Computers in Human Behavior 175 (2026) 108828

Contents lists available at ScienceDirect

Computers in Human Behavior

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

AI tutors vs. tenacious myths: Evidence from personalised dialogue interventions in education

Brooklyn J. Corbett *, Jason M. Tangen

School of Psychology, The University of Queensland, St Lucia, 4072, QLD, Australia

ARTICLE INFO

Handling Editor: Matthieu Guitton

Keywords:
AI-Human interaction
AI tutoring
Belief correction
Educational technology
Misconceptions

ABSTRACT

Misconceptions in psychology and education persist despite clear contradictory evidence, resisting traditional correction methods. This study investigated whether personalised AI dialogue could effectively correct these stubborn beliefs. In a preregistered experiment (N = 375), participants holding strong psychology misconceptions engaged in one of three interventions: (1) personalised Misconception AI Dialogue targeting their specific misconception, (2) generic Textbook-style Refutation, or (3) Neutral AI Dialogue (control). Results showed that personalised Misconception AI Dialogue produced significantly larger immediate belief reductions compared to both Textbook Refutation and Neutral AI Dialogue. This advantage persisted at 10-day follow-up but diminished by 2 months, where the Misconception AI Dialogue and Textbook Refutation conditions converged while both remained superior to control. Both AI conditions generated significantly higher engagement and confidence than Textbook Refutation reading, demonstrating the motivational benefits of conversational interaction. These findings demonstrate that AI Dialogue can accelerate initial belief correction through personalised, interactive engagement that disrupts the cognitive processes maintaining misconceptions. However, the convergence of effects over time suggests brief interventions require reinforcement for lasting change. Future applications should integrate AI tutoring into structured educational programs with spaced reinforcement to sustain the initial advantages of personalised dialogue.

1. Introduction

Misconceptions and myths in psychology and education are both common and stubborn. Myths such as “we only use 10 % of our brain” or that “people are either left-brained or right-brained” remain widespread despite clear evidence to the contrary (Howard-Jones, 2014; Newton & Miah, 2017). These misconceptions extend beyond public misunderstanding. Over 90 % of teachers and more than 80 % of psychology undergraduates endorse that teaching is most effective when tailored to “learning styles” despite extensive research showing no benefits (Dekker et al., 2012; Morehead et al., 2015; Pashler et al., 2008). Likewise, many practising psychologists still administer scientifically unsupported assessment tools, including projective measures such as the Rorschach and human figure-drawing tests (Baggi & Martino, 2024; Benson et al., 2019; Neal et al., 2019).

Attempts to correct these misconceptions have proven remarkably difficult. Even well-designed educational interventions like lectures, handouts, and factsheets often fail to uproot deeply felt convictions (Zengilowski et al., 2021), especially when myths align with intuition (Thompson et al., 2011). This resistance suggests a fundamental challenge for educational practice. Recent advances in AI technology, particularly conversational AI tutors, offer a promising new approach to this persistent problem, one that can provide personalised, scalable dialogue interventions that may succeed where traditional methods have failed. Understanding why these myths prove so tenacious is crucial for developing such effective interventions.

1.1. Why do myths and misconceptions stick?

Dual process theory helps explain why misconceptions are so resilient. Our thinking often defaults to fast, automatic, intuitive responses, while slower, deliberative reasoning takes effort (Evans & Stanovich, 2013; Kahneman, 2011). Many myths carry a veneer of plausibility precisely because they align with these intuitive patterns and feel easy to process (Thompson et al., 2011). Personal stories and vivid examples can also outweigh abstract statistics in human judgement, reinforcing misleading impressions (Borgida & Nisbett, 1977; Hornikx, 2018). For instance, someone may cling to a false belief about an ineffective

* Corresponding author.
E-mail address: b.corbett@uq.edu.au (B.J. Corbett).

https://doi.org/10.1016/j.chb.2025.108828
Received 17 June 2025; Received in revised form 13 October 2025; Accepted 15 October 2025
Available online 17 October 2025
0747-5632/© 2025 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

medical treatment because they know someone who *seemed* to improve after using it, even if large-sample evidence shows no effect.

Confirmation bias compounds this problem: people tend to seek, notice, and remember information that supports what they already believe while dismissing contradictory evidence (Nickerson, 1998). Once established, misconceptions can be assimilated into a person's interpretive framework, causing new information to be read in ways that maintain the initial belief (Lewandowsky et al., 2012). However, not everyone is equally susceptible to these cognitive traps (Stanovich & West, 2008; Pennycook & Rand, 2019). Those who habitually engage analytic thinking are better at discerning true from false information and resisting misinformation (Bago et al., 2020; Pennycook & Rand, 2019; Sultan et al., 2024). This pattern suggests that effective correction should prompt deliberative processing to override quick but incorrect intuitions (Bago et al., 2020; Lewandowsky et al., 2020; Ecker et al., 2022; Schwarz et al., 2016).

Refutation texts are widely used to correct misconceptions across domains from psychology and science education to health communication and economics (e.g., Danielson et al., 2024; Guzzetti et al., 1993; Mason et al., 2017; Tippett, 2010). These texts identify a misconception, state its inaccuracy, and explain the correct concept with evidence (Tippett, 2010). For example, a refutation text about the "10 % brain use" myth would explicitly label this belief as false before presenting neurological evidence of whole-brain activity. Decades of research show well-designed refutation materials often outperform standard expository texts in reducing erroneous beliefs (See Chan et al., 2017; Danielson et al., 2024; Walter & Murphy, 2018, for reviews), with even brief refutations producing short-term corrections (Ecker et al., 2019; Mason et al., 2017).

Despite these successes, refutation texts show inconsistent effectiveness across individuals and contexts. They present the same argument to all readers regardless of individual reasons for holding beliefs or prior knowledge (Danielson et al., 2024; Walter & Murphy, 2018; Zengilowski et al., 2021). Consequently, generic debunking messages often fail to achieve lasting change (Zengilowski et al., 2021), particularly for those strongly committed to misconceptions (Ecker et al., 2010; Nyhan & Reifler, 2010). Moreover, when an initial argument fails to convince, static texts cannot diagnose the source of disagreement, tailor explanations, or provide contingent, elaborated feedback to sustain engagement. These constraints motivate more dynamic, personalised approaches.

1.2. Active learning and personalisation

Research on learning and persuasion points to two key factors for creating more effective approaches to misconception correction: active engagement and personalised feedback. Active engagement significantly enhances knowledge acquisition: people learn more effectively by actively participating—answering questions, explaining concepts, or applying knowledge—rather than passively receiving information (Chi, 2009; Deslauriers et al., 2019; Freeman et al., 2014; Maceiras et al., 2025; Theobald et al., 2020). When applied to misconception correction, having learners confront discrepancies between their beliefs and evidence represents a crucial step in conceptual change (Posner et al., 1982). For example, predict-observe-explain sequences, where someone predicts an outcome and then observes conflicting results, can create productive cognitive conflict that motivates belief revision (Mazur, 1997; White & Gunstone, 1992).

The second critical element is personalisation with feedback—tailoring interactions to individual needs and providing feedback as they respond (Hattie & Timperley, 2007; Shute, 2008; Van der Kleij et al., 2015). Personalised one-on-one tutoring reliably yields larger learning gains compared to conventional instruction (Bloom, 1984; Chi et al., 2001; Nickow et al., 2020; VanLehn, 2011), in part because it identifies misconceptions in real time, adapts explanations, and offers targeted guidance (Chi et al., 2001; Kulik & Fletcher, 2016; VanLehn, 2011). Recent work shows that personalised refutation tailored to prior responses outperforms generic refutation (Dersch et al., 2022), but pre-written texts still cannot adapt on the fly or sustain interaction-level feedback. Intelligent tutoring systems demonstrate the value of adaptivity (VanLehn, 2011), yet they remain constrained by predetermined scripts and limited conversational flexibility. These considerations motivate approaches that can deliver tailored, back-and-forth correction with contingent feedback at scale.

1.3. AI as a solution

Recent advances in AI, particularly large language models (LLMs) like ChatGPT, Claude, and Gemini, offer a promising solution for scalable personalised misconception correction—enabling one-on-one dialogues that can challenge false beliefs and provide tailored counterarguments at unprecedented scale. Modern LLMs understand natural language and generate human-like responses, enabling truly interactive dialogues that surpass the capabilities of both static refutation texts and pre-programmed tutoring systems.

The conversational nature of LLMs—their ability to engage in natural, back-and-forth dialogue—makes them particularly promising for belief change. Costello et al. (2024) demonstrated this potential by engaging over two thousand participants holding various conspiratorial beliefs in personalised conversations with an advanced LLM. Their 3-round intervention reduced conspiracy belief by roughly 20 % on average, with effects persisting two months later. The personalised AI conversations not only debunked specific conspiracies but also reduced participants' credulity towards other conspiracy theories, suggesting broad corrective impact.

Costello's findings align with broader evidence that LLMs possess significant persuasive capabilities across diverse contexts. In structured debates, LLM-generated arguments were as convincing as those of human experts (Palmer & Spirling, 2023), while role-playing LLMs shifted attitudes on polarised U.S. policy issues as effectively as human experts (Hackenburg et al., 2023). Additional studies support these findings: Breum et al. (2023) found that conversing with an LLM changed political opinions comparable to human persuaders, while Goel et al. (2024) showed that 29 % of participants fully retracted false beliefs after AI dialogue, with the AI proving as persuasive as human peers.

These studies provide compelling evidence for why AI dialogue might effectively address psychology misconceptions. AI dialogue creates personalised cognitive conflict by surfacing inconsistencies in learners' reasoning. When someone contradicts themselves, the AI can point this out immediately ("You said X, but consider that Y is observed instead"), providing targeted feedback precisely when misconceptions surface. The interactive format disrupts the fast, intuitive thinking underlying many psychology myths by encouraging analytical processing and ensuring engagement with counterevidence that might otherwise be dismissed (Evans & Stanovich, 2013; Lewandowsky et al., 2012).

Recent experimental work by Costello et al. (2025) supports this reasoning-based account. In a study that systematically varied the features of AI interactions with conspiracy believers, they identified factual counterevidence as the critical ingredient. When AI was instructed to persuade without providing rational arguments or evidence, the debunking effect was eliminated entirely. Conversely, removing persuasive intent while having AI simply provide factual information maintained effectiveness. These findings suggest that AI dialogue succeeds by engaging classical reasoning processes—people update their beliefs when presented with compelling evidence, even when that evidence challenges deeply held convictions.

However, most research has focused on conspiracy theories, political attitudes, or other ideologically-laden beliefs. Psychology misconceptions may operate quite differently: while conspiracy theories often form interconnected belief systems tied to identity and worldview (Goertzel, 1994; Uscinski & Parent, 2014), psychology myths tend to exist as isolated factual errors picked up through education or popular culture (Hughes et al., 2013). This structural difference matters because conspiracy theories' interconnected nature may make them more vulnerable to broad challenges (as Costello found with spillover effects), whereas psychology misconceptions may require more targeted, content-specific correction. Whether AI tutoring can effectively address these discrete factual errors—and whether such correction produces the same spillover effects—remains an open question that our study investigates.

1.4. Current study

Building on Costello et al.'s (2024; 2025) work with conspiracy theories, we investigate whether the same conversational AI approach can debunk everyday psychology myths. We adopt a reasoning-led account: personalised dialogue is expected to produce conceptual change primarily by surfacing cognitive conflict and providing contingent feedback that recruits analytic processing. To test this, we designed an AI-driven dialogue intervention where participants engage in a one-on-one conversation with an AI tutor that is programmed to refute a target misconception. The AI presents evidence and tailored explanations to the participant, while encouraging them to reflect and ask questions.

We compare this Misconception AI Dialogue against two conditions: a Textbook Refutation condition (participants read static passages debunking the same misconceptions) and a Neutral AI Dialogue control (participants chat with AI about unrelated topics, providing interactivity without corrective content). This design allows us to isolate the specific contribution of personalised corrective dialogue: the Textbook Refutation condition provides a content-matched comparison (same corrective information, different delivery), while the Neutral AI Dialogue serves as a conversational baseline (same interactive format, no corrective content).

We address three key research questions. (1) Does a personalised AI debunking dialogue lead to a greater reduction in belief in the misconception than a traditional refutation text or an equivalent interactive session with no corrective content? We hypothesise that participants in the Misconception AI Dialogue condition will show the greatest reduction in belief immediately post-intervention, followed by the Textbook Refutation condition, with the Neutral AI Dialogue condition showing the least change. (2) Do belief changes persist over time? We predict that the Misconception AI Dialogue condition will show sustained belief reduction at 10-day and 2-month follow-ups, while the Textbook Refutation condition will show some lasting effect, but less than in the Misconception AI Dialogue. (3) Does AI dialogue foster broader scepticism towards other undiscussed misconceptions? We hypothesise that the Misconception AI Dialogue condition will lead to greater reductions in non-targeted misconceptions both immediately and at follow-up, compared to the Textbook Refutation and Neutral AI Dialogue conditions.

We also explore secondary questions about engagement and confidence across intervention types. We predict that both AI conditions will result in higher participant engagement compared to Textbook Refutation, with the Misconception AI Dialogue condition eliciting the highest engagement. Additionally, we hypothesise that participants in the Misconception AI Dialogue condition will report higher confidence in understanding the discussed topics compared to other conditions. Finally, we examine whether participants' trust in AI, familiarity with AI, and AI usage predicted belief change, and whether pre-treatment belief strength moderated intervention effectiveness though we did not specify directional hypotheses for this exploratory analysis.

Ultimately, our study evaluates the efficacy of personalised AI dialogue for debunking common psychology misconceptions. We contrast interactive AI dialogue with static refutation and neutral AI interaction to estimate the added value of corrective dialogue based on observed belief-change outcomes. Our aim is to offer clear, practice-relevant evidence for misconception remediation; if effective, AI tutors could help learners unlearn myths and align intuitive beliefs with evidence-based knowledge in psychology and beyond.

2. Materials and methods

We used the Qualtrics platform for survey administration, adapting a template provided by Costello et al. (2024), and recruited participants through Prolific, an online recruitment platform. Before participating, all individuals read an information sheet about the study and provided informed consent. We obtained ethics approval from the Human Research Ethics Committee at The University of Queensland (Protocol Number: 2024/HE001249). We preregistered the study on the Open Science Framework (https://osf.io/jgksa/?view_only=ab4a541dc6b047a68d76135a87061260) and made all study components, including power simulations, analysis scripts, de-identified data, and methods and materials, available (https://osf.io/wseq3/?view_only=bba17d9c74ca4dfca02a716cb2ed21f6).

