independence**  $P(C1|C2)$  vs.  $P(C1)$ ,  $P(C1|\neg C2)$  vs.  $P(C1)$ 
  - 独立性に関する質問に全て正解した参加者の割合は、チャンスレベルを上回った
    - グループ 1: 96%、グループ 2: 81%、グループ 3: 100%、グループ 4: 94%
    - チャンスレベルは、グループ 1, 2 が  $(1/3)^2 \doteq 11\%$ 、3, 4 が  $(1/3)^3 \doteq 4\%$
  - どのモデルに割り当てられたかにかかわらず、参加者は原因間の確率的独立性を正しく報告した
    - コイントスという、独立事象であることがわかりやすい題材だったから？
- ☆ **Diagnostic Reasoning** 【定性】  $P(C1|LB)$  vs.  $P(C1)$ ,  $P(C2|LB)$  vs.  $P(C2)$  【定量】  $P(C1|LB)$ ,  $P(C2|LB)$ 
  - 【正確確率検定】診断推論に関する定性的な質問に全て正解した参加者の割合は、グループ間で有意に異なった ( $p < .0001$ )
    - グループ 1: 6%、2: 42.5%、3: 20.4%、4: 30% (chance level:  $(1/3)^2 = 11\%$ )
    - 多重比較の結果、グループ 1・2 間、グループ 1・4 間で有意差あり
  - 【正確確率検定】診断推論に関する定量的な質問に全て正解した（正解の範囲は、規範解答の±5%まで）参加者の割合は、グループ間で有意差なし
    - グループ 1: 0%、2: 11%、3: 10%、4: 8%
  - 診断推論に関する定性的な質問において、事前確率の操作が回答の精度に与える影響の程度は、3-node モデルの方が、5-node モデルよりも高かった
    - グループ 1・2 間で差があるが、3・4 間で差が見られなかったため
    - なお、定量的な質問において、この効果は見られなかった
- **Explaining Away**
  - **Direct** 【定性】  $P(C1|LB)$  vs.  $P(C1)$ ,  $P(C2|LB)$  vs.  $P(C2)$  【定量】  $P(C1|LB, C2)$ ,  $P(C2|LB, C1)$ ,  $P(C2|LB, \neg C1)$ 
    - 【正確確率検定】Direct Explaining Away に関する定性的な質問に全て正解した参加者の割合は、4 モデル間で有意に異なった ( $p = .0004$ )
      - グループ 1: 2%、2: 25.5%、3: 6%、4: 4% (chance level:  $(1/3)^4 = 4\%$ )
      - 多重比較の結果、グループ 1・2 間、グループ 2・4 間で有意差あり
      - したがって、診断的推論の結果と同様に、3-node モデルを用いて推論したグループ間では、5-node モデルを用いて推論したグループ間に比べて、Direct Explaining Away に関する定性的な質問への回答の精度が異なった
        - つまり、3-node モデルの場合、事前確率が影響した

- 【ANOVA】 Direct Explaining Away に関する 3 つの質問を独立変数、その回答を従属変数として分析した結果、全グループ内で有意差あり ( $p_{all} < .0001$ )
  - 定量的な質問に対する回答を比較した結果 (表 2 の 95%CI 参照)、いずれも規範的な差 (Norm. diff.) が含まれていなかったため、どのグループにおいても Explaining away が十分でないことがわかった
  - しかし、グループ間で Explaining Away の程度が異なり (Emp. diff in Table 2 参照)、グループ 2 だけが、3 つの比較のすべてにおいてその程度がゼロより大きかった (他グループと比べて、規範的な差分が大きいからでは?)

