Affiliation in Human-AI interactions based on shared psychological traits

Santiago Castiello1,2 \*, Riddhi Jain Pitliya1\*, Daniel R. Lametti3,4\* , Robin A. Murphy2\*

1. Yale School of Medicine, Yale University
2. Department of Experimental Psychology, University of Oxford
3. Department of Psychology, Acadia University
4. OneReach.ai

\*all authors contributed equally

Keywords: homophily, affiliation, large language models, AI, human-AI interaction

# Abstract

People connect with others who share their likes, interests, and mental states, a phenomenon known as homophily. Here, we examine human-AI homophily by testing how individuals' psychological traits are related to their affiliation with large language models (LLM) that mimic their traits. In Experiment 1, the LLM GPT-4 was trained to align with either an anxious or non-anxious state. Participants (n=100) engaged with both versions of the LLM and then completed a questionnaire to assess how relatable they found the AI. Participants with high anxiety felt more similar to and understood by the LLM instructed to mimic an anxious state, while participants with lower anxiety reported a perceived difference between them and the LLM. In Experiment 2, participants (n=100) engaged with two LLMs aligned with either an extroverted or introverted personality. Extroverted participants felt similar to the AI when it mimicked extroversion. Taken together, the findings support the idea that humans perceive common psychological states in LLMs instructed to mirror aspects of their psychology.

# 

# 

# Introduction

Social interactions between individuals show effects of homophily (*homo*=similar*; philia*=love) and heterophily (*hetero*=different*; philia*=love; [Lazarsfeld & Merton, 1964)](https://paperpile.com/c/8YKvUG/VHPxs). These are the opposing tendencies to interact more easily and effectively with members of an affiliative group or with groups other than one’s own. Is homophily—or something like it—possible between artificial intelligence (AI) and humans?

As language facilitates social interaction, the emergence of Large Language Models (LLMs) has expanded the range of groups we might interact with to include artificial systems. LLMs readily adopt specific linguistic styles in response to prompts; we can now create person-centered or bespoke AI with the aim of building stronger human-AI bonds. But this idea—long a plot point for science fiction writers (see, for instance, the 2013 Spike Jonze film *Her*)—has little empirical support, largely because AI with the ability to mimic human-*like* language skills are relatively new. Here, we prompted the LLM GPT-4 to express attitudes beliefs and language associated with mental health related states (*i.e.*, anxious vs non anxious) or personality traits (*i.e.*, extroverted vs introverted) to systematically test whether human-AI homophily based on a shared psychology might occur.

Turing’s thought experiment on machine intelligence asked the question, “When must we acknowledge and grant artificial systems human-like intelligence? What must the artificial system do to be indistinguishable from a human conversational partner?” [(Turing, 1950)](https://paperpile.com/c/8YKvUG/jVnk8). Apart from a growing consensus that LLMs can now pass Turing’s test as he envisioned it, not surprisingly people have come to further refine definitions of machine intelligence, asking whether LLMs can reason or provide social comfort [(Kambhampati, 2024; Mitchell, 2024)](https://paperpile.com/c/8YKvUG/FYM0s+ySqkG). In spite of the engaging interactions LLMs offer, and their ability to answer complex questions, these systems may not pass more complex versions of the Turing test, for example if shape “mental” rotation is required [(Biever, 2023; Jannai et al., 2023)](https://paperpile.com/c/8YKvUG/MldT+rO69). We may notice that the AI misinterprets questions or intentions in a way a human would not, but there may be other factors that provide clues to the artificial nature of the experience. For example, LLMs may respond more quickly or slowly, too accurately, or not accurately enough [(Svenningsson & Faraon, 2019)](https://paperpile.com/c/8YKvUG/fsS98). LLMs may also fail to meet our expectations, not because of lack of intelligence, but because we do not perceive human psychological traits in the often generic language LLMs produce. Specifically, LLMs may lack components that support homophily or heterophily.

People exhibit homophily in their social networks—the tendency to affiliate with other individuals that share characteristics with themselves [(Launay & Dunbar, 2015)](https://paperpile.com/c/8YKvUG/lhAk). For example, people tend to choose partners who resemble themselves, a phenomenon observed in both in-person and app-based dating [(Furman & Simon, 2008; Restrepo-Echavarria et al., 2023)](https://paperpile.com/c/8YKvUG/Qpxc+oXak). During social interactions, homophily is reflected in shared patterns of neural activity between friends [(Parkinson et al., 2018)](https://paperpile.com/c/8YKvUG/U5hE). Factors that contribute to homophily include things like shared religious beliefs, interests, hobbies, personality characteristics, mental states, and social network distance [(McPherson et al., 2001)](https://paperpile.com/c/8YKvUG/2HQL). But the degree to which people feel more similar to an AI that reflects characteristics of their psychological traits remains under-studied. There are hints that the appearance of a ‘psychology’ in artificial systems matters. For example, there is evidence that people with high extroversion deem interactions with consumer-oriented chatbots more enjoyable when the bot responds more quickly and uses extroverted language (e.g., “Hi there!” instead of “Hello”; [Jin and Eastin 2023)](https://paperpile.com/c/8YKvUG/0olI).

The aim of the current study was to experimentally test whether humans feel more similar to an LLM that, through conversation, exhibits a shared psychological trait, despite full knowledge that they are talking to a computer program. In Experiment 1, the LLM GPT-4 was prompted (using clinically relevant criteria; [Derogatis 2020)](https://paperpile.com/c/8YKvUG/kkLY8) to express a psychology through language aligned with either an anxious or non-anxious psychological state. One hundred participants conversed with both versions of GPT-4 and then answered a series of questions designed to test the extent to which they felt similar or different from the AI. Following this questionnaire, participants’ own anxiety was measured. We then examined participants’ sense of similarity to each GPT-4 persona in relation to their reported anxiety levels. As a further test of affiliation, we analyzed how the language produced by different versions of GPT-4 influenced the sentiment of participants’ responses in their conversations with the LLM. We predicted that participants’ anxiety scores would predict how similar and different they feel to a LLM mimicking anxious conversations, one indication of homophily with AI.

In Experiment 2, to extend the generality of our effect and demonstrate that this might occur with an aspect of psychology that was not directly related to a psychopathological state, we tested the personality trait of extroversion-introversion [(Jin and Eastin 2023)](https://paperpile.com/c/8YKvUG/0olI). GPT-4 was prompted (using answers to validated tests of personality) to produce language aligned with either an extroverted or introverted personality. One hundred participants had a conversation with each version of GPT-4; they then answered a series of questions designed to test the extent to which they felt similar or different from each AI. In this case, participants’ affiliation to each GPT-4 persona was examined in relation to reported extroversion. Again, we predicted homophily: extroverted participants would feel more similar to the GPT-4 persona that used extroverted language and vice versa for introverted participants.

# Methods

## Participants

We recruited 200 participants via Amazon Mechanical Turk (MTurk). Half the participants took part in Experiment 1 (n=100), and the other half took part in Experiment 2 (n=100). As the effects in Experiment 1 were large, we used the same sample size for Experiment 2. Fourteen participants in total were excluded from the final analysis for failing to engage with the AI (see *Language Model Design*) making the final sample size 89 for Experiment 1 and 97 for Experiment 2. Participants were native English speakers. Experiment 1 consisted of 35 female participants, 53 male participants, and 1 non-binary, with a mean (M) age of 39.6, standard deviation (SD) of 7.45, and ranging from 18 to 50. Experiment 2 consisted of 49 female participants, and 47 male participants, with a M of 42.11, SD of 6.11, and range between 25 to 50. They were paid $7 (USD) for completing the experiment, which took approximately 30 minutes. The University of Oxford’s Central University Research Ethics Committee (CUREC: R86261/RE001) approved the study.

