

ステレオタイプ脅威の再現性

「女性は数学が苦手」を、女性のパフォーマンスを下げるのか？

ステレオタイプ脅威とはなにか？

ステレオタイプ脅威とは、「アイデンティティ付随条件の中でも（中略）目に見えない脅威」（スティール 訳・藤原, 2020）のことである。アイデンティティ付随条件とは、「ある状況下で、特定の社会的アイデンティティを持つがゆえに（中略）対処しなければならない物事」のことである。たとえば、女性(社会的アイデンティティ)が数学の試験を受ける(特定の状況下)で対処しなければならないこと(アイデンティティ付随条件)は、「女性は数学が苦手」という目に見えない脅威(ステレオタイプ脅威)である。『ステレオタイプの科学』はステレオタイプ脅威研究の第一人者であるスティール氏が著した書であり、ステレオタイプ脅威が様々な場面で大きな負の効果を持つことを多くの実験により示されている。今回紹介する論文は、ステレオタイプ脅威の再現性を概念的に検証した研究である。

ステレオタイプ脅威の問題点

ステレオタイプ脅威は、多くの研究で示されている。

- 黒人に「認知能力をテストする」と宣言して言語テストを行わせると、成績が低下したが、宣言しなかった方は白人と有意差なかった(Steele & Aronson, 1995)
- 数学ができる女性は、女性的特徴との関連を否定し、女性的特徴との関連が弱いと考えられるものは否定しない(ステレオタイプ驚異を顕現させるとこの効果は強まったが、数学ができない女性にはこの傾向見られなかった)(Pronin et al., 2004)

一方で、ステレオタイプ脅威の存在に疑問を投げかける研究もある。

- Steele & Aronson (1995)の実験を再現できたと思われる研究の内、再現できたものは55%であり、再現できたも研究をメタ分析したところ、数学成績を調整した研究しかステレオタイプ脅威は見られなかった(Stoet & Geary, 2012)
- メタ分析により、女子学生におけるステレオタイプ脅威は出版バイアスの可能性があることが示された(Flore & Wichers, 2015)

この論文では特に、Steele and Aronson (1995)を追試したStricker & Ward (2004)の論文を巡る議論に着目し、議論を進めている。

Stricker & Ward (2004)は、（黒人のステレオタイプを顕現させると黒人の認知成績が下がることを示した）Steele and Aronson (1995)の実験を追試した。AP Calculus AB(数学科目?)の問題を被験者に

回答させた。回答させる際は最初に人種と性別を回答させる群、と最後に回答させる群に分けて検討した。成績を分析した結果、黒人、女性、その他のグループの生徒の成績には有意差なかった。

この論文に対し、反対論文を提出したのがDanaher & Crandall (2008)である。彼らはStricker & Ward (2004)の結果を再分析し、人種と性別を回答させることが大きな効果を持つと主張した。その効果の大きさは、性別に関する質問の順番を変えるだけでAP Calculus AB(数学科目?)の単位を取得する米国人女性を毎年4,700人以上増加させるほどだという。さらに、Stricker & Ward (2004)の分析手法を「タイプIエラーではなく、タイプIIエラーを生じさせる手法だ」(p.1642)と批判した。