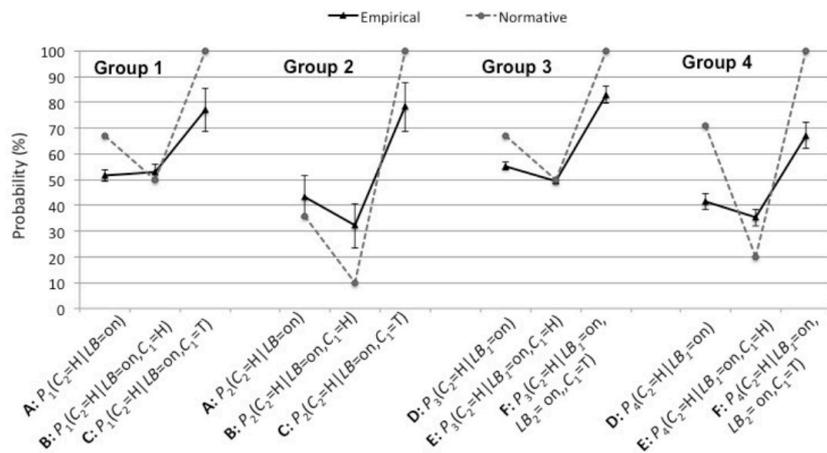


Figure 3: Explaining away inferences. Error bars are 95% confidence intervals.

Table 2: Within group explaining away.

Norm. diff. = normative difference, Emp. diff. = empirical difference, 95% CI of emp. diff. = 95% CI of empirical difference.

Inferences <sup>10</sup>	Norm. diff.	Emp. diff.	95% CI of emp. diff.
<i>Group 1</i>			
A - B	17	-1.2	[-5, 2.4]
C - B	50	24.2	[14.1, 34.3]
A' - B'	17	0.8	[-3.3, 4.9]
<i>Group 2</i>			
A - B	26	11.3	[4.2, 18.4]
C - B	90	46.2	[32.4, 59.9]
A' - B'	51	13.9	[7.3, 20.5]
<i>Group 3</i>			
D - E	17	5.7	[2.8, 8.5]
F - E	50	33.3	[26.6, 39.8]
D' - E'	13	2.1	[-0.7, 4.9]
<i>Group 4</i>			
D - E	51	6.3	[-0.1, 12.7]
F - E	80	31.9	[19.4, 44.4]
D' - E'	25	7.3	[3.2, 11.3]

【グループ1, 2】

- A :  $P(C_2 | LB)$
- A' :  $P(C_1 | LB)$
- B :  $P(C_2 | LB, C_1)$
- B' :  $P(C_1 | LB, C_2)$
- C :  $P(C_2 | LB, -C_1)$

【グループ3, 4】

- D :  $P(C_2 | LB1)$
- D' :  $P(C_1 | LB1)$
- E :  $P(C_2 | LB1, LB2, C_1)$
- E' :  $P(C_1 | LB, LB2, C_2)$
- F :  $P(C_2 | LB, LB2, -C_1)$

- **【ANOVA】** グループを独立変数、グループ 1, 2 の C、3, 4 の F における確率推定値を従属変数として分析した結果、有意差なし ( $p = 0.08$ )
  - グループ 1: 76.9%、2: 78.3%、3: 82.9%、4: 67.1%
  - node の数、事前確率による影響は見られなかった
  - また、いずれの値も、規範解である 100%よりも有意に低かった (Fig. 3)
  
- **Chained** 【定量】 P(C3|LB1, LB2, C1)、P(C3|LB1, LB2)
  - Chained Explaining Away に関する定性的な質問に全て正解した参加者の割合は、グループ 3 で 4%、4 で 8% (chance level:  $(1/3)^2 = 11\%$ )
    - **【正確確率検定】** グループ間で有意差なし ( $p = .36$ )
      - 事前確率は、Chained Explaining Away に関する質的な質問への回答の精度には影響しなかった
    - **【t 検定】** 各グループ内で、Diagnostic reasoning に関する定性的な質問に全て正しく答えた参加者と、そうでない参加者に分けて、(おそらく Direct・Chained を問わず) Explaining Away のスコアを比較した
      - グループ 1: 有意差なし ( $t(2) = 1.52, p = .26$ )
      - グループ 2: あり ( $t(29) = 4.76, p < .0001$ )
      - グループ 3: あり ( $t(10) = 3.5, p = .006$ )
      - グループ 4: あり ( $t(22) = 3.4, p = .002$ )
      - Diagnostic reasoning に関する定性的な質問の正答率が、その後の Explaining Away の程度を予測することを示唆
      - ただし、事前確率が中程度で、3-node 構造を持つ因果関係において確率推定を行ったグループ 1 の参加者には当てはまらなかった
        - グループ 1 の参加者のうち、診断推論の両質問に正解したのが 3 人だけだったから？