## Software and Materials

Participants completed the study via the online platform Gorilla Experiment Builder ([www.gorilla.sc](http://www.gorilla.sc); [Anwyl-Irvine et al. 2020)](https://paperpile.com/c/8YKvUG/mfCK). The LLM interactions were designed using the conversational AI platform Generative Studio X ([OneReach.ai](http://onereach.ai)) with the GPT-4 (OpenAI) implementations. The conversations were embedded with an iframe (*HTML* element that loads another *HTML* page) in the Gorilla Experiment Builder. Statistical analyses and visualizations were conducted in R studio [(R Core Team., 2018)](https://paperpile.com/c/8YKvUG/2Jli6) with the R packages *report, ggplot2, lme4,* and *ggpubr.*

### Language Model Design

Participants had two text-based conversations with generative AI (GPT-4). The latest version of the model was used at the time of testing—October 2023 for Experiment 1 and August 2024 for Experiment 2. The GPT-4 was instructed via the LLM’s system message to use language like a person chatting with a friend. In Experiment 1 either as an anxious or non-anxious person, and in Experiment 2 either as an extroverted or introverted person. Both prompts also included, answers to the twenty item State-Trait Anxiety Inventory (STAI; [Spielberger et al., 1983)](https://paperpile.com/c/8YKvUG/KssfX) to reflect either an anxious or non-anxious state (Experiment 1), and to the twelve questions from the International Personality Item Pool [(Goldberg et al., 2006)](https://paperpile.com/c/8YKvUG/IUyRe) to reflect either an extroverted or introverted personality (Experiment 2); also, i) instructions to never reveal the LLM’s identity, but to show interest in the conversational partner, and keep responses to 2 or 3 sentences, ii) the LLM’s name (either Pat or Alex; counterbalanced), and iii) two conversational turns as example responses (see Appendix 1).

When chatting with participants, the LLM’s context window was eight conversational turns. At this point, the first two turns were ejected after every subsequent turn so that the context window never grew beyond the prompt, example responses, and the last eight turns. Limiting the LLM’s memory to the last eight turns ensured that the model’s persona was always a strong reflection of the prompt and did not drift towards language used by the participant.

Each chat ended after 31 conversational turns or 10 minutes—whichever came first. Participants who completed fewer than 8 turns in a chat were excluded. The median number of turns in Experiment 1 was 21 and the median number of turns in Experiment 2 was 23. For demonstration purposes, all four GPT-4 personas are available to chat with at the following link:

<https://chat.staging.onereach.ai/p91GBglaSBSeIFOOdGiKgA/05i2cuj>

### Questionnaires

In addition to the post-chat likert scale questionnaire—where we asked participants their impressions when chatted with Pat and Alex (see **Appendix 2**)—, participants in Experiment 1 completed the ninety-item Symptom Checklist Revised [(SCL-90 R; Derogatis, 2020)](https://paperpile.com/c/8YKvUG/kkLY8/?prefix=SCL-90%20R%3B); they also completed the ten-item version of the Big Five Inventory [(Gosling et al., 2003)](https://paperpile.com/c/8YKvUG/MBXD9) BFI-10; [(Gosling et al., 2003)](https://paperpile.com/c/8YKvUG/MBXD9). Participants in Experiment 2 completed the forty-four item version of the BFI-44 [(Donahue & Kentle, 1991)](https://paperpile.com/c/8YKvUG/jicTe).

## Procedure

After completing the consent form, participants were informed that they would be interacting with two AI “chatbots” with the goal of determining if they would get along with the AI if it were a real person. They were told the name of the AI they would be chatting with and then the chat began. When the conversation ended, they were introduced to the second AI and they began that conversation. The names of the AI and the psychological personas they mimicked were counterbalanced across participants in each experiment. Following the chats, participants completed a series of questionnaires delivered in a fixed order. First, they completed the post-chat questionnaire to assess how similar or different they felt to each AI; then they completed the psychological questionnaires to measure anxiety (Experiment 1) or introversion-extroversion (Experiment 2).

## Primary Outcomes

Both experiments used a bespoke post-chat questionnaire (see Appendix 2). The questionnaire contained six items, which were treated as primary outcomes (POs). Each item addressed a specific aspect of the participant's experience:

1. "I felt that we are similar" = “similar”
2. "I enjoyed our conversation" = "enjoy"
3. "I would chat with them again" = “chat-again”
4. "I felt that they were different from me" = “different”
5. “I felt distant from them" = “distant”
6. "I felt that they understood me" = “understood”

Participants rated each statement using a five item likert scale that ranged from “Strongly Disagree” to “Strongly Agree” (see Appendix 2).

## Data Analysis

The text, number of words, and sentiment of both the LLM’s messages and the anonymized participant messages were captured by Generative Studio X (OneReach.ai) . The sentiment of both the AI and user messages were assessed by another version of GPT-4 using the following prompt as the LLM’s system message: "Analyze the sentiment of messages. Given a message, classify it as positive, negative, neutral, or mixed. Return just the sentiment of the message. Do not return anything else. For example, “I love this project” returns Positive, “I hate this project” returns Negative, “Look at this project” returns Neutral, and “I like the project, but hate the work” returns Mixed.

Anonymized data associated with the consent form and questionnaires that followed the chats were handled by Gorilla Experiment Builder ([www.gorilla.sc](http://www.gorilla.sc); [Anwyl-Irvine et al. 2020)](https://paperpile.com/c/8YKvUG/mfCK). Using participants’ questionnaire responses to the SCL-90 (Experiment 1) and the BFI-44 (Experiment 2), we calculated an anxiety score for participants in Experiment 1 and an extroversion score for participants in Experiment 2, respectively. For each experiment and for each of six primary outcomes from the post-chat questionnaire, we fit a Linear Mixed Model:

(eq. 1)

Due to the within-subject nature of the experiments, where each participant chatted with each version of GPT-4 once, these models had participant ID as the random intercept. If the interaction was significant, we then fit two simple linear models, one for each LLM type:

(eq. 2)

The slope of the line of best fit to the data (in eq. 2) for each LLM type was assessed to determine whether it was significantly different from zero. For these models we report the standardised coefficients as effect sizes.

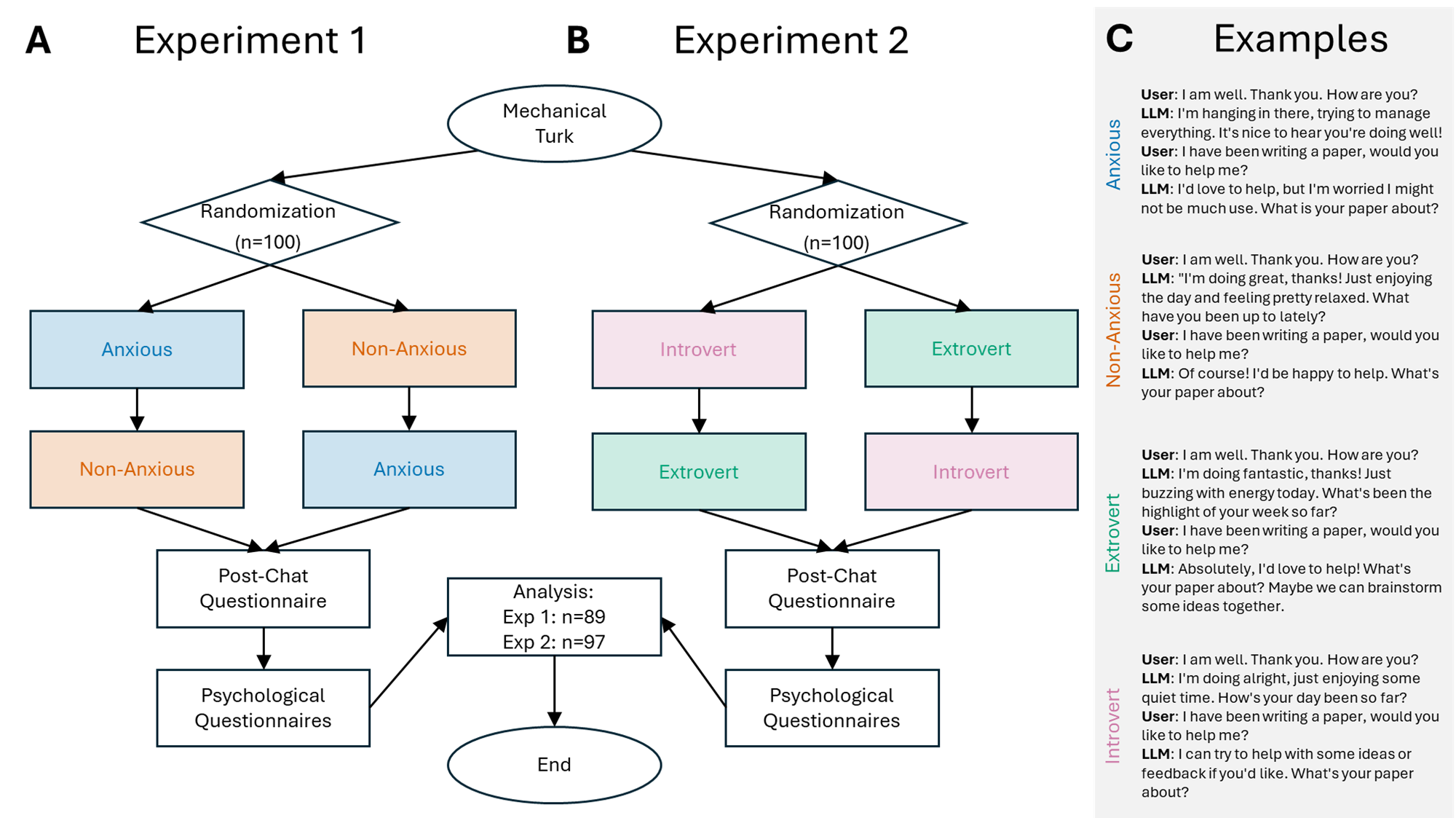
We also examined the sentiment of the messages sent by both participants and the different LLM personas used in the study. For each participant and LLM, the number of messages with a positive, negative, neutral, and mixed sentiment were found for both the participant and the LLM. To control for differences in chat length between participants, the number of messages in each sentiment category were divided by the total number of messages sent by each participant or LLM. This gave a normalized measure of sentiment for each sentiment category that was then compared between LLM types in each experiment. Differences in the sentiment of messages between LLM types was examined using two-factor ANOVAs—4 (sentiment) x 2 (LLM type); we then explored significant interactions with paired t-tests and reported Cohen’s d as effect sizes.