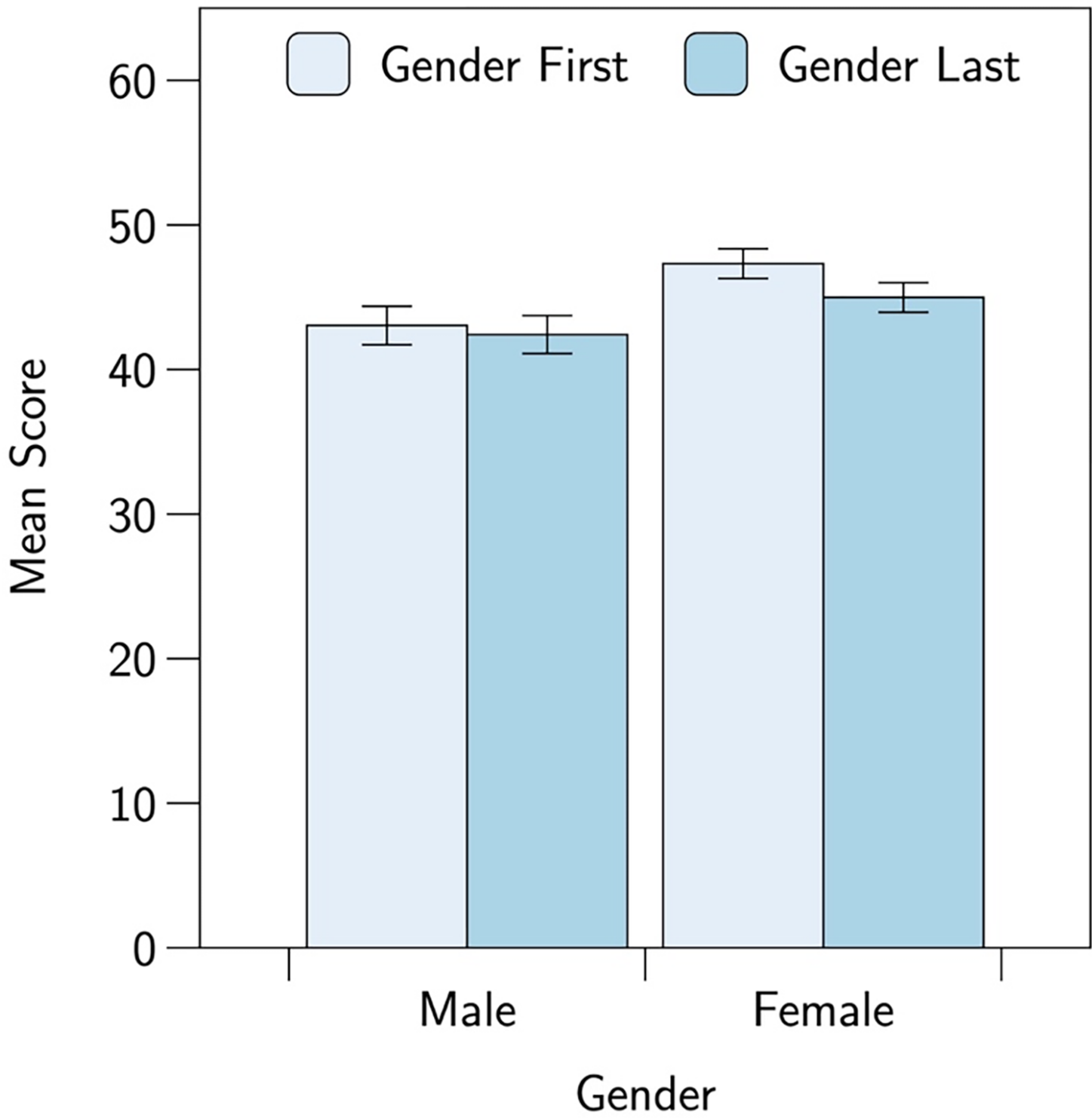
その後すぐにStricker & Ward (2008)は、このDanaher & Crandall (2008)について論文で再反論した。彼らはDanaher & Crandall (2008)について、一部を恣意的に指摘しているために方向性が歪んでおり、統計的手法に対する批判は間違っている(数学の成績が正規分布ではないのに、標準特典に変換して成績が3以上の割合を求めている、白人が59%と割合が等しくないのに、セルの割合が等しい場合に用いられる最小二乗法での分散分析で検討しているなど)と指摘した(Stricker & Ward, 2008)。では、どちらの主張が正しいのだろうか。今回紹介するInglis & O'Hagan (2022)の目的は、Stricker & Ward (2004)の実験を別のテストで概念的に再現することで、ジェンダーに関する質問を最初にしても女性の成績はしないことを示すことである。

方法

- イギリスで全国的に行われている数学のテスト(JMC)を用いてテスト
- 5つの学校に通う、11-13才の生徒を対象とした
- 2種類の解答用紙(genderを最後 vs. 最初に聞く)を配布
 - 用紙の順をランダムにし、配るときは順番に
- ANOVA(ベイズアンモ)、multi levelをプレレジした上で分析
 - 探索的に女子校(1校)を除外しても分析

結果

聞く順番で成績は変わらない(交互作用に有意差なし)



ベイズファクターも帰無仮説を採択[リンク]
(file:///C:/Users/Miffy/Downloads/All_1169_Participants-results.html)

Bayesian ANOVA

A Bayesian version of the main ANOVA, with a Cauchy prior width of 0.5.

Model Comparison

Models	P(M)	P(M data)	BF _M	BF ₀₁	error %
Gender	0.200	0.684	8.662	1.000	
Null model	0.200	0.144	0.675	4.736	8.280e -4
Form + Gender	0.200	0.125	0.572	<u>5.466</u>	1.792
Form	0.200	0.031	0.130	21.762	0.029
Form + Gender + Form * Gender	0.200	0.015	0.060	<u>46.052</u>	2.149

補足

成績の平均値は女性の方が高かったので($p=.004$)、女子校を除外したら、男性の方がわずかに高かった($d = .165$)。

しかし、どちらのサンプルでも交互作用は有意ではない。

この結果で80%の検出力を想定すると、 $\eta=.0065$ になる(小さすぎる)。

高橋の感想と質問

これぞ科学！な論文でした

この論文は、「ステレオタイプ脅威ある！→いや、ない！→ないなんてことはない！分析が間違ってる！→いやいや、その指摘が間違ってる！」という流れの中で、分析のようなテクニカルな水掛け論ではなく、概念的に再現されるかを検証した論文でした。概念的な妥当性はもちろん、分析の方法論も非常に精緻で、まさに科学の発展を実感させる内容でした。すごい。

再現実験は、数で殴るだけじゃない

再現実験といえば、TMTのようなMany lab + 提唱者を混ぜた有無を言わせない研究だという印象でした。しかし、この論文のように、概念的に検討することも立派な再現実験であり、心理学のメインストリートに逆らう方法には、こんなやり方もあるんだな〜と勉強になりました。

やっぱり、ベイズは強い

帰無仮説を積極的に採択できることで、再現可能性が低い理論や実験を科学的に淘汰できるのはものすごく強いなと。ベイズファクターがあることで「差があるとは言えないという証拠は見つけれなかった」などという言い訳は通じません。本当に心強い指標であるとともに、これからの心理学(を担う僕ら若者)の強い味方だなと。

ステレオタイプ脅威、やっぱり怪しい

今回の論文では「性別を回答させる順番の効果」が再現されなかっただけであり、ステレオタイプ脅威がないとは言っていないです。ただ、社会的不平等に社会的カテゴリーは必要ないこと(それに付随する社会的アイデンティティも必要ないこと)を考えると、理論的にも欠陥があるよなーと感じました。