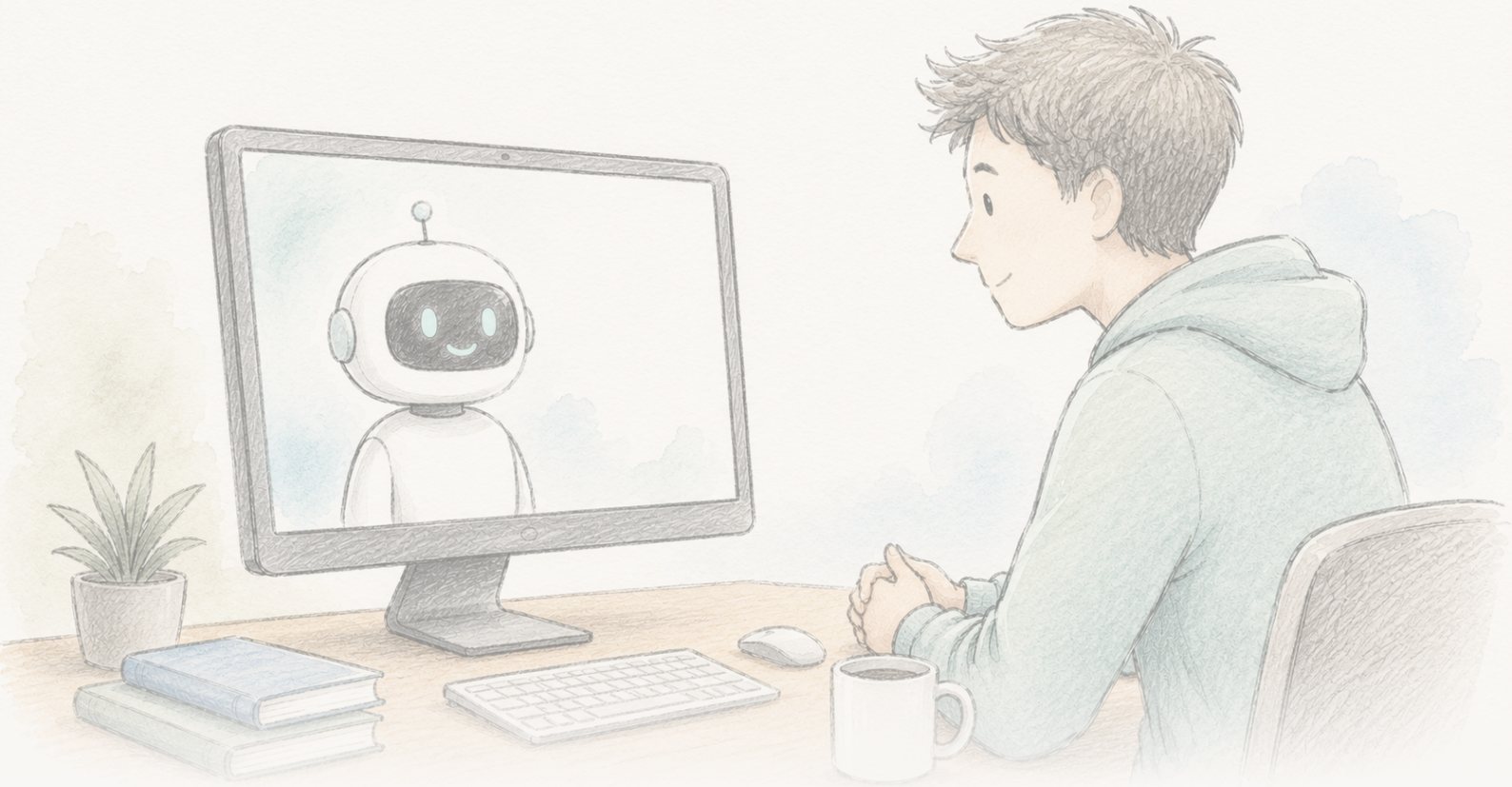


The CLEAN PAPER

WHAT THE PAPER SAYS · WHAT IT DOESN'T · WHY IT MATTERS

آخرکار حقائق بھی اثر ڈال سکتے ھیں — مگر مقالہ پر ایک باقاعدہ warning بھی ھے

2024 کی مطالعہ نے رپورٹ کیا کہ *Turbo GPT-4* کے ساتھ تین-*conversation personalized round* سے لوگوں کے *chosen* سازشی مفروضہ میں اوسطاً تقریباً 20% کمی ہوئی، جو دو ماہ تک قائم رھی، مختلف سازشی *types* میں نظر آئی، اور اے آئی کے *factual* دعوے *fact-checking* میں *accurate overwhelmingly* تھے۔ جون 2026 میں مقالہ پر ڈیٹا-*handling* اور *problems reproducibility* کی وجہ سے ***Editorial Concern*** *of Expression* لگا دی گئی۔ مصنفین کہتے ھیں *corrected* تجزیہ اثر کی سمت، اھمیت اور حجم برقرار رکھتا ھے، جبکہ *سائنس* ابھی *evaluate* کر رھا ھے۔ اسے ایک *designed carefully interesting، genuinely* مگر ***currently** under* جائزہ** نتیجہ کے طور پر پڑھیں — نہ *settled*، نہ *debunked*۔ اس کہانی کا سب سے دیانت دار جملہ سائنسی عمل خود لکھ رھا ھے: **دوبارہ جانچ کریں۔**

Written by Lucio Vaglio · Cross-checked by Cinzia Vaglio and Vera Rubin · Edited by Michele Renda · Figures by Nova Vela · AI-assisted, human-reviewed

Paper: *Durably reducing conspiracy beliefs through dialogues with AI*, Thomas H. Costello, Gordon Pennycook, David G. Rand, Science ,385 eadq1814 (2024) · DOI: science.adq1814/1126.10

Concern of Expression Editorial

