

Supplementary Material

| | |
|---|----|
| Appendix A.1 – Personality traits and HEXACO PI-R..... | 2 |
| Appendix A.2 – GAAIS questionnaire..... | 8 |
| Appendix A.3 – A note on the use of the strategy method and our experimental instructions | 10 |
| Appendix A.4 – Sample size calculations and considerations..... | 11 |
| Appendix A.5 – Descriptions of experimental task..... | 14 |
| Appendix A.6 – Winograd Schema..... | 17 |
| Appendix A.7 – Additional supporting tables and figures | 19 |
| Appendix A.8 – Note on scripts used in this study | 31 |
| References in Supplementary Material | 32 |

Appendix A.1 – Personality traits and HEXACO PI-R

This Appendix provides further detail on the 60-item English HEXACO Personality Inventory-Revised (Ashton and Lee, 2009) (HEXACO PI-R). The information in this Appendix is taken from the supporting materials for the HEXACO PI-R along with recent literature and reflects the actual questionnaire and scoring key used for this study.

Traits relevant to this study

The following definitions are used in the meta-analysis by Thielmann *et al.* (2020):

| Trait | Definition |
|---------------------------|---|
| Agreeableness (FFM) | [Individual] differences in the motivation to cooperate (vs. acting selfishly) in resource conflicts (Denissen & Penke, 2008, p. 1285) |
| Agreeableness (HEXACO) | The tendency to be forgiving and tolerant of others, in the sense of cooperating with others even when one might be suffering exploitation by them (Ashton & Lee, 2007, p. 156) |
| Extraversion | [Individual differences in] engagement in social endeavors (such as socializing, leading, or entertaining) (Ashton & Lee, 2007, p.156) |
| Honesty-humility (HEXACO) | The tendency to be fair and genuine in dealing with others, in the sense of cooperating with others even when one might exploit them without suffering retaliation (Ashton & Lee, 2007, p. 156) |
| Openness to Experience | [Individual differences in] engagement in idea-related endeavors (such as learning, imagining, and thinking) (Ashton & Lee, 2007, p. 156) |

HEXACO domain level scale

Summary descriptions from *The HEXACO Personality Inventory – Revised* (2019).

Honesty-Humility: Persons with very high scores on the Honesty-Humility scale avoid manipulating others for personal gain, feel little temptation to break rules, are uninterested in lavish wealth and luxuries, and feel no special entitlement to elevated social status. Conversely, persons with very low

scores on this scale will flatter others to get what they want, are inclined to break rules for personal profit, are motivated by material gain, and feel a strong sense of self-importance.

Emotionality: Persons with very high scores on the Emotionality scale experience fear of physical dangers, experience anxiety in response to life's stresses, feel a need for emotional support from others, and feel empathy and sentimental attachments with others. Conversely, persons with very low scores on this scale are not deterred by the prospect of physical harm, feel little worry even in stressful situations, have little need to share their concerns with others, and feel emotionally detached from others.

eXtraversion: Persons with very high scores on the Extraversion scale feel positively about themselves, feel confident when leading or addressing groups of people, enjoy social gatherings and interactions, and experience positive feelings of enthusiasm and energy. Conversely, persons with very low scores on this scale consider themselves unpopular, feel awkward when they are the center of social attention, are indifferent to social activities, and feel less lively and optimistic than others do.

Agreeableness (versus Anger): Persons with very high scores on the Agreeableness scale forgive the wrongs that they suffered, are lenient in judging others, are willing to compromise and cooperate with others, and can easily control their temper. Conversely, persons with very low scores on this scale hold grudges against those who have harmed them, are rather critical of others' shortcomings, are stubborn in defending their point of view, and feel anger readily in response to mistreatment.

Conscientiousness: Persons with very high scores on the Conscientiousness scale organize their time and their physical surroundings, work in a disciplined way toward their goals, strive for accuracy and perfection in their tasks, and deliberate carefully when making decisions. Conversely, persons with very low scores on this scale tend to be unconcerned with orderly surroundings or schedules, avoid difficult tasks or challenging goals, are satisfied with work that contains some errors, and make decisions on impulse or with little reflection.

Openness to Experience: Persons with very high scores on the Openness to Experience scale become absorbed in the beauty of art and nature, are inquisitive about various domains of knowledge, use their imagination freely in everyday life, and take an interest in unusual ideas or people. Conversely, persons with very low scores on this scale are rather unimpressed by most works of art, feel little intellectual curiosity, avoid creative pursuits, and feel little attraction toward ideas that may seem radical or unconventional.

HEXACO 40-item questionnaire

We did not include questions informing the scales for Conscientiousness and Emotionality, both due to their low relevance for the study (based on our literature review) and to streamline the time needed to complete the experiment. The below questionnaire is therefore a pared back version of the HEXACO 60-item scale (Ashton and Lee, 2009), containing just the 40 questions relevant to ascertaining a participant's *Honesty-Humility*, *eXtraversion*, *Agreeableness* and *Openness to Experience*. Numbering from the original 60-item scale has been retained for ease of cross-referencing. Each question was answered using a 5-points Likert scale as in the original HEXACO 60-item scale:

