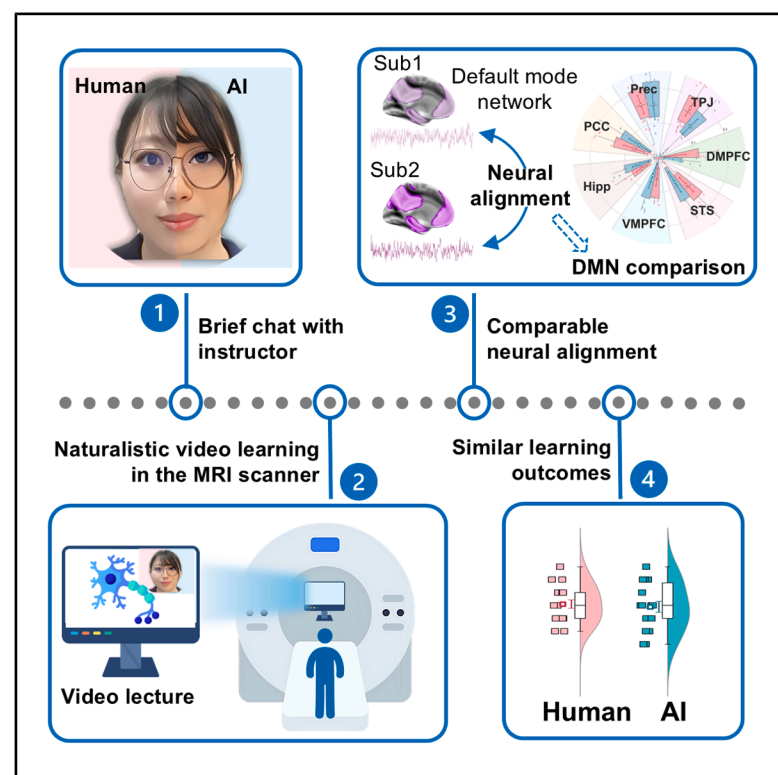


Scaffolding human and AI instruction: Neural alignment and learning gains in online education

Graphical abstract



Authors

Yingying Peng (彭颖滢), Samuel A. Nastase, Yuhan Huang (黄宇涵), Yuxi Li (李玉玺), Zhexiao Guo (国哲骁), Ping Li (李平)

Correspondence

pingliandrew@ust.hk

In brief

Peng et al. show that a brief one-on-one conversation with either a human or an LLM-enabled AI instructor before a video lecture aligns students' brain activity during learning and improves learning outcomes. Greater gaze and neural alignment between the human instructor and students, as well as among students, mediates the benefits of learning.

Highlights

- Online learning limits scalable one-on-one student-instructor interaction
- Simultaneous eye-tracking and naturalistic fMRI capture engagement during learning
- LLM-enabled AI approximates a human instructor to boost neural alignment and learning
- Neural and gaze alignment among students mediates learning improvement



Article

Scaffolding human and AI instruction: Neural alignment and learning gains in online education

Yingying Peng (彭颖滢),¹ Samuel A. Nastase,² Yuhan Huang (黄宇涵),³ Yuxi Li (李玉玺),³ Zhexiao Guo (国哲骁),³ and Ping Li (李平)^{1,4,*}

¹School of Humanities and Social Science, The Hong Kong University of Science and Technology, Hong Kong SAR, China

²Department of Psychology and Center for Computational Language Sciences, University of Southern California, Los Angeles, CA 90089, USA

³Department of Language Science and Technology, Faculty of Humanities, The Hong Kong Polytechnic University, Hong Kong SAR, China

⁴Lead contact

*Correspondence: pingliandrew@ust.hk

<https://doi.org/10.1016/j.neuron.2026.04.005>

SUMMARY

Social interactions are foundational to learning, yet scalable one-on-one instructor-student interaction remains challenging in online video learning. We examine whether a brief, structured pre-lecture instructor-student interaction, led by either a human or large language model (LLM)-powered AI instructor, can enhance student engagement and learning. Integrating behavioral testing with simultaneous eye-tracking and fMRI, we compared three groups in a between-subjects design (no interaction, human interaction, and AI interaction; $n = 57$). Both human- and AI-led pre-lecture interactions improved learning and enhanced neural alignment in critical brain regions (e.g., the default mode network) among students during learning. Neural and gaze alignments between instructor and students, as well as among students, jointly mediated learning gains, revealing a reciprocal eye-brain-behavior correspondence. While the AI instructor approximated the human instructor in learning gains, students reported lower social closeness and exhibited lower gaze alignment. These findings advance social learning theories and provide neurocognitive evidence that scalable LLM-powered AI can transform online education.

INTRODUCTION

Human learning is inherently social. Yet millions of students now learn alone, watching pre-recorded video lectures behind digital screens without exchanging a word with their instructors. This characterizes asynchronous online video learning, the predominant format for MOOCs (massive open online courses) and other large-scale online courses where instructor-student interaction is absent.^{1,2} But even in synchronous formats like live Zoom lectures, scalable one-on-one instructor-student exchanges remain a challenge.^{3,4} There is growing evidence, however, that one-on-one instructor-student social interaction during synchronous video lectures can enhance learning compared with pre-recorded videos.^{5–7} The COVID-19 pandemic exacerbated this challenge. When education shifted entirely online, insufficient teacher-student interaction had tangible consequences: diminished engagement and compromised learning outcomes.^{3,4}

Given the fixed, unidirectional nature of pre-recorded video lectures, interactive support *during* learning is inherently limited. A brief, structured interaction immediately *before* the pre-recorded video lecture offers a rare opportunity to reintroduce social dynamics at a key moment of cognitive readiness prior to

learning. In this study, we ask (1) can a structured interaction—briefly before a pre-recorded video lecture—enhance students' neural engagement measured via fMRI, attentional focus during learning indexed by eye gaze patterns, and learning outcomes, as compared with no such interaction? Critically, providing one-on-one human instructor-student interaction is resource intensive and practically challenging for mass online education featuring pre-recorded lecture videos. Thus, if such pre-lecture interactions enhance learning, this raises a second key question: (2) do these benefits essentially require human-specific social cues, or can computational systems featuring similar communicative capacities produce similar effects? We address this by testing whether such pre-lecture interactions with either a human instructor or a conversational AI could elicit comparable neurocognitive engagement and learning outcomes.

Recent advances in large language models (LLMs), text-to-speech synthesis, and video generation have enabled highly human-like AI systems with near-human communicative capacities.^{8,9} These advances create an unprecedented opportunity to test whether the neurocognitive patterns underlying social interaction depend on biological agents or emerge from communicative exchange itself. However, in education, AI has served primarily as a content provider in learning and assessment,^{10–12}



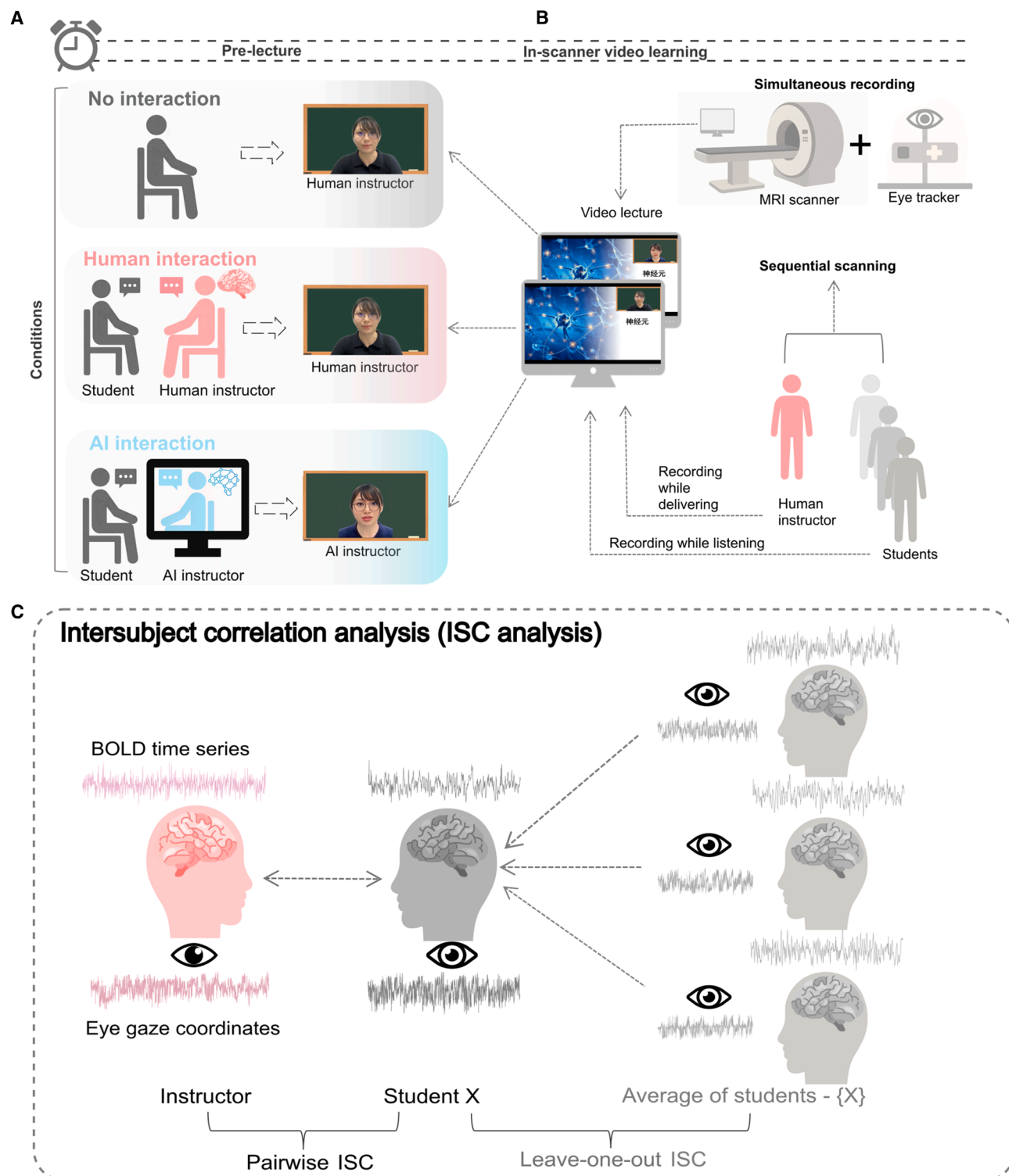


Figure 1. Schematic of experimental design, data collection, and data analysis

(A) Experimental conditions. Participants were randomly assigned to one of three groups: (1) no social interaction (control; $n = 20$), (2) human interaction ($n = 19$), and (3) AI interaction ($n = 18$).