Science نے اس paper کے ساتھ Editorial Expression of Concern کی منسلک کہ جب ہے، data handling اور نہیں؛ بھی finding کا misconduct اور نہیں retraction یہ ہیں۔ رہے جا کیے evaluate سوالات کے reproducibility چاہیے۔ پڑھنا کر سمجھ provisional کو reported result تک ہونے review resolve کہ ہے مطلب کا اس

→ Concern of Expression Editorial

آخرکار حقائق بھی اثر ڈال سکتے ہیں — مگر مقالہ پر ایک باقاعدہ تنبیہ بھی ہے

2024 میں ایک مطالعہ نے wisdom conventional کے ایک آرام دہ خیال کے خلاف نتیجہ رپورٹ کیا۔ اگر کسی شخص کو اے آئی — Turbo GPT-4 — کے ساتھ مختصر تین-مرحلہ گفتگو دی جائے، اور ماڈل کو خاص طور پر اس شواھد کا جواب دینے کے لیے ھدایت کیا جائے جسے وہ شخص اپنی مانی ہوئی سازشی نظریہ کے حق میں پیش کرتا ہے، تو اوسطاً اس نظریہ پر اس کا یقین تقریباً **20% کم** ہو جاتا ہے۔ یہ کمی دو ماہ بعد بھی باقی تھی۔ اثر روایتی assassination، Kennedy F. John — conspiracies — اور موجودہ issues — election US 2020 COVID-19، — دونوں میں نظر آیا، حتیٰ کہ ان لوگوں میں بھی جن کا مفروضہ مضبوط اور identity سے جڑا ہوا تھا۔ Professional حقیقت-checker نے اے آئی کے 128 factual دعوے grade کیے تو 127 درست، 1 misleading اور کوئی بھی غلط نہیں نکلا۔

جو سرخی پھیلی وہ یہ تھی کہ لوگوں کو **hole rabbit سے حقائق کے ذریعے باھر لایا جا سکتا ہے** — سازشی believers شواھد سے reach beyond نہیں؛ شاید انہیں بس ان کے اپنے مخصوص شواھد کا مخصوص جواب چاہیے۔ یہ genuinely دلچسپ دعویٰ ہے، اور persuasion تحقیق کے عام معیاری کے مقابلے میں مطالعے کافی carefully تیار کیا گیا تھی۔