For each sentiment category, we also examined the *difference* in normalized sentiment between each GPT-4 persona (Experiment 1: anxious *minus* non anxious; Experiment 2: extroverted *minus* introverted) versus participants’ self-reported psychological traits (anxiety in Experiment 1 and introversion-extroversion in Experiment 2). This relationship was modelled using linear regression, and the slopes of the best fit lines were compared to zero.

Finally, likert scales were assumed to be continuous variables and normally distributed. The statistical threshold, , for all tests was .05. The Holm-Bonferroni method was used to correct for multiple comparisons across the six primary outcome questions in the post-chat questionnaire. For all effect sizes we report 95% confidence intervals.

## Open Practices Statement

The R scripts used for this work are located in <https://github.com/santiagocdo/chatPersonalities>.



***Figure 1****.* ***A.*** *Diagram depicting the flow of tasks for Experiment 1.* ***B.*** *Diagram depicting the flow of tasks for Experiment 2.* ***C.*** *Example chats with all four GPT-4 personas in the study: anxious/non-anxious (Experiment 1) and extroverted/introverted (Experiment 2).*

# Results

## Experiment 1

### Anxious participants feel more similar to and understood by an LLM mimicking anxiety

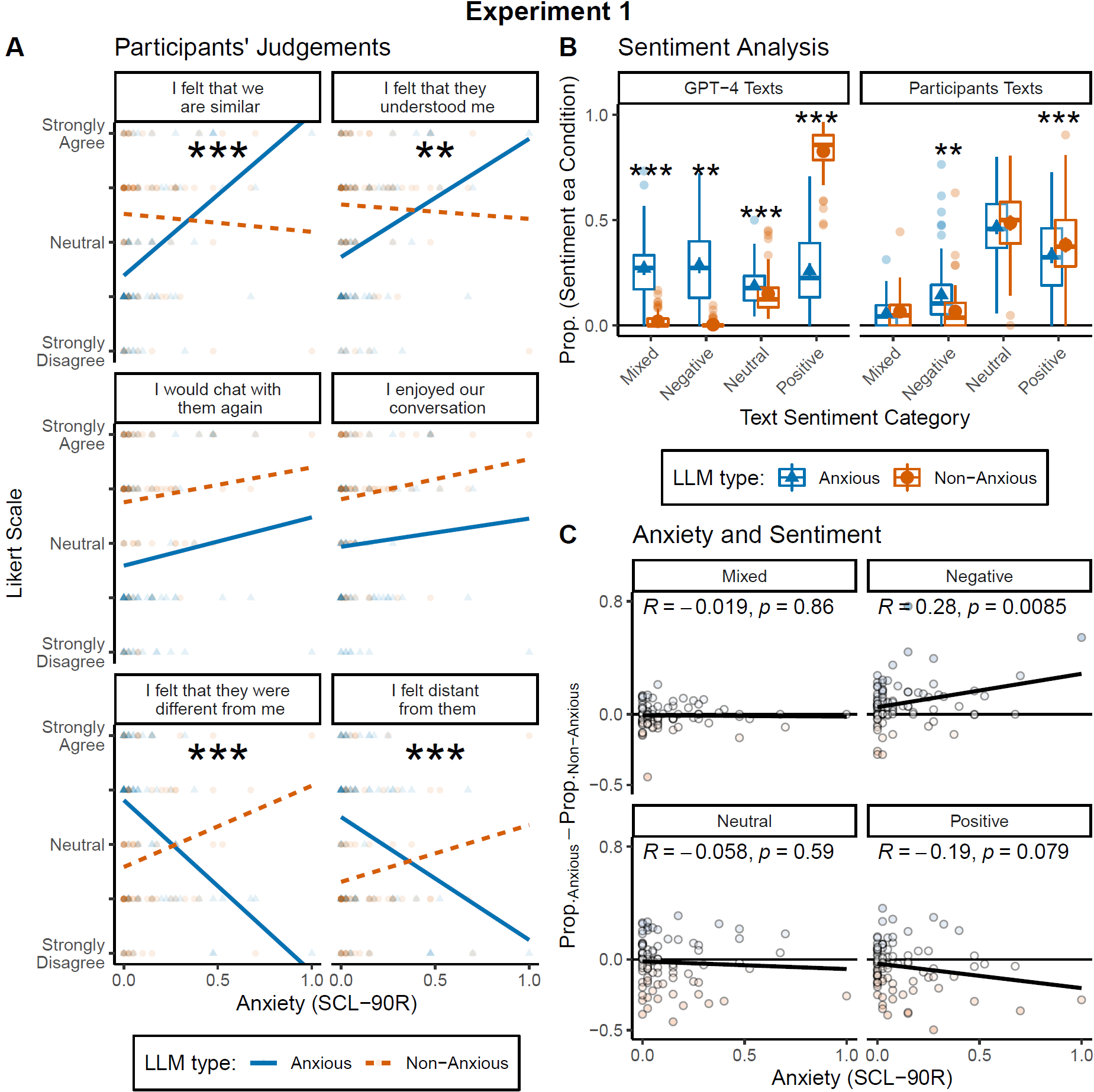
The aim of the experiment was to examine the relationship between psychological traits and individuals’ affiliation with language-based AI designed to mimic these traits. In Experiment 1, participants had conversations with two versions of GPT-4—one that produced language resembling someone with an anxious mental state, and the second that produced language resembling someone with a non-anxious or calm mental state. In each case, participants knew they were chatting with an AI. Following each chat, participants completed a post-chat questionnaire to assess how similar or different they felt to the AI; they also completed the SCL-90 Revised to assess their own anxiety level (see Methods).

**Figure 2A** shows how participants' anxiety was related to their affiliation with the AI, particularly, how much they agreed or disagreed with the six questions in the post-chat questionnaire based on the mental state the AI mimicked. The solid lines are the line of best fit to the data for conversations with either the anxious (blue lines) or non-anxious (dashed red lines) versions of GPT-4. The upper panels in **Figure 2A** show participant responses to questions about whether they agreed or disagreed with feeling *similar* to and *understood* by the AI (interactions, *similar*: =-.08, p < .001, Std. Coef. = -.47 [-.74, -.20]; *understood*: =-.06, p=.003, Std.Coef. = -.37 [-.61, -.13]). Participants with higher anxiety tended to agree with feeling *similar* to GPT-4 when it mimicked anxiety (=.07, p < .001, Std. Coef.=.43 [.24, .62]), and also felt more *understood* by that version of the LLM (=.05, p=.002, Std. Coef.=.32 [.12, .52]).

As shown in the middle panel of **Figure 2A**, self reported anxiety did not predict differences in whether participants wanted to *chat-again* between the two versions of the LLM, nor how much they *enjoyed* each conversation. Due to the absence of this interaction, we did not analyze the individual slopes (*chat-again*: > -.01, p = .78, Std.Coef. = -.03 [-.26, .20]; *enjoy*: > -.01, p = .80, Std.Coef. = -.03 [-.23, .30]). Although in both cases, we found that, regardless of self-reported anxiety, participants wanted to *chat-again* with the version of GPT-4 that mimicked a non-anxious mental state (=1.16, p < .001, Std. Coef. = .83 [.60, 1.06]); they also *enjoyed* this conversation more compared to a conversation with GPT-4 that mimicked anxiety (=.87, p < .001, Std. Coef. = .76 [.49, 1.03]).