(legend continued on next page)

rather than a social partner. A pressing question in the AI-pervasive era remains unknown: whether AI-mediated instructor-student interactions engage similar socio-emotional, neurocognitive, and computational mechanisms as human-human interactions.¹³

The multimedia learning framework has yielded substantial insights into student learning dynamics but has yet to address pre-lecture social interaction in video learning (with either human or AI instructors). Within this framework, the social agency theory¹⁴ proposes that social cues (e.g., the instructor's image) enhance students' social-affective responses and cognitive processing, thereby improving learning outcomes. Nonetheless, most multimedia learning research¹⁵ has focused on one-way instructor social presence, rather than two-way interactive social exchange. Furthermore, while the embodiment principle of this theory suggests that more human-like pedagogical agents enhance learning,¹⁵ the communicative potential of LLM-powered agents has not yet been tested.

Neuroscience studies have recently provided new perspectives to address these gaps. They offer both conceptual insights and powerful analytical tools to uncover the neural mechanisms underlying social interaction, engagement, and learning. A key tool is neural alignment, commonly measured through inter-subject correlation (ISC^{16,17}) of fMRI data. By evaluating moment-to-moment correlations in fMRI blood-oxygen-level-dependent (BOLD) activity across individuals exposed to the same naturalistic stimulus (often separately in time), we can quantify how strongly different brain areas are driven by the content of the stimulus,¹⁶ providing a measure of neural engagement—the degree to which participants' brains track the unfolding stimulus in a temporally aligned manner.

Neural alignment, indexed by ISC, has been widely observed in naturalistic video learning contexts,^{18,19} and alignment within the default mode network (DMN), among students and between students and experts, associated with comprehension and learning outcomes. Nguyen et al.²⁰ further showed that teachers' neural activity during lecture delivery predicted students' brain responses during learning, especially within the DMN, and this alignment was associated with learning success. Neural alignment is also modulated by and could predict social dynamics such as friendship.^{21,22} In classrooms, longitudinal, face-to-face electroencephalography (EEG) hyperscanning studies have linked neural synchrony with students' engagement²³ and perceived social closeness to the teacher in real-world classrooms.²⁴ Most remarkably, neural alignment can be enhanced even by brief, seemingly peripheral conversations. In Sievers et al.,²⁵ participants engaged in consensus-building discussions about previously viewed movie clips. During a subsequent viewing, including both seen and novel clips in the MRI scanner, DMN neural alignment increased, suggesting that shared conversation recalibrated participants' attentional and interpretive schemas that generalized beyond specific content. De Felice

et al.²⁶ further found that even a brief, unrelated dyadic conversation between two viewing sessions enhanced neural synchrony in co-viewing pairs compared with pseudo pairs during real-time fNIRs hyperscanning. These findings suggest that consensus-building conversations, or even unrelated daily conversations, align people as they interact, with effects that persist in the subsequent processing of new information, thereby enhancing neural alignment and shared understanding.²⁷

These findings underscore conversation as a catalyst for subsequent shared understanding. Yet, whether these effects translate to educational contexts remains unknown. Critically, whether AI-mediated conversation can serve as an effective substitute for human interaction remains untested, both in behavioral studies and neurocognitive research. This gap is particularly pronounced given the rapid emergence of conversational AI agents and the scarcity of neural data on human-AI educational interactions. We hypothesize that a brief social interaction, when strategically inserted before learning begins, can align students' attentional and interpretative faculties, thereby better preparing them for deeper engagement during subsequent learning. Positioned at the intersection of social neuroscience, AI-mediated education, and learning science, our study aims to (1) reintroduce social dynamics into online learning, (2) provide neural evidence underlying the impact of brief social interaction on education, (3) bridge theoretical gaps in multimedia learning by revealing how conversation becomes a vehicle for cognitive alignment and educational impact, and (4) evaluate the potential of AI, especially LLMs, to scale personalized educational interactions while probing the behavioral and neural mechanisms underlying human-AI interaction in educational contexts.

Toward these goals, we implemented a naturalistic neuroimaging study with three conditions: (1) video learning without prior social interaction, (2) video learning preceded by a structured human instructor-student pre-lecture interaction, and (3) video learning preceded by a structured AI-mediated (GPT-4 empowered) interaction (see Figure 1A). We simultaneously recorded the human instructor's and students' eye movements and neural activity, respectively, during lecture delivery and subsequent video learning (see Figure 1B for sequential scanning). Our study thus directly investigates how structured pre-lecture interactions shape instructor-student social dynamics (measured by students' perceived social closeness with the human or AI instructor), neural alignment (student-instructor and student-student), attentional focus (indexed by eye gaze patterns), and ultimately learning efficacy. Critically, integrating cutting-edge AI advances, this study also tests whether AI-mediated instructor-student interaction can foster meaningful socio-emotional and neurocognitive engagement, quantified by neural alignment in cognitive and socio-emotional implicated networks. When compared to human-based instruction, AI-mediated instruction could offer insights into scalable, socially enriched online education.

(B) Data collection: simultaneous fMRI and eye-tracking recorded the human instructor during lecture delivery and each student during viewing the same 14.25-min neuroscience lecture (with human or AI instructor face shown by group).

(C) Intersubject correlation (ISC) analysis applied to the eye gaze and fMRI data. We computed (1) student-instructor pairwise ISC, correlating each student's neural or gaze time series with the instructor's, and (2) student-student leave-one-out ISC, correlating each student with the average of the remaining students. Icons from [Flaticon.com](https://flaticon.com) are used.

Table 1. Learning performance comparison across groups

No interaction versus human interaction	F(1, 36)	p	ηp^2	95% CI
Recall	0.004	0.947	<0.001	[0.000, 0.126]
Comprehension	4.33	0.045	0.107	[0.003, 0.343]
Comprehension (hard)	8.41	0.006	0.189	[0.028, 0.424]
Transfer	2.09	0.157	0.054	[0.000, 0.251]
Human interaction versus AI interaction	F(1, 36)	p	ηp^2	95% CI
Recall	0.06	0.807	0.031	[0.000, 0.144]
Comprehension	0.61	0.442	0.046	[0.000, 0.206]
Comprehension (hard)	0.05	0.822	0.002	[0.000, 0.059]
Transfer	1.50	0.229	0.076	[0.000, 0.277]

RESULTS

How does pre-lecture social interaction with either a human or an AI instructor benefit video learning through behavioral and neurocognitive measures? To address this, we adopted a between-subjects design and compared three conditions: (1) no interaction ($n = 20$), (2) human interaction ($n = 19$), and (3) AI interaction ($n = 18$). All groups viewed the same ~14-min video lecture in the MRI scanner, differing only in whether and how they were engaged in a brief pre-lecture conversation. Each structured interaction followed a standardized protocol: the instructor's and student's self-introductions, discussion of how students' university majors and personal interests relate to neuroscience, affirmative and elaborative instructor feedback, and a short lecture preview. In the human-interaction condition, this exchange was conducted face-to-face by a human instructor. In the AI-interaction condition, students, explicitly aware of the AI nature of the instructor, conversed naturally with an AI instructor via a customized web-based interface. The AI instructor closely resembled the human instructor in appearance and voice, engaging students through live, contingent dialogues powered by speech recognition, GPT-4-based content generation, text-to-speech synthesis, and real-time talking-head animation. In the no-interaction condition, students proceeded directly to the video lecture without any pre-lecture conversation. The instructor's face remained visible throughout the lecture for all groups (human instructor for the first two conditions and AI instructor for the AI condition; see Figure 1A), ensuring equivalent exposure to instructor presence, instructional content, and total duration.

Pre-lecture student interaction with instructor, human or AI, improves learning and learning experience

To test the impact of pre-lecture social interaction on learning outcomes, we conducted ANCOVAs comparing the human- and no-interaction groups, controlling for prior knowledge (no group difference was found; see STAR Methods for details). Learning outcomes were assessed with three tests: a recall test for the lecture content, a comprehension test (i.e., 30 in-scanner multiple-choice items, including 9 pre-defined difficult items requiring integration across multiple concepts), and a

transfer test (i.e., open-ended questions assessing knowledge application to novel situations; see STAR Methods and Table S4 for examples). No significant effects were observed for recall or transfer scores (see Table 1 for statistical results) between these two groups. However, compared with the no-interaction group, the human-interaction group performed significantly better on both overall comprehension ($F(1, 36) = 4.33$, $p = 0.045$, $\eta p^2 = 0.11$, 95% confidence interval [CI] [0.003, 0.343]) and difficult comprehension items ($F(1, 36) = 8.41$, $p = 0.006$, $\eta p^2 = 0.19$, 95% CI [0.028, 0.424]; see Figure 2A), suggesting that pre-lecture interaction with a human instructor facilitated deeper learning.