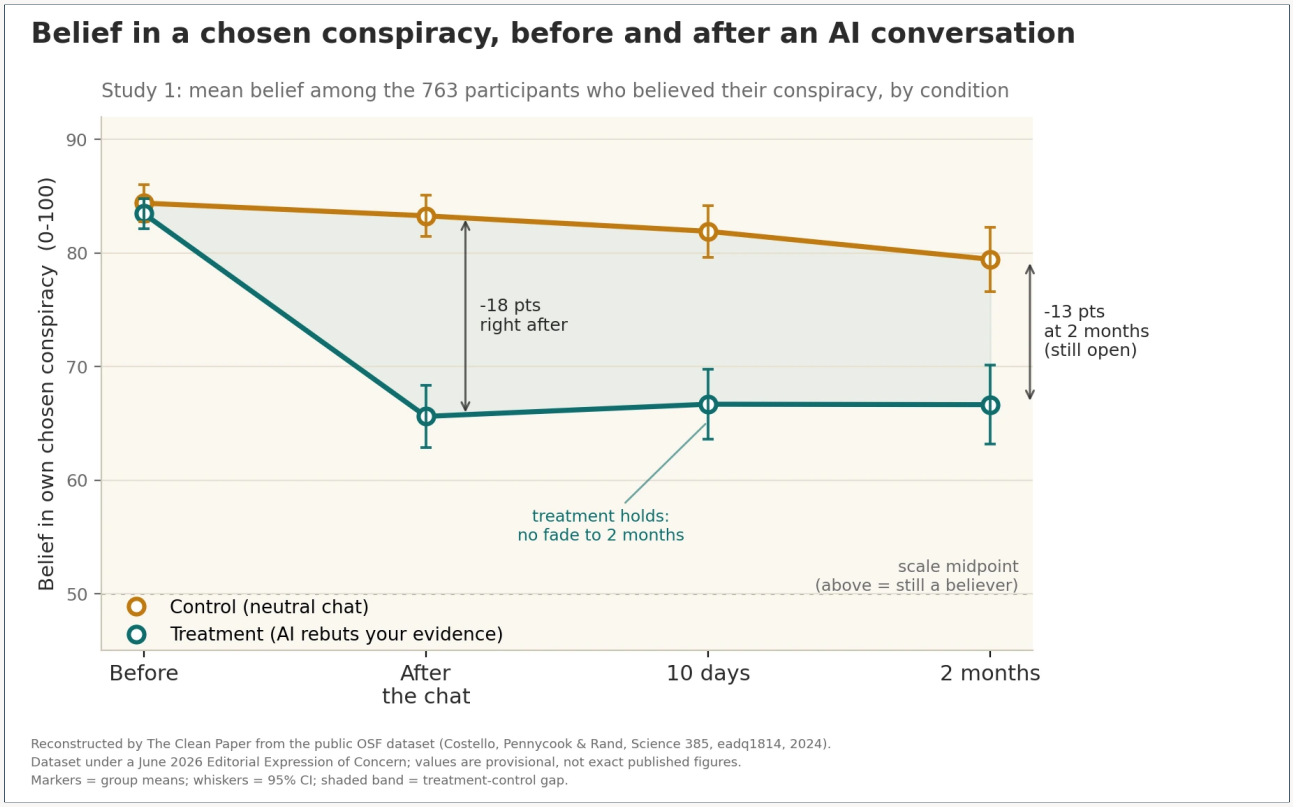
پھر جون 2026 میں سائنس نے مقالہ پر **Editorial ادارتی اظہارِ تشویش** جاری کر دی۔ زیادہ تر retellings یہی حصہ چھوڑ دیں گی، حالانکہ ابھی نتیجہ پر کتنا وزن رکھا جا سکتا ہے، یہ طے کرنے میں یہی حصہ اہم ہے۔

Concern of Expression کیا ہے — اور کیا نہیں

کہ جائے بتایا کو readers تا کہ ہے، ہوتا نشان صوری گیا لگایا پر مقالہ سے طرف کی جریدہ تشویش اظہارِ ادارتی Editorial بخود خود یہ اور — گیا لیا نہیں واپس مقالہ — **نہیں retraction** یہ ہیں۔ رہے ہو investigate ابھی اور ہیں اٹھے سوال کچھ ریکارڈ سائنسی کیونکہ پڑھیں، کر رکھ میں ذھن کو caveat اس ہے: یہ صرف مطلب نہیں۔ بھی فیصلہ کا misconduct اور کیں رپورٹ کو سائنس تفصیلات کی، investigation نے مصنفین بعد کے آنے سامنے issues یہاں ہے۔ سکتا بدل کیا۔ فراہم تجزیہ corrected