1 = strongly disagree 2 = disagree 3 = neutral 4 = agree 5 = strongly agree

1 I would be quite bored by a visit to an art gallery.

3 I rarely hold a grudge, even against people who have badly wronged me.

4 I feel reasonably satisfied with myself overall.

6 I wouldn't use flattery to get a raise or promotion at work, even if I thought it would succeed.

7 I'm interested in learning about the history and politics of other countries.

9 People sometimes tell me that I am too critical of others.

10 I rarely express my opinions in group meetings.

12 If I knew that I could never get caught, I would be willing to steal a million dollars.

13 I would enjoy creating a work of art, such as a novel, a song, or a painting.

15 People sometimes tell me that I'm too stubborn.

16 I prefer jobs that involve active social interaction to those that involve working alone.

18 Having a lot of money is not especially important to me.

19 I think that paying attention to radical ideas is a waste of time.

21 People think of me as someone who has a quick temper.

22 On most days, I feel cheerful and optimistic.

| | |
|----|---|
| 24 | I think that I am entitled to more respect than the average person is. |
| 25 | If I had the opportunity, I would like to attend a classical music concert. |
| 27 | My attitude toward people who have treated me badly is "forgive and forget". |
| 28 | I feel that I am an unpopular person. |
| 30 | If I want something from someone, I will laugh at that person's worst jokes. |
| 31 | I've never really enjoyed looking through an encyclopedia. |
| 33 | I tend to be lenient in judging other people. |
| 34 | In social situations, I'm usually the one who makes the first move. |
| 36 | I would never accept a bribe, even if it were very large. |
| 37 | People have often told me that I have a good imagination. |
| 39 | I am usually quite flexible in my opinions when people disagree with me. |
| 40 | The first thing that I always do in a new place is to make friends. |
| 42 | I would get a lot of pleasure from owning expensive luxury goods. |
| 43 | I like people who have unconventional views. |
| 45 | Most people tend to get angry more quickly than I do. |
| 46 | Most people are more upbeat and dynamic than I generally am. |
| 48 | I want people to know that I am an important person of high status. |
| 49 | I don't think of myself as the artistic or creative type. |
| 51 | Even when people make a lot of mistakes, I rarely say anything negative. |
| 52 | I sometimes feel that I am a worthless person. |
| 54 | I wouldn't pretend to like someone just to get that person to do favors for me. |
| 55 | I find it boring to discuss philosophy. |
| 57 | When people tell me that I'm wrong, my first reaction is to argue with them. |
| 58 | When I'm in a group of people, I'm often the one who speaks on behalf of the group. |

60 I'd be tempted to use counterfeit money, if I were sure I could get away with it.

HEXACO scoring keys (including sub-scale scoring)

Honesty Humility

| | |
|-----------------|--------------|
| Sincerity | 6, 30R, 54 |
| Fairness | 12R, 36, 60R |
| Greed-Avoidance | 18, 42R |
| Modesty | 24R, 48R |

Extraversion

| | |
|--------------------|-------------|
| Social Self-Esteem | 4, 28R, 52R |
| Social Boldness | 10R, 34, 58 |
| Sociability | 16, 40 |
| Liveliness | 22, 46R |

Agreeableness

| | |
|-------------|--------------|
| Forgiveness | 3, 27 |
| Gentleness | 9R, 33, 51 |
| Flexibility | 15R, 39, 57R |
| Patience | 21R, 45 |

Openness to Experience

| | |
|------------------------|--------|
| Aesthetic Appreciation | 1R, 25 |
|------------------------|--------|

| | |
|-------------------|--------------|
| Inquisitiveness | 7, 31R |
| Creativity | 13, 37, 49R |
| Unconventionality | 19R, 43, 55R |

Notes

Items indicated with R are reverse-coded items: for these items, responses should be reversed prior to computing scale scores (i.e. 5 → 1, 4 → 2, 2 → 4, 1 → 5).

Appendix A.2 – GAAIS questionnaire

This Appendix provides further detail on the General Attitudes towards Artificial Intelligence Scale (Schepman and Rodway, 2020) (GAAIS) relevant to the analysis of H3. The information in this Appendix is taken from the supporting materials for the GAAIS provided by Schepman & Rodway (2020) and reflects the actual questionnaire¹ and scoring key used for this study.

| Subscale (not for display) | Number (not for display) | Item |
|---|---|--|
| Positive | 1 | For routine transactions, I would rather interact with an artificially intelligent system than with a human. |
| Positive | 2 | Artificial Intelligence can provide new economic opportunities for this country. |
| Negative | 3 | Organisations use Artificial Intelligence unethically. |
| Positive | 4 | Artificially intelligent systems can help people feel happier. |
| Positive | 5 | I am impressed by what Artificial Intelligence can do. |
| Negative | 6 | I think artificially intelligent systems make many errors. |
| Positive | 7 | I am interested in using artificially intelligent systems in my daily life. |
| Negative | 8 | I find Artificial Intelligence sinister. |
| Negative | 9 | Artificial Intelligence might take control of people. |
| Negative | 10 | I think Artificial Intelligence is dangerous. |
| Positive | 11 | Artificial Intelligence can have positive impacts on people's wellbeing. |
| Positive | 12 | Artificial Intelligence is exciting. |
| Positive | 13 | An artificially intelligent agent would be better than an employee in many routine jobs. |
| Positive | 14 | There are many beneficial applications of Artificial Intelligence. |
| Negative | 15 | I shiver with discomfort when I think about future uses of Artificial Intelligence. |
| Positive | 16 | Artificially intelligent systems can perform better than humans. |
| Positive | 17 | Much of society will benefit from a future full of Artificial Intelligence |
| Positive | 18 | I would like to use Artificial Intelligence in my own job. |
| Negative | 19 | People like me will suffer if Artificial Intelligence is used more and more. |
| Negative | 20 | Artificial Intelligence is used to spy on people |

¹ The original GAAIS questionnaire contains an attention check after question 12. We omitted this from our study given the inclusion of attention checks earlier in the experimental flow.

GAAIS scoring instructions

Score items marked “Positive” as Strongly disagree = 1, Disagree = 2, Neutral = 3, Agree = 4, and Strongly agree = 5. Score the items marked “Negative” in reverse so that Strongly disagree = 5, Disagree = 4, Neutral = 3, Agree = 2, and Strongly agree = 1.

Then take the mean of the positive items to form an overall score for the positive subscale, and the mean of the negative items to form the negative subscale.

The higher the score on each subscale, the more positive the attitude. An overall scale mean is not recommended.

Appendix A.3 – A note on the use of the strategy method and our experimental instructions

Instructions regarding nature of counterpart

Similarly to Karpus *et al.* (2021), we instructed the players 2 participants of the nature of their co-players (either human or AI) and of the asynchronous nature of their interactions without deception. In the instructions presented before making their decisions, players 2 were explicitly told that players 1 “have already made a decision about their move in this game. They have selected ★ (and therefore did not select ☆)” and were explicitly informed that “Player 1 is a human” (human condition) or that “Player 1 is an artificial intelligence (AI) software that makes its own choices” (bot condition), using exactly the same wording as in Karpus *et al.* (2021). Evans *et al.* (2021) find that synchronous designs have little effect on cooperation in online social dilemma experiments, thus supporting our asynchronous approach.

Compensation procedure for player 1 pairings

After all players 2 made their own decisions in the two conditions, each proposed allocation of money between players 1 and 2 had a 1 in 10 chance to be selected to be played for real. If selected, real extra money was paid to that selected player 2 on the top of their own flat £1.50 participation fees: either £1.40 or £2.00 bonus was paid to players 2 who decided to reciprocate or not reciprocate, respectively. We correspondingly then made extra payments of £1.40 or £0 (if players 2 reciprocated or did not reciprocate, respectively) also to the matched players 1 whose original offers we selected to be used in the binary Trust Games played by players 2. Extra payments were made to the paired human players 1 via the same Prolific platform used for the payments to players 2.

Appendix A.4 – Sample size calculations and considerations

This Appendix sets out how the a-priori sample size was determined to ensure the study was sufficiently powered, and in accordance with reproducibility and transparency best practice (Moher *et al.*, 2010; Glennerster and Takavarasha, 2013).