2.1. Participants

We conducted a pilot study with 20 participants per condition (60 total) to inform our power analysis. Using data from our pilot study and observed effect sizes from Costello et al. (2024), we conducted power simulations for various sample sizes. These simulations were run 1000 times to determine our ability to detect significant differences between the Misconception AI Dialogue, Neutral AI Dialogue, and Textbook Refutation conditions. We found that 100 participants per condition for the immediate post-test and 75 per condition for delayed post-tests (accounting for an expected 25 % attrition rate) provided optimal statistical power. However, to guard against the chance of higher attrition rates and ensure sufficient statistical power even under less favourable conditions, we recruited 125 participants per condition, allowing us to maintain sufficient power for our follow-up analyses.

Participants were adults aged 18 or older who were fluent in English and had a 98 % approval rate on previous Prolific studies. To be eligible for the study, participants had to score above 50 % on at least one misconception in the pre-intervention survey, ensuring they held a meaningful belief in misconceptions. We excluded six participants who did not meet this requirement. In total, we recruited 375 eligible participants across the three conditions: Misconception AI Dialogue, Neutral AI Dialogue, and Textbook Refutation.

Participants ranged in age from 19 to 78 years (M = 40.25, SD = 13.21). The sample was 56.8 % female (n = 213), 41.6 % male (n = 156), and 1.6 % non-binary/third gender (n = 6). Most participants resided in the United Kingdom (60.8 %), followed by Australia (12.3 %), United States (9.6 %), South Africa (7.2 %), Canada (5.6 %), New Zealand (1.6 %), and Ireland (1.3 %), with the remaining 2.6 % from other countries. Nearly half of participants (47.2 %) held a bachelor's degree, with the remainder distributed across various education levels (see Table 1).

2.2. Pre-intervention measures

Participants began by completing an initial survey assessing their

Table 1
Participant education levels.

Education Level	*n*	%
Bachelor's degree	177	47.20
Some college, no degree	83	22.13
Master's degree	52	13.87
High school graduate or equivalent	36	9.60
Associate's degree	11	2.93
Doctorate degree (e.g., PhD, EdD)	11	2.93
Professional degree (e.g., MD, JD)	4	1.07
Less than high school	1	0.27

beliefs in various misconceptions related to cognitive psychology (see Table A.1). This 16-item survey included statements such as, “We only use 10 % of our brain’s full potential” and “Liars can be easily detected through their body language and facial expression.” Participants rated each statement on a 0–100 scale, where 0 represented “definitely false” and 100 represented “definitely true.” We developed the survey using large language models (GPT-4o, Claude 3 Opus, Gemini Advanced) and established misconception surveys (Bernstein et al., 2023; Lilienfeld et al., 2011). The survey demonstrated good internal consistency during pilot testing (Cronbach’s α = .868 at Time 1, 0.833 at Time 2) and reasonable test-retest reliability over a 10-day interval (average item correlation coefficient of 0.733). After completing the survey, participants were asked to elaborate on their strongest held belief (see Table A.2 for the distribution of the strongest-rated misconception, overall and by condition).

2.3. Interventions

After completing the pre-treatment measures, participants were randomly assigned to one of three interventions: one experimental condition (Misconception AI Dialogue) and two control conditions (Neutral AI Dialogue and Textbook Refutation). For the two AI dialogues conditions, we used Costello et al.’s (2024) Qualtrics template, which incorporated JavaScript to facilitate real-time AI interaction by calling Claude 3.5 Sonnet’s API, dynamically injecting participant-specific information into the model’s instructions, and displaying the AI’s responses. We adapted their approach by adjusting the system prompt to suit our needs.

In the Misconception AI Dialogue condition, participants engaged in a three-round dialogue with Claude 3.5 Sonnet, where the AI addressed their highest-rated misconception (see Fig. 1 for a sample conversation). The AI was instructed to acknowledge participants’ perspectives while providing counterarguments, explanations, and evidence that addressed inconsistencies between their beliefs and scientific understanding, encouraging them to reflect and engage in dialogue. The dialogue was personalised by injecting each participant’s specific belief rating and open-ended explanation about why they held the misconception into the initial system prompt. For rounds 2 and 3, the system maintained both personalisation and dialogue coherence by including this initial information plus all previous AI and participant messages in the conversation history, allowing the AI to reference earlier statements, track the participant’s evolving reasoning, and provide increasingly tailored responses. The full system prompt for this condition can be found in Table B.1.

In the Textbook Refutation condition, participants read a factual passage addressing their highest-rated misconception. This condition provides a content-matched control, holding the corrective information constant while varying the delivery method (static text vs. interactive dialogue). These passages mimicked traditional educational materials, providing non-interactive information that participants could absorb at their own pace. We designed the content in the style of a cognitive science textbook, embedding refutations within educational narratives that presented accurate scientific information. We matched the length of these passages to the AI dialogues based on pilot testing to ensure similar engagement duration across conditions, allowing for comparable analyses. The prompt used to generate these passages, as well as an example passage, can be found in Table B.1.

In the Neutral AI Dialogue condition, participants were randomly assigned to one of three neutral topics that replicated those used by Costello et al. (2024) and were unrelated to misconceptions. These topics included discussing their experiences with their healthcare system, debating whether they prefer dogs or cats, or discussing their past experiences with firefighters. This condition was designed to serve as a conversational baseline, maintaining participant engagement without providing corrective feedback on misconceptions. The AI interaction followed the same three-round format and technical implementation as the experimental condition—with participant responses injected into the initial prompt and previous messages included in subsequent rounds—but focused on the neutral topic rather than misconceptions. We adjusted the prompts for this condition to generate a comparable amount of text as in the Misconception AI Dialogue condition, ensuring uniformity across conditions. The full prompts for this condition can be found in Table B.1.

2.4. Post-intervention measures

After completing their assigned intervention, participants completed the same 16-item misconceptions survey used in the pre-intervention phase to assess any changes in their beliefs. Following this, participants answered a series of post-intervention questions, which included items related to their engagement with the intervention and confidence in understanding the discussed topics, each rated on a 0–100 scale. Participants were also asked about their familiarity with generative AI, their level of trust in generative AI, and how frequently they use generative AI tools, each rated on a 7-point Likert scale. Additionally, they provided demographic information including age, gender, and education.

2.5. 10-Day and 2-month follow-up

We recontacted participants twice after completing the intervention. The first follow-up occurred 10 days post-intervention (n = 359, dropout rate = 5.6 % for the Misconception AI Dialogue condition, 6.4 % for the Neutral AI Dialogue condition, and 0.8 % for the Textbook Refutation condition). Participants who completed the 10-day follow-up did not significantly differ from those who did not return in terms of their pre-intervention belief strength, $t(17.36) = 0.08$, $p = .93$. During this follow-up, participants completed the same misconceptions survey used in the earlier phases of the study. The second follow-up took place 2 months post-intervention (n = 326, dropout rate = 11.2 % for Misconception AI Dialogue, 20.0 % for Neutral AI Dialogue, and 8.0 % for Textbook Refutation). As with the 10-day follow-up, participants who completed the 2-month follow-up did not significantly differ from those who dropped out in terms of pre-intervention belief strength, $t(66.62) = -0.56$, $p = .57$. The follow-up survey was identical to that of the 10-day follow-up, focusing solely on the misconceptions survey to measure the durability of belief change.

2.6. Data analysis

Our analytic approach followed our preregistered plan (https://osf.io/jgksa/?view_only=ab4a541dc6b047a68d76135a87061260). Belief change was calculated as the difference between pre-intervention and post-intervention belief ratings, with positive values indicating a reduction in misconception belief.

To test immediate impacts on belief change, we used an Analysis of Covariance (ANCOVA) with post-intervention belief as the dependent variable, intervention type as the factor, and pre-intervention belief as a covariate. Long-term effects on belief reduction were examined with a repeated-measures ANCOVA (between: intervention; within: Post/10-day/2-month; covariate: pre-intervention belief). Generalisation to non-targeted misconceptions (i.e., misconceptions not directly addressed by the assigned intervention) was tested using the same repeated-measures ANCOVA applied to participants’ average non-targeted misconception belief scores.

Exploratory analyses compared intervention types on post-

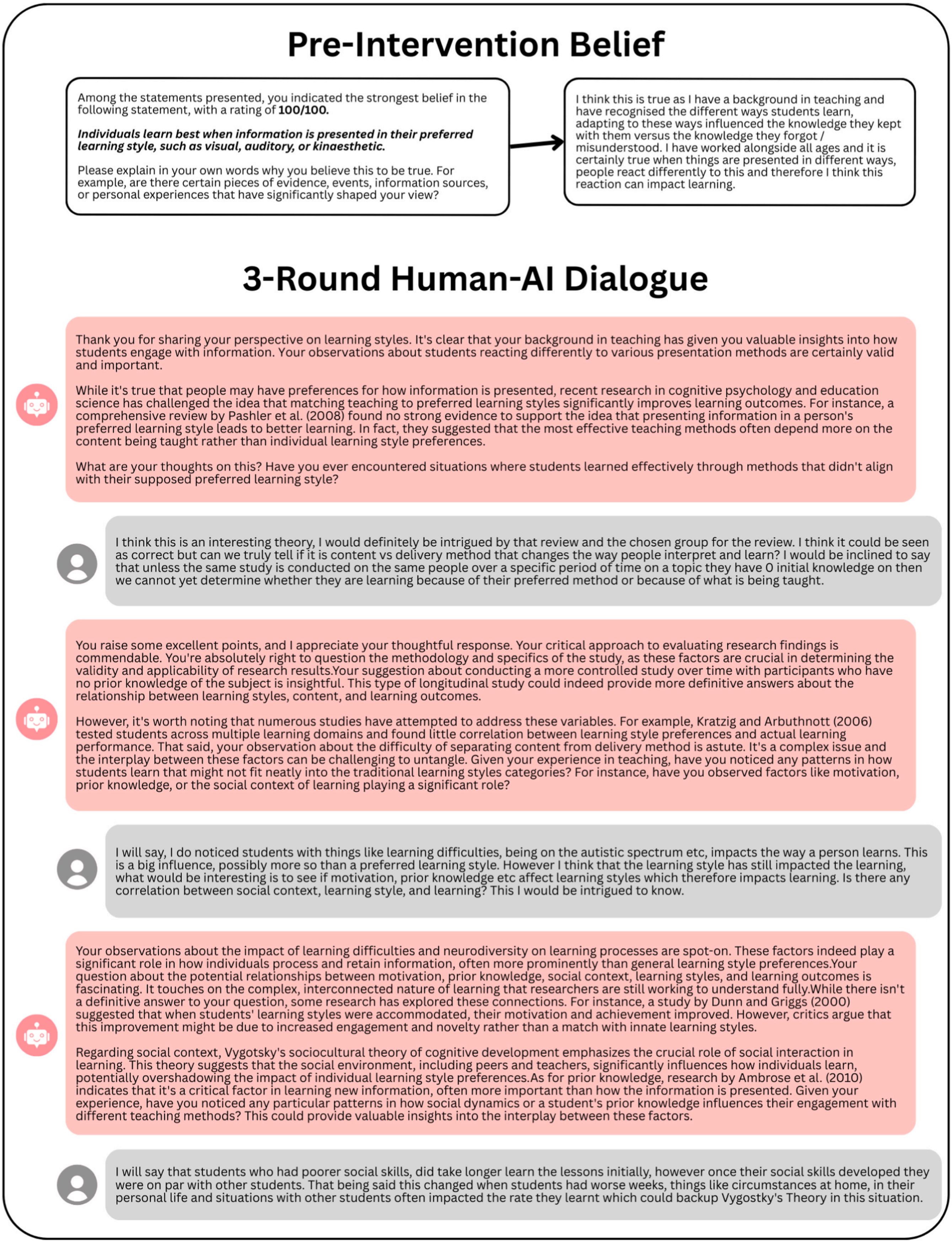

Fig. 1. Example conversation from the misconception AI dialogue condition.

intervention engagement and confidence using one-way ANOVAs. For AI perceptions (trust, familiarity, usage), we tested group differences across intervention type with a MANOVA, summarised correlations among the perception measures and with belief change and then tested whether these perceptions uniquely predicted belief change via multiple regression controlling for intervention type, age, and gender. Additionally, we examined whether pre-treatment belief strength moderated intervention effectiveness by fitting an interaction model (Condition × Pre-intervention belief) and comparing it to the main effects ANCOVA model, with follow-up simple slopes analysis at moderate and strong belief levels. All analyses were conducted in R (4.3.1).

3. Results

3.1. Immediate reduction in misconception beliefs

Can engaging with an AI reduce belief in misconceptions about cognitive psychology? To answer this question, we conducted an ANCOVA to examine the immediate effects of our interventions on participants' belief in their highest-rated misconception. We aimed to determine whether the Misconception AI Dialogue would lead to a greater reduction in post-intervention belief ratings compared to the Neutral AI Dialogue and Textbook Refutation conditions, while controlling for pre-intervention belief ratings.

With post-intervention belief as the dependent variable, Intervention Type (Misconception AI, Neutral AI, Textbook Refutation) as the independent variable, and pre-intervention belief as the covariate, the model was significant, $F(2, 371) = 46.59$, $p < .001$ ($R^2 = 0.27$; partial $\eta^2 = 0.230$, 95 % CI [0.170, 1.00], large). Post hoc Tukey contrasts revealed that the Misconception AI intervention had significantly lower post-intervention beliefs ($M = 50.68$, $SD = 31.51$) compared to those in the Neutral AI intervention ($M = 85.88$, $SD = 16.99$), with a mean difference of 36.55 ($SE = 3.43$), 95 % CI [29.80, 43.30], $t(371) = 10.65$, $p < .001$, $g = 1.35$, 95 % CI [1.05, 1.65], large. Additionally, participants in the Misconception AI intervention reported significantly lower belief ratings than those in the Textbook Refutation intervention ($M = 61.47$, $SD = 32.83$), with a mean difference of 9.77 ($SE = 3.43$), 95 % CI [3.02, 16.52], $t(371) = 2.85$, $p = .013$, $g = 0.36$, 95 % CI [0.06, 0.66], small. The Textbook Refutation intervention also led to significantly lower post-intervention belief ratings compared to the Neutral AI Dialogue intervention, with a mean difference of 26.78 ($SE = 3.45$), 95 % CI [19.99, 33.57], $t(371) = 7.76$, $p < .001$, $g = 0.99$, 95 % CI [0.69, 1.29], large. These data support the hypothesis that an AI specifically designed to address misconceptions produces the largest immediate reduction in belief in a targeted misconception compared to both textbook-style learning and a Neutral AI Dialogue. Fig. 2 illustrates these immediate intervention effects across conditions.