نشان کیے گئے مسائل مخصوص ہیں۔ مصنفین کو معلوم ہوا کہ مسودہ اور شائع شدہ تجزیہ کوڈ میں شریک-اسکریننگ criteria apply کرنے کے طریقے میں inconsistencies تھیں، اور عوامی ڈیٹا سیٹ میں کوڈ-merging غلطی سے غیر متعلق قطاریں splice ہو گئی تھیں۔ ان problems کی وجہ سے released ڈیٹا/کوڈ سے بعض رپورٹ شدہ اعداد بڑھانا مشکل تھا۔ مصنفین نے سائنس کو corrected تجزیہ pipeline اور updated نتائج دیے، اور کہتے ہیں کہ corrected نتائج اصل کے ساتھ **سمت، شماریاتی اہمیت اور substantive حجم** میں مطابقت کرتی ہیں۔ سائنس ابھی انہیں evaluate کر رہا ہے۔ جب تک جریدہ کی جائزہ ختم نہیں ہوتی، عین figures عارضی ہیں۔

اس لیے یہاں دو کہانیوں ایک ساتھ چل رہی ہیں: کیا حقائق beliefs entrenched کو حرکت کر سکتے ہیں، اور سائنسی ریکارڈ اپنی غلطیوں کو عوامی طور پر کیسے درست کرتا ہے۔ اہم بات دونوں کو الگ رکھنا ہے۔



مطالعہ میں شریک کے اپنے stated شواہد کو rebut کرنے والی تین-مرحلہ اے آئی گفتگو سے اس کے منتخب سازشی مفروضہ میں اوسطاً تقریباً 20% کمی رپورٹ ہوئی؛ unrelated-topic کنٹرول chat کے برعکس یہ drop 10 دن اور 2 مہینوں بعد بھی قابلِ پیمائش تھا۔ اثر پائیدار مگر partial تھا: زیادہ تر شرکاء afterward سازشی پر اب بھی یقین رکھتے تھے — یہ کمی تھی، علاج نہیں۔ مقالہ پر جون 2026 میں سائنس کی Editorial ادارتی اظہارِ تشویش موجود ہے، inconsistencies reporting اور عوامی ڈیٹا سیٹ کی غلطی کی وجہ سے؛ مصنفین کہتے ہیں corrected تجزیہ اثر کی سمت، اہمیت اور حجم برقرار رکھتا ہے، جبکہ سائنس جائزہ جاری رکھے ہوئے ہے۔ Chart عوامی ڈیٹا سیٹ سے The صاف مقالہ نے بازسازی کرنا کیا ہے، اس لیے displayed قدریں عارضی ہیں، مقالے کے حتمی figures corrected نہیں۔

Original figure — The Clean Paper CC BY 0.4

مصنفین نے کیا کیا

- دو تجربات میں **2,190 شرکاء** شامل کیے، جن میں ہر شخص نے اپنے الفاظ میں ایک سازشی نظریہ بتائی جس پر وہ یقین رکھتا تھا اور وہ شواہد بھی بتایا جسے وہ اس کے حق میں سمجھتا تھا۔
- ہر شریک نے GPT-4 Turbo کے ساتھ تین-مرحلہ گفتگو کی۔ **علاج** حالت میں اے آئی کو شریک کے مخصوص شواہد rebut کرنے کے لیے ہدایت کیا گیا؛ **کنٹرول** حالت میں topic unrelated پر بات ہوئی۔
- منتخب سازشی پر مفروضہ گفتگو سے پہلے، فوراً بعد، پھر **10 دن اور 2 مہینوں** بعد پیمائش کیا گیا۔
- Professional حقیقت-checker سے نمونہ میں اے آئی کے تمام **128 factual دعوے** کی درستگی evaluate کرائی گئی۔
- یہ بھی دیکھا گیا کہ اثر دوسری، غیر متعلقہ سازشی عقائد تک پھیلتا ہے یا نہیں؛ اور تخصیص کے امتحان کے طور پر یہ بھی جانچا گیا کہ مداخلت ایسے دعووں پر بھی یقین کم کرتی ہے یا نہیں جو حقیقت میں **درست** ہوں۔