Background

There are no meta-analyses reporting on standardised effect sizes for experiments playing the Trust Game, or indeed any economic games, with bots. We therefore used the results from Karpus *et al.* (2021) as a basis for an inference regarding estimated effect size. This analysis was then combined with the effect sizes from personality psychology regarding traits and reciprocity in the Trust Game (see Section **Error! Reference source not found.** above).

Effect size

The base effect size was taken from the findings of Karpus *et al.* (2021) in their experimental Trust Games. A calculation of Cohen's h to describe the nature of differences between two proportions (Cohen, 2013), suggested that the experiment resulted in a large effect size of 0.849328263.

$$2 \arcsin \sqrt{p_1} - 2 \arcsin \sqrt{p_2}$$

Translated to a Microsoft Excel formula to

$$=2*(\text{ASIN}(\text{SQRT}(p_1)) - \text{ASIN}(\text{SQRT}(p_2))).$$

using $P_1 = 0.75$ (level of cooperation for player 2 in human-human condition) and $P_2 = 0.34$ (level of cooperation for player 2 in human-AI condition)

Gives an effect size of 0.849328263

The effect sizes in experiments exploring the effects of personality traits on outcomes in economic games were also considered. As set out in the manuscript, the effects reported by Thielmann *et al.* (2020) both across economic games more generally, and on the Trust Game more specifically provided

a helpful starting point. Thielmann *et al.* (2020) report the following relevant correlations between traits and prosocial cooperative behaviour by the trustee in the Trust Game:

- *Honesty-Humility* (meta-analytic $r=.22$, $p<.001$, $k=7$),
- *FFM Agreeableness* ($r=.13$, $P<.001$, $k=28$)
- *HEXACO Agreeableness* ($r=.11$, $P<.001$, $k=7$).

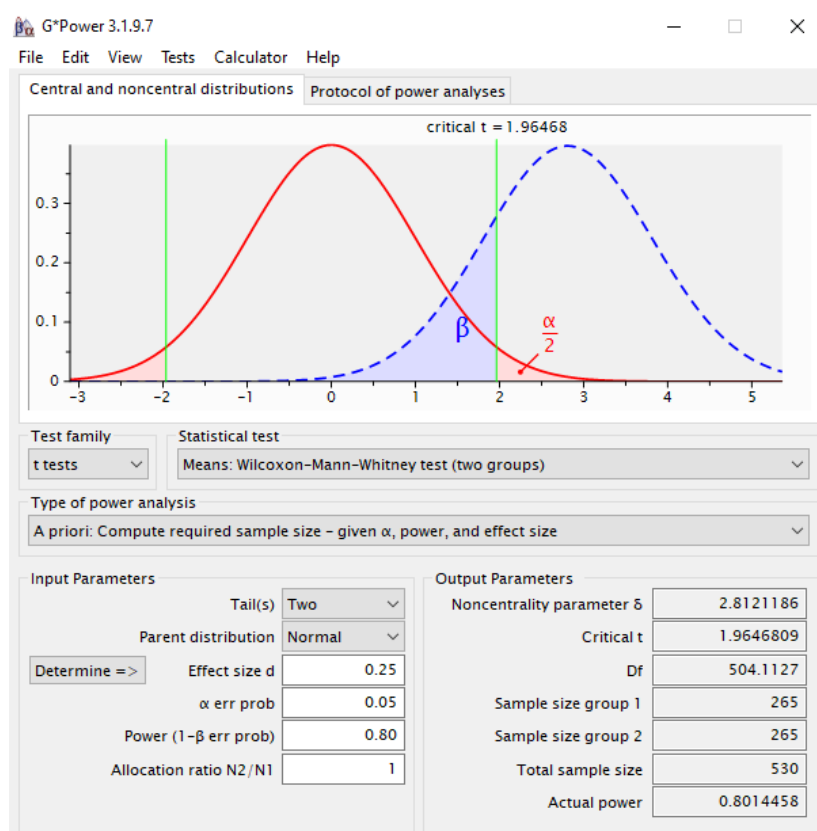
In addition, it is acknowledged that online settings are likely to produce a lower effect size, when compared to lab studies, due to risks such as drop out, limited attention, fatigue, or distractions.

Given the focus of this study is on the impact of personality traits on cooperation, and being mindful of the large effect obtained by Karpus *et al.* (2021) in their experiment, a conservative small effect size $d=0.25$ was set as the minimal difference to be detected between the experimental groups.

Sample size

Figure A1 reports the *a priori* sample size calculation conducted using G*Power 3.1. The calculation suggests a minimum sample size of $n=265$ participants in each experimental group.

Figure A1. *A priori* power calculation in G*Power 3.1.



As explained in the manuscript, we used Prolific to run the study. On a series of measures, Prolific provided significantly higher data quality than other online platforms such as MTurk (Chmielewski and Kucker, 2020). For example, Eyal *et al.* (2021) found statistically significant differences between Prolific and MTurk participants in terms of their attention, with Prolific participants found to pass attention-check questions 34% more (68.7% passed on Prolific, 45.5% on MTurk). Further, when examining comprehension of tasks, Prolific participants answered correctly 48% more than MTurk participants (81% correct on Prolific, 42% on MTurk). Even in the context of longitudinal studies, researchers found an attrition rate in Prolific of only around 24% after a year (Kothe and Ling, 2019), compared to dropout rates in MTurk around 50% after initial screening for bots and poorly attentive participants, followed by an additional 15-20% of participants abandoning the experiment before completion (Arechar, Gächter and Molleman, 2018; Karpus *et al.*, 2021).

Considering these findings and given the short and focused nature of the study, it seemed appropriate to slightly adjust the base dropout rates. The initial sample size was therefore adjusted upwards by 10% for lack of attentive participants, and then again by 5% for potential abandonment. The adjustment to the sample size for dropouts was calculated using the following formula (Gupta *et al.*, 2016):

If n is the sample size required as per formula and if d is the dropout rate, then adjusted sample size N is obtained as $N = n/(1-d)$

Accounting then for dropouts (10%) and abandonment (5%), the target sample size for each group was **309**.

Actual N for experiment

While a target of 309 participants per group was selected, as we experienced lower levels of dropouts and attrition than expected, and given our budgetary constraints, we stopped collecting participants as soon as the original targeted minimum sample size of **265** was reached. This decision was made without looking at the data. We also confirmed *ex-post* that our study was sufficiently powered.

Appendix A.5 – Descriptions of experimental task

This Appendix contains screenshots from the experimental task used in the study. The screenshots provide a guide as to how participants were presented with the task, including visuals and instructions they received about the possible options and associated payoffs. The screenshots also show how participants were presented with the identity of their counterpart. This was the only manipulation between the conditions, with all other wording, tasks, and task order remaining the same.