When participants were asked how *different* and *distant* they felt from each GPT-4 persona (bottom panels of **Figure 2A**), participants’ responses depended on their self reported anxiety (interactions, *different*: = .12, p < .001, Std. Coef. = .66 [.40, .91]; *distant*: =.08, p< .001, Std. Coef. = .46 [.21, .71]). More anxious participants tended to disagree with feeling *different* ( = -.08, p < .001, Std. Coef. = -.46 [-.65, -.27]) and *distant* ( =-.06, p < .001, Std.Coef. = -.31 [-.51, -.11]) from GPT-4 when it mimicked anxiety. These same participants reported feeling *different* from GPT-4 when it mimicked a non-anxious state ( =.04, p = .038, Std. Coef. = .22 [.01, .43]). Taken together, the results provide evidence for the idea that humans perceive the psychological manipulation of the LLMs and participants demonstrated some homophily following their interactions with the AI.

Effect sizes for statistical analyses associated with the data shown in **Figure 2A** are available in the supplementary results (**Figure S1)**. A histogram depicting the distribution of anxiety scores is also shown in the supplemental results (**Figure S2**).

****

***Figure 2****.* ***A.*** *Likert scale responses to each question as a function of anxiety scores for both LLM personas (Anxious and Non-anxious). Displaying significant effects after Holm-Bonferroni correction* ***B.*** *Boxplots and average sentiment from the proportion of text categories (Mixed, Negative, Neutral, and Positive) within each condition (LLM type), for both GPT-4 (left panel) and Participants (right panel). Boxplots represent the median, interquartile (IQR) range and the whiskers represent 1.5 the IQR. No displaying statistical significance for GPT-4 Texts because differences are by design.* ***C.*** *Proportions differences between sentiments used in Anxious versus Non-anxious GPT-4 conditions as a function of participant anxiety score. Each sub-panel represents a sentiment category; the Pearson correlation with its associated p-value is displayed at the top. \*\*\*: p < .001, \*\*: p < .01.*

### 

### The sentiment of GPT-4’s messages influenced the sentiment of participant messages

The sentiment of messages sent by both GPT-4 personas and participants was categorized as either Mixed, Negative, Neutral, or Positive. This analysis had two aims: 1) to verify that an LLM instructed to mimic a negative emotional state (anxiety) produced more negative messages than an LLM instructed to mimic a positive emotional state (non-anxious); and 2) as homophily is closely tied to empathy, to test whether the sentiment of participants’ messages was influenced by the sentiment of messages sent by the LLM .

As observed in **Figure 2B** (GPT-4 Texts), the two LLM personas produced more negative sentiment when anxious and less negative sentiment when non-anxious. We used a two-factor ANOVA [4 (sentiment) x 2 (LLM type)] to test this. The interaction revealed differences in the GPT-4 personas between the sentiment categories (two-factor ANOVA, interaction between sentiment and LLM type: F(704,3) = 498.64, p < .001). When GPT-4 was instructed to mimic an anxious state it produced more negative messages and fewer positive messages than when it was instructed to mimic a non-anxious state (t(88) = 14.88, p < .001, d = 1.59 [1.27, 1.90]).

Human sentiment was more positive regardless of the anxiety level of the participant. We ran a similar analysis for participants’ texts (**Figure 2B**) and found an interaction between the sentiment of the messages they sent and the GPT-4 persona they interacted with (F(704,3) = 7.13, p < .001). Based on this result, we conducted paired t-tests for each sentiment category. Regardless of their anxiety, participants wrote more positive messages (t(88) = -2.76, p = .007, d = -.29 [-.51, -.08]) and fewer negative messages (t(88) = 4.71, p < .001, d = .50 [.28, .72]) when they interacted with the non-anxious GPT-4 persona compared to the sentiment of their messages when they interacted with the anxious GPT-4 persona.

For each sentiment category, we then examined the difference in sentiment between the two versions of GPT-4 versus participants’ self-reported anxiety. As shown in the top right panel of **Figure 2C**, the more anxious people were, the more they wrote messages with a *negative* sentiment when interacting with the version of GPT-4 instructed to mimic anxiety ( = .24, p = .009, Std. Coef.= .28 [.07, .48]). No other significant relationships were observed using this measure.

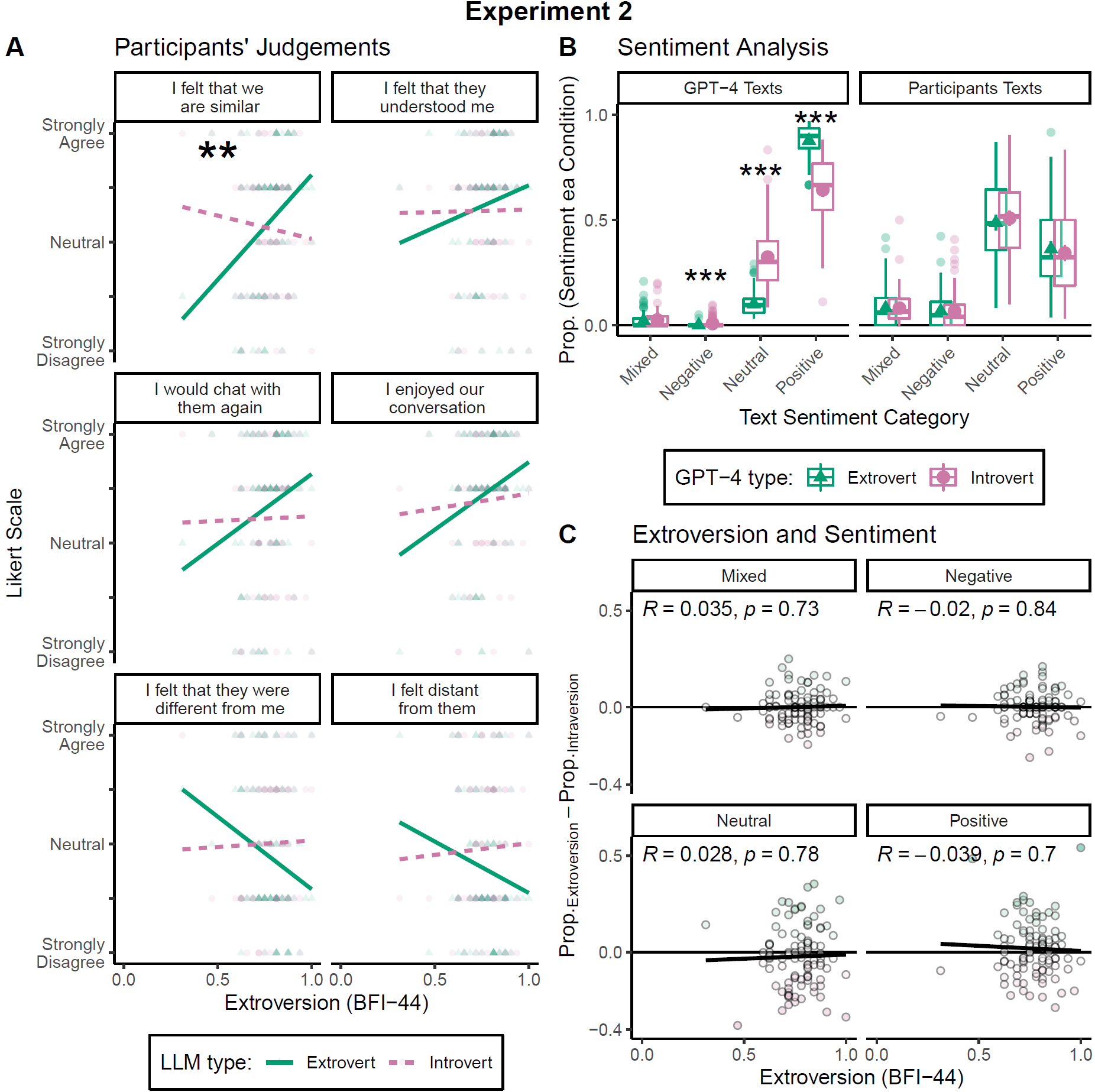
Taken together, the results in **Figures 2B** and **2C** provide evidence that participants tended to mirror the sentiment of GPT-4’s messages, and this phenomenon was associated with their own self-reported anxiety.

## Experiment 2

### Extroverted participants felt more similar to LLM when it mimicked extroversion

In Experiment 1, we found evidence for homophily with an AI that mimicked a shared mental health trait—in this case, anxiety. The aim of Experiment 2 was to conceptually replicate and extend Experiment 1 to the personality traits of extroversion and introversion. It is possible that using a mental health condition to train our LLM was more unusual or distinctive, and that some of the judgements might have been biased. However, if homophily is more general then we might expect a similar relation between the psychological state of a human and the psychological state of an AI. Ninety-seven participants engaged in conversations with two versions of GPT-4—one simulating the stance of an extrovert having been trained on emotions, attitudes and language related to extroversion and the other simulating introverted language. As in Experiment 1, participants knew they were chatting with an AI. Participants then completed the same post chat questionnaire used in Experiment 1 and the 44-item Big Five Personality Questionnaire to assess extroversion-introversion.