کیا ملا

- علاج گروپ میں کنٹرول کے مقابلے میں منتخب سازشی مفروضہ میں **تقریباً 20% اوسط کمی**، مطالعہ 1 میں 100-نقطہ پیمانہ پر 95% اعتماد وقفہ [8.13, 7.19] نقاط، P < 0.001۔
- **اثر برقرار رہا۔** علاج گروپ میں مفروضہ 10 دن اور 2 مہینوں پر بھی just-بعد chat- سطح کے قریب رہی، واپس اوپر نہیں گئی؛ کنٹرول زیادہ بلند رہا۔ مقالہ اسے منتخب سازشی پر اثر کا کم از کم دو ماہ undiminished رہنا بیان کرتا ہے۔
- **اثر مختلف سازشی اقسام میں generalize ہوا** اور مضبوط believers initial میں بھی نظر آیا۔
- **اے آئی کے factual دعوے اس ترتیب میں بہت درست تھے:** 128 میں سے 127 (99.2%) درست، 1 (0.8%) misleading، صفر غلط۔
- **cynicism Blanket نہیں آیا۔** مداخلت نے درست conspiracies میں مفروضہ کم نہیں کیا؛ یہ beliefs unsupported حرکت کرنے کے مطابق ہے، ہر چیز پر doubt پیدا کرنے کے نہیں۔
- **محدود spillover** بھی ہوا: conspiracies unrelated میں مفروضہ تقریباً 12% کم ہوئی اور شرکاء نے سازشی دعوے چیلنج کرنے کی زیادہ intention رپورٹ کی۔ اہم اثر مطالعہ 2 میں بھی رہا، اور ایک smaller نہیں-کنٹرول نمونہ میں ایک ہی سمت میں تھا — کمزور شواہد، مگر ہم آہنگ۔

یہ کیا ثابت نہیں کرتا

- یہ نتیجہ ابھی **unquestioned نہیں۔** مقالہ ڈیٹا-handling اور issues reproducibility کی وجہ سے Editorial ادارتی اظہارِ تشویش کے تحت ہے، اور corrected اعداد سائنس ابھی evaluate کر رہا ہے۔ مخصوص قدریں کو جائزہ مکمل ہونے تک عارضی سمجھیں۔
- یہ نہیں دکھاتا کہ اے آئی سازشی مفروضہ کو **"علاج"** کرتا ہے۔ تقریباً 20% اوسط کمی معنی خیز تبدیلی ہے، erasure نہیں؛ زیادہ تر شرکاء afterward نظریہ پر اب بھی یقین رکھتے تھے، صرف کم۔
- یہ **حقیقی دنیا میں تعیناتی** کا نتیجہ نہیں۔ شرکاء رضاکارانہ طور پر مطالعے میں آئے اور اے آئی کے ساتھ سنجیدگی سے گفتگو کی۔ تجربہ گاہ سے باہر سب سے پختہ یقین رکھنے والے لوگ شاید ایسے چیٹ بوٹ سے بات ہی نہ کریں جو ان کی بات کی تائید نہ کرے۔
- 99.2% درستگی **اس کام، ماڈل اور محتاط prompting** کے لیے ہے؛ یہ زبان ماڈلز کی عمومی truth ضمانت نہیں۔ مصنفین خود واضح ہیں کہ یہی شخصی persuasive طاقت غلط push beliefs کرنے کے لیے بھی استعمال ہو سکتی ہے۔
- یہاں "پائیدار" کا مطلب **دو ماہ** ہے، permanent نہیں۔

شواہد کتنے مضبوط ہیں

- **ڈیزائن واقعی مضبوط ہے۔** علاج-بمقابلہ-کنٹرول تجربات، پلیسبو-like کنٹرول، دو مطالعات میں replication، آزاد نمونہ، durability پیروی-ups اور اے آئی دعوے کی واضح درستگی جانچ — persuasion تحقیق کے بہت سے مطالعات سے زیادہ محتاط۔ جہاں ٹیسٹ کیا گیا اثر کی سمت ہم آہنگ رہی۔
- **ادارتی اظہارِ تشویش footnote نہیں۔** Released ڈیٹا اور کوڈ سے رپورٹ شدہ اعداد دوبارہ بننا کرنا trust کا بنیادی حصہ ہے، اور یہی چیز ٹوٹی: اسکریننگ-criteria inconsistencies اور کوڈ-merging غلطی سے duplicated/extraneous قطاریں۔ مصنفین کا کہنا کہ pipeline corrected نتیجہ برقرار رکھتا ہے encouraging ہے، مگر ابھی مصنفین کا اپنا account ہے؛ آزاد فیصلہ نہیں۔ اہم اگلا قدم سائنس کی جائزہ ہے۔