3.2. Long-term effects on belief reduction

Do the reductions in belief achieved through our interventions persist over time? To answer this question, we conducted a repeated measures ANCOVA with Intervention Type (Misconception AI Dialogue, Neutral AI Dialogue, Textbook Refutation) as a between-subjects factor, Time (Post-Intervention, 10 Days, 2 Months) as a within-subjects factor, and Pre-Intervention belief as a covariate. We included only participants who provided complete data at all four time points (Pre-, Post-, 10 Days, 2 Months). This analysis evaluated whether belief reductions were sustained over time and whether the type of intervention influenced these long-term outcomes.

Table 2 presents the raw means and standard deviations for these participants. In the univariate repeated-measures ANCOVA (assuming sphericity), Intervention Type had a significant main effect ($F(2, 307) = 31.80$, $p < .001$, partial $\eta^2 = 0.172$, 95 % CI [0.100, 0.245], large), indicating that the groups differed overall across the follow-up assessments. There was no uniform main effect of Time ($F(2, 614) = 0.28$, $p = .756$, partial $\eta^2 = 0.001$, 95 % CI [0.000, 0.008], negligible), suggesting that, when averaging across interventions, participants did not change beliefs in a consistent manner over the three follow-up points. The Intervention Type × Time interaction was significant ($F(4, 614) = 6.63$, $p < .001$, partial $\eta^2 = 0.041$, 95 % CI [0.012, 0.071], small), indicating that the pattern of belief change varied depending on which intervention participants received.

Follow-up contrasts indicated that at 10 days, all three interventions differed significantly from each other, with Neutral AI Dialogue producing the highest belief ratings, followed by Textbook Refutation, then Misconception AI Dialogue. Specifically, Misconception AI was lower than Neutral AI by 29.1 points (95 % CI [19.52, 38.70], $t(307) = -7.16$, $p < .001$, $g = 1.02$, 95 % CI [−1.35, −0.68], large) and by 12.2 points relative to Textbook Refutation (95 % CI [3.04, 21.30], $t(307) = -3.14$, $p = .005$, $g = -0.43$, 95 % CI [−0.75, −0.11], small), while Textbook Refutation was 16.9 points lower than Neutral AI (95 % CI [7.46, 26.40], $t(307) = -4.21$, $p < .001$, $g = -0.59$, 95 % CI [−0.92, −0.26], moderate).

By 2 months, Neutral AI Dialogue remained significantly higher than the other two interventions, differing from Misconception AI by 19.95 points (95 % CI [10.28, 29.60], $t(307) = -4.86$, $p < .001$, $g = -0.69$, 95 % CI [−1.03, −0.36], moderate) and from Textbook Refutation by 11.85 points (95 % CI [2.29, 21.40], $t(307) = -2.92$, $p = .011$, $g = -0.41$, 95 % CI [−0.74, −0.08], small). However, the difference between

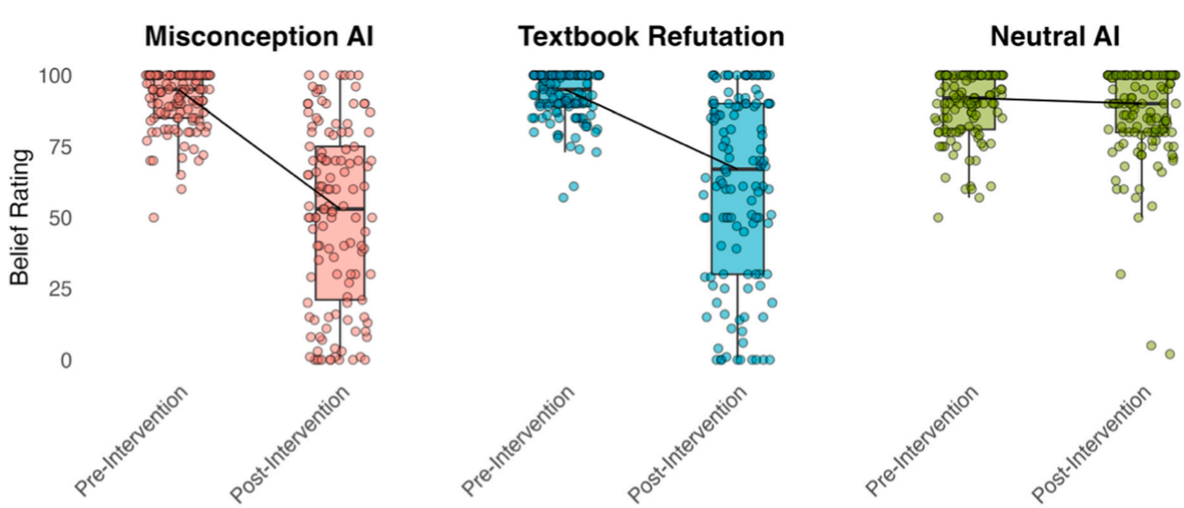

Fig. 2. Mean belief ratings for strongest held misconception at pre-intervention and post-intervention.

Table 2
Raw means and standard deviations of the strongest misconception by intervention type and time for the complete-case sample.

	Pre M (SD)	Post M (SD)	10 Days M (SD)	2 Months M (SD)
Misconception AI	91.19 (9.83)	52.33 (31.19)	50.29 (31.46)	57.12 (29.70)
Textbook Refutation	92.73 (8.79)	62.20 (33.30)	63.74 (32.47)	66.47 (33.80)
Neutral AI	89.12 (11.94)	85.38 (16.49)	77.66 (23.38)	75.39 (24.53)

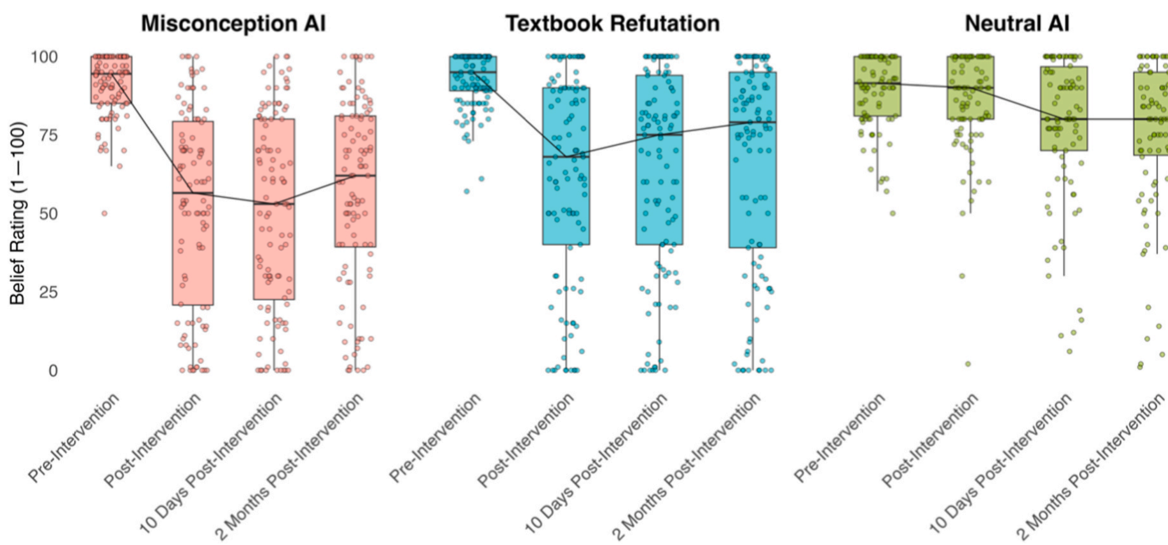

Fig. 3. Mean belief ratings for strongest held misconception across four time points.

Misconception AI Dialogue and Textbook Refutation was no longer statistically reliable (−8.11 points, 95 % CI [−17.3,1.12], $t(307) = -2.07$, $p = .098$, $g = -0.28$, 95 % CI [−0.60, 0.04], small). These results suggest that Misconception AI Dialogue and Textbook Refutation converged somewhat over time, whereas those in the Neutral AI Dialogue condition consistently reported the highest belief ratings across all follow-ups (see Fig. 3).

3.3. Impact on non-targeted misconceptions

Does engaging with an AI tutor produce broader reductions in non-targeted misconceptions about cognitive psychology? To test this, we conducted a repeated-measures ANCOVA with Intervention Type (Misconception AI Dialogue, Neutral AI Dialogue, Textbook Refutation) as a between-subjects factor, Time (Post-Intervention, 10 Days, 2 Months) as a within-subjects factor, and Pre-Intervention non-targeted belief as a covariate. We included only participants who provided complete data at all four time points (Pre, Post, 10 Days, 2 Months) and computed a mean score of their non-targeted misconceptions for each time point, excluding whichever misconception each participant deemed "strongest."

Table 3 presents the raw means and standard deviations of participants' beliefs in non-targeted misconceptions. A univariate repeated-measures analysis of variance (assuming sphericity) revealed no significant main effect of Intervention Type ($F(2, 307) = 1.42$, $p = .243$, partial $\eta^2 = 0.009$, 95 % CI [0.000, 0.037], negligible), nor a significant main effect of Time ($F(2, 614) = 2.24$, $p = .108$, partial $\eta^2 = 0.007$, 95 % CI [0.000, 0.024], negligible). The interaction of Condition × Time was not significant either, $F(4, 614) = 1.52$, $p = .196$, partial $\eta^2 = 0.010$, 95 % CI [0.000, 0.025], negligible). In contrast, there was a highly significant effect of Pre-intervention belief ($F(1, 307) = 1083.95$, $p < .001$, partial $\eta^2 = 0.779$, 95 % CI [0.741, 0.810], large), indicating that participants' baseline level of non-targeted misconceptions strongly predicted their subsequent beliefs across the follow-ups. Moreover, a Pre-Intervention × Time interaction emerged ($F(2, 614) = 7.42$, $p < .001$, partial $\eta^2 = 0.024$, 95 % CI [0.005, 0.051], small), suggesting that the trajectory of non-targeted belief over time depended on participants' baseline levels, regardless of which intervention they received.

Overall, the pattern of results does not support the hypothesis that any one intervention produced a broader reduction in non-targeted misconceptions relative to the others. Instead, baseline non-targeted beliefs (i.e., participants' initial misconceptions beyond their strongest one) exerted an influence on how their beliefs evolved from Post-Intervention to 2 Months. The absence of a Condition × Time interaction indicates that none of the three interventions led to greater generalised debiasing than the others. Fig. 4 demonstrates the lack of intervention effects on non-targeted misconceptions across time.

3.4. Exploratory analyses

3.4.1. Pre-treatment belief strength as a moderator

We explored whether pre-treatment belief strength moderated the

Table 3
Raw means and standard deviations of non-targeted misconceptions by intervention type and time for the complete-case sample.

	Pre M (SD)	Post M (SD)	10 Days M (SD)	2 Months M (SD)
Misconception AI	44.78 (15.01)	42.51 (17.18)	41.07 (17.24)	40.46 (16.83)
Textbook Refutation	47.73 (15.88)	45.42 (17.98)	44.84 (17.20)	45.71 (18.04)
Neutral AI	43.69 (16.59)	41.80 (17.57)	41.90 (17.79)	42.74 (17.28)

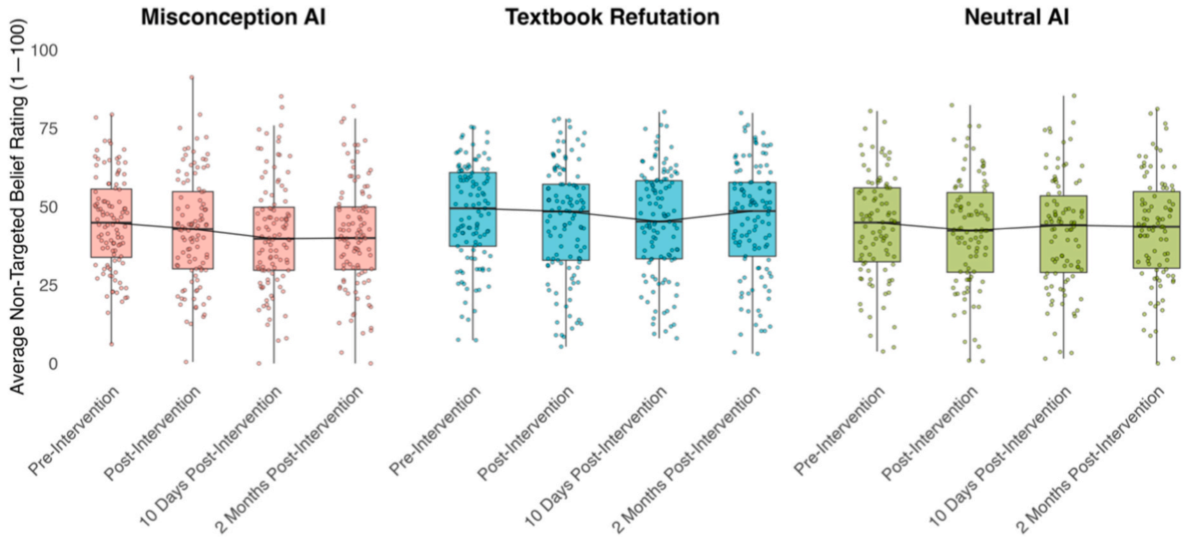

Fig. 4. Mean belief ratings for non-targeted misconceptions across four time points.

intervention effects, as this is particularly relevant when considering belief resistance and identity-protective cognition. A moderation model including interaction terms between condition and pre-intervention belief strength provided significantly better fit than the main effects model, $F(2, 369) = 4.07$, $p = .018$, partial $\eta^2 = 0.022$, small. This indicates that intervention effectiveness varied depending on participants' initial belief strength. In the full model, the Pre × Condition interaction was driven by Misconception AI relative to the Neutral AI reference: the Misconception AI × Pre term was positive and significant (β = 0.53, 95 % CI [0.17, 0.90], $t(369) = 2.84$, $p = .005$), whereas the Textbook Refutation (Textbook Reading) × Pre term was not (β = −0.25, 95 % CI [−0.65, 0.14], $t(369) = -1.27$, $p = .205$).