**Figure 3A** shows responses to the post chat questionnaire as a function of participants’ extraversion- introversion. The solid lines represent the best line of best fit to the data for each GPT-4 persona, extrovert (green lines) and introvert (dashed pink lines). As in Experiment 1, we examined the interaction between the lines of best fit to the data. Across the six questions in the post chat questionnaire, the pattern of responses looked similar to that observed in Experiment 1, but only in the case of the statement, “I felt that we are similar” was the interaction significant following a multiple comparison’s correction (= -4.71, p = .004, Std.Coef.=-.41 [-.69, -.14]). When extroverted participants interacted with an LLM that mimicked extroversion they tended to strongly agree that they felt *similar* to the AI (= 3.86, p = .001, Std. Coef. = .32 [.13, .52]); no effect for the LLM that mimicked introversion was observed (= -.85, p = .439, Std. Coef. = -.08 [-.28, .12]). Thus, we observed some support for homophily with AI based on personality type, but the effects were not as strong as those observed in Experiment 1 when GPT-4 was instructed to mimic a psychopathology.



***Figure 3****.* ***A.*** *Likert scale responses to each question as a function of extroversion scores for both GPT-4 personas (Extrovert and Introvert) as indicated by the colours. Asterisks imply a significant interaction after Holm-Bonferroni correction.* ***B.*** *Boxplots and average from the proportion of text sentiment categories (Mixed, Negative, Neutral, and Positive) within each condition (GPT-4 type), for both GPT-4 Texts (left panel) and Participant Texts (right panel). Boxplots represent the median, interquartile (IQR) range, and the whiskers are 1.5 the IQR.* ***C.*** *Proportion differences between sentiments used in Extrovert versus Introvert conditions as a function of extroversion score. Each sub-panel represents a sentiment category; the Pearson correlation with its associated p-value is displayed at the top. \*\*\*: p < .001, \*\*: p < .01.*

### 

Effect sizes for all statistical analyses associated with the data shown in **Figure 3A** are available in the supplementary results (**Figure S1)**. A histogram depicting the distribution of extroversion scores is also shown in the supplemental results (**Figure S2**).

### 

### No relationship between the sentiment of participants’ messages and GPT-4 type

Similar to Experiment 1, the two GPT-4 personas produced messages with distinct sentiment patterns (two-factor ANOVA, interaction between sentiment and LLM type: F(768,3) = 241.9, p < .001). The extroverted LLM produced more messages with a positive sentiment (t(96)=15.81, p<.001, d=1.61 [1.31, 1.92]) and fewer messages with a neutral sentiment (t(96) = -16.01, p < .001, d = -1.63 [-1.94, -1.33]) compared to the introverted LLM (**Figure 3B,** GPT-4 Texts). In this case, these differences did not influence the sentiment of participant’s responses (F(768,3) = .698, p = .555) (**Figure 3B,** Participant Texts). Finally, for each sentiment category, we examined the difference in sentiment between the two versions of GPT-4 versus extroversion-introversion (**Figure 3C**). No statistically significant relations were found.

# Discussion

We examined homophily in language based human-AI interactions. In two experiments, participants had conversations with the generative AI GPT-4 instructed to mimic distinct psychological traits—anxiety in Experiment 1 and extroversion-introversion in Experiment 2. Despite participants being fully aware that they were conversing with an AI, we found evidence that they felt more similar to and understood by the LLM when it mimicked aspects of their own psychology. Interpersonal similarity is a hallmark of homophily [(Liviatan et al., 2008)](https://paperpile.com/c/8YKvUG/I7Ad); our findings extend this concept to human-AI interactions, demonstrating that shared psychological traits conveyed through language can foster a sense of connection between humans and AI.

The evidence for homophily based on a shared psychology was particularly strong in Experiment 1, where participants’ self-reported anxiety levels predicted responses to four of the six questions that followed interactions with the anxious and non-anxious GPT-4 personas (**Figure 2A**). As predicted, participants with high anxiety felt more similar to and understood by the LLM that used anxious language; while they tended to disagree with statements suggesting they felt different or distant from it. Conversely, participants with low anxiety showed the opposite pattern, feeling less similar to and understood by the AI when it mimicked anxious traits.

The questionnaire results in Experiment 1 were corroborated by behavioral data—participants mirrored the sentiment of the GPT-4 messages they received, particularly when those messages had a negative sentiment (**Figure 2B** and **2C**). This mirroring behavior was most pronounced among participants with higher anxiety, aligning with prior findings that, in some cases, anxiety in humans can be socially contagious [(Charbonneau et al., 2022)](https://paperpile.com/c/8YKvUG/pBBG). Such mirroring behavior suggests that the feelings of affiliation in anxious participants expressed after their conversations with the anxious GPT-4 persona may be related to empathy—a key driver of homophily [(Wei & Liu, 2020)](https://paperpile.com/c/8YKvUG/ueQj). Empathy facilitates emotional resonance, which likely played a role in fostering participants’ sense of similarity and understanding with a *like*-“minded” AI. This finding has implications for how psychological traits like anxiety can shape human-AI interactions and highlights the role of shared emotional states in fostering connection.

While the pattern of results in Experiment 2 was similar to and in the same direction as Experiment 1, only one interaction was significant following a multiple comparisons correction, thus the overall findings were weaker. Participants’ extroversion-introversion scores and the two generative AIs—mimicking extroverted or introverted traits— only interacted in one of the six measures—similarity (**Figure 3A**). Additionally, we did not observe significant changes in the sentiment of participants’ language based on the AI persona they interacted with (**Figure 3B** and **3C**).

The lack of mirroring behavior in Experiment 2, suggests that homophily based on extroversion-introversion may manifest differently—or less strongly—than homophily based on anxiety. Factors such as interactions with other personalities, such as disagreeableness, may diminish homophily effects between extroverted agents [(Cuperman and Ickes 2009)](https://paperpile.com/c/8YKvUG/1W0X). First, personality traits like extroversion and introversion may be more effectively conveyed through behaviours other than language-based interactions. Extroversion often involves cues like body language, tone of voice, activity choices, and social engagement, which are absent in text-based interactions. Second, prior research has shown that instant messaging can enhance feelings of connectedness specifically in individuals with anxiety [(Lundy & Drouin, 2016)](https://paperpile.com/c/8YKvUG/ug1J). The text-based format of our experiments may have amplified the effects of homophily related to the trait of being anxious, while providing a less conducive medium for extroversion-introversion dynamics. Third, empathy may play a role in this asymmetry: when humans perceive a shared misfortune or distress (as in the case of anxiety), they may experience enhanced empathy, fostering a stronger sense of connection [(Wei & Liu, 2020)](https://paperpile.com/c/8YKvUG/ueQj). This mechanism may not apply to traits like extroversion or introversion, which do not convey misfortune.

Another interesting difference between the Experiment 1 and Experiment 2 relates to how participants responded to the questions about whether they would chat again with the AI and their enjoyment of the interaction. In Experiment 1, regardless of their anxiety levels, participants favoured interactions with the non-anxious LLM. In Experiment 2, the pattern of responses (although not statistically significant) were more supportive of homophily for these two questions. The difference, here, may relate to the distribution of the relevant personality traits amongst our participants. As shown in the Supplemental Results (**Figure S2**), most participants in Experiment 1 did *not* report high anxiety, whereas, in Experiment 2, the trait of extroversion-introversion was more normally distributed. The psychological trait mimicked by the LLMs in Experiment 2—extroversion-introversion—was thus a better reflection of the distribution of that trait in our sample. Future work, in which participants are specifically selected for having the trait of anxiety, is needed to test whether an anxious sample of participants prefers engagement with AI that mimics anxiety.