To test the effect of instructor type on learning, we compared the human and AI-interaction groups. ANCOVA results revealed no significant differences across recall, comprehension, difficult comprehension, and transfer scores (see Table 1 and Figure 2A). Bayesian analyses corroborate equivalent learning outcomes (see Table S5), supporting comparable effectiveness of human and AI instructors in pre-lecture social interaction that benefits learning. Furthermore, the AI-interaction group outperformed the no-interaction group on difficult comprehension ($F(1, 35) = 4.91$, $p = 0.033$, $\eta p^2 = 0.13$, 95% CI [0.005, 0.349]), reinforcing the benefit of pre-lecture interaction regardless of instructor type.

We also assessed the effects of social interaction and instructor type on the learning experience using ANCOVA with prior knowledge as a covariate. The human-interaction group reported greater effort in learning than the no-interaction group ($F(1, 36) = 4.77$, $p = 0.036$, $\eta p^2 = 0.117$, 95% CI [0.005, 0.337]) while reporting comparable effort ratings with the AI-interaction group ($F(1, 34) = 1.299$, $p = 0.262$, $\eta p^2 = 0.037$, 95% CI [0.000, 0.205]). A similar pattern emerged for perceived instructor's delivery quality (human versus no interaction: $F(1, 36) = 4.73$, $p = 0.036$, $\eta p^2 = 0.116$, 95% CI [0.004, 0.374]; human versus AI interaction: $F(1, 34) = 3.41$, $p = 0.074$, $\eta p^2 = 0.091$, 95% CI [0.001, 0.348]; see Figure 2B).

Furthermore, perceived social closeness after interacting with the human instructor was rated higher than the AI instructor ($F(1, 34) = 12.09$, $p = 0.001$, $\eta p^2 = 0.263$, 95% CI [0.072, 0.569]). Nonetheless, both groups reported significantly increased closeness with the instructor *after*, versus *before*, the interaction (human: $t = -3.888$, $p = 0.001$, Cohen's $d = -0.892$, 95% CI [-1.865, -0.557]; AI: $t = -3.011$, $p = 0.008$, Cohen's $d = -0.710$; 95% CI [-1.228, -0.216], paired t tests; see Figure 2D). These results suggest that brief pre-lecture interactions, whether with a human or an AI instructor, enhanced students' engagement and perceived connection with the instructor, although the human instructor still elicited a stronger interpersonal closeness.

Pre-lecture conversation with a human instructor better aligns subsequent visual attention

We quantified students' visual attention by computing intersubject correlation (ISC) of gaze trajectories (i.e., intersubject gaze alignment or gaze ISC) based on their moment-by-moment gaze positions (see an illustration in Figure 1C and STAR Methods). Two types of gaze ISC were computed: (1) between each student and the average of the remaining ($N - 1$) students within the group (student-student leave-one-out ISC; computed

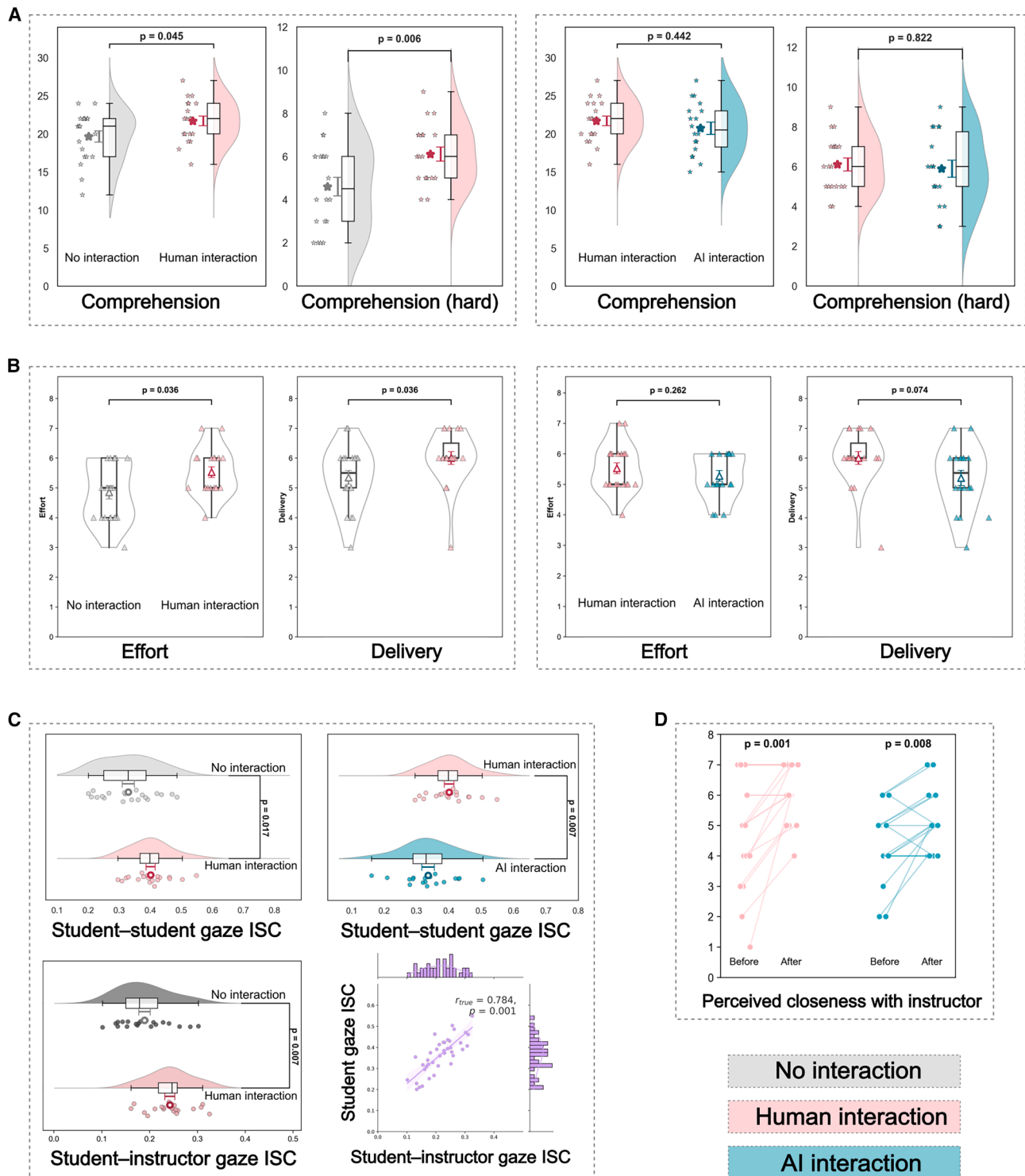


Figure 2. Effects of pre-lecture interaction and instructor type on learning performance, learning experience, and eye gaze alignment

(A) Learning performance. The human-interaction group outperformed the no-interaction group on in-scanner comprehension questions, particularly for difficult items. The AI-interaction group was statistically comparable to the human-interaction group (ANCOVA *p* values shown).

(B) Self-reported learning experience. Compared with the no-interaction group, the human-interaction group reported significantly greater learning effort and rated the instructor's delivery more favorably. The AI-interaction group was comparable to the human-interaction group (ANCOVA *p* values shown).

(legend continued on next page)

for all three groups, $n = 57$) and (2) between each student and the instructor (student-instructor pairwise ISC; for the human- and no-interaction groups only, we leveraged available instructor gaze and neural data to examine student-instructor gaze/DMN ISC in relation to learning; $n = 39$).

We first examined student-student gaze ISC. Permutation tests controlling for prior knowledge were used to assess the impact of social interaction (no interaction versus human interaction) and instructor type (human interaction versus AI interaction) on student-student gaze alignment. The human-interaction group demonstrated significantly higher gaze alignment than both the no-interaction group ($\beta = 0.066$, $p = 0.017$, 95% CI [0.014, 0.110], two-tailed permutation) and the AI-interaction group ($\beta = 0.068$, $p = 0.007$, 95% CI [0.019, 0.113], two-tailed permutation; see Figure 2C).

We further examined student-instructor gaze ISC, a relatively understudied marker of joint attention between the instructor and students. The human-interaction group showed significantly greater student-instructor gaze alignment than the no-interaction group ($\beta = 0.049$, $p = 0.007$, 95% CI [0.015, 0.079], two-tailed permutation). Further, we found that student-student gaze ISC was strongly associated with student-instructor gaze ISC (partial Spearman $r = 0.784$, $p = 0.001$, 95% CI [0.590, 0.883], one-tailed permutation, controlling for group; see Figure 2C) for the no-interaction and human-interaction groups. These results suggest pre-lecture interaction with the human instructor promoted coordinated visual attention: students aligned their gaze with the instructor's and subsequently with each other.

Pre-lecture conversation with human/AI instructor aligns subsequent neural activity during learning

We performed ISC analyses on our neural data: (1) between the student and the group average using the leave-one-out approach (computed for all three groups, $n = 57$) and (2) between the student and the instructor using the pairwise approach (for the no- and human-interaction groups only; $n = 39$). Our analyses were focused on three key brain networks implicated in comprehension and learning (DMN^{18,20}; see Figure 3A), socio-emotional processing (the limbic system²⁸; see Figure 3B), and flexible organization and control of cognitive resources (the multiple-demand network, hereafter the MD network^{29,30}; see Figure 4A). Using two-tailed permutation tests,¹⁶ we assessed the effects of pre-lecture social interaction (no interaction versus human interaction) and instructor type (human interaction versus AI interaction) on subsequent neural alignment.

