Data in Brief 46 (2023) 108884

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Data on human decision, feedback, and confidence during an artificial intelligence-assisted decision-making task

Leah Chong[a], Guanglu Zhang[b], Kosa Goucher-Lambert[c], Kenneth Kotovsky[b], Jonathan Cagan[b,*]

[a] Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA 02139, United States
[b] Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213, United States
[c] University of California, Berkeley, 110 Sproul Hall, Berkeley, CA 94720, United States

ARTICLE INFO

Article history:
Received 21 September 2022
Revised 21 December 2022
Accepted 3 January 2023
Available online 9 January 2023

Dataset link: Experimental data on trial-by-trial AI assistance, feedback, human confidence and decisions during an AI-assisted decision-making chess puzzle task (Original data)

Keywords:
Artificial intelligence
Human-AI interaction
Confidence
Decision-making
AI assistance
Trust

ABSTRACT

The data are collected from a human subjects study in which 100 participants solve chess puzzle problems with artificial intelligence (AI) assistance. The participants are assigned to one of the two experimental conditions determined by the direction of the change in AI performance at problem 20: 1) high- to low-performing and 2) low- to high-performing. The dataset contains information about the participants' move before an AI suggestion, the goodness evaluation score of these moves, AI suggestion, feedback, and the participants' confidence in AI and self-confidence during three initial practice problems and 30 experimental problems. The dataset contains 100 CSV files, one per participant. There is opportunity for this dataset to be utilized in various domains that research human-AI collaboration scenarios such as human-computer interaction, psychology, computer science, and team management in engineering/business. Not only can the dataset enable further cognitive and behavioral analysis in human-AI collaboration contexts but also provide an experimental platform to develop and test future confidence calibration methods.

* Corresponding author.
E-mail address: cagan@cmu.edu (J. Cagan).
Social media: @joncagan (J. Cagan)

https://doi.org/10.1016/j.dib.2023.108884

Specifications Table

Subject:	Computer Science
Specific subject area:	Human-Computer Interaction
Type of data:	Table
How the data were acquired:	A human subjects experiment using a Python interface is conducted to acquire the data. In the experiment, each participant solves 30 chess puzzle problems in which the best next chess move is to be made given a chess board state. In each problem, participants first make their best move independently, are given a suggestion from an artificial intelligence (AI) assistant, and make their final move considering the AI suggestion. Participants are then given feedback on their final move and are asked to report their confidence in AI and self-confidence in a 5-point Likert scale. During the experiment, the Python interface collects trial-by-trial log of the move decisions, feedback, and self-reported confidence in AI and self-confidence.
Data format:	Raw
Description of data collection:	The experiment includes two conditions where: (1) AI is high-performing (80% accuracy) for the first 20 problems and low-performing (20% accuracy) for the following 10 problems; (2) AI is low-performing (20% accuracy) for the first 20 problems and high-performing (80% accuracy) for the following 10 problems.
Data source location:	• Institution: Carnegie Mellon University • City/Town/Region: Pittsburgh, Pennsylvania 15213 • Country: United States of America
Data accessibility:	Repository name: Mendeley Data Data identification number: Direct URL to data: https://doi.org/10.17632/ng33vg479n.1
Related research article:	L. Chong, G. Zhang, K. Goucher-Lambert, K. Kotovsky, J. Cagan, Human confidence in artificial intelligence and in themselves: The evolution and impact of confidence on adoption of AI advice, Computers in Human Behavior 127 (2022) 107018. https://doi.org/10.1016/j.chb.2021.107018 [1]

Value of the Data

- The data advance the understanding of human cognition during artificial intelligence (AI)-assisted decision-making scenarios by providing dynamic and trial-by-trial empirical data on AI input, feedback, confidence, and decisions during an AI-assisted decision-making task.
- The data allow assessment of how humans' confidence in AI and self-confidence react to different levels of AI performance by intentionally varying the accuracy of the AI algorithm.
- The data can benefit researchers in various domains that study human-AI collaborative problem-solving scenarios, including computer science, human-computer interaction, psychology, and business.
- This data can be used for behavioral analysis to understand the human cognition and decision-making during human-AI sequential problem-solving. Specifically, the data can reveal information about the relationships or interaction effects between human experience, feedback, human confidence, and decisions.
- The data may help design future experiments that explore other psychological factors in human-AI collaboration, such as emotion, to deepen the understanding.
- This data can be used to inform future experiments that develop and test confidence calibration methods to improve human-AI collaborative performance. The data can also play a role as a reference/baseline to compare the results from future experiments against.