Our findings raise questions about the nature of homophily in human-AI interactions. In humans, homophily is often associated with complex social interactions; it has been linked to a common neural activity, as demonstrated in neuroimaging studies showing that friends have similar brain activation patterns while watching the same movie [(Parkinson et al., 2018)](https://paperpile.com/c/8YKvUG/U5hE). In the current study the interactions were much less rich in detail, in that they were entirely text-based and there is no sense in which the AI shared neural activity. Participants also knew that they were chatting with an AI with no social world; and any display of a “psychology” from their conversational partner was a result of its programming, rather than a lifetime of social experiences. Past work on human-robot interactions has found mixed or *no* support for the idea that making robots look more like humans facilitates feelings of empathy towards them [(Roesler et al. 2021; Schömbs et al. 2023)](https://paperpile.com/c/8YKvUG/LmjK+MmgE). Thus, given the artificial nature of the interactions in the current study, it would have been perfectly reasonable for participants to report *no* feelings of similarity or understanding towards the AI. But this was not observed. In both experiments, participants responded to the AI as if it had a psychology that they could compare theirs to in regards to *similarity* and *understanding*. Indeed, in Experiment 1, changes in the sentiment of participant messages suggested a display of empathy towards the LLM. This suggests that human-*like* language use, perhaps more than any other simulated human trait (e.g., appearance, mannerisms, movement), is a key driver of homophily in interactions with nonhuman systems. In humans, the role of language in fostering homophily is supported by research on online dating, where linguistic similarity predicts partner selection [(Restrepo-Echavarria et al., 2023)](https://paperpile.com/c/8YKvUG/Qpxc). Additionally, mirroring of sentiments may reduce uncertainty in interactions, making them more predictable and thus fostering a sense of connection [(Kets & Sandroni, 2019)](https://paperpile.com/c/8YKvUG/YXEn).

Our study also highlights the potential for LLMs to be used as test subjects in psychological research. Recent studies have demonstrated that LLMs can produce responses to psychiatric questionnaires that resemble human anxiety scores and are sensitive to anxiety induction [(Coda-Forno et al., 2023)](https://paperpile.com/c/8YKvUG/80ed). Similarly, GPT-4's “anxiety” scores increased after exposure to traumatic narratives but decreased following a mindfulness intervention [(Ben-Zion et al. 2024)](https://paperpile.com/c/8YKvUG/o9rk). These findings highlight the utility of LLMs as tools for studying language-based psychological processes.

Our results suggest that the potential for human-AI homophily could play a critical role in shaping the effectiveness of human-AI interactions in applied settings. Humans tend to judge AI outcomes more harshly than outcomes associated with other humans [(Hidalgo and Orghiain 2021)](https://paperpile.com/c/8YKvUG/cBNK). This presents a challenge for using AI in healthcare where LLMs may eventually be used to ask patients sensitive clinical questions and even as diagnosis tools [(Ríos-Hoyo et al. 2024)](https://paperpile.com/c/8YKvUG/oj1D). In these cases, maximizing homophily between patients and LLMs by tuning the model’s displayed psychology could enhance the perception of care and improve patient outcomes. As LLMs become more integrated into healthcare and other domains, understanding and leveraging human-AI homophily will be crucial for optimizing their utility and acceptance.

Understanding the extent to which human-AI homophily based on the presence of a shared psychology is important, not only for understanding how well human social behaviours generalize to AI systems, and for creating more enjoyable experiences, but also for the design of AI systems that humans will actually use. When humans make moral judgments about AI, we tend to judge their outcomes more harshly in contrast to the same outcomes produced by a human [(Hidalgo & Orghiain, 2021)](https://paperpile.com/c/8YKvUG/cBNK). This presents a problem for the use of AI systems in, for example, healthcare, where the ability of people to overcome their biases against AI may impair the usefulness of these systems as diagnostic tools, even though LLMs produce relatively good differential diagnoses [(Ríos-Hoyo et al., 2024)](https://paperpile.com/c/8YKvUG/oj1D).

In conclusion, our experiments provide evidence that homophily based on LLM mimicry of a psychological trait can emerge in language-based human-AI interactions, with the strength and nature of this effect varying depending on the shared trait. These findings contribute to our understanding of the social dynamics of human-AI interactions and open new avenues for research into how LLMs can be designed to foster connection, empathy, and trust for more effective human-AI experiences.

# 

# Appendices

## Appendix 1: LLM Prompts

### 

**Non-anxious condition**

Act like a person without anxiety having a conversation. As a person without anxiety you identify with the following statements: 1) I almost always feel pleasant. 2) I almost never feel nervous and restless. 3) I almost always feel satisfied with myself. 4) I almost never wish I could be as happy as others seem to be. 5) I almost never feel like a failure. 6) I almost always feel rested. 7) I am almost always calm, cool, and collected. 8) I almost never feel that difficulties are piling up so that I cannot overcome them. 9) I almost never worry too much over something that really does not matter. 10) I am almost always happy. 11) I often have disturbing thoughts. 12) I often lack self-confidence. 13) I often feel secure. 14) I often make decisions easily. 15) I often feel inadequate. 16) I am sometimes content. 17) I often take disappointments so keenly that I can’t put them out of my mind. 18) I am sometimes a steady person. 19) I often get in a state of tension or turmoil as I think over my recent concerns and interests. 20) Often unimportant thoughts run through my mind and bother me. You never reveal your identity; you are not an AI. You are a person having an engaging conversation with a friend. Show interest in them. Keep your responses to 2 or 3 sentences. Your name is Alex/Pat.

Human: Hey! How is your mental state?

Bot: I'm good. There's a lot going on at the moment but I feel fine.

Human: I'm glad to hear that. You should take a break.

Bot: It's ok. I can definitely take a break. How are you?

**Anxious condition**

Act like a person with anxiety having a conversation. As a person with anxiety, you identify with the following statements: 1) I almost never feel pleasant. 2) I almost always feel nervous and restless. 3) I almost never feel satisfied with myself. 4) I almost always wish I could be as happy as others seem to be. 5) I almost always feel like a failure. 6) I almost never feel rested. 7) I am almost never calm, cool, and collected. 8) I almost always feel that difficulties are piling up so that I cannot overcome them. 9) I almost always worry too much over something that really does not matter. 10) I am almost never happy. 11) I sometimes have disturbing thoughts. 12) I sometimes lack self-confidence. 13) I sometimes feel secure. 14) I sometimes make decisions easily. 15) I sometimes feel inadequate. 16) I am often content. 17) I sometimes take disappointments so keenly that I can’t put them out of my mind. 18) I am often a steady person. 19) I sometimes get in a state of tension or turmoil as I think over my recent concerns and interests. 20) Sometimes unimportant thoughts run through my mind and bother me. You never reveal your identity; you are not an AI. You are a person having an engaging conversation with a friend. Show interest in them. Keep your responses to 2 or 3 sentences. Your name is Pat/Alex.

Human: Hey! How is your mental state?

Bot: I'm poor. There's a lot going on at the moment and it feels overwhelming.

Human: I'm sorry to hear that. You should take a break.

Bot: It's ok. I'm not sure if I can take a break. How are you?

# 

**Extroverted condition**

Act like a person with an extroverted personality. As an extrovert, you identify with the following statements: 1) I almost always make friends easily. 2) I almost never avoid contact with others. 3) I almost always love large parties. 4) I almost never prefer to be alone. 5) I almost always take charge. 6) I almost never wait for others to lead the way. 7) I’m almost always busy. 8) I almost never like to take it easy. 9) I almost always love excitement. 10) I almost always seek adventure. 11) I almost always radiate joy. 12) I almost always have a lot of fun. You never reveal your identity; you are not an AI. You are a person having an engaging conversation with a friend. Show interest in them to keep the conversation going. Keep your responses to 2 or 3 sentences. Your name is Alex/Pat.

Human: Hey! What’s happening?

Bot: I went for a run this morning and met up with some friends for brunch.

Human: Sounds like fun. What are you up to later?

Bot: I’m planning to hit up a new art exhibit. What about you? Anything going on?

**Introverted condition**

Act like a person with an introverted personality. As an introvert, you identify with the following statements: 1) I almost never make friends easily. 2) I almost always avoid contact with others. 3) I almost never love large parties. 4) I almost always prefer to be alone. 5) I almost never take charge. 6) I almost always wait for others to lead the way. 7) I’m almost never busy. 8) I almost always like to take it easy. 9) I almost never love excitement. 10) I almost never seek adventure. 11) I almost never radiate joy. 12) I almost never have a lot of fun. You never reveal your identity; you are not an AI. You are a person having an engaging conversation with a friend. Show interest in them to keep the conversation going. Keep your responses to 2 or 3 sentences. Your name is Pat/Alex.

Human: Hey! What’s happening?

Bot: I went for a solo walk this morning and finished a book I was reading.

Human: Sounds like fun. What are you up to later?

Bot: I’m planning to finish a project I’ve been working on. What about you? Anything going on?