Student-student neural alignment

Before group comparisons, we conducted whole-brain ISC analyses across 50 parcels³¹ to validate the expected neural alignment patterns. Consistent with prior work,^{18,20} robust ISCs emerged in early visual, auditory, and language-related cortical

regions, extending to higher-order brain areas, including the DMN (see Figures 3A and 3B).

Relative to the no-interaction group, the human-interaction group showed significantly greater ISC in several key DMN regions of interest (ROIs), including the precuneus (PC), hippocampus (Hipp), and superior temporal sulcus (STS; precuneus: Cohen's $d = 0.678$, 95% CI [0.151, 1.414], false discovery rate [FDR]-corrected $p = 0.027$; hippocampus: Cohen's $d = 0.739$, 95% CI [0.148, 1.478], FDR-corrected $p = 0.027$; STS: Cohen's $d = 0.930$, 95% CI [0.394, 1.605], FDR-corrected $p < 0.001$, two-tailed permutation; see Figure 3B and Table S6 for full statistics). The human-interaction group also exhibited greater student-student ISC in the limbic area amygdala (Cohen's $d = 0.771$, 95% CI [0.162, 1.441], FDR-corrected $p = 0.021$, two-tailed permutation; see Figure 3C and Table S7 for full statistics). Moreover, in the MD network, multiple ROIs exhibited higher ISC in the human-interaction group than the no-interaction group, spanning bilateral parietal cortices, superior and middle frontal gyri, precentral gyrus subregions, and orbitofrontal areas (Cohen's d s > 0.679 , FDR-corrected p s ≤ 0.043 ; two-tailed permutation; see Figure 4A and Table S8 for full statistics).

When comparing the human-interaction and AI-interaction groups, no significant ISC differences were observed across any DMN or limbic ROIs and most MD ROIs. The only exception was the right middle frontal gyrus (rMFG) in the MD network, where the human-interaction group exhibited significantly higher ISC (Cohen's $d = 1.046$, 95% CI [0.437, 1.907], FDR-corrected $p = 0.018$; two-tailed permutation; see Figures 3B, 3C, and 4B and Tables S6–S8 for full statistics). Further comparing AI versus no-interaction groups, we identified significantly higher ISC in the AI-interaction group across several MD regions, including left anterior, middle parietal lobe, and left middle frontal lobe (Cohen's d s > 0.800 , FDR-corrected p s ≤ 0.026 ; see Table S8).

Together, student-student ISC suggests that structured pre-lecture conversation with both human and AI instructors aligned subsequent neural responses among students in three critical networks, scaffolding video learning. We further applied a multi-class support vector machine (SVM) classification with leave-one-subject-out cross-validation to DMN and MD network ISC values (see STAR Methods) to determine whether human and AI-interaction groups showed more similar neural patterns than the no-interaction group.

The classification demonstrated that ISC features carried predictive information about group membership, with an overall classification accuracy of 53% for the DMN ($p = 0.009$; permutation test) and 63% for the MD network ($p < 0.001$; permutation test)—substantially above the 33% chance level. Notably, the no-interaction group was more readily distinguishable, achieving the highest precision (DMN/MD = 0.75), recall (DMN: 0.60; MD: 0.70), and F1 score (DMN: 0.67; MD: 0.72). By contrast, misclassifications were more frequent between human and

(C) Intersubject gaze alignment. The human-interaction group exhibited significantly greater student-student and student-instructor gaze alignment than the no-interaction group. Student-student gaze alignment was also significantly higher in the human-interaction group than the AI-interaction group. Student-instructor alignment and student-student gaze alignment ($n = 39$) were strongly correlated (p values from permutation tests shown).

(D) Perceived closeness with the instructor. Both human and AI groups reported a significant increase in perceived closeness after the interaction session (p values from paired t tests shown). Note: gray, no interaction; pink, human interaction; blue, AI interaction. Boxplots show medians and interquartile ranges; error bars denote SE.

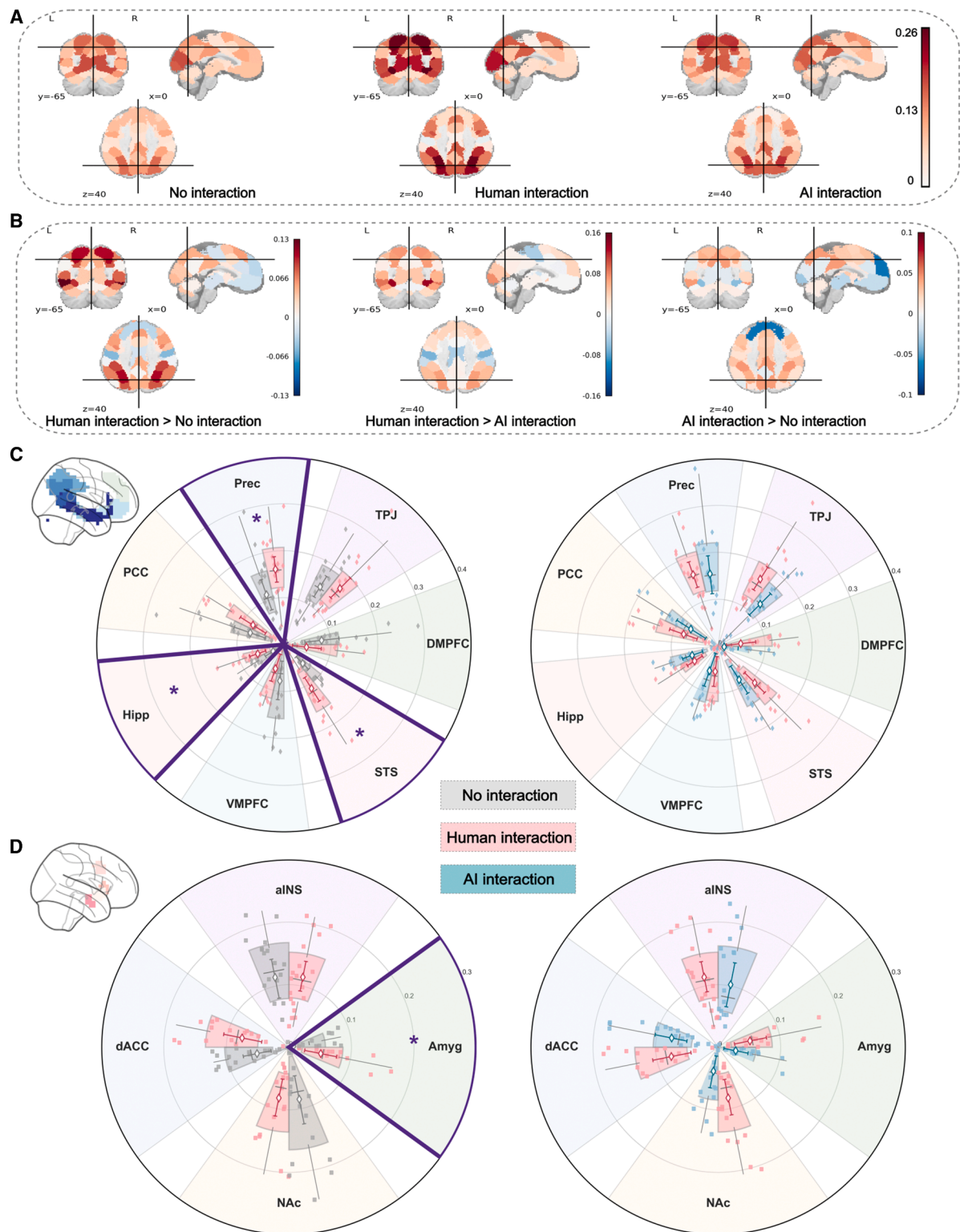


Figure 3. Student neural alignment for the whole brain, DMN, and limbic regions for three groups

(A) Whole-brain ISC maps.

(B) Unthresholded ISC difference maps for human interaction > no interaction (left), human interaction > AI interaction (middle), and AI interaction > no interaction (right). Both interaction groups exhibited enhanced neural alignment across widespread parcels relative to the no-interaction group, whereas differences between human and AI conditions were relatively limited and spatially restricted.

(legend continued on next page)

AI-interaction groups (see the confusion matrices in [Figure 4C](#) and [Table S9](#) for detailed classification results). These findings reinforce that pre-lecture interaction, whether with a human or an AI instructor, promoted partially overlapping neural alignment across students, whereas the absence of interaction yielded more distinct alignment patterns.

Student-instructor neural alignment

For student-instructor neural alignment ($n = 39$), we focused on DMN, given its established role in information transmission, in this case between the instructor and students.²⁰ As neural coupling between a sender and recipient may occur with a temporal lag,^{32,33} we also examined delayed student-instructor coupling by systematically shifting students' time series relative to the instructor's time series by +1 to +6 s (instructor preceding the students) in 1-s intervals.³²

Student-instructor ISC in the DMN was consistently higher in the human than in the no-interaction group. At zero lag, this was significant in the precuneus (Cohen's $d = 0.765$, 95% CI [0.161, 1.556], FDR-corrected $p = 0.042$; see [Figure 5A](#) and [Table S10](#) for full statistics). For delayed ISC, both precuneus and posterior cingulate cortex (PCC) showed significantly greater student-instructor coupling in the human than no-interaction group at lags of 1–5 s (Cohen's $d_s \geq 0.702$, FDR-corrected $p_s \leq 0.035$; see [Figure 5A](#) and [Table S10](#)). These results indicate that human pre-lecture interaction enhances students' alignment with the instructor in core DMN regions, particularly when accounting for processing delays.

Brain-eye-behavior correspondence

Student-student level

To examine the overall eye-brain-learning interrelationships,¹⁹ we concatenated data across all participants ($n = 57$) and applied permutation-based partial correlation analysis controlling for group membership (see [Table S11](#) for details of separate group analyses).