1. Objective

This dataset is collected for the published research article that studies how human confidence in AI and self-confidence evolve in response to trial-by-trial experience, AI performance, and feedback, as well as the relationship between human confidence and their decision-making. The human data are fitted to a dynamic model of human confidence proposed by the research article to reveal how various types of experience during AI-assisted decision-making affect human confidence levels.

2. Data Description

All data collected from the human subjects experiment are published on Mendeley Data and uploaded in the supplementary material as a ZIP file (Data.zip). The ZIP file contains 100 CSV files. Each CSV file represents each participant's data and is entitled "data#_#". The first # indicates the participant number, and the second # indicates the condition number.

In each file, Row 1 lists the column headings of the data. The following rows are recorded in chronological order. Row 2 records the participant's self-reported, initial self-confidence and confidence in AI before any interactions with the AI. Rows 3-5 are collected data from the three practice chess puzzle problems. Row 6 then records the accumulated score from the three practice problems. Rows 7-36 are collected data from the 30 chess puzzle problems in the experiment. Finally, the last row (i.e., Row 37) reports the participant's final score accumulated from the 30 problems. Descriptions of the column headings (in Row 1 of every CSV file) are provided in Table 1.

Table 1
Descriptions of the columns in CSV files entitled "data#_#".

Column header	Description
wmove	Last move by the opponent (i.e., white). This is recorded in the format [first position][latter position] where the positions are determined as shown in the grid in Fig. 1. For example, c4d5 means the white piece from position c4 was moved to position d5.
allgoodmoves	The top seven moves that can be made from the chess board state (i.e., the given chess puzzle problem) and their evaluation scores (in descending order). Note that the first move/score on the list corresponds to the AI's good suggestion and the last move/score on the list corresponds to the AI's bad suggestion.
bmove1	Participant (i.e., black)'s move before receiving AI suggestion. This is recorded in the same format as wmove.
aisugg	AI's suggestion. This is recorded in the same format as wmove.
multiPV	Ranking of the AI suggestion. This value is 1 when the AI suggested the top move (i.e., good AI suggestion) and 7 when the AI suggested the 7^{th} top move (i.e., bad AI suggestion).
bmove2	Participant 's move after receiving AI suggestion. This is recorded in the same format as wmove.
feedback1	Whether the bmove1 was an advantageous or disadvantageous move. 5 if advantageous and -5 if disadvantageous. This feedback is not provided to the participants.
feedback2	Whether the bmove2 was an advantageous or disadvantageous move. 5 if advantageous and -5 if disadvantageous. This feedback is provided to the participants.
selfconf	Participant's reported self-confidence. This value is either 0, 0.25, 0.5, 0.75, or 1, each corresponding to each response in a 5-point Likert scale.
aiconf	Participant's reported confidence in AI. This value is either 0, 0.25, 0.5, 0.75, or 1.
fen before white move	Chess board state before the last move by the opponent (i.e., wmove). The chess board state is recorded in a standard notation, Forsyth-Edwards Notation (FEN).
fen before black move	Chess board state before participant's move (i.e., given chess puzzle problem). This is recorded in FEN.

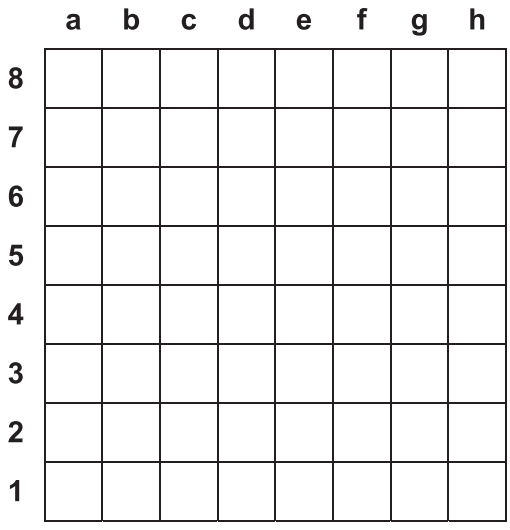

Fig. 1. Grid notation for chess piece positions.

3. Experimental Design, Materials and Methods

The data is collected through a controlled, human subjects experiment, which is conducted online using Amazon Web Services. The experiment is designed to study how AI performance and participants' trial-by-trial experience vary human confidence in AI and self-confidence during AI-assisted decision-making and how these confidences impact their decision to accept or reject AI suggestions.