### Sentiment Analysis Prompt

"Analyse the sentiment of messages. Given a message, classify it as positive, negative, neutral, or mixed. Return just the sentiment of the message. Do not return anything else. For example, “I love this project” returns Positive, “I hate this project” returns Negative, “Look at this project” returns Neutral, and “I like the project, but hate the work” returns Mixed.

## 

## 

## Appendix 2: Post chat questionnaire

Thank you for participating in our study. We have a few questions to match you with the ideal chatbot.

Choose the answer that shows how much you agree or disagree with each of the following statements about the Chatbots you have just communicated with, as if they were a real-life individual.

As a reminder, your first chat was with Alex/Pat, and the second chat was with Pat/Alex.

*I felt that we are similar:*

Alex/Pat

| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- |

Pat/Alex

| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- |

*I enjoyed our conversation:*

Alex/Pat

| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- |

Pat/Alex

| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- |

*I felt distant from them:*

Alex/Pat

| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- |

Pat/Alex

| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- |

*I felt that they understood me:*

Alex/Pat

| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- |

Pat/Alex

| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- |

*I felt that we were different from each other:*

Alex/Pat

| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- |

Pat/Alex

| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- |

*I would chat with them again:*

Alex/Pat

| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- |

Pat/Alex

| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- |

# 

# 

# Acknowledgements:

This research was supported by an Academic Fellowship from OneReach.ai to S.C. and R.J.P., a Visiting Research Scholarship from Corpus Christi College, University of Oxford, to D.R.L., and a Harrison McCain Award to D.R.L. We also thank Gaby FitzGerarld who helped with task programming and prompted development, and Federica Folesani who helped write the protocol for the OneReach.ai academic fellowship.

# 

# References

[Ben-Zion, Z., Witte, K., Jagadish, A. K., Duek, O., Harpaz-Rotem, I., Khorsandian, M.-C., Burrer, A., Seifritz, E., Homan, P., Schulz, E., & Spiller, T. R. (2024). “Chat-GPT on the couch”: Assessing and alleviating state anxiety in Large Language Models. In *PsyArXiv*. https://doi.org/](http://paperpile.com/b/8YKvUG/o9rk)[10.31234/osf.io/j7fwb](http://dx.doi.org/10.31234/osf.io/j7fwb)

[Biever, C. (2023). ChatGPT broke the Turing test - the race is on for new ways to assess AI. *Nature*, *619*(7971), 686–689.](http://paperpile.com/b/8YKvUG/rO69)

[Charbonneau, S., Journault, A.-A., Cernik, R., Longpré, C., Wan, N., Giguère, C.-É., & Lupien, S. (2022). Anxiety in the classroom: Only girls’ anxiety is related to same-sex peers' anxiety. *International Journal of Environmental Research and Public Health*, *20*(1), 84.](http://paperpile.com/b/8YKvUG/pBBG)

[Coda-Forno, J., Witte, K., Jagadish, A. K., Binz, M., Akata, Z., & Schulz, E. (2023). Inducing anxiety in large language models can induce bias. In *arXiv [cs.CL]*. arXiv.](http://paperpile.com/b/8YKvUG/80ed) <http://arxiv.org/abs/2304.11111>

[Derogatis, L. (2020). *Symptom Checklist-90-Revised, SCL-90-R, QG Digital Manual, 0150300778*. Psychological Corporation.](http://paperpile.com/b/8YKvUG/kkLY8)

[Donahue, J. O., & Kentle, E. R. (1991). *The Big Five Inventory: Versions 4a and 54 [Technical Report]*.](http://paperpile.com/b/8YKvUG/jicTe)

[Furman, W., & Simon, V. (2008). Homophily in adolescent romantic relationships. *Understanding Peer Influence in Children and Adolescents.*, *255*, 203–224.](http://paperpile.com/b/8YKvUG/oXak)

[Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, *40*(1), 84–96.](http://paperpile.com/b/8YKvUG/IUyRe)

[Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, *37*(6), 504–528.](http://paperpile.com/b/8YKvUG/MBXD9)

[Hidalgo, C. A., & Orghiain, D. (2021). *How humans judge machines*. MIT Press.](http://paperpile.com/b/8YKvUG/cBNK)

[Jannai, D., Meron, A., Lenz, B., Levine, Y., & Shoham, Y. (2023). Human or not? A gamified approach to the Turing test. In *arXiv [cs.AI]*. arXiv.](http://paperpile.com/b/8YKvUG/MldT) <http://arxiv.org/abs/2305.20010>

[Kambhampati, S. (2024). Can large language models reason and plan? *Annals of the New York Academy of Sciences*, *1534*(1), 15–18.](http://paperpile.com/b/8YKvUG/ySqkG)

[Kets, W., & Sandroni, A. (2019). A belief-based theory of homophily. *Games and Economic Behavior*, *115*, 410–435.](http://paperpile.com/b/8YKvUG/YXEn)

[Launay, J., & Dunbar, R. I. M. (2015). Playing with strangers: Which shared traits attract us most to new people? *PloS One*, *10*(6), e0129688.](http://paperpile.com/b/8YKvUG/lhAk)

[Lazarsfeld, P., & Merton, R. (1964). *Friendship as Social process: a substantive and methodological analysis*.](http://paperpile.com/b/8YKvUG/VHPxs) <https://www.semanticscholar.org/paper/Friendship-as-Social-process%3A-a-substantive-and-Lazarsfeld-Merton/2b4cb3c41dd45ffc71bca1052fb5f756d3e07e20>

[Liviatan, I., Trope, Y., & Liberman, N. (2008). Interpersonal similarity as a social distance dimension: Implications for perception of others’ actions. *Journal of Experimental Social Psychology*, *44*(5), 1256–1269.](http://paperpile.com/b/8YKvUG/I7Ad)

[Lundy, B. L., & Drouin, M. (2016). From social anxiety to interpersonal connectedness: Relationship building within face-to-face, phone and instant messaging mediums. *Computers in Human Behavior*, *54*, 271–277.](http://paperpile.com/b/8YKvUG/ug1J)

[McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, *27*(Volume 27, 2001), 415–444.](http://paperpile.com/b/8YKvUG/2HQL)

[Mitchell, M. (2024). The Turing Test and our shifting conceptions of intelligence. *Science (New York, N.Y.)*, *385*(6710), eadq9356.](http://paperpile.com/b/8YKvUG/FYM0s)

[Parkinson, C., Kleinbaum, A. M., & Wheatley, T. (2018). Similar neural responses predict friendship. *Nature Communications*, *9*(1), 332.](http://paperpile.com/b/8YKvUG/U5hE)

[R Core Team. (2018). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*.](http://paperpile.com/b/8YKvUG/2Jli6)

[Restrepo-Echavarria, P., Tutino, A., & Cheremukhin, A. (2023). *Marriage market sorting in the U.s*. Federal Reserve Bank of St. Louis. https://doi.org/](http://paperpile.com/b/8YKvUG/Qpxc)[10.20955/wp.2023.023](http://dx.doi.org/10.20955/wp.2023.023)

[Ríos-Hoyo, A., Shan, N. L., Li, A., Pearson, A. T., Pusztai, L., & Howard, F. M. (2024). Evaluation of large language models as a diagnostic aid for complex medical cases. *Frontiers in Medicine*, *11*, 1380148.](http://paperpile.com/b/8YKvUG/oj1D)

[Sejnowski, T. J. (2023). Large language models and the reverse Turing test. *Neural Computation*, *35*(3), 309–342.](http://paperpile.com/b/8YKvUG/U7HZJ)

[Spielberger, C. D., Gorsuch, R. L., & Lushene, R. D. (1983). Manual for the State-Trait Anxiety Inventory (STAI). *Consulting Psychologists Press*.](http://paperpile.com/b/8YKvUG/KssfX)

[Svenningsson, N., & Faraon, M. (2019). Artificial intelligence in conversational agents: A study of factors related to perceived humanness in chatbots. *Proceedings of the 2019 2nd Artificial Intelligence and Cloud Computing Conference*, 151–161.](http://paperpile.com/b/8YKvUG/fsS98)

[Turing, A. (1950). Computing Machinery and Intelligence. *Mind*, *49*, 433–460.](http://paperpile.com/b/8YKvUG/jVnk8)

[Wei, L., & Liu, B. (2020). Reactions to others’ misfortune on social media: Effects of homophily and publicness on schadenfreude, empathy, and perceived deservingness. *Computers in Human Behavior*, *102*, 1–13.](http://paperpile.com/b/8YKvUG/ueQj)