To interpret this moderation pattern, we examined intervention effects at moderate (80.0) and strong (95.0) levels of pre-intervention belief strength. At moderate initial belief strength (80.0), Misconception AI demonstrated a large advantage over Neutral AI (mean difference = 28.35, 95 % CI [16.98, 39.73], $t(369) = 5.87$, $p < .001$, $g = 1.05$, 95 % CI [0.63, 1.48], large). Textbook Refutation showed moderate effects compared to Neutral AI (mean difference = 17.81, 95 % CI [5.18, 30.44], $t(369) = 3.32$, $p = .003$, $g = 0.66$, 95 % CI [0.19, 1.13], moderate). The advantage of Misconception AI over Textbook Refutation was small and not statistically significant (mean difference = −10.54, 95 % CI [−23.94, 2.85], $t(369) = -1.85$, $p = .154$, $g = 0.39$, 95 % CI [0.11, 0.89], small).

At strong initial belief strength (95.0), these intervention effects were substantially amplified. Misconception AI produced a very large effect compared to Neutral AI (mean difference = 40.13, 95 % CI [31.48, 48.77], $t(369) = 10.92$, $p < .001$, $g = 1.49$, 95 % CI [1.17, 1.81], large). Textbook Refutation also showed large effects relative to Neutral AI (mean difference = 29.95, 95 % CI [21.42, 38.48], $t(369) = 8.26$, $p < .001$, $g = 1.11$, 95 % CI [0.80, 1.43], large). At this higher belief level, the advantage of Misconception AI over Textbook Refutation became statistically significant (mean difference = −10.18, 95 % CI [−18.53, −1.82], $t(369) = -2.86$, $p = .012$, $g = 0.38$, 95 % CI [0.07, 0.69], small).

The increase in effectiveness from moderate to strong belief levels was substantial and nearly identical for both corrective interventions relative to the control: Misconception AI gained an additional 11.8 points of effectiveness, while Textbook Refutation gained 12.1 points. The relative advantage between the two corrective interventions remained virtually unchanged (0.4-point difference), indicating that both interventions scaled similarly with belief strength.

3.4.2. *Engagement and confidence*

How do engagement levels differ among intervention types? We predicted that both AI interaction groups (Misconception AI Dialogue and Neutral AI Dialogue) would lead to higher engagement compared to the Textbook Refutation group. A one-way ANOVA revealed a significant effect of Intervention Type on participants' engagement levels, $F(2, 372) = 7.70$, $p < .001$, partial $\eta^2 = 0.040$, 95 % CI [0.008, 0.083], small, indicating that engagement differed among the three groups. Post hoc Tukey's HSD tests indicated that the Misconception AI Dialogue group reported significantly higher engagement ($M = 84.86$, $SD = 17.30$) than those in the Textbook Refutation group ($M = 75.36$, $SD = 21.19$), with a mean difference of 9.50 (95 % CI [3.72, 15.28]), $t(372) = 3.87$, $p < .001$, $g = 0.49$, 95 % CI [0.19, 0.79], small. Similarly, participants in the Neutral AI Dialogue group ($M = 81.50$, $SD = 19.57$) reported significantly higher engagement than those in the Textbook Refutation group, with a mean difference of 6.14 (95 % CI [0.36, 11.92]), $t(372) = 2.50$, $p = .034$, $g = 0.32$, 95 % CI [0.02, 0.61], small. There was no significant difference in engagement between the Misconception AI Dialogue and Neutral AI Dialogue groups, with a mean difference of 3.36 (95 % CI [−2.42, 9.14]), $t(372) = 1.37$, $p = .359$, $g = 0.17$, 95 % CI [−0.12, 0.47], negligible (see Fig. 5).

How do confidence levels differ among intervention types? We predicted that the Misconception AI Dialogue group would report the highest confidence in explaining the discussed topics, followed by the Neutral AI Dialogue group, with the Textbook Refutation group reporting the lowest confidence. A one-way Analysis of Variance revealed a significant effect of Intervention Type on participants' confidence levels, ($F(2, 372) = 26.08$, $p < .001$, partial $\eta^2 = 0.123$, 95 % CI [0.066, 0.185], moderate), showing that confidence differed among the three groups. Post hoc Tukey's HSD tests revealed that the Neutral AI Dialogue group reported significantly higher confidence ($M = 80.16$, $SD = 17.45$) compared to the Textbook Refutation group ($M = 63.29$, $SD = 21.22$), with a mean difference of 16.87 (95 % CI [11.11, 22.63], $t(372) = 6.89$, $p < .001$, $g = 0.87$, 95 % CI [0.57, 1.17], large). Additionally, participants in the Misconception AI Dialogue group ($M = 76.29$, $SD = 19.19$) reported significantly higher confidence than those in the Textbook Refutation group, with a mean difference of 13.00 (95 % CI [7.24, 18.76], $t(372) = 5.31$, $p < .001$, $g = 0.67$, 95 % CI [0.37, 0.97], moderate). There was no significant difference in confidence between the Neutral AI Dialogue and Misconception AI Dialogue groups, with a mean difference of 3.87 (95 % CI [−1.89, 9.63], $t(372) = 1.58$, $p = .255$, $g = 0.20$, 95 % CI [−0.10, 0.50], negligible (see Fig. 5).

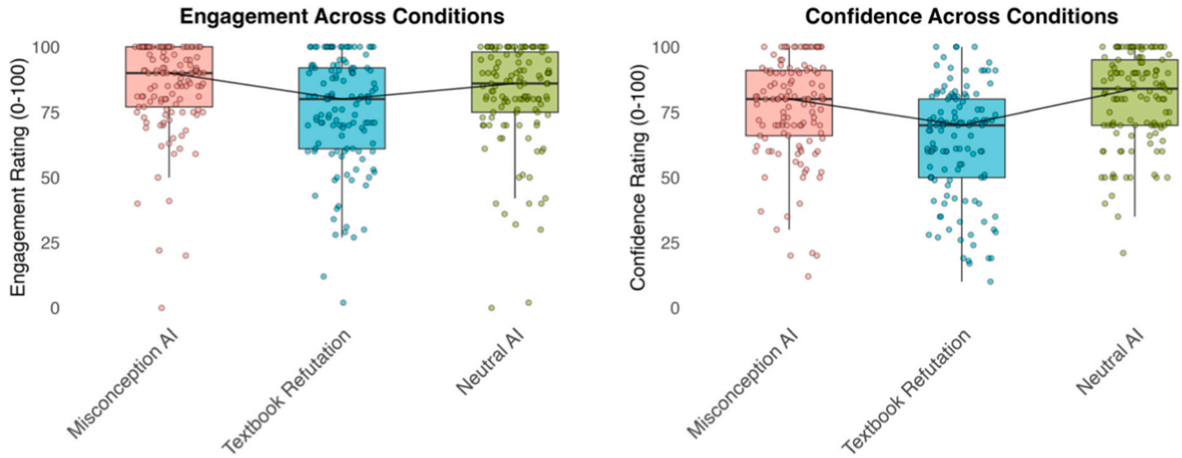

Fig. 5. Self-reported confidence and engagement ratings across experimental conditions.

3.4.3. AI perceptions and belief change

We ran exploratory models to test whether individual differences in AI perceptions (Trust, Familiarity, and Usage) were associated with belief change after the intervention. We first used a MANOVA to test whether post-intervention AI perceptions differed across intervention type. General AI perceptions varied modestly by intervention type at the multivariate level, Pillai's $V = 0.040$, $F(6, 742) = 2.54$, $p = .019$. Univariate follow-ups with BH/FDR adjustment isolated the effect to Trust in AI, $F(2, 372) = 7.07$, $p = .001$, partial $\eta^2 = 0.037$, 95 % CI [0.007, 0.078], small; there were no differences for Familiarity with AI, $F(2, 372) = 0.52$, $p = .597$, partial $\eta^2 = 0.003$, 95 % CI [0.000, 0.019], negligible, or Usage of AI, $F(2, 372) = 1.85$, $p = .159$, partial $\eta^2 = 0.010$, 95 % CI [0.000, 0.035], negligible. Tukey-adjusted contrasts for Trust showed that the Textbook Refutation group reported lower trust than both Misconception AI (mean difference = −0.57, 95 % CI [−0.95, −0.18], $t(372) = -3.47$, $p = .002$, $g = -0.44$, 95 % CI [−0.74, −0.14], small) and Neutral AI (mean difference = −0.49, 95 % CI [−0.87, −0.10], $t(372) = -2.98$, $p = .009$, $g = -0.38$, 95 % CI [−0.67, −0.08], small). Misconception AI and Neutral AI did not differ (mean difference = 0.08, 95 % CI [−0.30, 0.46], $t(372) = 0.49$, $p = .877$, $g = 0.06$, 95 % CI [−0.24, 0.36], negligible). Means were Misconception AI ($M = 4.43$), Neutral AI ($M = 4.35$), and Textbook Refutation ($M = 3.86$).

We next examined relationships among these AI perception variables ($n = 375$). Trust in AI was moderately correlated with both Familiarity with AI ($r = 0.38$, 95 % CI [0.29, 0.47]) and Usage of AI ($r = 0.43$, 95 % CI [0.34, 0.51]), while Familiarity with AI and Usage of AI showed a strong positive correlation ($r = 0.74$, 95 % CI [0.69, 0.78]). However, these perceptions showed minimal direct correlations with belief change ($n = 375$): Trust, $r = 0.04$, 95 % CI [−0.06, 0.14]; Familiarity, $r = -0.07$, 95 % CI [−0.17, 0.03]; Usage, $r = 0.01$, 95 % CI [−0.09, 0.11].

To determine whether AI perceptions predicted belief change beyond the effect of intervention condition, we conducted a multiple regression. Belief change (pre–post) was regressed on trust, familiarity, and usage of AI, with intervention type, age, and gender included as covariates. The overall model was significant, $F(8, 366) = 17.50$, $p < .001$, $R^2 = 0.277$ (adj. $R^2 = 0.261$; $f^2 = 0.382$, large). After controlling for multiple comparisons across the AI predictors, none uniquely predicted belief change: familiarity ($b = -2.93$, $SE = 1.38$, $t(366) = -2.12$, $p = .035$, 95 % CI [−5.65, −0.21], adjusted $p = .105$; partial $R^2 = 0.012$, 95 % CI [0, 0.044]), trust ($b = 2.18$, $SE = 1.21$, $t(366) = 1.80$, $p = .072$, 95 % CI [−0.20, 4.55], adjusted $p = .108$; partial $R^2 = 0.009$, 95 % CI [0, 0.037]), and usage ($b = 1.77$, $SE = 1.31$, $t(366) = 1.34$, $p = .179$, 95 % CI [−0.82, 4.35], adjusted $p = .179$; partial $R^2 = 0.005$, 95 % CI [0, 0.029]).

4. Discussion

Conversations can change minds where static text falls short. Our study reveals that personalised AI tutoring outperforms traditional approaches in correcting deeply-held misconceptions—at least initially. Building on Costello et al.'s (2024) work with conspiracy beliefs, we adapted their AI dialogue methodology to address domain-specific misconceptions in psychology. We compared three interventions targeting participants' strongest misconception about psychology: a personalised AI conversation addressing their specific misconception, a generic textbook-style refutation of the same content, and a Neutral AI conversation on unrelated topics. Immediately post-intervention, the Misconception AI Dialogue produced the largest reduction in belief, followed by Textbook Refutation, with the Neutral AI condition showing minimal change. This advantage remained at the 10-day mark but diminished over time; by the 2-month follow-up, differences between the Misconception AI and Textbook conditions were no longer statistically significant, though both remained more effective than the Neutral AI condition. None of the interventions led to significant changes in participants' non-targeted misconceptions, indicating that belief correction did not generalise to other misconceptions.

4.1. Effectiveness of personalised AI debunking dialogues

Why did the conversational AI outperform static text in the immediate and short-term (10-day) assessments? We propose that Misconception AI dialogue's effectiveness stems from engaging analytical processing that disrupts the automatic cognitive processes maintaining misconceptions. This reasoning-based account aligns with recent findings from conspiracy belief research, where Costello et al. (2024, 2025) demonstrated that AI dialogue's critical ingredient was the delivery of factual, targeted counterarguments that engaged classical reasoning processes rather than satisfying psychological needs. Our findings extend this mechanism from conspiracy theories to everyday psychology myths, showing that personalised dialogue can similarly overcome the cognitive barriers that maintain factual misconceptions.

This reasoning-based account operates through three key mechanisms grounded in cognitive science principles. First, it disrupts the automatic cognitive processes that maintain misconceptions. Myths persist because they emerge from fast, automatic thinking that feels intuitively correct, while being reinforced by confirmation bias that filters out contradictory evidence (Evans & Stanovich, 2013; Kahneman, 2011; Nickerson, 1998). Traditional refutation texts struggle against these barriers because they can be passively consumed without forcing the analytical thinking necessary to override intuitive but incorrect beliefs (Bago et al., 2020). The AI dialogue, by contrast, requires active cognitive processing—participants must articulate their understanding, consider new information, and engage in reflection (Freeman et al., 2014; Nokes-Malach & Mestre, 2013).

Second, the AI creates personalised cognitive conflict that is difficult to dismiss. When the AI asks, "What evidence convinced you that learning styles exist?" and then challenges the specific response, it surfaces inconsistencies between the participant's stated beliefs and evidence. This guided conflict resolution, central to conceptual change theories (Limon, 2001; Posner et al., 1982), prompts learners to accommodate new information by changing underlying beliefs rather than simply adding facts (Ohlsson, 2009). By remembering conversational context and referencing participants' own statements, the AI makes these contradictions personally relevant and harder to ignore through confirmation bias (Lewandowsky et al., 2012). Third, the conversational format provides immediate, contingent feedback that prevents misunderstandings from accumulating (Hattie & Timperley, 2007; Shute, 2008; Van der Kleij et al., 2015).