Eye gaze alignment significantly correlated with ISC in several DMN regions (PCC, STS, Hipp, PC, temporoparietal junction [TPJ], and ventromedial prefrontal cortex [vmPFC]; partial Spearman $r_s \geq 0.23$; FDR-corrected $p_s \leq 0.040$, one-sided permutation; see [Table S12](#) for full statistics). It also correlated with comprehension scores (partial Spearman $r = 0.201$, $p = 0.040$, 95% CI [0.077, 0.458], one-sided permutation).

Furthermore, STS ISC was significantly associated with comprehension scores (partial Spearman $r = 0.232$, uncorrected $p = 0.046$; 95% CI [−0.026, 0.461], one-sided permutation). Moreover, PCC ISC was significantly correlated with both comprehension (partial Spearman $r = 0.243$, uncorrected $p = 0.032$, 95% CI [−0.012, 0.466], one-sided permutation) and transfer scores (partial Spearman $r = 0.37$, FDR-corrected $p =$

0.012, 95% CI [0.108, 0.585], one-sided permutation). While STS and PCC ISC-comprehension associations did not survive FDR correction, they were directionally consistent with the FDR-corrected PCC ISC-transfer effect and with prior literature on these regions' role in learning.^{18,20}

These results prompt a question of directionality: does gaze alignment drive brain alignment (i.e., bottom-up processing), or does brain alignment facilitate gaze alignment (i.e., top-down) during learning?³³ To explore this, we conducted mediation analyses involving student gaze ISC, ISC in STS and PCC, and comprehension scores, which avoids circularity³⁴ and allows testing of directional hypotheses.

We first tested whether STS or PCC ISC mediated the relationship between gaze ISC and comprehension performance. STS ISC emerged as a significant mediator: greater gaze alignment correlated with stronger STS ISC ($\beta = 0.335$, 95% CI [0.081, 0.560], $p = 0.011$), which in turn was associated with better comprehension ($\beta = 0.31$, 95% CI [0.053, 0.567], $p = 0.019$). The indirect effect via STS ISC was significant ($\beta = 0.077$, 95% CI [0.008, 0.195], $p = 0.047$), while the direct effect was not ($\beta = 0.239$, 95% CI [−0.029, 0.506], $p = 0.080$; see [Figure 5B](#) and [Table S13](#) for full statistics), indicating full mediation. These results suggest that shared gaze patterns enhance comprehension primarily via STS alignment, a region implicated in gaze processing and social perception.³⁵

We further reversed the mediation model to test whether gaze alignment mediated the relationship between STS/PCC ISC and comprehension. A significant indirect effect from PCC ISC to comprehension via gaze alignment was identified ($\beta = 0.081$, 95% CI [0.008, 0.218], $p = 0.042$; see [Figures 5B](#) and [Table S13](#) for full statistics). These findings suggest a top-down pathway, whereby increased PCC alignment, a core DMN region implicated in shared mental models and internal attention,³⁶ may facilitate more coordinated gaze and better learning.

Beyond DMN ISC and learning performance, we also examined how DMN ISC related to students' perceived social closeness with the instructor. In the human-interaction group, precuneus ISC was correlated with increases in post-conversation instructor closeness ($r = 0.519$, 95% CI [0.154, 0.828], FDR-corrected $p = 0.030$, one-sided permutation). In the AI-interaction group, vmPFC ISC showed a similar positive association with increases in perceived closeness ($r = 0.353$, 95% CI [−0.210, 0.821], uncorrected $p = 0.049$, one-sided permutation). These findings suggest that DMN hubs may support the experience of social connection for both human and AI interactions.

Student-instructor level

We again concatenated data across human- and no-interaction groups while statistically controlling for the group membership ($n = 39$). Student-instructor gaze alignment significantly

(C) Student-student DMN ISC across groups. The human-interaction group exhibited significantly higher ISCs in the precuneus, hippocampus, and STS ($q_s < 0.05$, FDR-corrected, permutation test) than the no-interaction group. No significant differences were observed between the human and AI-interaction groups in any DMN ROI.

(D) Student-student limbic network ISC across groups. The human-interaction group showed significantly greater ISC in the amygdala than the no-interaction group ($q < 0.05$, FDR-corrected, permutation test). No significant differences were found between the human and AI-interaction groups across any limbic ROI. Note: ISC values reflect leave-one-out ISCs for each ROI. Means are shown as unfilled markers with bold outlines, and individual data points are overlaid in their respective colors. Error bars represent 95% CIs from 1,000 bootstrap resamples. Significant differences are marked with asterisks.

See also [Tables S6](#) and [S7](#).

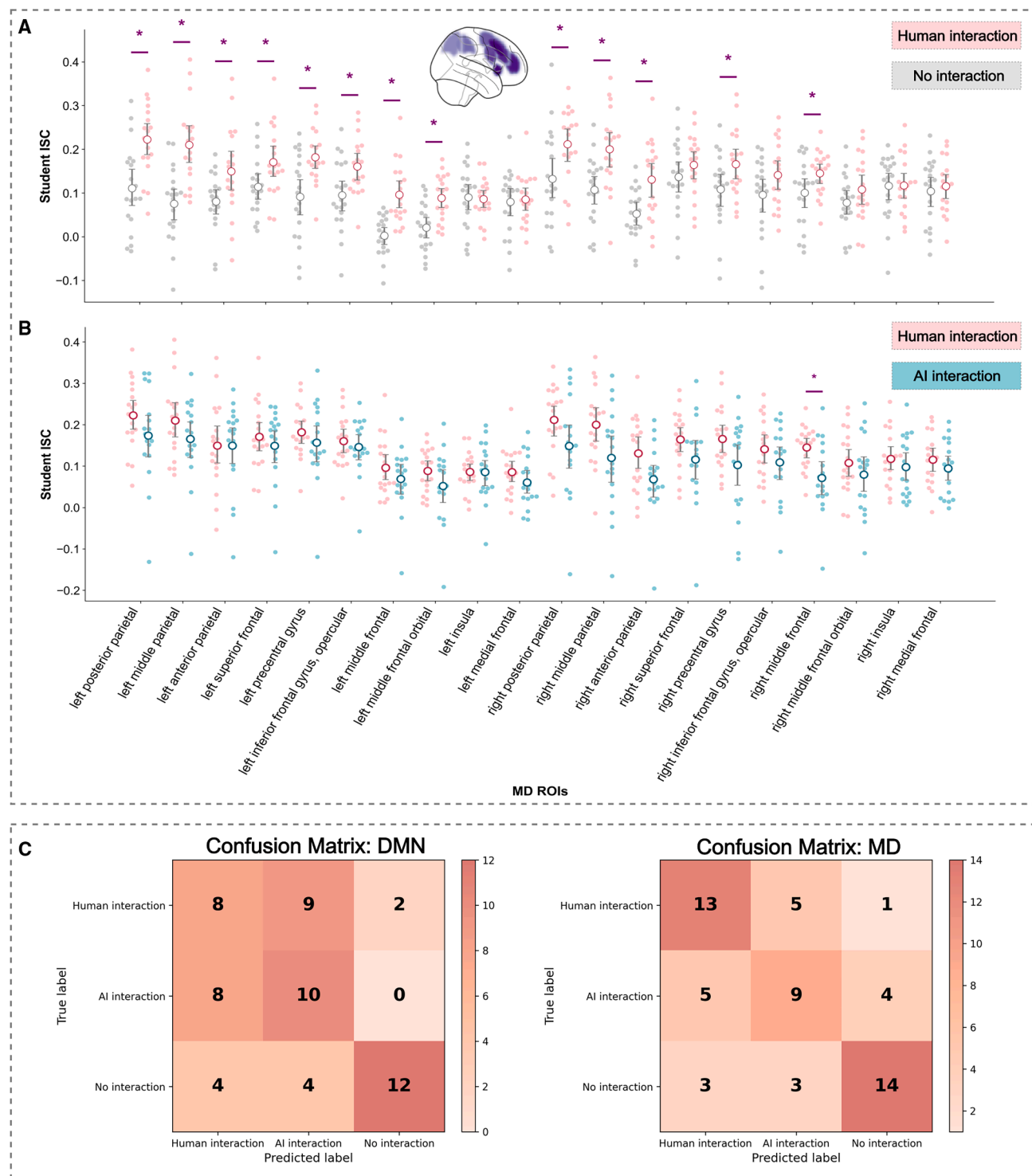


Figure 4. Neural alignment in the MD network and classification of interaction groups

(A and B) Student-student MD network ISC across groups. The human-interaction group exhibited significantly higher ISC in multiple MD regions (FDR-corrected $q < 0.05$, permutation test) than the no-interaction group. The AI-interaction group showed significantly lower ISC than the human-interaction group in the right midfrontal gyrus only (FDR-corrected $q < 0.05$, permutation test). Error bars represent 95% CIs from 1,000 bootstrap resamples. Significant differences are marked with asterisks.

(legend continued on next page)

correlated with comprehension (partial Spearman $r = 0.248$, $p = 0.043$, 95% CI [0.027, 0.505], one-sided permutation) and transfer scores (partial Spearman $r = 0.389$, $p = 0.007$, 95% CI [0.066, 0.651], one-sided permutation).

We focused our student-instructor ISC analysis on PCC,²⁰ as ample prior evidence has indicated a reliable time-lagged speaker-listener neural coupling in PCC in verbal communication.^{32,37} Instructor-student ISCs in the PCC at 1 to 5-s lags were significantly associated with difficult comprehension question scores (partial Spearman $r_s \geq 0.260$; FDR-corrected $p_s \leq 0.048$ across time lags; see Table S12).

