nature portfolio

Corresponding author(s): Aamir Sohail

Last updated by author(s): Oct 21, 2026

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	For human data, Rock-Paper-Scissors task was programmed in jsPsych (v6.3.1; de Leeuw, 2015), and hosted online on Cognition.run. The Inspection Game was presented using MATLAB 2012a (MathWorks) and the Cogent 2000 v125 toolbox (http://www.vislab.ucl.ac.uk/Cogent/). For large language model data, the two experiments were recoded as experiments using oTree (v5.11.1) running under Python (v3.11). API calls were routed through LiteLLM (v1.73.1).
Data analysis	Behavioural analyses were conducted in R (v4.4.1). Packages included tidyverse (v2.0.0), lme4 (v1.1-35.4), lmerTest (v3.1-3), emmeans (v2.0.3), effectsize (v1.0.0), rstatix (v0.7.3), PMCMRplus (v1.9.12) and MuMIn (v1.48.19). Hierarchical Bayesian modelling of the Inspection Game was carried out in JAGS (v4.3.2) via the rjags (v4-16) and runjags packages. Computational modelling of Rock-Paper-Scissors behavioural data was performed in MATLAB (R2022a), where models were fitted by maximum likelihood estimation with the MERLIN toolbox, and compared with random-effects Bayesian model selection using the VBA toolbox. Custom R and MATLAB scripts were written for data preprocessing, model fitting, statistical analysis and plotting. Computationally intensive analyses, including model fitting, were ran on the University of Birmingham's BlueBEAR high-performance computing cluster.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Data Availability

All data required to reproduce the results reported in this study are available in two Code Ocean capsules (Capsule A: https://codeocean.com/capsule/5768243/tree; Capsule B: https://codeocean.com/capsule/1348625/tree). These capsules contain:
- the trial-level behavioural data from large language model (LLM) agents and human participants in the Rock-Paper-Scissors and Inspection Game tasks;
- the comprehension-check data;
- the pre-computed computational model fits.

Human participant data were de-identified before release. On publication, all data, together with the processed inputs needed to refit all models, will be deposited in a permanent Zenodo archive under a CC BY 4.0 licence.

Code Availability

All analysis code is available in the same two containerized Code Ocean capsules, which reproduce the reported figures and statistics from the pre-computed model fits:
- Capsule A covers the R/JAGS behavioural and Inspection Game modeling analyses
- Capsule B covers the MATLAB Rock-Paper-Scissors modeling analyses

On publication, the complete analysis code, including the model-fitting pipelines, will be archived on Zenodo under the MIT licence. The code used to implement the tasks and collect data from LLM agents and human participants are available from the corresponding author on request.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Biological sex (male/female) was self-reported by human participants. The inspection game sample comprised 67 participants (25 males, 42 females). The Rock-Paper-Scissors sample comprised 185 participants (76 males, 106 females, 3 not reported). Sex was not a focus of the study, and analyses were not stratified by sex.
Reporting on race, ethnicity, or other socially relevant groupings	No race or ethnicity variables were collected, as they were not relevant to the study aims.
Population characteristics	Human participants were healthy young adults in both samples. For the inspection game (67 participants; mean age = 23.3 ± 4.21 years), individuals who studied psychology, had a history of neurological or psychiatric disorders, or had abnormal vision were excluded a priori. For the Rock-Paper-Scissors game (185 participants; mean age = 24.91 ± 3.57 years), subjects were all UK residents and native English speakers.
Recruitment	For the inspection game, human participants were recruited through Oxford University participant databases, social media, email lists and adverts in local newspapers. For the Rock-Paper-Scissors task, human participants were recruited from the general population via Prolific.
Ethics oversight	All participants provided informed consent. The inspection game study was approved by the Oxford University Medical Sciences Inter Divisional Research Ethics Committee and National Health Service Ethics. The Rock-Paper-Scissors study was approved by the University of Birmingham STEM Ethics Committee (ERN 22-1192,ERN 09-719F, ERN 2311).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Quantitative experimental study comparing humans and large language model (LLM) agents on two repeated-trial economic games (Inspection Game and Rock-Paper-Scissors). The LLM conditions manipulated prompting strategy (standard vs Social Chain-of-Thought), and the Rock-Paper-Scissors task manipulated opponent sophistication. Analyses combined behavioural statistics with computational modeling of trial-by-trial choices.
Research sample	Humans constituted two samples of healthy young adults (total N = 252). Large language models totaled 2,099 simulated participants. Models were selected for cost and feasibility, and on the basis of comprehension checks run across a wider set of candidate models.
Sampling strategy	Human subjects were convenience samples. The inspection game sample size was set by the available dataset, and the Rock-Paper-Scissors sample size set by recruitment for a larger parent study. LLM sample sizes were set by a priori power analyses (80% power, $\alpha = 0.05$, Bonferroni-corrected where applicable). For all analyses, both samples exceeded the required sample sizes for all planned tests.
Data collection	For the inspection game, human data were collected using a computer-based task programmed in MATLAB 2012a (MathWorks) with the Cogent 2000 v125 toolbox. Data were collected from human participants in the Rock-Paper-Scissors task which was programmed in jsPsych 6.3.1 and hosted on Cognition.run. In both tasks, participants believed they were playing against a real opponent but were playing against computer algorithms. LLM data were collected automatically through API calls using custom code integrating oTree 5.11.1, botex 0.2.0 and LiteLLM 1.73.1.
Timing	Human inspection game data were collected between September 2016 and November 2019. Human Rock-Paper-Scissors data were collected in May 2025. LLM data were collected between August 2025 and September 2025 for the Inspection Game, and between September 2025 and October 2025 for the Rock-Paper-Scissors task.
Data exclusions	Exclusion criteria were based on missing responses. In the Inspection Game, participants with more than 5 missing trials were excluded. 8 of 75 participants met this criterion, leaving 67 for both behavioural and modeling analyses. In the Rock-Paper-Scissors task, 13 participants with 5 or more missing trials were similarly excluded from behavioural analyses, leaving 172 participants. All 185 participants were retained in modeling analyses. For LLMs in the Inspection Game, a target sample of 50 agents per model and prompting condition was set, doubled to 100 for GPT-5. As 53 DeepSeek SCoT agents were collected, 3 were randomly removed to balance the groups. For Rock-Paper-Scissors modeling, individual 40-trial LLM runs were combined into 120-trial participants within each group to match the human participants.
Non-participation	No human partcipants dropped out after consenting to participate.
Randomization	Human participants were not allocated to experimental groups. In the Rock-Paper-Scissors task, the order of the three algorithmic opponents was randomised and counterbalanced. LLM instances were similarly assigned to a single prompting condition and, in the Rock-Paper-Scissors task, to a single opponent level, with equal numbers per condition where possible.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

Authentication

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.