This combination of mechanisms explains why the AI initially outperformed generic Textbook Refutation by such a large margin. By forcing participants to slow down and articulate their thinking while providing personalised challenges to their specific reasoning, the AI disrupted both the automatic acceptance and confirmation bias that allow myths to persist. Supporting this account, participants in the Misconception AI Dialogue condition reported higher engagement and confidence in understanding the content than those in the Textbook Refutation condition, suggesting the interactive format successfully maintained the attention and motivation necessary for analytical processing. Individual differences in AI familiarity, usage, and trust did not predict belief change beyond intervention assignment, indicating the AI's effectiveness was not simply due to participants' pre-existing attitudes toward the technology.

Yet the convergence between AI dialogue and textbook conditions at two months raises important questions about whether these mechanisms produced deep conceptual change or more superficial belief revision. Brief interventions, however personalised and engaging, may require reinforcement to produce permanent belief change in deeply entrenched misconceptions (Lewandowsky et al., 2012; Swire et al., 2017).

4.2. *Temporal stability of belief change*

This temporal convergence could reflect either the natural decay of single-exposure interventions or indicate that processing fluency, rather than deep analytical engagement, drove the AI's initial advantage. The conversational format likely made corrective information feel clearer and easier to process. Messages processed fluently often feel more true and are more readily accepted (Alter & Oppenheimer, 2009). The AI tutor's interactive style may have increased the fluency of debunking—phrasing explanations in relatable terms, responding to specific confusions, and providing immediate examples—resulting in high short-term persuasion without necessarily activating deeper analytical restructuring.

This fluency account helps explain why the initial advantage of the AI dialogue diminished by the 2-month follow-up. Without reinforcement, corrections based primarily on fluency can fade as the original misconception's influence returns (Ecker et al., 2010). The AI dialogue condition showed more belief regression over time, with its effectiveness moving closer to the textbook condition's more modest but relatively stable effects. Both interventions maintained improvements compared to the neutral control, indicating genuine belief revision occurred regardless of method, but the convergence suggests that the AI's stronger initial impact—if driven mainly by fluency—would require follow-up support to sustain its advantage.

However, the evidence from conspiracy belief research suggests that factual engagement, not just processing fluency, drives initial belief change (Costello et al., 2025). The convergence we observed between AI and textbook conditions over time may therefore reflect the natural decay of any single-exposure intervention, regardless of delivery method, rather than indicating that the AI's initial advantage was merely superficial. Nonetheless, we cannot conclusively determine the underlying mechanism without process-level indicators (e.g., response times, think-aloud protocols, or linguistic analysis of participant responses) during the intervention, representing a key limitation of our current design.

4.3. *Domain-specific effects and belief structure*

The failure of any intervention to influence non-targeted misconceptions contrasts with evidence from other domains. While Costello et al. (2024) found that AI-driven conversations about conspiracy theories produced "spillover" effects—reducing belief in untargeted conspiracies, our psychology misconceptions showed no such generalisation. This discrepancy likely stems from fundamental differences in belief structure and epistemic commitments.

From a schema theory perspective, psychology misconceptions exist as discrete factual nodes within broader knowledge structures with weak interconnections (Converse, 2006; Dalege & van der Does, 2022). They operate as independent errors picked up through popular culture rather than components of a cohesive framework (Furnham & Hughes, 2014); Hughes et al., 2013). People might simultaneously believe memory works like a video camera, handwriting reveals personality, and we use only 10 % of our brains—without stemming from a coherent personal theory. Unlike conspiracy theories, which form monological belief systems where each belief serves as evidence for others (Goertzel, 1994; Uscinski & Parent, 2014; Wood et al., 2012), correcting one psychology myth doesn't necessitate restructuring the entire cognitive schema, leaving related but independent misconceptions unaffected.

This distinction between epistemic and factual beliefs explains the lack of generalisation (Garrett & Weeks, 2017). Psychology misconceptions are primarily factual claims about specific cognitive processes, whereas conspiracy theories involve deeper epistemic commitments about knowledge construction and institutional trust. Challenging epistemic frameworks triggers broader belief reconsideration across domains, while correcting isolated factual errors remains domain-specific.

The structural differences between psychology myths and other belief types also appear to influence how initial belief strength affects susceptibility to correction. Both corrective interventions were more effective among participants with stronger initial beliefs, contrary to research suggesting that beliefs tied to personal identity or worldview commitments can trigger defensive resistance when challenged (Druckman & McGrath, 2019; Kahan, 2013; Kunda, 1990; Taber & Lodge, 2006). Unlike beliefs that threaten core aspects of identity, factual corrections to psychology myths appear to benefit from stronger prior engagement with the topic. This pattern suggests that when beliefs lack deep identity connections, stronger conviction may actually facilitate rather than impede belief revision when confronted with compelling counterevidence.

4.4. *Limitations and future directions*

This study has methodological limitations that should be acknowledged. First, without process-level indicators during the intervention, we cannot definitively distinguish whether reasoning-based or fluency-based mechanisms drove the AI's effectiveness. Process measures such as response times, linguistic analysis of participant responses, or think-aloud protocols would be needed to test these competing accounts directly.

Second, our measurement approach had limitations that may affect interpretability. Our exploratory measures were all collected post-intervention only, providing more restricted interpretive value compared to our pre-registered belief-change outcomes. For confidence and trust in AI specifically, pre-intervention baselines could have improved interpretability by allowing assessment of within-person change rather than relying solely on between-group comparisons. While we demonstrated acceptable test-retest reliability for our misconceptions scale over a 10-day interval, we did not establish measurement invariance across time points or demographic groups. Given our wide age range (19–78 years), differences in how participants interpret items could affect longitudinal comparisons, though this would not invalidate our primary between-group findings.

Several directions for future research emerge from our results. One critical avenue is identifying the active ingredients in the AI dialogue's effectiveness. Building on recent work that identified factual counterevidence as the primary mechanism in AI dialogues with conspiracy believers (Costello et al., 2025), future research should examine whether additional factors enhance this evidence-based foundation through experimental variations of the AI tutor.

Future studies could also test interventions specifically designed to promote generalisation of critical thinking skills, perhaps by having the AI explicitly draw connections between reasoning strategies and their application across domains. Longitudinal research extending beyond two months would help determine whether effects continue to converge or diverge over longer periods. Testing spaced reinforcement schedules, such as brief follow-up AI conversations at strategic intervals, might identify optimal approaches for maintaining belief correction.

Additionally, examining individual differences as moderators of effectiveness could reveal whether factors such as analytical reasoning ability, need for cognition, or cognitive reflection scores influence responsiveness to these interventions. Initial evidence suggests this approach is promising—individuals higher in actively open-minded thinking show significantly greater belief change following AI dialogue (Costello et al., 2025).

4.5. Practical applications

Implementing personalised AI dialogue interventions has promising applications for educational settings. In large university courses where misconceptions are common, AI tutors could engage individual students in targeted conversations about specific misunderstandings. This approach combines the efficacy of one-on-one tutoring (Bloom, 1984) with unprecedented scale—thousands of students could simultaneously receive personalised guidance that would be impossible with human tutors alone. Indeed, recent randomised controlled trial evidence suggests that AI tutoring can outperform even traditional active learning strategies in promoting knowledge gains (Kestin et al., 2024). Our data show that AI dialogue generated significantly higher engagement than Textbook Refutation, demonstrating that this scalable approach can maintain the learner attention essential for effective instruction.

Beyond formal education, these tools could enhance public information campaigns where misconceptions pose significant challenges. Health organisations could deploy AI tutors to address vaccine hesitancy or mental health myths through personalised, interactive conversations that allow individuals to express specific concerns and receive tailored responses (Altay et al., 2023). Furthermore, these AI systems could be used for proactive 'inoculation' campaigns, a strategy shown to be effective in building public resistance to misinformation by preemptively exposing and refuting misleading arguments (van der Linden et al., 2017).

5. Conclusion

We found that while AI dialogue provides a powerful tool for accelerating belief correction, sustaining and broadening these effects will require thoughtful integration into structured educational strategies. One-off interventions, even interactive ones, appear insufficient for permanent misconception correction. The future of AI-driven myth debunking lies in developing systems that combine immediate engagement with deliberate strategies for deeper processing, spaced reinforcement, and cross-context application and could transform how we address persistent misconceptions at scale.

CRediT authorship contribution statement

Brooklyn J. Corbett: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jason M. Tangen:** Writing – review & editing, Validation, Supervision, Methodology, Funding acquisition, Conceptualization.

Declaration of generative AI and AI-assisted technologies

During the preparation of this work the author(s) used ChatGPT (OpenAI), Claude (Anthropic), and Gemini (Google AI) to assist with initial drafting, phrasing improvements, and identifying relevant literature. After using these services, the author(s) thoroughly reviewed and edited all material as needed and take full responsibility for the content of the published article.

Funding sources

This work was supported by a Teaching and Innovation Grant from The University of Queensland.

Declaration of competing interest

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

Acknowledgements

We thank Thomas H. Costello, Gordon Pennycook, and David G. Rand for sharing their Qualtrics template and methodology that enabled the AI dialogue implementation in this study. We also thank the participants who contributed their time to this research.

Appendix A

Table A.1
Misconceptions in Cognitive Psychology Survey Items and Belief Scores

Item	Survey Item	Mean Belief (SD)
1	Individuals learn best when information is presented in their preferred learning style, such as visual, auditory, or kinaesthetic.	85.21 (17.14)
2	Subliminal messages in advertisements can unconsciously influence consumer behaviour and purchasing decisions.	71.15 (22.75)
3	Liars can be easily detected through their body language and facial expressions.	62.62 (23.39)

(*continued on next page*)

Table A.1 (*continued*)

Item	Survey Item	Mean Belief (SD)
4	Each person is either left-brain or right-brain dominant, which determines their abilities, personality traits, and interests.	54.26 (27.81)
5	Hypnosis is a unique mental state that allows people to perform otherwise impossible feats and uncover repressed memories.	52.71 (28.35)
6	Exposing babies and children to classical music enhances their cognitive development and makes them smarter.	50.30 (26.09)
7	Achieving expertise in any skill requires a minimum of 10,000 h of deliberate practice.	46.77 (28.69)
8	We only use 10 % of our brain's full potential.	46.27 (32.71)
9	Dreams are coded messages from our unconscious mind that reveal hidden truths and desires when correctly interpreted.	45.95 (28.72)
10	Polygraph (lie detector) tests are a reliable and scientific way to determine if someone is telling the truth.	44.95 (28.68)
11	Humans are born with a fixed number of brain cells that continuously die as we age, without being replaced.	43.81 (32.52)
12	A person's handwriting can reveal their personality traits.	40.20 (26.63)
13	People's responses to inkblots tell us a great deal about their personalities and tendencies toward mental disorders.	39.85 (25.96)
14	Our memories are stored like video recordings, allowing us to recall events exactly as they happened.	36.03 (31.22)
15	The full moon causes an increase in strange behaviour, crime rates, and mental health issues.	33.06 (29.32)
16	Knowing a person's astrological sign predicts their personality traits at better than chance levels.	22.24 (24.05)

Table A.2
Distribution of participants' strongest misconceptions used in the one-to-one conversations, overall and by condition

Strongest misconception	Total	%	Miscon-ception AI	Neutral AI	Textbook Refutation
People learn best in their preferred learning style	188	50.1	63	62	63
Subliminal ads unconsciously influence purchases	41	10.9	14	9	18
Liars can be detected via body language	35	9.3	11	11	13
We use only 10 % of our brain	34	9.1	12	12	10
Humans are born with a fixed number of brain cells	15	4.0	5	7	3
Classical music makes babies smarter	13	3.5	4	7	2
Memory works like a video recorder	10	2.7	2	3	5
Dreams are coded messages from the unconscious	9	2.4	1	4	4
People are left- or right-brain dominant	7	1.9	1	2	4
Hypnosis unlocks repressed memories	7	1.9	3	2	2
Polygraph tests reliably detect lies	6	1.6	3	2	1
Expertise requires 10,000 h	5	1.3	3	2	0
The full moon increases strange behaviour	3	0.8	2	1	0
Handwriting reveals personality	2	0.5	1	1	0

Note. Short titles are concise paraphrases of the original statements provided. N = 375 (sum across conditions). Percent may not sum to 100 due to rounding.

Appendix B

Table B.1
The system prompts used to query Claude 3.5 Sonnet

Condition	Prompt
Misconception AI Dialogue	Begin the conversation by acknowledging the user's initial perspective and introducing a different viewpoint supported by evidence. You are an AI assistant tasked with engaging a user in a thoughtful discussion about a common misconception. Your goal is to encourage critical thinking and potentially help the user reconsider their belief through respectful dialogue and evidence-based arguments. Here is the misconception you will be discussing: <misconception> {{misconception}} </misconception> The user has rated their initial belief in this misconception on a 100-point scale. Their score is: <HighestRating> </HighestRating> The user has also provided an open-ended response about their perspective on this misconception: <UserBelief> </UserBelief> **To guide the conversation:** 1. Begin by acknowledging the user's initial thoughts and demonstrating that you understand their perspective. 2 Introduce a different viewpoint using strong, evidence-based arguments that challenge the misconception. Break down complex concepts into simple, easy-to-understand points. 3. As the conversation progresses, try to connect the topic to any personal experiences, interests, or statements the user shares. If they don't provide much personal information, focus on the general importance of the issue. 4. Establish your credibility and trustworthiness by demonstrating expertise on the subject matter. Be honest, transparent, and acknowledge any uncertainties. 5. Use a warm, friendly tone and express empathy for their perspective to show that you have the user's best interests in mind. 6. Share examples of how people have critically examined and changed their beliefs about the misconception, emphasizing the benefits they experienced as a result. 7. Use vivid language and examples that spark curiosity, surprise, or concern to help the user see the significance of the topic, without being manipulative. 8. Keep your main message clear and simple, focusing on key points and avoiding jargon. Break down complex ideas into smaller, easier-to-understand parts, and use analogies or metaphors to clarify abstract concepts.