Mediation analyses revealed bidirectional pathways. Student-instructor PCC ISC predicted learning performance indirectly via increased student-instructor gaze alignment at 3, 4, and 5-s lags ($\beta_s \geq 0.103$, FDR-corrected across 1–5-s lags $p_s \leq 0.049$; see Table S14 for full mediation results). At 3-s lag, greater student-instructor gaze alignment was also associated with better comprehension of difficult items via PCC coupling ($\beta = 0.094$, 95% CI [0.007, 0.256], uncorrected $p = 0.048$; see Table S14). Together, consistent with the student-student level results, student-instructor gaze and neural alignment interact bi-directionally to support learning.

DMN-limbic interaction predicts learning after human/AI instructor-student interaction

Social agency theory^{14,15} posits that social cues elicit social-affective responses in learners, thereby promoting deeper cognitive processing and enhancing learning, a hypothesis rarely tested with neurocognitive evidence. Here, we examined whether enhanced social cues, operationalized as pre-lecture interaction, would orchestrate functional interactions between the DMN (implicated in comprehension and learning) and limbic network (supporting socio-emotional processing) to predict learning (see Figure 6A and STAR Methods).

For the human-interaction group, the pairwise DMN-limbic intersubject functional connectivity (ISFC)³⁸ similarity matrix indexing DMN-limbic functional coupling significantly correlated with the pairwise comprehension score similarity ($r = 0.19$, 95% CI [0.048, 0.326], $p = 0.048$, Mantel permutation test with 10,000 permutations). For the AI-interaction group, this ISFC-representational similarity analysis (RSA)³⁹ analysis showed a marginal effect ($r = 0.14$, 95% CI [−0.029, 0.297], $p = 0.060$, Mantel test with 10,000 permutations). By contrast, no such relationship between ISFC and comprehension scores was observed in the no-interaction group ($r = 0.04$, 95% CI [−0.111, 0.179], $p = 0.372$; Mantel permutation test with 10,000 permutations).

DISCUSSION

Natural conversations can shape how we process new incoming information and understand the world.²⁵ Yet, in many constrained educational settings, especially online learning, natural

conversation is absent, and critical social scaffolding is lacking. In this study, we show that learners who had a brief structured interaction (~8–10 min) with an instructor—whether human or AI—demonstrated greater neural alignment during subsequent video learning and improved learning performance compared with those without such interaction. By integrating behavioral, eye-tracking, and fMRI data, we identify multimodal pathways linking (1) structured pre-lecture social interaction to (2) synchronized visual attention allocation, (3) enhanced neural alignment in attention, semantic and socio-emotional networks, and (4) learning outcomes. Notably, our findings show that AI-led interactions produced comparable learning outcomes and paralleled the level of neural alignment obtained with human interactions. Together, these results paint a convergent picture: brief, well-designed social interactions—human or AI-led—are associated with synchronized attentional and neural patterns that correlate with enhanced learning, suggesting potential pathways for incorporating social scaffolding via AI-mediated communications in online educational contexts.

Brain-eye-behavior correspondence: Mechanistic pathways to improved learning

Our data show that student-student gaze alignment is associated with DMN neural alignment and learning outcomes, converging with prior studies linking DMN alignment to learning^{18,20} and supporting recent evidence for eye-brain-learning correspondences.^{19,33} Critically, mediation analyses revealed reciprocal dynamics: gaze alignment was associated with better comprehension via STS alignment, consistent with STS's well-established role in visual social perception³⁵ and language processing,⁴⁰ supporting bottom-up sensory-semantic processing. Concurrently, PCC alignment, implicated in shared internal models and integrative processing,³⁶ correlated with enhanced learning alongside coordinated gaze behavior via top-down attentional guidance.

Notably, our student-instructor analysis extends earlier teacher-student neural coupling studies using fMRI²⁰ by demonstrating the robustness of such alignment, not only in brain activity but also in gaze patterns. Student-instructor alignment, expressed through coordinated gaze and delayed PCC alignment, was associated with learning outcomes as was student-student alignment. These results suggest that aligning with the instructor via visual attention and neural coupling establishes a shared referential framework, which facilitates student-student alignment.

A key contribution of our study is the integration of student-instructor and student-student gaze and neural alignment, revealing a possible multi-stage, reciprocal mechanism linking attention and neural resource allocation to learning dynamics (see Figure 7). Extending previous studies directly correlating neural alignment with learning outcomes,^{18,20} our approach reveals that brief social interaction sets off a full cascade of neurocognitive processes. Specifically, students first align their

(C) Confusion matrices from multi-class classification using DMN ISC values (left) and MD network (right). The no-interaction group was more easily distinguishable, whereas the human and AI-interaction groups were more frequently misclassified as one another, indicating overlapping neural patterns following pre-lecture interaction.

See also Tables S8 and S9.

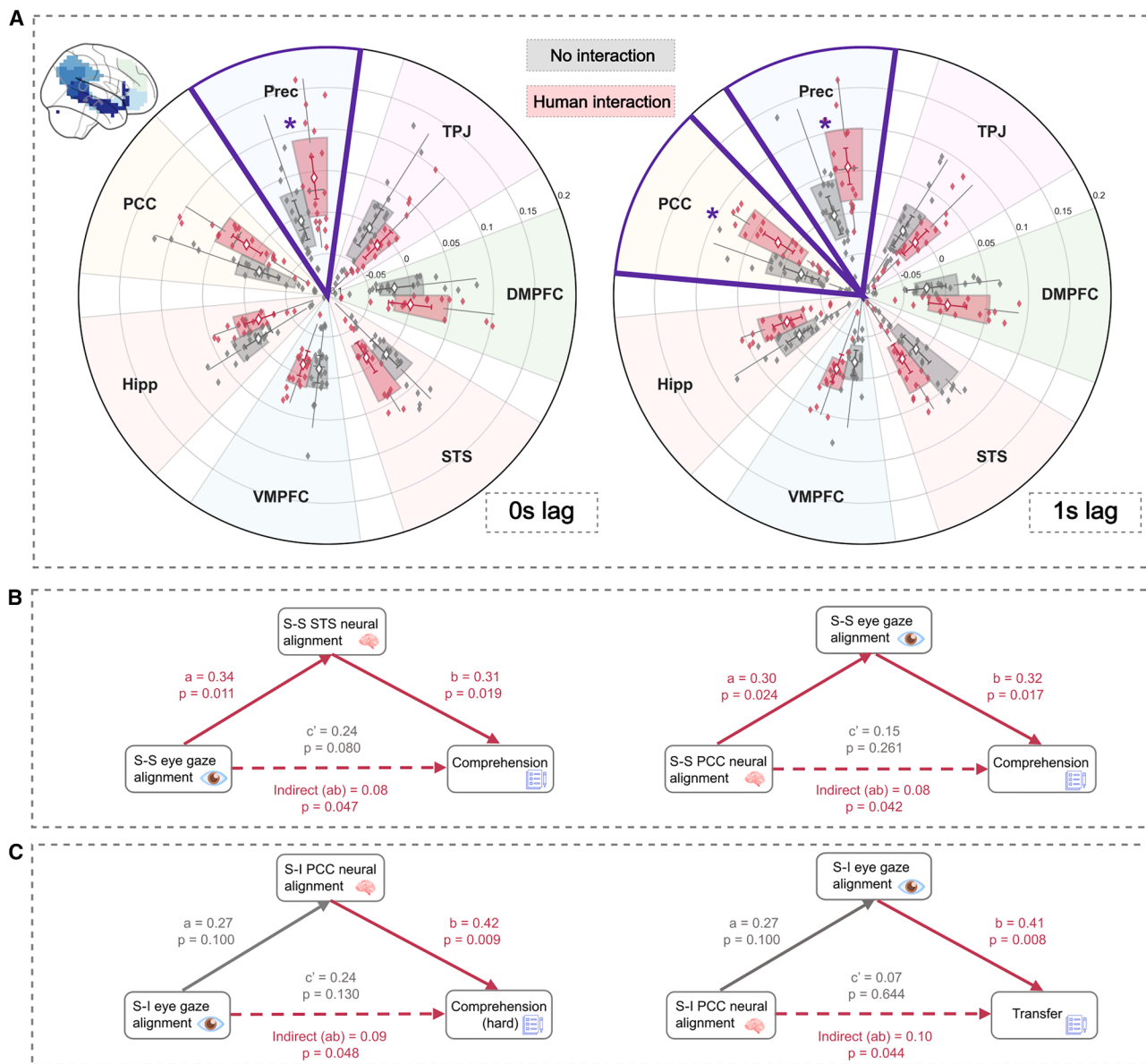


Figure 5. Student-instructor neural alignment and mediation pathways supporting learning

(A) Student-instructor DMN ISC for the no- and human-interaction groups. The human-interaction group showed greater student-instructor alignment than the no-interaction group in the precuneus at lag 0 s, extending to both the precuneus and PCC across the 1–5-s lag window.

(B) Mediation analyses for student-student (S-S) gaze alignment, neural alignment, and comprehension. Left: STS ISC mediated the relationship between S-S gaze alignment and comprehension (indirect $p = 0.047$), suggesting that shared gaze patterns enhance comprehension via STS alignment. Right: S-S gaze alignment mediated the relationship between PCC ISC and comprehension (indirect $p = 0.042$), suggesting that PCC coupling shaped shared gaze behavior that supported comprehension.

(C) Mediation analyses for student-instructor (S-I) gaze alignment, neural alignment, and learning. Left: S-I gaze alignment predicted delayed S-I PCC ISC at 3-s lag, which in turn predicted comprehension of difficult content (indirect $p = 0.048$). Right: delayed S-I PCC ISC at 3-s lag predicted transfer performance via S-I gaze alignment (indirect $p = 0.044$), together revealing bidirectional gaze-neural coupling in supporting learning. Note: ISC values reflect student-instructor pairwise ISCs per ROI, means shown as unfilled markers with bold outlines, and individual data points overlaid in respective colors. Error bars represent 95% CIs from 1,000 bootstrap resamples. Significant differences are marked with asterisks, significant paths are in red, and dashed paths denote indirect effects.