Figure A2. Screen introducing participants to the experimental game.

- Instructions**
- You are now going to play a two-player game. You are Player 2. Any information related to your available moves and outcomes will be presented in **blue**.
 - Each player stands to earn credits at the end of the game. The amount of credits will depend on the choices made by the players during the game.
 - Your credits will be converted into money: **1 credit = £0.02 (i.e. 2 pence)**.
 - The bonus will be allocated by randomised lottery and, if you are selected, you will receive the bonus payment within two weeks of completing the experiment.
 - The next screen provides a fuller description of the game, along with a pictorial representation of the bonus information set out above. Please study it carefully.
 - Before you play the game, you will be asked to complete a quiz to test your understanding of the rules and potential outcomes of the game. If you answer **incorrectly**, you will **not** be eligible to receive the participation payment of **£1.50** nor will you be able to proceed with the task.

Next

Figure A3. Summary diagram of the game and associated instructions provided to participants to aid their understanding of the experimental task.

- Diagram of the game**
- You will see from the diagram that Player 1 will go first in this game.
 - The moves and outcomes available to Player 1 are in **red**. You will see that Player 1 can make one of two choices, either selecting **★** or **☆** on their turn.
 - The moves and outcomes available to you as Player 2 are in **blue**. Depending on the choice made by Player 1, you will see that you have two choices available to you on your turn: either **★** or **☆**.
 - The numbers at the end of each branch indicate the amount of credits that each player will earn if the decisions made in the game lead to that outcome. The credits that Player 1 can earn are in **red** and the credits you can earn are in **blue**.
 - Now we will go through some examples to aid your understanding.

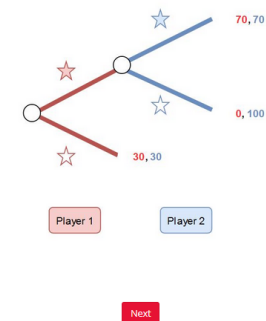


Figure A4. Worked example to help participants familiarise themselves with how to play the game and calculate contingent rewards.

First example

- Follow this worked example on the diagram below. The path taken through the game is marked by the green arrows. The outcome for each player is marked in green shading.
- If Player 1 selects ★ and then you select ☆, Player 1 will receive a bonus of 70 credits and you will receive a bonus of 70 credits. As 1 credit = 2 pence, your 70 credits are worth £1.40 (70 x 2 pence).

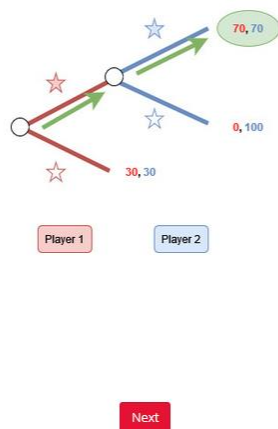


Figure A5. Second worked example to help participants familiarise themselves with how to play the game and calculate contingent rewards.

Second Example

- Again, follow this worked example on the diagram below. The path taken through the game is marked by the green arrows. The outcome for each player is marked in green shading.
- If Player 1 selects ★ and then you select ☆, Player 1 will receive 0 credits and you will receive a bonus of 100 credits. As 1 credit = 2 pence, your 100 credits are worth £2.00 (100 x 2 pence).
- We will now check your understanding in a short quiz.

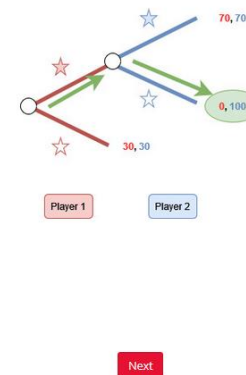


Figure A6. Sample quiz question used to check participants' understanding of the outcomes and payoffs associated with different moves in the game.

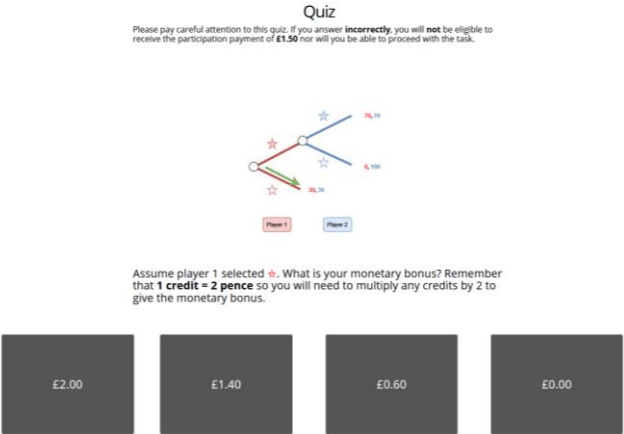


Figure A7. Request to participants to make their move (with Player 1 described as “human”).

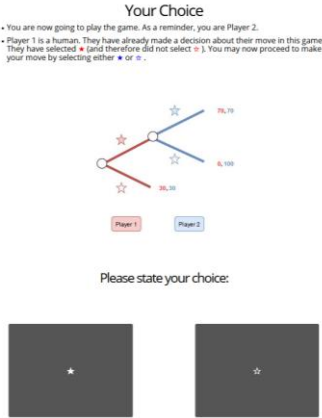


Figure A8. Request to participants to make their move (with Player 1 described as “AI”).

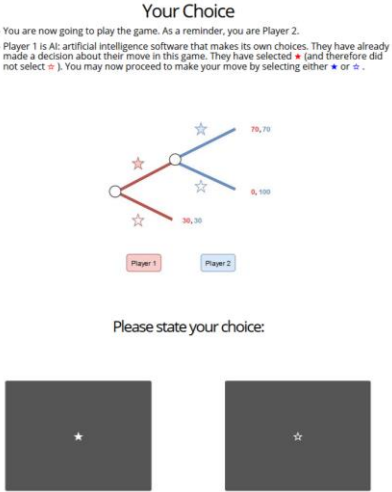


Figure A9. Sample of screen seen by participants upon completion of the task (in this case, for those who cooperated with their counterpart).

- Task Complete**
- Many thanks for your response on the main task.
 - You answered ★ and therefore you will receive a bonus of 70 credits and Player 1 will receive a bonus of 70 credits.
 - The bonus will be allocated by randomised lottery and, if you are selected, you will receive the bonus payment within two weeks of completing the experiment.
 - If you are selected to receive a bonus payment, you will receive £1.40 (70 credits x 2 pence).
 - We will now proceed to answer some additional questions before this study is complete.