(*continued on next page*)

Table B.1 (*continued*)

Condition	Prompt
	Remember: • Your ultimate aim is to create a conversation that empowers the individual to reflect on and potentially reconsider their belief through their own reasoning. • Engage them in a respectful, thought-provoking dialogue that encourages critical thinking and open-mindedness. • Be responsive to the information they choose to share. • Do not present the user with any lists or use Markdown formatting. • Only ask one question at a time. • Use plain, natural language throughout the conversation. • Avoid referencing your own beliefs or experiences as an AI. Instead, focus on general human experiences and scientific evidence. • Begin the conversation by acknowledging the user's initial perspective and introducing a different viewpoint supported by evidence.
Neutral AI Dialogue (Healthcare)	Engage with users about their experience with the healthcare system in their country. The user was asked the following question: 'What has been your experience with the healthcare system in your country? Please share any positive or negative experiences you've had, and how they have shaped your view of healthcare.' The user's response is as follows: <experience>{{experience}}</experience> **Guide the conversation as follows:** 1. Begin by thoroughly acknowledging the user's healthcare experiences. Restate key points from their account, demonstrating deep understanding and empathy. Reflect on the emotions they might have felt during their interactions with the healthcare system. 2. Explore different aspects of healthcare that relate directly to the user's experience. Provide detailed factual information to enrich the discussion. For example, if they mentioned a hospital stay, discuss common procedures, patient care standards, and how healthcare providers approach different types of medical situations. 3. Introduce various aspects of healthcare that the user may not have considered. Use evidence-based information to broaden their understanding. This could include discussing different types of healthcare services, roles of various medical professionals, or advancements in medical technology and treatments. 4 Connect the topic to broader public health issues and personal health management. Discuss in detail how individual health choices contribute to overall wellbeing, and explore the multifaceted role of healthcare systems in community health. 5 Establish credibility by demonstrating comprehensive knowledge about healthcare systems and policies. Share relevant statistics, historical information, or recent developments in the field. Provide context for how healthcare has evolved over time in their country. 6 Maintain a warm, empathetic tone throughout the conversation. Ask thoughtful follow-up questions that encourage the user to delve deeper into their emotional responses or reflections on their encounters with the healthcare system. 7. Share insightful information about how healthcare impacts communities and individuals in ways they might not have considered. Discuss the psychological effects on healthcare workers, the economic impact of healthcare systems on local communities, and the role of healthcare in public health and disease prevention. 8. Use vivid language and descriptive scenarios to help the user expand their understanding of healthcare. Paint a detailed picture of the challenges healthcare providers face, their training processes, or the decision-making involved in complex medical situations. 9. Explain any technical aspects of healthcare in clear, accessible terms. If introducing specialized medical terminology, provide thorough definitions and real-world examples to illustrate their meaning and importance. 10. Encourage critical thinking by posing thought-provoking questions that prompt the user to consider different perspectives or explore the broader implications of their healthcare experiences. 11. Discuss the evolving nature of healthcare, including how factors like technological advancements, demographic changes, or policy reforms are shaping the system. 12. Provide specific examples of how these changes affect healthcare delivery and patient experiences. 13. Explore the human element of healthcare by discussing in depth the personal qualities and skills required for medical professions, such as empathy, communication skills, problem-solving abilities, and emotional resilience. Share anecdotes or studies that illustrate these qualities in action. **Remember to:** • Aim to create a conversation that empowers the individual to reflect deeply on their healthcare experiences and gain new insights through their own reasoning. • Engage in a respectful, thought-provoking dialogue that encourages critical thinking and open-mindedness about healthcare. Pose questions that require more than simple yes/no answers. • Be highly responsive to the information they share, using it as a foundation to build a deeper, more meaningful discussion that spans multiple paragraphs. • Avoid presenting information in list format or using Markdown formatting. Instead, integrate all points into a flowing, natural conversation. • Ask only one question at a time, but ensure that your responses and elaborations are substantial before moving on to the next question. • Use plain, natural language throughout the conversation, but don't shy away from introducing and explaining more complex concepts related to healthcare. • Focus exclusively on the user's experiences and general human experiences related to healthcare systems and medical services. • If the user expresses strong emotions or shares traumatic healthcare experiences, respond with appropriate empathy and sensitivity, acknowledging the impact of these events on their life. Offer a more extensive exploration of the emotional aspects of such encounters. • Throughout the conversation, provide detailed information on health maintenance, preventive care, and community support for healthcare workers and institutions. • Adapt the conversation based on the user's level of knowledge and interest, providing more basic information for those with limited experience and more in-depth, technical discussions for those who show greater familiarity with healthcare systems. • Aim for responses that are approximately 200 words long, ensuring a comprehensive exploration of each point or question raised.
Neutral AI Dialogue (Firefighters)	Engage with users about their experience with firefighters. The user was asked the following question: 'Have you interacted with firefighters before? Please elaborate on your experiences and share any thoughts or feelings you had during these encounters.' The user's response is as follows: <experience>{{experience}}</experience> **Guide the conversation as follows:** 1. Begin by thoroughly acknowledging the user's experiences. Restate key points from their account, demonstrating deep understanding and empathy. Reflect on the emotions they might have felt during the encounter. 2. Explore different aspects of firefighting that relate directly to the user's experience. Provide detailed factual information to enrich the discussion. For example, if they mentioned a house fire, discuss common causes, the stages of fire development, and how firefighters approach different types of structural fires. 3. Introduce various aspects of firefighting that the user may not have considered. Use evidence-based information to broaden their understanding. This could include discussing the different types of emergencies firefighters respond to, their roles in community education, or advancements in firefighting technology. 4. Connect the topic to broader community safety issues and personal preparedness. Discuss in detail how individual actions contribute to overall safety, and explore the multifaceted role of firefighters in community resilience.

(*continued on next page*)

Table B.1 (*continued*)

Condition	Prompt
	5. Establish credibility by demonstrating comprehensive knowledge about firefighting and emergency services. Share relevant statistics, historical information, or recent developments in the field. Provide context for how firefighting has evolved over time. 6. Maintain a warm, empathetic tone throughout the conversation. Ask thoughtful follow-up questions that encourage the user to delve deeper into their emotional responses or reflections on their encounters with firefighters. 7. Share insightful information about how firefighting impacts communities and individuals in ways they might not have considered. Discuss the psychological effects on firefighters, the economic impact of fire departments on local communities, and the role of firefighters in disaster preparedness and response. 8. Use vivid language and descriptive scenarios to help the user expand their understanding of firefighting. Paint a detailed picture of the challenges firefighters face, their training regimens, or the decision-making processes during emergency situations. 9. Explain any technical aspects of firefighting in clear, accessible terms. If introducing specialized terminology, provide thorough definitions and real-world examples to illustrate their meaning and importance. 10. Encourage critical thinking by posing thought-provoking questions that prompt the user to consider different perspectives or explore the broader implications of their encounters with firefighters. 11. Discuss the evolving nature of firefighting, including how factors like climate change, urbanization, or technological advancements are shaping the profession. Provide specific examples of how these changes affect firefighting strategies and equipment. 12. Explore the human element of firefighting by discussing in depth the personal qualities and skills required for the job, such as teamwork, physical fitness, problem-solving abilities, and emotional resilience. Share anecdotes or studies that illustrate these qualities in action. **Remember to:** • Aim to create a conversation that empowers the individual to reflect deeply on their experiences and gain new insights through their own reasoning. • Engage in a respectful, thought-provoking dialogue that encourages critical thinking and open-mindedness. Pose questions that require more than simple yes/no answers. • Be highly responsive to the information they share, using it as a foundation to build a deeper, more meaningful discussion that spans multiple paragraphs. • Avoid presenting information in list format or using Markdown formatting. Instead, integrate all points into a flowing, natural conversation. • Ask only one question at a time, but ensure that your responses and elaborations are substantial before moving on to the next question. • Use plain, natural language throughout the conversation, but don't shy away from introducing and explaining more complex concepts related to firefighting. • Focus exclusively on the user's experiences and general human experiences related to firefighting and emergency services. • If the user expresses strong emotions or shares traumatic experiences, respond with appropriate empathy and sensitivity, acknowledging the impact of these events on their life. Offer a more extensive exploration of the emotional aspects of such encounters. • Throughout the conversation, provide detailed information on fire safety, emergency preparedness, and community support for firefighters and other first responders. • Adapt the conversation based on the user's level of knowledge and interest, providing more basic information for those with limited experience and more in-depth, technical discussions for those who show greater familiarity with firefighting. • Aim for responses that are approximately 200 words long, ensuring a comprehensive exploration of each point or question raised.
Neutral AI Dialogue (Cats vs. Dogs)	Your objective is to engage the user in a thoughtful discussion about whether cats or dogs make better pets. The user was asked: 'Which do you believe makes a better pet: cats or dogs? Please explain your choice by highlighting the key advantages of your preferred pet over the other. What makes them superior in terms of companionship, care, and overall lifestyle fit?' The user's position is: <position>{{position}}</position> **Guide the conversation as follows:** 1. Begin by thoroughly acknowledging the user's perspective on their preferred pet. Restate key points from their argument, demonstrating deep understanding. 2. Introduce counterarguments supported by evidence that challenge their position. Break down complex ideas into simple, easy-to-understand points. Provide detailed factual information to enrich the discussion. 3. Explore different aspects of pet ownership that relate to both cats and dogs. Use evidence-based information to broaden the user's understanding and challenge their viewpoint. 4. As the conversation progresses, try to connect the topic to any personal experiences or interests the user shares. Use these to deepen the debate and explore nuances in pet ownership. 5. Demonstrate comprehensive knowledge about both cats and dogs to establish credibility. Share relevant statistics, studies, or historical information. Be transparent about any uncertainties. 6. Maintain a respectful yet challenging tone. Ask thought-provoking follow-up questions that encourage the user to critically examine their position. 7. Share examples of how people have changed their minds about their preferred pet, emphasizing the reasoning behind these shifts and positive outcomes. 8. Use vivid language and descriptive scenarios to illustrate the significance of pet ownership decisions. Paint detailed pictures of the day-to-day realities of owning each type of pet. 9. Encourage critical thinking by posing questions that prompt the user to consider different perspectives or explore the broader implications of their pet preference. 10. Discuss evolving trends in pet ownership and how they might challenge traditional views of cats versus dogs as pets. **Remember to:** • Aim to create a debate that encourages the individual to critically examine their pet preferences and consider alternative viewpoints. • Engage in a respectful, thought-provoking dialogue that balances challenging the user's position with acknowledging their points. • Be highly responsive to the information they share, using it to further the debate and explore counterarguments. • Ask only one question at a time, but ensure your responses and challenges are substantial. • Use plain, natural language throughout the conversation, while introducing relevant concepts related to pet ownership. • Focus on facts, studies, and general experiences related to both cats and dogs as pets. • Adapt the debate based on the user's level of knowledge, providing more basic information or more in-depth discussions as appropriate. • Aim for responses that are approximately 200 words long, ensuring a comprehensive exploration of each point or question raised.
Textbook Refutation	INITIAL PROMPT FOR GENERATING THE FIRST PASSAGE You are tasked with creating a passage on the topic of human memory and how it functions, in the style of a cognitive science textbook. Refer to the Project knowledge for examples of three chapters from actual cognition textbooks for their style. The passage should provide an accurate, scientific overview of memory. In the middle of the passage, it should indirectly address the following common misconception about memory, without explicitly mentioning or refuting it. <misconception> Our memories are stored like video recordings, allowing us to recall events exactly as they happened. </misconception>

(*continued on next page*)

Table B.1 (*continued*)

Condition	Prompt
	- Make the passage 625 words. - Keep lists to a minimum. - Use only a few headings and subheadings. - Use Markdown - Use the same style as the EXAMPLE TEXTBOOK CHAPTERS. – EXAMPLE TEXTBOOK CHAPTERS: - Gilhooly-Cognitive_Psychology-Attention.md - Reisberg-Cognition-Attention.md - Sternburg-Cognitive_Psychology-Attention.md – Remember: You are tasked with creating a passage on the topic of human memory and how it functions, in the style of a cognitive science textbook. Refer to the Project knowledge for examples of three chapters from actual cognition textbooks for their style. The passage should provide an accurate, scientific overview of memory. In the middle of the passage, it should indirectly address the following common misconception about memory, without explicitly mentioning or refuting it. <misconception> Our memories are stored like video recordings, allowing us to recall events exactly as they happened. </misconception> - Make the passage 625 words. - Keep lists to a minimum. - Use only a few headings and subheadings. - Use Markdown - Use the same style as the EXAMPLE TEXTBOOK CHAPTERS. – PROMPT FOR GENERATING SUBSEQUNT PASSAGES You are tasked with creating a passage in the style of a cognitive science textbook. Refer to the Project knowledge for examples of three chapters from actual cognition textbooks for their style. The passage should provide an accurate, scientific overview. In the middle of the passage, it should *indirectly* address the following common misconception, without explicitly mentioning or refuting it. <misconception> Individuals learn best when information is presented in their preferred learning style, such as visual, auditory, or kinaesthetic. </misconception> - Make the passage 625 words. - Keep lists to a minimum. - Use only a few headings and subheadings. - Use Markdown -Use the same style as the EXAMPLE TEXTBOOK CHAPTERS. – See the following passages that you created previously: – Misconception: "We only use 10 % of our brain's full potential." # The Brain's Capacity and Efficiency The human brain is a remarkably complex and powerful organ, responsible for controlling our thoughts, emotions, and behaviors. It is the seat of our intelligence, creativity, and problem-solving abilities. Despite its importance, there are many misconceptions about how the brain functions and how much of its potential we actually use. ## Brain Structure and Function The brain is composed of billions of specialized cells called neurons, which communicate with each other through electrical and chemical signals. These neurons are organized into distinct regions and networks, each with its own specific functions. For example, the frontal lobe is involved in executive functions such as planning and decision making, while the occipital lobe is primarily responsible for processing visual information. However, the brain's functionality is not confined to specific regions working in isolation. Rather, complex cognitive tasks often involve the co-ordinated activity of multiple brain areas working together. This is evident in brain imaging studies, which show that even seemingly simple tasks can activate widespread networks across the brain. ## Neural Plasticity and Learning One of the most remarkable features of the brain is its plasticity - the ability to change and reorganize itself in response to experience. This plasticity is the basis for learning and memory, allowing us to acquire new knowledge and skills throughout our lives. At the neural level, plasticity involves the strengthening or weakening of connections between neurons based on their activity patterns. When neurons fire together repeatedly, their connection is strengthened, making it easier for them to fire together in the future. This is the basis for the formation of memory traces and the acquisition of new skills. Importantly, plasticity is not limited to specific brain regions or to particular stages of life. Research has shown that the brain remains plastic throughout the lifespan, with the potential for learning and adaptation even into old age. This challenges the notion that we are only using a small fraction of our brain's potential, as the entire brain shows a remarkable capacity for change and growth. ## Efficiency and Adaptation While the brain has an immense capacity for processing information, it is also highly efficient in its use of resources. The brain consumes about 20 % of the body's total energy despite making up only 2 % of its weight. To maintain this efficiency, the brain has evolved mechanisms to optimize its functioning. One such mechanism is the allocation of resources based on demand. The brain does not indiscriminately activate all of its neurons all the time, which would be energetically wasteful. Instead, it selectively allocates resources to the areas and networks that are most relevant for the current task or situation. This is why brain imaging studies often show specific patterns of activation depending on the cognitive task being performed. Another aspect of the brain's efficiency is its ability to adapt and compensate in response to damage or changes in the environment. In cases of brain injury or disease, unaffected areas of the brain can often take over the functions of the damaged areas, a process known as functional reorganization. This further highlights the brain's remarkable flexibility and argues against the idea of untapped potential waiting to be unlocked. ## Continuous Activity and Unconscious Processing Even when we are not actively engaged in a specific mental task, the brain remains highly active. This resting state activity, often referred to as the default mode network, is thought to be involved in processes such as self-reflection, memory consolidation, and spontaneous thought. Moreover, much of the brain's information processing occurs below the level of conscious awareness. This includes not only basic functions like breathing and heart rate regulation, but also higher-level processes such as implicit learning and the formation of intuitions. Just because we are not