- **دیانت دار حیثیت "flagged" ہے، "debunked" یا "تصدیق شدہ" نہیں۔** اچھی طرح تیار کیا گیا اور striking مطالعہ، مگر عین اعداد صوری جائزہ کے تحت ہیں۔ اچھی کہانی کو یہاں روکنا uncomfortable ہے — مگر فی الحال درست یہی ہے۔

یہ کیوں اہم ہے

کئی برسوں سے ایک influential منظر یہ رہا کہ سازشی needs psychological believers اور identity سے driven ہوتے ہیں، اس لیے حقائق سے انہیں ان beliefs سے دلیل دیتے out نہیں کیا جا سکتا۔ اگر پائیدار شواھد-پر مبنی کمی جائزہ کے بعد قائم رہتی ہے تو یہ اس pessimism کے خلاف معنی خیز counterweight ہوگی: شاید عمومی debunking اس لیے ناکام ہونا ہوتی رہی کیونکہ وہ **مخصوص شواھد** کو حل کرنا نہیں کرتی جسے believer خود cite کرتا ہے، جبکہ LLM اسے پیمانہ پر کر سکتا ہے۔

یہ سائنس کے طریقۂ کار کی کیس مطالعہ بھی ہے۔ ادارتی اظہارِ تشویش کا lesson یہ نہیں کہ celebrated نتائج worthless ہیں؛ بلکہ یہ کہ غلطیاں سطح ہوتے ہیں، ڈیٹا دوبارہ جانچا جاتا ہے، اور دعوے عوامی طور پر دوبارہ جانچ کیے جاتے ہیں۔ یہ مشینری چلنا سائنس کی خصوصیت ہے، scandal نہیں۔ اسی لیے درست posture نہ فوری applause ہے نہ dismissal، بلکہ patience۔

اور اے آئی کے لحاظ سے technique دونوں طرف کاٹتی ہے۔ Tailored دلیل غلط مفروضہ کو deflate کر سکتا ہے تو غلط مفروضہ کو inflate بھی کر سکتا ہے۔ Technique غیر جانب دار ہے؛ مقصد غیر جانب دار نہیں۔

صاف خلاصہ

2024 کے مطالعے نے رپورٹ کیا کہ Turbo GPT-4 کے ساتھ تین مرحلوں کی شخصی گفتگو سے لوگوں کے منتخب سازشی عقیدے میں اوسطاً تقریباً 20% کمی ہوئی، جو دو ماہ تک برقرار رہی اور مختلف قسم کی سازشی سوچ میں نظر آئی۔ حقیقت جانچنے والوں نے اے آئی کے دعووں کو مجموعی طور پر بہت درست پایا۔ جون 2026 میں ڈیٹا سنبھالنے اور نتائج دوبارہ پیدا کرنے کے مسائل کی وجہ سے مقالے پر **ادارتی اظہارِ تشویش** لگا دیا گیا۔ مصنفین کہتے ہیں کہ درست کیے گئے تجزیے میں اثر کی سمت، اہمیت اور حجم برقرار ہے، جبکہ *Science* ابھی اس کا جائزہ لے رہا ہے۔ اس نتیجے کو واقعی دلچسپ اور احتیاط سے ڈیزائن شدہ، مگر **اس وقت زیرِ جائزہ** سمجھنا چاہیے — نہ مکمل طور پر طے شدہ، نہ رد شدہ۔ اس کہانی کا سب سے دیانت دار جملہ خود سائنسی عمل لکھ رہا ہے: **دوبارہ جانچیں۔**

ماخذ

eadq1814 ,385 Science AI, with dialogues through beliefs conspiracy reducing Durably Rand, & Pennycook Costello, (2024)
https://doi.org/10.1126/science.adq1814

DOI Editor-in-Chief), Thorp, Holden H. ;2026 June 11 (posted Science Concern, of Expression Editorial science.aej2383/1126.10
https://www.science.org/doi/10.1126/science.aej2383