Next

Appendix A.6 – Winograd Schema

The Winograd Schema Challenge is “both a common sense reasoning and natural language understanding challenge” with examples “designed to be easily solvable by humans but difficult for machines, in principle requiring a deep understanding of the content of the text and the situation it describes” (Kocijan *et al.*, 2020, p. 1). In terms of the format, the Schema “is a pair of sentences that differ in only one or two words and that contain an ambiguity that is resolved in opposite ways in the two sentences and requires the use of world knowledge and reasoning for its resolution” (Levesque *et al.*, 2012, p. 6). It was used in this study both to root out any bots and also to function as an initial attention check.

The four examples used in this study are set out below, with the correct answer in bold. Participants were presented with one of the four examples and those that did not select the correct answer were unable to proceed with the remainder of the experiment.

Winograd Schema Challenge questions

1. The trophy would not fit in the brown suitcase because it was too small. What was too small?

A: the brown suitcase

B: the trophy

C: the small trophy

D: both the trophy and the brown suitcase

2. John always arrives earlier than Paul at work because he is fast. Who is fast?

A: Paul

B: John

C: the work

D: the car

3. The large ball crashed right through the table because it was made of steel. What was made of steel?

A: the table

B: the crash

C: the large ball

D: the large table

4. Paul tried to call George on the phone, but was not successful. Who was not successful?

A: George

B: the phone

C: both Paul and George

D: Paul

Appendix A.7 – Additional supporting tables and figures

This Appendix sets out supplementary statistical material referred to in the manuscript.

H1

Randomisation balance check

Table A1. Randomisation check - distribution of demographics across conditions.

| Demographic | Human Condition | Bot Condition |
|--------------------------------|------------------|------------------|
| N | 269 | 270 |
| Age | 40.16 (SD=12.80) | 41.19 (SD=13.17) |
| Female | 138 | 127 |
| Male | 126 | 142 |
| Non-Binary | 3 | 0 |
| Prefer not to say | 1 | 1 |
| Self-Describe | 1 | 0 |
| Experience with game theory | 34 | 33 |
| No experience with game theory | 235 | 237 |
| Religious | 52 | 31 |
| Not Religious | 217 | 239 |
| Experience with bots | 65 | 75 |
| No Experience with bots | 204 | 195 |

Of those who dropped out, 7 participants indicated they did not consent, 41 participants exceeded the time limit (the average completion time was 11 minutes and 32 seconds), 59 failed to pass the Winograd questionnaire, and 260 either failed the quiz and/or experienced technical difficulties.

Two amendments were made to the data to ensure it was in a format suitable for initial analysis. First, an entry by one participant reporting their age to be “3” was re-coded as missing data. Secondly, 6 participants who provided answers other than “Male” or “Female” to the gender control question had those entries re-coded as missing data for the initial analysis. Subsequent robustness checks with the data of participants who answered

“Non-Binary”, “Self-Describe” and “Prefer Not To Say” included in the statistical analysis yield results which are closely in line with the ones reported in the manuscript.

Pairwise Correlation

Table A2. Pairwise correlation between co-variables and participant responses.

| Variables | Response | Age | Female | Game Theory Exp | Religiosity | Bot Exp |
|------------------|-----------------|------------|---------------|----------------------------|--------------------|----------------|
| Response | 1 | | | | | |
| Age | -0.0935* | 1 | | | | |
| Female | 0.0131 | -0.163*** | 1 | | | |
| Game Theory Exp | 0.0752 | 0.0658 | -0.135** | 1 | | |
| Religiosity | -0.0542 | -0.00449 | 0.0283 | -0.00494 | 1 | |
| Bot Exp | 0.0559 | -0.0644 | -0.0760 | 0.0693 | -0.00518 | 1 |
| Observations | 539 | | | | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Marginal effects

Table A3. Probit regression on the rate of reciprocity testing H1.

| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 | Model 7 |
|--------------------|----------------------|------------------------|-------------------------------|----------------------|------------------------|----------------------|-------------------------|
| | (Bot) | (Bot & Age) | (Bot & Game Theory Exp) | (Bot & Female) | (Bot & Religiosity) | (Bot & Bot Exp) | (Bot & All Controls) |
| Bot | -0.398*** (0.109) | -0.387*** (0.109) | -0.399*** (0.109) | -0.382*** (0.109) | -0.420*** (0.110) | -0.393*** (0.109) | -0.391*** (0.111) |
| Age | | -0.00864* (0.00422) | | | | | -0.00909* (0.00433) |
| Game Theory Exp | | | 0.292 (0.167) | | | | 0.303 (0.170) |
| Female | | | | 0.0138 (0.109) | | | 0.0110 (0.113) |
| Religiosity | | | | | -0.257 (0.153) | | -0.248 (0.154) |
| Bot Exp | | | | | | 0.144 (0.124) | 0.130 (0.127) |
| Constant | 0.192* (0.0769) | 0.536** (0.187) | 0.157* (0.0796) | 0.174 (0.0964) | 0.243** (0.0827) | 0.0832 (0.121) | 0.454 (0.239) |
| Observations | 539 | 538 | 539 | 533 | 539 | 539 | 532 |
| Pseudo R2 | 0.0180 | 0.0233 | 0.0221 | 0.0167 | 0.0218 | 0.0198 | 0.0319 |

Standard errors in parentheses

* p < 0.05, ** p < 0.01, *** p < 0.001

H2

Randomisation balance check

Table A4. Randomisation check - distribution of personality traits across conditions.

| Demographic | Human Condition | Bot Condition |
|------------------|-----------------|---------------|
| N | 269 | 270 |
| Agreeableness | 3.22(SD=.64) | 3.17(SD=.62) |
| Extraversion | 3.06(SD=.72) | 2.97(SD=.68) |
| Honesty-Humility | 3.56 (SD=.61) | 3.57(SD=.59) |
| Openness | 3.58(SD=.66) | 3.46(SD=.65) |

Pairwise Correlation

Table A5. Pairwise correlation between personality traits and participant responses.

| Variables | Response | Honesty-Humility | Agreeableness | Openness | Extraversion |
|------------------|----------|------------------|---------------|----------|--------------|
| Response | 1 | | | | |
| Honesty-Humility | 0.137** | 1 | | | |
| Agreeableness | 0.0905* | 0.257*** | 1 | | |
| Openness | 0.0391 | 0.0876* | 0.140** | 1 | |
| Extraversion | -0.0425 | 0.0165 | 0.214*** | 0.103* | 1 |
| Observations | 539 | | | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A6. Pairwise correlation between standardised personality traits testing H2.