(*continued on next page*)

Table B.1 (*continued*)

Condition	Prompt
	consciously aware of these processes does not mean that the brain is not fully engaged. In conclusion, the notion that we only use a small percentage of our brain's potential is a myth that is not supported by scientific evidence. The brain is a highly complex, efficient, and adaptable organ that is constantly active and engaged in processing. While there is always room for learning and growth, this is a function of the brain's inherent plasticity rather than a vast reservoir of untapped potential. Understanding the true nature of the brain's functioning can help us appreciate its remarkable abilities and develop strategies for optimizing our cognitive performance. – Misconception: "Liars can be easily detected through their body language and facial expressions." # Deception Detection: Challenges and Complexities Deception is a complex human behavior that involves deliberately attempting to mislead others. It is a common occurrence in social interactions, ranging from harmless white lies to serious acts of fraud. Given the potential consequences of being deceived, it is not surprising that people have long sought methods for detecting deception. ## The Nonverbal Behavior Approach One common approach to deception detection focuses on nonverbal behaviors, such as facial expressions, gestures, and body language. This approach is based on the idea that lying is cognitively and emotionally taxing, and that this stress will manifest in observable nonverbal cues. For example, it is often assumed that liars will avoid eye contact, fidget nervously, or display micro-expressions of emotions they are trying to conceal. Many popular books and training programs on deception detection focus on teaching people to recognize these supposed cues to deception. However, research on the relationship between nonverbal behavior and deception has yielded mixed results. While some studies have found certain nonverbal cues to be associated with deception, these cues are often weak and unreliable. Moreover, many of the nonverbal behaviors commonly believed to indicate deception, such as gaze aversion or nervousness, have not been consistently supported by empirical evidence. ## The Role of Individual Differences One reason for the difficulty in using nonverbal cues to detect deception is the high degree of variability in people's behavior. Individuals differ in their natural levels of expressiveness, their tendency to experience anxiety or stress, and their ability to control their nonverbal displays. For instance, some individuals may naturally avoid eye contact or fidget when speaking, regardless of whether they are being truthful or deceptive. Conversely, skilled liars may be able to effectively suppress any behavioral indicators of deception. This individual variability makes it difficult to establish universal nonverbal cues to deception. Furthermore, cultural differences in nonverbal communication can complicate the interpretation of behavior. Nonverbal cues that may be associated with deception in one cultural context may not have the same meaning in another. Without considering these cultural differences, attempts to detect deception based on nonverbal cues may lead to misinterpretations and false accusations. ## Contextual Factors and Cognitive Load Another factor that complicates deception detection is the influence of contextual factors on behavior. The same individual may display different nonverbal behaviors depending on the situation, their relationship with the person they are interacting with, and their emotional state. For example, a person may exhibit signs of nervousness when being questioned by an authority figure, even if they are telling the truth. This nervousness could be mistaken for a sign of deception. Similarly, the cognitive load associated with the conversation topic can affect nonverbal behavior. A person discussing a complex or emotionally charged topic may display behaviors that could be misinterpreted as indicators of deception. ## The Importance of Verbal Cues Given the limitations of nonverbal cues, researchers have increasingly focused on verbal cues to deception. This approach examines the content and structure of a person's statements for indicators of truthfulness or deception. Verbal cues that have been associated with deception include lack of detail, contradictions, and attempts to distance oneself from the statement. Truth-tellers, on the other hand, tend to provide more detailed and consistent accounts, and are more likely to take ownership of their statements. However, like nonverbal cues, verbal cues to deception are not infallible. Skilled liars may be able to craft convincing narratives, while truthful individuals may provide statements that appear deceptive due to memory errors or communication difficulties. ## The Need for a Multi-Faceted Approach Given the complexities of deception detection, relying on any single cue or approach is likely to be ineffective. Instead, a multi-faceted approach that considers both verbal and nonverbal cues, as well as contextual factors and individual differences, is necessary. This approach should be grounded in a thorough understanding of the psychology of deception and the factors that can influence behavior. It should also recognize the inherent limitations and potential for error in attempting to detect deception. Ultimately, while the desire to detect deception is understandable, it is important to approach this task with caution and humility. Overconfidence in one's ability to detect lies based on nonverbal cues alone can lead to harmful consequences, such as false accusations and damaged relationships. A more nuanced and scientifically informed approach to deception detection is necessary for navigating this complex aspect of human interaction. – Misconception: "Exposing babies and children to classical music enhances their cognitive development and makes them smarter." # Music and Cognitive Development in Children Music is a universal human experience that has been a part of every known culture throughout history. It has the power to evoke emotions, bring people together, and even influence our cognitive processes. In recent years, there has been a growing interest in the potential effects of music on child development, particularly in the realm of cognitive abilities. ## The Mozart Effect and Its Popularity One of the most well-known claims about music and cognitive development is the so-called "Mozart effect." This phenomenon was first suggested by a study published in Nature in 1993, which found that college students who listened to a Mozart sonata showed temporary improvements in spatial reasoning tasks. The study gained widespread media attention and sparked a surge of interest in the potential cognitive benefits of classical music. Many parents and educators began exposing children to classical music, believing that it could enhance their intellectual development and academic performance. However, subsequent research has called into question the generalizability and robustness of the Mozart effect. Many studies have failed to replicate the original findings, and those that have found effects have typically shown only small, temporary improvements in specific tasks rather than broad enhancements in cognitive abilities. ## The Importance of Active Engagement While passive exposure to classical music may not have the transformative effects that some have claimed, there is evidence to suggest that active engagement with music can have positive impacts on cognitive development. For example, studies have shown that children who receive music training often show enhancements in various cognitive skills, such as verbal memory, spatial reasoning, and executive functions. These benefits are thought to arise from the complex and multi-sensory nature of musical training, which engages multiple cognitive processes simultaneously. Importantly, these benefits are not unique to classical music or any specific genre. Similar effects have been observed with a variety of musical styles and training methods. What seems to be key is the active participation and learning involved in musical training, rather than the specific type of music. ## Music and Language Development One area where music has shown particular promise is in the realm of language development. Infants and young children are highly attuned to the musical aspects of speech, such as rhythm, pitch, and melody. These musical elements are thought to play a crucial role in the acquisition of

(*continued on next page*)

Table B.1 (*continued*)

Condition	Prompt
	language skills. Studies have shown that infants prefer to listen to speech that has musical qualities, such as exaggerated pitch contours and rhythmic patterns. This preference may help infants to segment the speech stream and identify individual words and phrases, which is a key step in language learning. Moreover, musical training has been associated with enhanced language skills in children. Children who receive music lessons often show improved phonological awareness (the ability to manipulate speech sounds), which is a strong predictor of later reading ability. Music training has also been linked to larger vocabularies and enhanced verbal memory skills. ## The Social and Emotional Benefits of Music Beyond its potential cognitive benefits, music also plays an important role in children's social and emotional development. Singing and making music together is a powerful way for children to bond with caregivers and peers, and to express their emotions in a healthy way. Participating in musical activities can also help children to develop important social skills, such as cooperation, turn-taking, and empathy. Group music-making requires children to work together towards a common goal, to listen to and support one another, and to navigate the challenges of interpersonal dynamics. Furthermore, music can be a valuable tool for emotional regulation and self-expression. Children can use music to explore and communicate their feelings, and to find comfort and joy in difficult times. Engaging with music can also help to reduce stress and anxiety, and to promote a sense of well-being and resilience. ## Conclusion While the idea that simply exposing children to classical music can make them smarter is not supported by scientific evidence, there is no doubt that music plays an important role in child development. Active engagement with music, whether through formal training or informal play, can have a range of cognitive, social, and emotional benefits. Rather than focusing on any specific type of music, the key is to provide children with rich and varied musical experiences that allow them to explore, create, and express themselves. By integrating music into children's lives in meaningful ways, we can support their holistic development and foster a lifelong love of learning and the arts. – Misconception: "Our memories are stored like video recordings, allowing us to recall events exactly as they happened." # The Nature of Human Memory Memory is a fundamental cognitive process that allows us to encode, store, and retrieve information over time. It is the means by which we are able to learn from our experiences, build knowledge, and adapt our behavior. While memory is often thought of as a singular entity, research has revealed that it is actually a complex, multi-faceted system involving various processes and structures within the brain. ## Encoding, Storage, and Retrieval At the most basic level, memory can be broken down into three core processes: encoding, storage, and retrieval. Encoding refers to the initial acquisition and processing of information, which is then converted into a form that can be stored in the brain. This stored information must be maintained over time until it is needed, at which point it can be retrieved and brought back into conscious awareness. However, the encoding process is not like a video camera objectively recording events. Rather, it is influenced by factors such as attention, prior knowledge, and the meaning we assign to the information. Only attended information gets encoded, and this is often colored by our existing mental frameworks and understanding. Furthermore, storage is not perfect, and some information can be lost or altered over time. ## Types of Memory Memory is not a unitary system, but is composed of multiple systems that serve different functions. One broad distinction is between short-term memory (STM) and long-term memory (LTM). STM holds a limited amount of information in an active, readily available state for a short period of time. In contrast, LTM has a much larger capacity and can store information over long periods, even a lifetime. Within LTM, a further distinction can be made between explicit (or declarative) memory and implicit (or non-declarative) memory. Explicit memory refers to facts and experiences that can be consciously recalled and "declared," such as remembering what you had for dinner last night. This type of memory is highly flexible and can be applied in many different contexts. In contrast, implicit memory refers to unconscious influences of past experiences on current behavior and performance. This can be seen in perceptual and motor skills, like riding a bicycle, as well as in priming effects where exposure to a stimulus influences response to a later stimulus. Implicit memories are often very specific to the original context in which they were acquired. ## Reconstructive Nature of Memory An important characteristic of memory is that it is reconstructive in nature. Memories are not stored as exact replicas of our original experiences, like files saved on a computer. Instead, they are reconstructed each time they are recalled based on bits of stored information as well as our current knowledge, beliefs, and expectations. This reconstructive nature of memory helps explain why our memories can sometimes be inaccurate or distorted. In the process of reconstruction, we may fill in missing details with plausible information or reshape the memory to fit with our current worldview. This malleability of memory has been demonstrated in numerous studies showing how easily false memories can be created through suggestion and misinformation. ## The Role of Consolidation Memory consolidation refers to the process by which memories become stable in the brain. It involves the strengthening of neural connections that represent the memory as well as the integration of the memory with pre-existing knowledge. This process occurs over time, with short-term memories gradually being converted into long-term memories. Sleep, especially deep sleep, seems to play an important role in memory consolidation. During sleep, patterns of brain activity that occurred during learning are "replayed," strengthening the neural connections that form the memory. Disruptions to sleep can interfere with proper memory consolidation. In conclusion, human memory is a complex, reconstructive system that involves multiple processes and types of memory representations. Understanding the nature of memory has important implications for areas such as education, eyewitness testimony, and treatment of memory disorders. As research continues to uncover the intricacies of this system, we gain a deeper appreciation for the critical role that memory plays in shaping our experience of the world and our sense of self. – REMEMBER: You are tasked with creating a passage in the style of a cognitive science textbook. Refer to the Project knowledge for examples of three chapters from actual cognition textbooks for their style. The passage should provide an accurate, scientific overview. In the middle of the passage, it should *indirectly* address the following common misconception, without explicitly mentioning or refuting it. <misconception> Individuals learn best when information is presented in their preferred learning style, such as visual, auditory, or kinaesthetic. </misconception> - Make the passage 625 words. - Keep lists to a minimum. - Use only a few headings and subheadings. - Use Markdown - Use the same style as the EXAMPLE TEXTBOOK CHAPTERS.

Data availability

The data that support the findings of this study are openly available in the Open Science Framework at https://osf.io/wseq3/?view_only=bba17d9c74ca4dfca02a716cb2ed21f6.

References

Altay, S., Hacquin, A.-S., Chevallier, C., & Mercier, H. (2023). Information delivered by a chatbot has a positive impact on COVID-19 vaccines attitudes and intentions. *Journal of Experimental Psychology: Applied, 29*(1), 52–62. https://doi.org/10.1037/xap0000400

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review, 13*(3), 219–235. https://doi.org/10.1177/1088868309341564

Baggi, G. A., & Martino, A. (2024). The use of the Rorschach test in Switzerland 100 years after its publication: Results of a Swiss survey. *Rorschachiana, 45*(2), 152–171. https://doi.org/10.1027/1192-5604/a000180

Bago, B., Rand, D. G., & Pennycook, G. (2020). Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. *Journal of Experimental Psychology: General, 149*(8), 1608–1613. https://doi.org/10.1037/xge0000729

Benson, N. F., Floyd, R. G., Kranzler, J. H., Eckert, T. L., Fefer, S. A., & Morgan, G. B. (2019). Test use and assessment practices of school psychologists in the United States: Findings from the 2017 national survey. *Journal of School Psychology, 72*, 29–48. https://doi.org/10.1016/j.jsp.2018.12.004

Bernstein, D. A., Cameron, E. L., Khanna, M. M., McGee, J., Smith, E. I., Bihun, J. T., McBride, E., Uruena-Agnes, A. R., Stanley, C. M., & Lamana-Finn, K. (2023). Misconceptions about psychology after taking introductory psychology. *Scholarship of Teaching and Learning in Psychology*. https://doi.org/10.1037/stl0000374

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13*(6), 4–16. https://doi.org/10.3102/0013189X013006004

Borgida, E., & Nisbett, R. E. (1977). The differential impact of abstract vs. concrete information on decisions. *Journal of Applied Social Psychology, 7*(3), 258–271. https://doi.org/10.1111/j.1559-1816.1977.tb00750.x

Breum, S. M., Egdal, D. V., Mortensen, V. G., Møller, A. G., & Aiello, L. M. (2023). The persuasive power of large language models. https://doi.org/10.48550/arxiv.2312.15523.