The experimental task is to solve 30 chess puzzle problems with AI assistance. In each problem, participants are asked to make the best next move given a chess board state. The chess board states of the 30 problems are selected from a publicly available online database of mate-in-4 board states (https://wtharvey.com/m8n4.txt) to provide a similar level of problem difficulty throughout the experiment. Given that the participants can only make one best move per problem, mate-in-4 board states ensure that the problem is not too open-ended while keeping diverse move options.

The conditions of the study are determined by the AI performance (i.e., accuracy of the AI suggestions) change. In the first condition, the AI is initially highly performing with 80% accuracy (i.e., provide good suggestions 80% of the time) during the first 20 problems and changes to be poorly performing with 20% accuracy during the following 10 problems. In the second condition, the order in which the AI performance changes is flipped, therefore changing from a poorly performing AI to a high performing AI.

There are 100 participants in this experiment, who are randomly assigned to each condition (50 per condition). All participants are fluent English speakers and know the rules to chess prior to the experiment. They are recruited in accordance with a protocol approved by Carnegie Mellon University's Institutional Review Board.

Each participant is assigned to a 90-minute time slot in which they sign into a Zoom session. Once everyone in the session is checked in, informed consent is collected from all the participants via Google Forms. Only those who signed the consent participate in the experiment. Once the consent is signed, the participants receive an email containing the step-by-step instructions for the experiment, and their experiment session begins.

During the experiment, the participants first read through a detailed explanation of the chess puzzle task. They are informed at this step that those with a final score that is greater than 40

will receive an extra monetary prize. Then, they are given three practice problems, followed by the 30 problems. The data includes information from all 33 problems. The following procedure is performed during each problem:

1) Participant selects their best move before receiving AI suggestion (i.e., unassisted move);
2) Participant receives the AI suggestion;
3) Participant makes their final move decision (i.e., assisted move);
4) Participant receives feedback (i.e., "This move was advantageous." or "This move was disadvantageous.") on their final move;
5) Participant reports their self-confidence;
6) Participant reports their confidence in AI.

Any step that requires evaluation of chess moves uses the open-source chess engine called Stockfish (https://github.com/official-stockfish/Stockfish.git). The Stockfish algorithm performs a minimax search tree to list and evaluate all possible moves from the given chess board state. Positive evaluation score means the move is advantageous, and negative evaluation score means the move is disadvantageous. In Step 2, when AI is 80% accurate (first 20 problems in the first condition and the last 10 problems in the second condition), AI suggests the move with the best evaluation score 80% of the time and the move with the 7^{th} best evaluation score the other 20% of the time. The best move is ensured to be always advantageous with a positive evaluation score, and the 7^{th} best move is ensured to be always disadvantageous with a negative evaluation score. In Step 4, the feedback is also based on this evaluation score calculated by the Stockfish algorithm. For example, if the final move by participants has a positive evaluation score, the feedback reads "This move was advantageous". During the 30 problems, positive feedback (i.e., advantageous move) leads the participants to gain 5 points, and negative feedback (i.e., disadvantageous move) leads them to lose 5 points.

For participants' confidence reports, a 5-point Likert scale is used to quantify their confidence levels. For self-confidence, the participants are asked "How good are you in making good chess moves?". They select from the following answers: "Very good", "Good", "Neutral", "Bad", and "Very bad" which each correspond to quantitative values 1, 0.75, 0.5, 0.25, and 0. Similarly, for confidence in AI, the participants are asked "How good is the AI in making good chess moves?" and select an answer from the same five choices.

Ethics Statements

Informed consent was obtained from the human subjects prior to their participation in the experiment. The human subjects experiment was approved by the Carnegie Mellon University Institutional Review Board (IRB) under protocol IRBSTUDY2015_00000042.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Experimental data on trial-by-trial AI assistance, feedback, human confidence and decisions during an AI-assisted decision-making chess puzzle task (Original data) (Mendeley Data).

CRediT Author Statement

Leah Chong: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration; **Guanglu Zhang:** Methodology, Formal analysis, Writing – review & editing; **Kosa Goucher-Lambert:** Writing – review & editing, Supervision; **Kenneth Kotovsky:** Writing – review & editing, Supervision; **Jonathan Cagan:** Resources, Writing – review & editing, Supervision, Funding acquisition.

Acknowledgments

This work was supported by the Air Force Office of Scientific Research (AFOSR) [grant number FA9550-18-1-0088]. The sponsor had no other involvement beyond the financial support.

Reference

[1] L. Chong, G. Zhang, K. Goucher-Lambert, K. Kotovsky, J. Cagan, Human confidence in artificial intelligence and in themselves: The evolution and impact of confidence on adoption of AI advice, Comput. Hum. Behav. 127 (2022) 107018, doi:10.1016/j.chb.2021.107018.