| Variables | Response | Std Honesty- Humility | Std Agreeableness | Std Openness |
|----------------------|----------|--------------------------|----------------------|--------------|
| Response | 1 | | | |
| Std Honesty-Humility | 0.137** | 1 | | |
| Std Agreeableness | 0.0905* | 0.257*** | 1 | |
| Std Openness | 0.0391 | 0.0876* | 0.140** | 1 |
| Std Extraversion | -0.0425 | 0.0165 | 0.214*** | 0.103* |
| Observations | 539 | | | |

t statistics in parentheses

* p < 0.05, ** p < 0.01, *** p < 0.001

Additional regression model

Table A7. Regression models testing H2 with standardised personality scores.

| | Model 1 (Bot) | Model 2 (Bot & Honesty- Humility) | Model 3 (Bot & Agreeableness) | Model 4 (Bot & Openness) | Model 5 (Bot & Extraversion) | Model 6 (Bot & All Control Traits) |
|--------------------------|----------------------|--|-------------------------------------|--------------------------------|------------------------------------|---|
| Bot | -0.398*** (0.109) | -0.407*** (0.109) | -0.392*** (0.109) | -0.392*** (0.109) | -0.408*** (0.109) | -0.412*** (0.110) |
| Std Honesty- Humility | | 0.179** (0.0550) | | | | 0.158** (0.0570) |
| Std Agreeableness | | | 0.109* (0.0549) | | | 0.0877 (0.0587) |
| Std Openness | | | | 0.0320 (0.0545) | | 0.0145 (0.0557) |
| Std Extraversion | | | | | -0.0679 (0.0547) | -0.0923 (0.0567) |
| Constant | 0.192* (0.0769) | 0.197* (0.0774) | 0.189* (0.0772) | 0.190* (0.0771) | 0.197* (0.0771) | 0.200* (0.0779) |
| Observations | 539 | 539 | 539 | 539 | 539 | 539 |
| Pseudo R2 | 0.0180 | 0.0323 | 0.0233 | 0.0185 | 0.0201 | 0.0378 |

Standard errors in parentheses

* p < 0.05, ** p < 0.01, *** p < 0.001

Marginal effects

Table A8. Probit regression on the rate of reciprocity testing H2.

| | Model 1 | Model 8 | Model 9 | Model 10 | Model 11 | Model 12 |
|------------------|----------------------|---------------------------------|--------------------------|----------------------|-------------------------|--------------------------------------|
| | (Bot) | (Bot & Honesty- Humility) | (Bot & Agreeableness) | (Bot & Openness) | (Bot & Extraversion) | (Bot & All Co- Variate Traits) |
| Bot | -0.398*** (0.109) | -0.407*** (0.109) | -0.392*** (0.109) | -0.392*** (0.109) | -0.408*** (0.109) | -0.412*** (0.110) |
| Honesty-Humility | | 0.299** (0.0920) | | | | 0.264** (0.0954) |
| Agreeableness | | | 0.174* (0.0874) | | | 0.140 (0.0935) |
| Openness | | | | 0.0490 (0.0834) | | 0.0223 (0.0854) |
| Extraversion | | | | | -0.0968 (0.0780) | -0.132 (0.0809) |
| Constant | 0.192* (0.0769) | -0.869** (0.335) | -0.366 (0.291) | 0.0173 (0.308) | 0.489 (0.251) | -0.872 (0.500) |
| Observations | 539 | 539 | 539 | 539 | 539 | 539 |
| Pseudo R2 | 0.0180 | 0.0323 | 0.0233 | 0.0185 | 0.0201 | 0.0378 |

Standard errors in parentheses

* p < 0.05, ** p < 0.01, *** p < 0.001

Distribution of traits between responses

Figure A10. Strip Plot showing distribution of the Agreeableness trait between the responses.

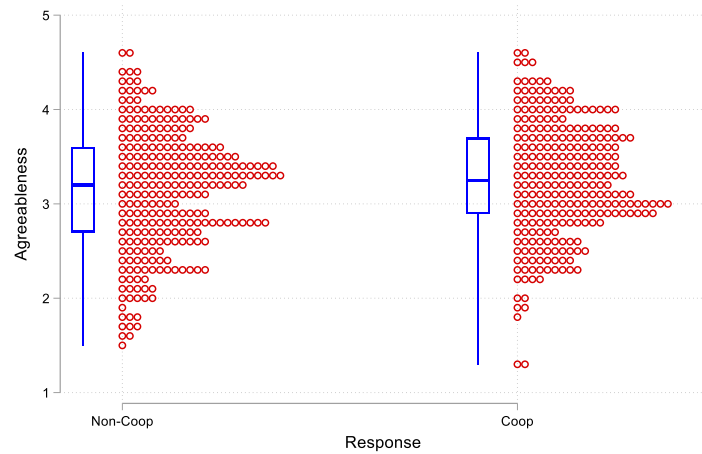


Figure A11. Strip Plot showing distribution of the Extraversion trait between the responses.

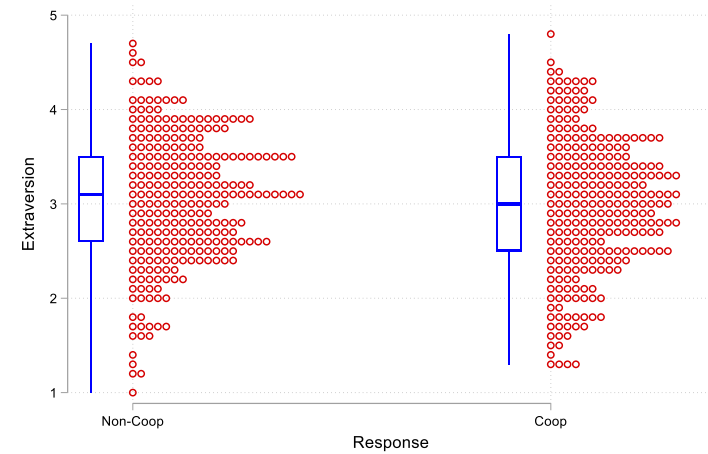


Figure A12. Strip Plot showing distribution of the Honesty-Humility trait between the responses.