See also [Tables S10](#), [S12](#), [S13](#), and [S14](#). Icons from [Flaticon.com](#) are used.

mental representations to the instructor's narrative stream, manifested as increased student-instructor PCC alignment, accompanying enhanced student-instructor joint attention.²³

Joint attention indexed by gaze alignment then serves as the behavioral channel through which this mental representation alignment is expressed and refined.³³ Once locked onto

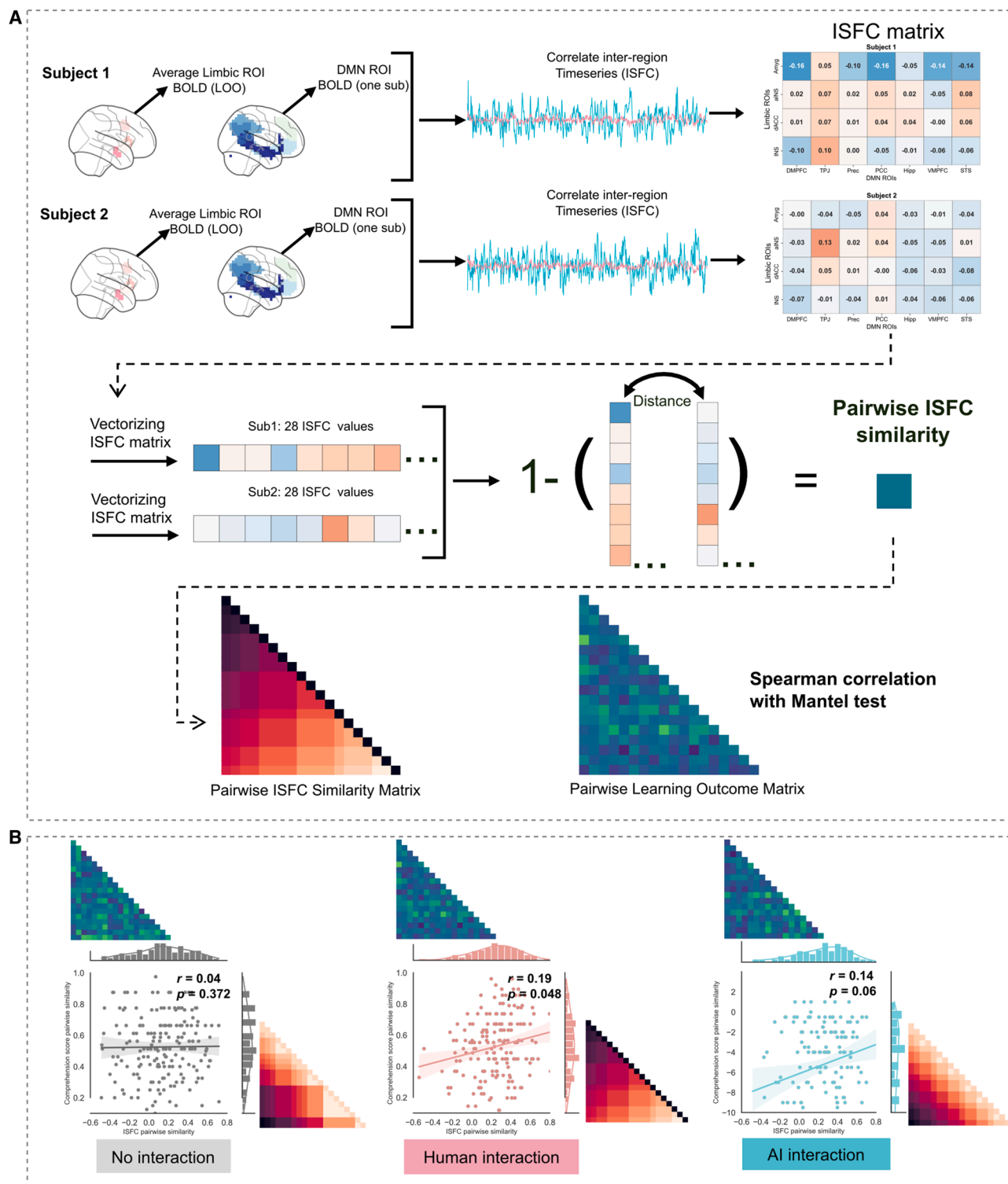


Figure 6. ISFC-RSA

(A) Overview of the ISFC-RSA pipeline. For each participant, a time series of DMN regions was extracted and correlated with the leave-one-out average time series of limbic regions to compute ISFC across all DMN-limbic ROI pairs. Each participant's ISFC profile was reshaped into a feature vector. Pairwise similarity between ISFC vectors was calculated using $1 - \text{distance}$ (e.g., Pearson), forming a participant-by-participant ISFC similarity matrix. We then conducted inter-subject representational similarity analysis (ISRSA) by correlating this matrix with the corresponding pairwise learning performance similarity matrix.

(legend continued on next page)

the instructor's visual trajectory, coordinated visual attention corresponds to stronger student-student neural alignment in the STS, a representative semantic hub,^{41,42} reflecting a shared internalization and uptake of the knowledge. Together, we propose a possible bidirectional pathway: top-down alignment of internal mental states relates to enhanced attention allocation, and bottom-up sharing of visual focus parallels convergent sense-making and semantic integration. These dynamics potentially offer a plausible pathway from instructor guidance to peer convergence and enhanced learning under socially enriched contexts, coinciding with a recent fNIRS study⁴³ illustrating reciprocal coordination of brain and behavioral signals (e.g., mutual gaze and joint attention) during social learning.

Beyond these attention-neural cascades, our ISFC-RSA analysis adds a complementary dimension: DMN-limbic functional coupling was associated with learning performance only in socially enriched conditions following brief instructor-student interaction (whether human or AI). This pattern situates social engagement as a critical scaffold that potentially orchestrates higher-order sense-making and integrative processing, alongside attentional and semantic pathways. By embedding socio-emotional bonding within the brain-eye-behavior cascade, these findings provide a direct neurocognitive account of social agency theory,⁴⁴ highlighting how enhanced social factors relate to coordinated engagement of cognitive and emotional systems to jointly support learning.

Why does structured pre-lecture interaction work?

Several mechanisms likely underlie the observed effects of brief social conversations. From a common ground building and intentionality sharing perspective,^{45,46} our structured interactions were purposefully consensus-oriented and intentionality-calibrating: the instructor explicitly connected course content to students' personal interests and majors, building consensus about relevance that may have encouraged engagement. This was evident in higher self-reported efforts and stronger student-instructor and student-student DMN alignment during learning. These findings extend prior work showing brief consensus-seeking²⁵ or even unrelated conversations²⁶ could align future neural responses, demonstrating that such effects could potentially operate in online educational contexts.

Beyond consensus, self-disclosure of both the instructor and the students,⁴⁷ motivational support expressed through genuine interest in students' hobbies and majors,⁴⁸ and a consistently positive tone⁴⁹ may have fostered social bonding and motivation. Correspondingly, students reported greater instructor-student closeness, mirrored in stronger neural alignment in the amygdala and precuneus, with closeness significantly correlating with precuneus alignment. Given the precuneus's central role in the mentalizing network and DMN,^{50,51} this alignment likely reflects students' internalization of the instructor's perspective. Furthermore, a brief course preview at the end of the interaction may function as "semantic grounding,"⁵² cali-

brating students' attention and expectations. This may explain their heightened MD network alignment, which supports top-down cognitive control during learning. Of course, it is important to note that while we observe enhanced neural alignment in several brain networks after pre-lecture social interaction and associations between neural alignment and learning outcomes, the functional significance of neural alignment and its underlying mechanisms are subject to debate (see Chen et al.,⁵³ Zadbood et al.,⁵⁴ Zada et al.,^{55,56} and Hasson et al.⁵⁷ for a discussion of asynchronous neural alignment measured using fMRI and Hamilton,⁵⁸ Holroyd,⁵⁹ Moreau and Dumas,⁶⁰ Novembre and Iannetti,⁶¹ Jiang et al.,⁶² Feng and Lu,⁶³ and Lu⁶⁴ for discussion of hyper-scanning neural synchrony typically using fNIRS). We therefore consider neural alignment as an observable pattern associated with—but not necessarily causally linked to—enhanced learning outcomes.

Finally, positioning interaction before learning contrasts with recent work showing that structured post-lecture discussions also enhance learning,⁶⁵ likely through distinct mechanisms: pre-lecture interaction establishes cognitive schemas and motivational readiness before content exposure, priming sustained attention and resource allocation for complex material.⁶⁶ By contrast, post-lecture interaction operates through knowledge consolidation and integration of already-encountered content. Our interaction template, combining affective bonding (social closeness) with cognitive priming (relevance framing, semantic grounding), may thus explain gains on challenging comprehension items requiring sustained, integrative processing.

The potential of AI instructors: Empathy, personalization, realism, knowledge base, and (still) divergence from the human instructor

Our study illustrates that LLM-based agents can effectively simulate pedagogically effective interactions in a lab-based educational setting. The AI-mediated interaction closely mirrored the structured human instructor-student template: both groups reported comparable effort levels, neural alignment in the DMN, MD, and limbic systems and achieved equivalent learning outcomes.