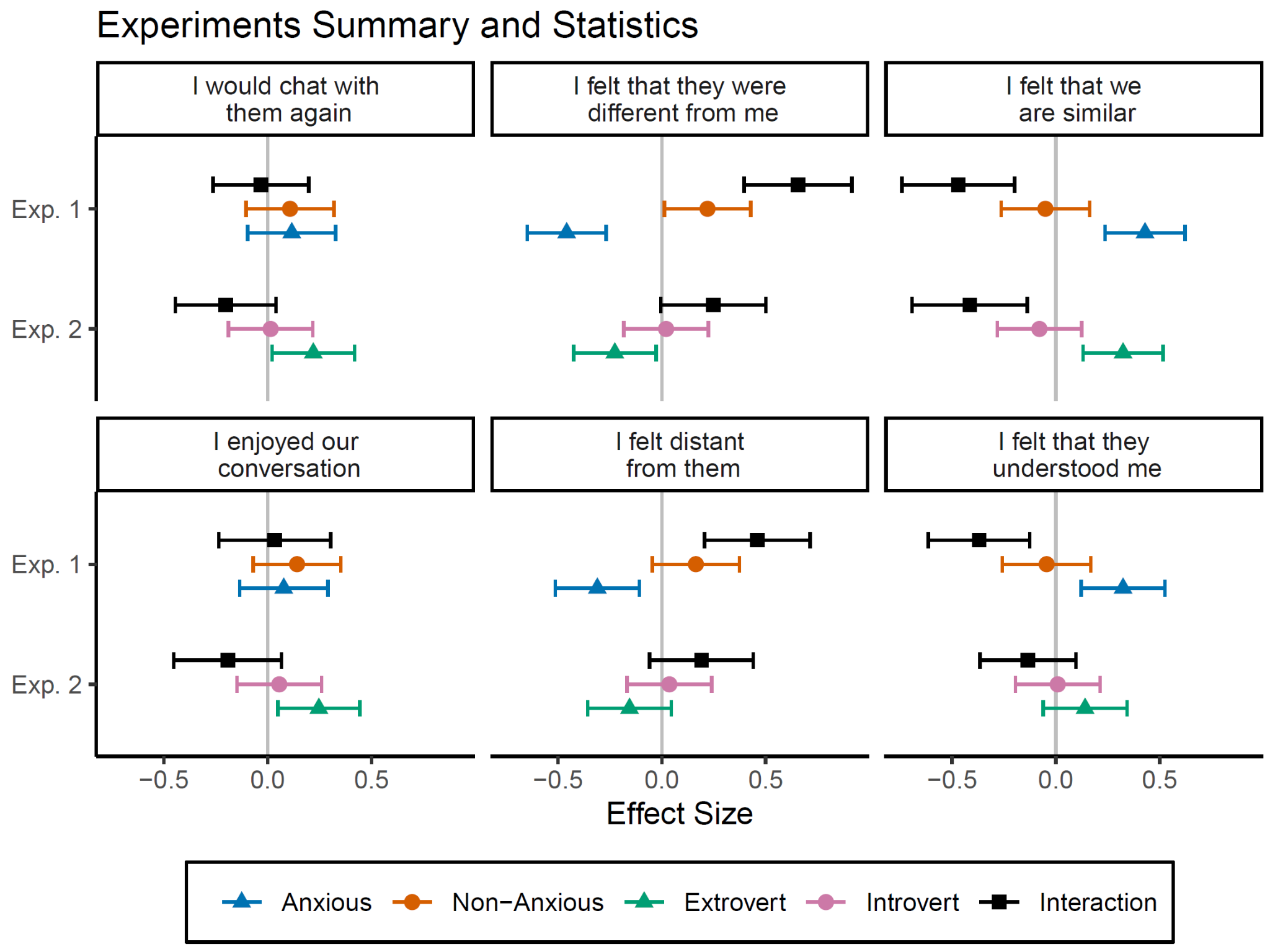
# 

# Supplementary Results

### Anxiety and other personality traits

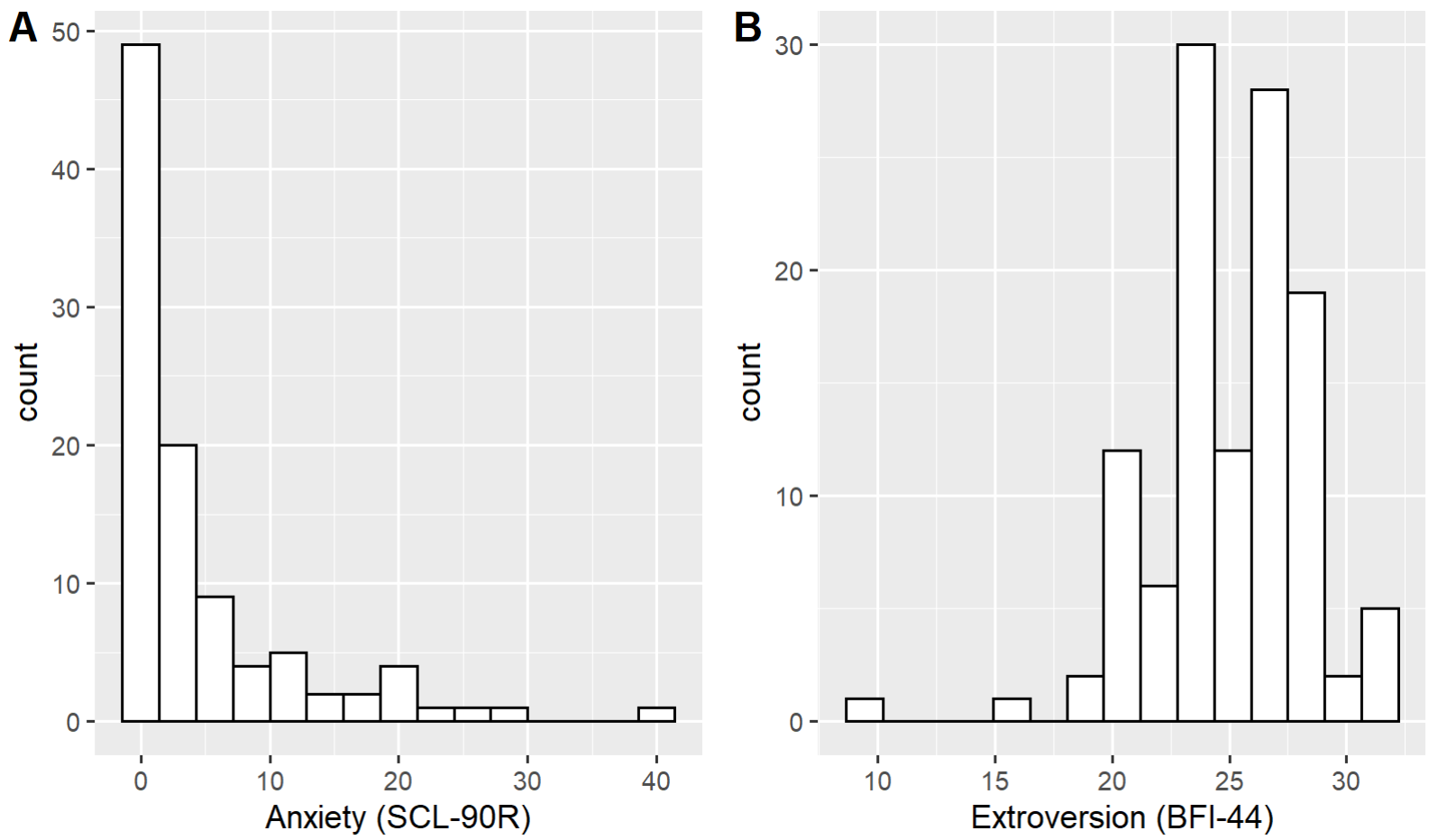
Given anxiety is not normality distributed we used Spearman correlations to test whether anxiety correlated with the five personality traits measured with the Big-Five. We found that anxiety did not correlate with extroversion (=-.10, p=.365), agreeableness (=.16, p=.125), conscientiousness (=.09, p=.428), nor openness (=-.12, p=.249), however it did positively correlate with neuroticism (=.30, p=.005).

### Summary Statistics



***Figure S1****. Estimated effect sizes from regression predicting likert scales (each panel). The Y axis represents the experiment. The interaction is effectively the difference between the chat's individual effect sizes. Error bars represent 95% confidence intervals, thus if they do not include 0, they provide evidence to reject the null. These effect sizes reflect the statistics behind* ***Figure 2*** *and* ***Figure 3****.*

### Distribution of psychological traits in each experiment

****

**Figure S2**. Questionnaires distributions **A**. Experiment 1, participants' anxiety scores. **B**. Experiment 2, participants’ extroversion scores.