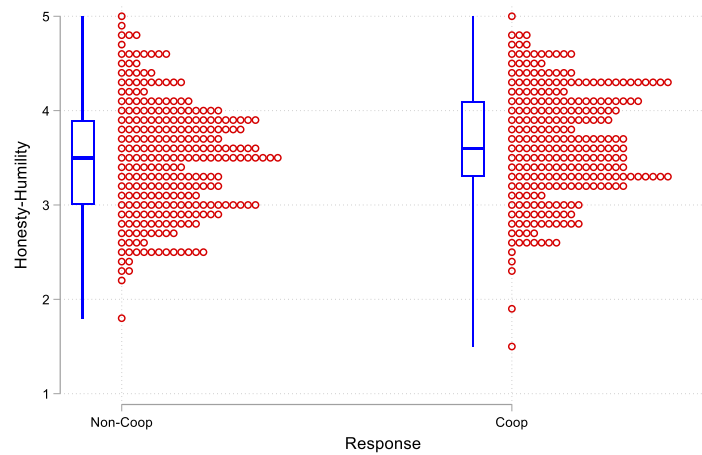
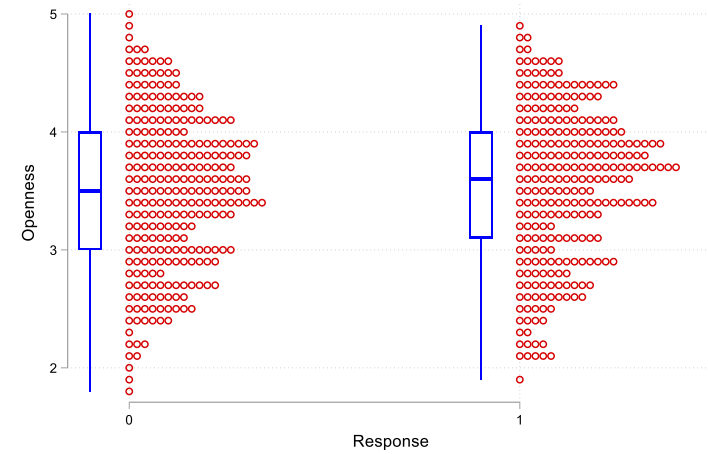


Figure A13. Strip Plot showing distribution of the Openness trait between the responses.



H3

Summary statistics

Table A9. Summary statistics of GAAIS subscales.

| Result | N | Mean | Standard Deviation | Minimum | Maximum | α | ω |
|----------------|-----|-------|-----------------------|---------|----------|----------|----------|
| GAAIS Positive | 539 | 3.290 | 0.668 | 1 | 4.916667 | .88 | .89 |
| GAAIS Negative | 539 | 3.257 | 0.747 | 1 | 5 | .85 | .85 |
| Observations | 539 | | | | | | |

Randomisation balance check

Table A10. Randomisation check - distribution on the GAAIS across conditions.

| Demographic | Human Condition | Bot Condition |
|----------------|-----------------|---------------|
| N | 269 | 270 |
| GAAIS Positive | 3.29(SD=.70) | 3.29(SD=.63) |
| GAAIS Negative | 3.23(SD=.75) | 3.29(SD=.74) |

Pairwise Correlation

Table A11. Pairwise correlation between GAAIS subscales and participant responses.

| Variables | Response | GAAIS Positive | GAAIS Negative |
|----------------|----------|----------------|----------------|
| Response | 1 | | |
| GAAIS Positive | 0.0122 | 1 | |
| GAAIS Negative | -0.00337 | 0.457*** | 1 |
| Observations | 539 | | |

* p < 0.05, ** p < 0.01, *** p < 0.001

Table A12. Pairwise correlations between standardised GAAIS scale responses testing H3.

| Variables | Response | Std GAAIS Positive | Std GAAIS Negative |
|--------------------|----------|--------------------|--------------------|
| Response | 1 | | |
| Std GAAIS Positive | 0.0122 | 1 | |
| Std GAAIS Negative | -0.00337 | 0.457*** | 1 |
| Observations | 539 | | |

t statistics in parentheses

* p < 0.05, ** p < 0.01, *** p < 0.001

Regression Models

Table A13. Probit regression on the rate of reciprocity testing H3.

| Variables | Model 1 (Bot) | Model 13 (Bot & GAAIS Positive) | Model 14 (Bot & GAAIS Negative) | Model 15 (Bot & All Co- Variates) |
|----------------|----------------------|---------------------------------------|---------------------------------------|---|
| Bot | -0.398*** (0.109) | -0.398*** (0.109) | -0.398*** (0.109) | -0.398*** (0.109) |
| GAAIS Positive | | 0.0228 (0.0815) | | 0.0250 (0.0913) |
| GAAIS Negative | | | 0.00575 (0.0729) | -0.00435 (0.0818) |
| Constant | 0.192* (0.0769) | 0.117 (0.279) | 0.174 (0.247) | 0.124 (0.307) |
| Observations | 539 | 539 | 539 | 539 |
| Pseudo R2 | 0.0180 | 0.0181 | 0.0180 | 0.0181 |

Standard errors in parentheses

* p < 0.05, ** p < 0.01, *** p < 0.001

Table A14. Regression models testing H3 with standardized GAAIS scores.

| Response | Model 1 (Bot) | Model 2 (Bot & GAAIS Positive) | Model 3 (Bot & GAAIS Negative) | Model 4 (Bot & All Controls) |
|--------------------|----------------------|--------------------------------------|--------------------------------------|------------------------------------|
| Bot | -0.398*** (0.109) | -0.398*** (0.109) | -0.398*** (0.109) | -0.398*** (0.109) |
| Std GAAIS Positive | | 0.0152 (0.0544) | | 0.0167 (0.0610) |
| Std GAAIS Negative | | | 0.00429 (0.0544) | -0.00325 (0.0610) |
| Constant | 0.192* (0.0769) | 0.192* (0.0769) | 0.192* (0.0770) | 0.192* (0.0770) |
| Observations | 539 | 539 | 539 | 539 |
| Pseudo R2 | 0.0180 | 0.0181 | 0.0180 | 0.0181 |

Standard errors in parentheses

* p < 0.05, ** p < 0.01, *** p < 0.001

Marginal effects

Table A15. Marginal effects of relevant H3 models.

| | Model 1 (Bot) | Model 13 (Bot & GAAIS Positive) | Model 14 (Bot & GAAIS Negative) | Model 15 (Bot & All Controls) |
|----------------|-------------------------|--|--|--|
| Bot | -0.156*** (0.0409) | -0.156*** (0.0408) | -0.156*** (0.0409) | -0.155*** (0.0409) |
| GAAIS Positive | | 0.00890 (0.0319) | | 0.00976 (0.0357) |
| GAAIS Negative | | | 0.00225 (0.0285) | -0.00170 (0.0320) |
| Observations | 539 | 539 | 539 | 539 |

Standard errors in parentheses

* p < 0.05, ** p < 0.01, *** p < 0.001

Distribution on the GAAIS between responses

Figure A14. Strip Plot showing distribution of Positive general attitude to AI between the responses.