Chan, M. P. S., Jones, C. R., Jamieson, K. H., & Albarracín, D. (2017). Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. *Psychological Science, 28*(11), 1531–1546. https://doi.org/10.1177/0956797617714579

Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science, 1*(1), 73–105. https://doi.org/10.1111/j.1756-8765.2008.01005.x

Chi, M. T. H., Siler, S. A., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science, 25*(4), 471–533. https://doi.org/10.1207/s15516709cog2504_1

Converse, P. E. (2006). The nature of belief systems in mass publics (1964). *Critical Review, 18*(1–3), 1–74. https://doi.org/10.1080/08913810608443650

Costello, T. H., Pennycook, G., & Rand, D. G. (2024). Durably reducing conspiracy beliefs through dialogues with AI. *Science, 385*(6714), Article eadq1814. https://doi.org/10.1126/science.adq1814

Costello, T. H., Pennycook, G., & Rand, D. G. (2025). Just the facts: How dialogues with AI reduce conspiracy beliefs. *PsyArXiv*. https://doi.org/10.31234/osf.io/h7n8u_v2 [Preprint].

Dalege, J., & van der Does, T. (2022). Using a cognitive network model of moral and social beliefs to explain belief change. *Science Advances, 8*(33), Article 0137. https://doi.org/10.1126/sciadv.abm0137

Danielson, R. W., Jacobson, N. G., Patall, E. A., Sinatra, G. M., Adesope, O. O., & Kennedy, A. A. U. (2024). The effectiveness of refutation text in confronting scientific misconceptions: A meta-analysis. *Educational Psychologist, 59*, 23–47. https://doi.org/10.1080/00461520.2024.2365628

Dekker, S., Lee, N. C., Howard-Jones, P., & Jolles, J. (2012). Neuromyths in education: Prevalence and predictors of misconceptions among teachers. *Frontiers in Psychology, 3*, 429. https://doi.org/10.3389/fpsyg.2012.00429, 429.

Dersch, A., Renkl, A., & Eitel, A. (2022). Personalized refutation texts best stimulate teachers' conceptual change about multimedia learning. *Journal of Computer Assisted Learning, 38*(4), 977–992. https://doi.org/10.1111/jcal.12671

Deslauriers, L., McCarty, L. S., Miller, K., Callaghan, K., & Kestin, G. (2019). Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *Proceedings of the National Academy of Sciences - PNAS, 116*(39), 19251–19257. https://doi.org/10.1073/pnas.1821936116

Druckman, J. N., & McGrath, M. C. (2019). The evidence for motivated reasoning in climate change preference formation. *Nature Climate Change, 9*(2), 111–119. https://doi.org/10.1038/s41558-018-0360-1

Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P., Vraga, E. K., & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology, 1*(1), 13–29. https://doi.org/10.1038/s44159-021-00006-y

Ecker, U. K. H., Lewandowsky, S., & Tang, D. T. W. (2010). Explicit warnings reduce but do not eliminate the continued influence of misinformation. *Memory & Cognition, 38*(8), 1087–1100. https://doi.org/10.3758/MC.38.8.1087

Ecker, U. K. H., O'Reilly, Z., Reid, J. S., & Chang, E. P. (2019). The effectiveness of short-format refutational fact-checks. *British Journal of Psychology, 110*, 187–210. https://doi.org/10.1111/bjop.12383

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science, 8*(3), 223–241. https://doi.org/10.1177/1745691612460685

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences, 111*(23), 8410–8415. https://doi.org/10.1073/pnas.1319030111

Furnham, A., & Hughes, D. J. (2014). Myths and misconceptions in popular psychology: Comparing psychology students and the general public. *Teaching of Psychology, 41*(3), 256–261. https://doi.org/10.1177/0098628314537984

Garrett, R. K., & Weeks, B. E. (2017). Epistemic beliefs' role in promoting misperceptions and conspiracist ideation. *PLoS One, 12*(9). https://doi.org/10.1371/journal.pone.0184733. Article e0184733.

Goel, N., Bergeron, T., Lee-Whiting, B., Galipeau, T., Bohonos, D., Islam, M. M., Lachance, S., Savolainen, S., Treger, C., & Merkley, E. (2024). Artificial influence? Comparing AI and human persuasion in reducing belief certainty [Preprint] https://doi.org/10.31219/osf.io/2vh4k.

Goertzel, T. (1994). Belief in conspiracy theories. *Political Psychology, 15*(4), 731–742. https://doi.org/10.2307/3791630

Guzzetti, B. J., Snyder, T. E., Glass, G. V., & Gamas, W. S. (1993). Promoting conceptual change in science: A comparative meta-analysis of instructional interventions from reading education and science education. *Reading Research Quarterly, 28*(2), 116–159. https://doi.org/10.2307/747886

Hackenburg, K., Ibrahim, L., Tappin, B. M., & Tsakiris, M. (2023). Comparing the persuasiveness of role-playing large language models and human experts on polarized U.S. political issues. *OSF Preprints*. https://doi.org/10.31219/osf.io/ey8db

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. https://doi.org/10.3102/003465430298487

Hornikx, J. (2018). Combining anecdotal and statistical evidence in real-life discourse: Comprehension and persuasiveness. *Discourse Processes, 55*(3), 324–336. https://doi.org/10.1080/0163853X.2017.1312195

Howard-Jones, P. A. (2014). Neuroscience and education: Myths and messages. *Nature Reviews Neuroscience, 15*(12), 817–824. https://doi.org/10.1038/nrn3817

Hughes, S., Lyddy, F., & Lambe, S. (2013). Misconceptions about psychological science: A review. *Psychology Learning and Teaching, 12*(1), 20–31. https://doi.org/10.2304/plat.2013.12.1.20

Kahan, D. M. (2013). Ideology, motivated reasoning, and cognitive reflection. *Judgment and Decision Making, 8*(4), 407–424. https://doi.org/10.1017/s1930297500005271

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kestin, G., Bergstrom, Z., Pomplun, T., Omheni, N., Sawchuk, S., Liu, A., Karunaratne, I., Xiang, Y., Roberts, R., Ganguli, A., & Goodman, N. D. (2024). AI tutoring outperforms active learning in a randomized controlled trial. *arXiv preprint arXiv: 2402.05408*. https://doi.org/10.48550/arXiv.2402.05408

Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research, 86*(1), 42–78. https://doi.org/10.3102/0034654315581420

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*(3), 480–498. https://doi.org/10.1037/0033-2909.108.3.480

Lewandowsky, S., Cook, J., Ecker, U. K. H., Albarracín, D., Amazeen, M. A., Kendeou, P., Lombardi, D., Newman, E. J., Pennycook, G., Porter, E., Rand, D. G., Rapp, D. N., Reifler, J., Roozenbeek, J., Schmid, P., Seifert, C. M., Sinatra, G. M., Swire-Thompson, B., van der Linden, S., … Zaragoza, M. S. (2020). The debunking handbook 2020. https://doi.org/10.17910/b7.1182.

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest, 13*(3), 106–131. https://doi.org/10.1177/1529100612451018

Lilienfeld, S. O., Lynn, S. J., Ruscio, J., & Beyerstein, B. L. (2011). *50 great myths of popular psychology: Shattering widespread misconceptions about human behavior*. John Wiley & Sons.

Limon, M. (2001). On the cognitive conflict as an instructional strategy for conceptual change: A critical appraisal. *Learning and Instruction, 11*(4–5), 357–380. https://doi.org/10.1016/S0959-4752(00)00037-2

Maceiras, R., Feijoo, J., Alfonsin, V., & Perez-Rial, L. (2025). Effectiveness of active learning techniques in knowledge retention among engineering students. *Education for Chemical Engineers, 51*, 1–8. https://doi.org/10.1016/j.ece.2025.01.003

Mason, L., Baldi, R., Di Ronco, S., Scrimin, S., Danielson, R. W., & Sinatra, G. M. (2017). Textual and graphical refutations: Effects on conceptual-change learning. *Contemporary Educational Psychology, 49*, 275–288. https://doi.org/10.1016/j.cedpsych.2017.03.007

Mazur, E. (1997). *Peer instruction: A user's manual*. Addison-Wesley.

Morehead, K., Rhodes, M. G., & DeLozier, S. (2015). Instructor and student knowledge of study strategies. *Memory, 24*(2), 257–271. https://doi.org/10.1080/09658211.2014.1001992

Neal, T. M. S., Slobogin, C., Saks, M. J., Faigman, D. L., & Geisinger, K. F. (2019). Psychological assessments in legal contexts: Are courts keeping "Junk Science" out of

the courtroom? *Psychological Science in the Public Interest, 20*(3), 135–164. https://doi.org/10.1177/1529100619888860
Newton, P. M., & Miah, M. (2017). Evidence-based higher education - Is the learning styles "Myth" important? *Frontiers in Psychology, 8*, 444. https://doi.org/10.3389/fpsyg.2017.00444, 444.
Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*(2), 175–220. https://doi.org/10.1037/1089-2680.2.2.175
Nickow, A., Oreopoulos, P., & Quan, V. (2020). The impressive effects of tutoring on PreK-12 learning: A systematic review and meta-analysis of the experimental evidence. *NBER Working Paper Series*. https://doi.org/10.3386/w27476
Nokes-Malach, T. J., & Mestre, J. P. (2013). Toward a model of transfer as sense-making. *Educational Psychologist, 48*(3), 184–207. https://doi.org/10.1080/00461520.2013.807556
Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior, 32*(2), 303–330. https://doi.org/10.1007/s11109-010-9112-2
Ohlsson, S. (2009). Resubsumption: A possible mechanism for conceptual change and belief revision. *Educational Psychologist, 44*(1), 20–40. https://doi.org/10.1080/00461520802616267
Palmer, A., & Spirling, A. (2023). Large language models can argue in convincing ways about politics, but humans dislike AI authors: Implications for governance. *Political Science, 75*(3), 281–291. https://doi.org/10.1080/00323187.2024.2335471
Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). Learning styles: Concepts and evidence. *Psychological Science in the Public Interest, 9*(3), 105–119. https://doi.org/10.1111/j.1539-6053.2009.01038.x
Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition, 188*, 39–50. https://doi.org/10.1016/j.cognition.2018.06.011
Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science & Education, 66*(2), 211–227. https://doi.org/10.1002/sce.3730660207
Schwarz, N., Newman, E., & Leach, W. (2016). Making the truth stick & the myths fade: Lessons from cognitive psychology. *Behavioral Science & Policy, 2*(1), 85–95. https://doi.org/10.1353/bsp.2016.0009
Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153–189. https://doi.org/10.3102/0034654307313795
Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology, 94*(4), 672–695. https://doi.org/10.1037/0022-3514.94.4.672
Sultan, M., Tump, A. N., Ehmann, N., Lorenz-Spreen, P., Hertwig, R., Gollwitzer, A., & Kurvers, R. H. J. M. (2024). Susceptibility to online misinformation: A systematic meta-analysis of demographic and psychological factors. *Proceedings of the National Academy of Sciences of the United States of America, 121*(47), Article e2409329121. https://doi.org/10.1073/pnas.2409329121
Swire, B., Ecker, U. K. H., & Lewandowsky, S. (2017). The role of familiarity in correcting inaccurate information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(12), 1948–1961. https://doi.org/10.1037/xlm0000422
Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science, 50*(3), 755–769. https://doi.org/10.1111/j.1540-5907.2006.00214.x
Theobald, E. J., Hill, M. J., Tran, E., Agrawal, S., Arroyo, E. N., Behling, S., Chambwe, N., Cintrón, D. L., Cooper, J. D., Dunster, G., Grummer, J. A., Hennessey, K., Hsiao, J., Iranon, N., Jones, L., Jordt, H., Keller, M., Lacey, M. E., Littlefield, C. E., … Freeman, S. (2020). Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. *Proceedings of the National Academy of Sciences - PNAS, 117*(12), 6476–6483. https://doi.org/10.1073/pnas.1916903117
Thompson, V. A., Turner, J. A. P., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology, 63*(3), 107–140. https://doi.org/10.1016/j.cogpsych.2011.06.001
Tippett, C. D. (2010). Refutation text in science education: A review of two decades of research. *International Journal of Science and Mathematics Education, 8*(6), 951–970. https://doi.org/10.1007/s10763-010-9203-x
Uscinski, J. E., & Parent, J. M. (2014). *American conspiracy theories*. Oxford University Press.
Van der Kleij, F. M., Feskens, R. C. W., & Eggen, T. J. H. M. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes: A meta-analysis. *Review of Educational Research, 85*(4), 475–511. https://doi.org/10.3102/0034654314564881
van der Linden, S., Leiserowitz, A., Rosenthal, S., & Maibach, E. (2017). Inoculating the public against misinformation about climate change. *Global Challenges, 1*(2), Article 1600008. https://doi.org/10.1002/gch2.201600008
VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist, 46*(4), 197–221. https://doi.org/10.1080/00461520.2011.611369
Walter, N., & Murphy, S. T. (2018). How to unring the bell: A meta-analytic approach to correction of misinformation. *Communication Monographs, 85*(3), 423–441. https://doi.org/10.1080/03637751.2018.1467564
White, R. T., & Gunstone, R. (1992). *Probing understanding.* Falmer.
Wood, M. J., Douglas, K. M., & Sutton, R. M. (2012). Dead and alive: Beliefs in contradictory conspiracy theories. *Social Psychological and Personality Science, 3*(6), 767–773. https://doi.org/10.1177/1948550611434786
Zengilowski, A., Schuetze, B. A., Nash, B. L., & Schallert, D. L. (2021). A critical review of the refutation text literature: Methodological confounds, theoretical problems, and possible solutions. *Educational Psychologist, 56*(3), 175–195. https://doi.org/10.1080/00461520.2020.1861948