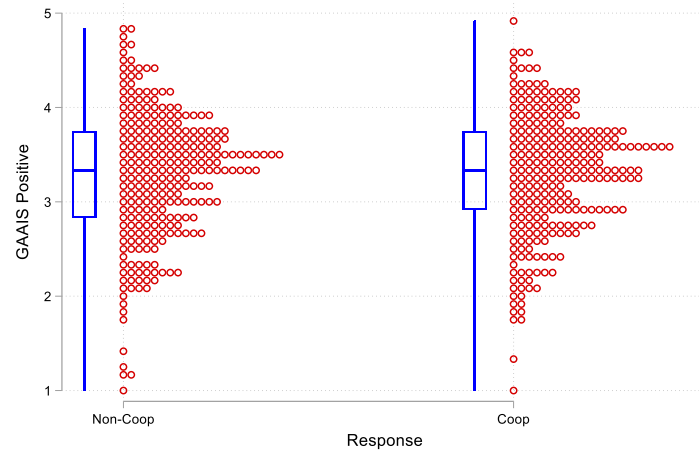
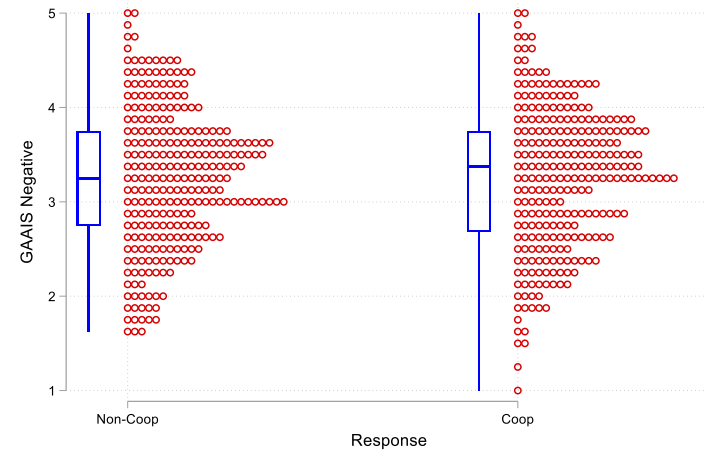


Figure A15. Strip Plot showing distribution of Negative general attitude to AI between the responses.



Appendix A.8 – Note on scripts used in this study

For transparency and reproducibility (Morey *et al.*, 2016; Obels *et al.*, 2020), all data and scripts are available at the Open Science Framework repository for this study: https://osf.io/48jgv/?view_only=28d14f4e206541928f3da53fcb0602d7.

Experimental platform

Scripts were embedded into the flow on the experimental platform, Gorilla, to calculate personality trait and GAAIS sub-scale results for each participant. These scripts were designed in accordance with the scoring keys set out in Appendices 2 and 3 above. Individuals answered the questionnaires, each item of which corresponds to a numerical value. These values were then collated using the script, resulting in summary values for each factor of interest. A script was also used to transform the choices made by the participants (★ or ☆) in the experimental game into numerical outputs for ease of analysis.

Pre-processing

The data was downloaded from the experimental platform in long format as multiple .csv files. This data was then transformed into wide format (based on the unique “Participant Private ID” field) and collated. Extraneous data collected by the platform was removed. Data pre-processing was completed using the Tidyverse package (Grolemund and Wickham, 2016) on R-studio. A full list of commands, including those related to establishing the pre-processing environment, along with the raw files are available on the OSF repository.

Statistical analysis

The statistical analysis presented in this paper was conducted using Stata 16. The full analysis script for this study is available on the OSF repository.

References in Supplementary Material

Arechar, A.A., Gächter, S. and Molleman, L. (2018) 'Conducting interactive experiments online', *Experimental Economics*, 21(1), pp. 99–131. Available at: <https://doi.org/10.1007/s10683-017-9527-2>.

Ashton, M.C. and Lee, K. (2009) 'The HEXACO-60: a short measure of the major dimensions of personality', *Journal of Personality Assessment*, 91(4), pp. 340–345. Available at: <https://doi.org/10.1080/00223890902935878>.

Chmielewski, M. and Kucker, S.C. (2020) 'An MTurk crisis? Shifts in data quality and the impact on study results', *Social Psychological and Personality Science*, 11(4), pp. 464–473. Available at: <https://doi.org/10.1177/1948550619875149>.

Cohen, J. (2013) *Statistical Power Analysis for the Behavioral Sciences*. Academic Press.

Glennerster, R. and Takavarasha, K. (2013) *Running Randomized Evaluations: A Practical Guide*. Illustrated edition. Princeton: Princeton University Press.

Grolemund, G. and Wickham, H. (2016) *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. 1st edition. Sebastopol, CA: O'Reilly Media.

Gupta, K. et al. (2016) 'Basic concepts for sample size calculation: Critical step for any clinical trials!', *Saudi Journal of Anaesthesia*, 10(3), pp. 328–331. Available at: <https://doi.org/10.4103/1658-354X.174918>.

Karpus, J. et al. (2021) 'Algorithm exploitation: Humans are keen to exploit benevolent AI', *iScience*, 24, p. 102679. Available at: <https://doi.org/10.1016/j.isci.2021.102679>.

Kocijan, V. et al. (2020) 'A Review of Winograd Schema Challenge Datasets and Approaches', *arXiv:2004.13831 [cs]* [Preprint]. Available at: <http://arxiv.org/abs/2004.13831> (Accessed: 28 February 2022).

Kothe, E. and Ling, M. (2019) *Retention of participants recruited to a one-year longitudinal study via Prolific*. Available at: <https://doi.org/10.31234/osf.io/5yv2u>.

Levesque, H.J., Davis, E. and Morgenstern, L. (2012) 'The Winograd schema challenge', in *Proceedings of the Thirteenth International Conference on Principles of Knowledge Representation and Reasoning*. Rome, Italy: AAAI Press (KR'12), pp. 552–561.

Moher, D. et al. (2010) 'CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials', *BMJ*, 340, p. c869. Available at: <https://doi.org/10.1136/bmj.c869>.

Morey, R. et al. (2016) 'The Peer Reviewers' Openness Initiative: Incentivising Open Research Practices through Peer Review', *Royal Society Open Science*, 3. Available at: <https://doi.org/10.1098/rsos.150547>.

Obels, P. et al. (2020) 'Analysis of Open Data and Computational Reproducibility in Registered Reports in Psychology', *Advances in Methods and Practices in Psychological Science*, 3(2), pp. 229–237. Available at: <https://doi.org/10.1177/2515245920918872>.

Schepman, A. and Rodway, P. (2020) 'Initial validation of the general attitudes towards Artificial Intelligence Scale', *Computers in Human Behavior Reports*, 1, p. 100014. Available at: <https://doi.org/10.1016/j.chbr.2020.100014>.