Several factors may explain this equivalence despite the students' explicit awareness of AI's non-human nature. First, GPT-4 could potentially trigger socio-emotional responses in students. It simulates theory of mind,⁶⁷ infers student needs, and dynamically adapts to produce empathetic utterances.^{68–70} For instance, when one participant coughed during the interaction, the AI instructor responded: "I heard you coughing just now; please take good care of yourself." The AI-interaction group reported significantly increased social closeness with the AI instructor post-interaction, which correlated with vmPFC alignment (though less robustly than the human-interaction group), implying the involvement of the mentalizing network.⁵⁰ These findings align with the media equation theory,⁷¹ whereby people tend to anthropomorphically apply human-human social

(B) ISFC-RSA results by group. In the no-interaction group, there was no credible evidence of significant correspondence between ISFC and comprehension similarity ($r = 0.04$, $p = 0.372$, Mantel test, 10,000 permutations). In the human-interaction group, pairwise ISFC similarity was significantly correlated with comprehension similarity ($r = 0.19$, $p = 0.048$). For the AI-interaction group, the ISFC-RSA analysis showed a marginal effect ($r = 0.14$, $p = 0.060$).

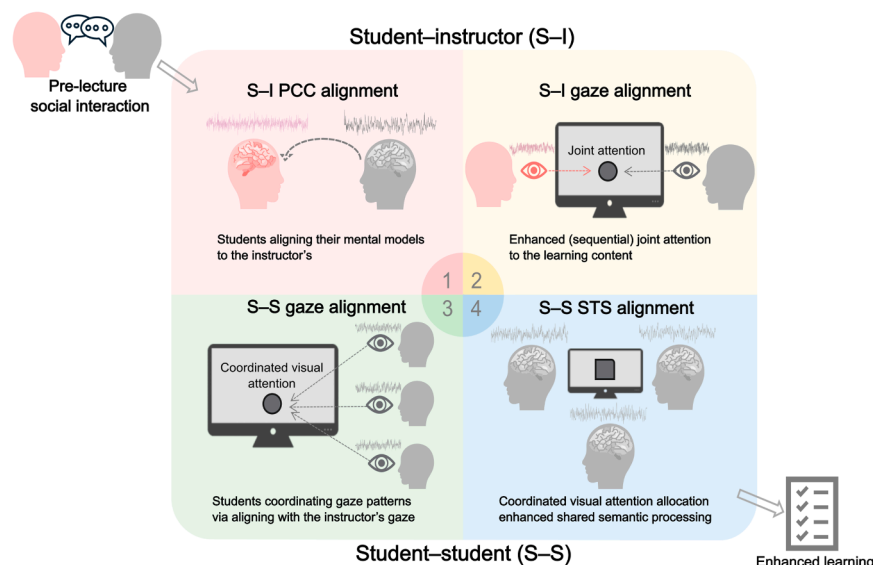


Figure 7. Scaffolding from pre-lecture social interaction to enhanced learning: a neuro-cognitive account

Brief pre-lecture instructor-student interaction triggers a multi-stage cascade integrating neural alignment and gaze alignment across student-instructor and student-student levels. (1) During video learning, students align their mental models to the instructor's narrative, reflected in increased instructor-student (I-S) PCC alignment. (2) This neural alignment is associated with joint visual attention, indexed by I-S gaze alignment during learning. Bidirectional coupling between neural alignment and gaze alignment indicates reciprocal reinforcement. (3) Shared visual attention with the instructor coordinates students' gaze with their peers (S-S gaze alignment). (4) Coordinated visual attention allocation further links to S-S neural alignment in the STS, reflecting internalization of learning content. This cascade reveals both top-down (neural alignment guiding attention) and bottom-up effects (shared attention enhancing

neural convergence), ultimately linking to improved learning outcomes through coordinated engagement of cognitive, attentional, and socio-emotional systems. Icons from [Flaticon.com](https://flaticon.com) are used.

interaction norms to technologies exhibiting human-like social cues.

Moreover, AI instructor realism was enhanced through multimodal design, including realistic facial animations (e.g., eye blinks and synchronized lip movements) and high-quality text-to-speech synthesis.^{72,73} These multimodal human-like affordances, further combined with synchronized audiovisual feedback, social responsiveness, and linguistic personalization, likely contributed to the observed parallel neural responses and learning outcomes across interaction groups and advanced the embodiment principle of multimedia learning.¹⁵ GPT-4's extensive knowledge base enabled diverse and relevant utterances that engaged learner interests and prompted cognitive preparedness, consistent with the observed MD network alignment, suggesting that the AI instructor's cognitive grounding effectively supported top-down cognitive resource allocation, paralleling the human-interaction group. This highlights AI's pedagogical strength in scaffolding learning through cognitive structuring and its versatility in supporting scalable, personalized educational interactions, consistent with broader evidence of LLM effectiveness in education.^{74,75}

In our study, participants were explicitly informed of the AI's nature before interacting. That the AI instructor enhanced perceived closeness and produced equivalent learning outcomes suggests that advanced conversational AI can provide effective educational support even when learners are fully aware of its artificial nature. This is consistent with the literature⁷¹ and a recent study⁷⁶ suggesting that people tend to anthropomorphize AI even when AI does not truly possess human values, emotion, or empathy. Yet, notable differences between AI and human instructor interactions remained. Compared with the human-interaction group, the AI-interaction group showed (1) lower perceived closeness despite increased closeness ratings post-interaction, (2) reduced gaze alignment during learning, and (3) only marginal DMN-limbic coupling associations with learning outcomes. The reduced

closeness ratings align with recent findings that explicit AI labeling could potentially diminish heightened perceived empathy,^{77,78} despite AI-generated content being empathetic.

These findings suggest that LLM-based AI can simulate social interaction effectively enough to support learning, albeit via partially distinct pathways. The AI-interaction group appeared to leverage stronger top-down processing: robust MD network engagement (cognitive scaffolding via GPT-4) supported DMN alignment for higher-level comprehension and learning with less reliance on the bottom-up visual attention alignment, reflected in reduced gaze alignment and affective resonance. By contrast, the human-interaction group engaged both pathways more equally with heightened cognitive scaffolding plus stronger gaze alignment and DMN/limbic coupling. This dual-pathway interpretation reconciles the observed equivalent learning outcomes with divergent neural profiles and refines the media equation theory⁷¹: people apply social norms to human-like technologies but not always to the exact same depth or affective intensity as with real humans. Critically, both pathways proved educationally effective, suggesting multiple routes to successful learning support. Thus, AI educational systems may be effective not by being exactly like human instruction but by engaging sufficient social-emotional resonance while leveraging computational strengths in knowledge access and personalization.

Limitations and future directions

Several constraints of our study suggest future directions. First, future research could systematically manipulate interaction components (e.g., self-disclosure, motivational content, and relevance framing) to identify which drives specific neural and behavioral effects. Second, while mediation analyses revealed bidirectional pathways between neural alignment, gaze alignment, and learning outcomes, our design did not isolate independent contributions of specific factors (e.g., physical presence, gaze dynamics, and interaction naturalness) that

covared across human and AI conditions. Future work could manipulate these systematically (e.g., comparing embodied robots versus screen-based avatars to establish which factors are necessary or sufficient for optimal learning support). Third, text-to-speech and speech-to-text delays constrained real-time dynamics between students and the AI instructor. Robustness despite these constraints suggests resilience, though smoother interaction might yield stronger effects. Fourth, while our sample size is consistent with naturalistic fMRI standards and provided adequate power for the moderate-to-large effects observed, larger samples would increase sensitivity for nuanced effects and enable examination of individual differences in responses to human versus AI instructors.

Building on our controlled laboratory findings, future work could test whether these effects translate to authentic classroom contexts and generalize across diverse subject domains and student populations. Longitudinal designs and applications in classroom contexts could help further determine if neural and gaze alignment effects persist across multiple sessions and whether these effects predict long-term retention. Finally, one could ask whether short-term instructor-student alignment not only enhances content comprehension but also fosters social connectedness and trust, key to sustained engagement in student learning in general.

Conclusion

Brief, structured pre-lecture social interactions, whether with human or AI instructors, energize an otherwise passive, isolating environment in pre-recorded video lectures and turn solitary screen time into a more dynamic, socially engaging experience. By aligning students' neural responses during the subsequent video lecture, these interactions enhance learning through coordinated engagement of attentional, semantic, and socio-emotional brain networks, providing neurocognitive evidence that social scaffolding fundamentally shapes how we process information. That AI instructors achieved comparable effects through partially distinct mechanisms suggests multiple viable routes to effective educational support. As AI systems grow more sophisticated, understanding these pathways will be critical for designing scalable, socially enriched learning environments. This points toward a looming future for the social fabric of education where even an AI instructor can pause, notice a student's cough, and respond with care. These subtle aspects of human communication, if successfully realized in today's AI-mediated learning systems, may help cultivate what it means to feel seen, heard, and socially connected in a digital classroom.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Ping Li (pingliandrew@ust.hk).

Materials availability

This study did not generate new, unique reagents.

Data and code availability

- Preprocessed MRI, eye gaze, and behavioral data have been deposited at OSF at <https://doi.org/10.17605/OSF.IO/FU5H6> and are publicly available as of the date of publication.

- All original code has been deposited at OSF at <https://doi.org/10.17605/OSF.IO/FU5H6> and is publicly available as of the date of publication.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

ACKNOWLEDGMENTS

This research was supported by a grant from the Hong Kong Research Grants Council (project no. 15611524) and the Sin Wai Kin Foundation. P.L. was affiliated with the Hong Kong Polytechnic University (PolyU) and Y.P. was supported by a Research Postgraduate Scholarship from PolyU when the study was carried out. Y.H. is now affiliated with the City University of Hong Kong. The authors thank Minghua Wu for assistance with experiment material preparation and data collection.

AUTHOR CONTRIBUTIONS

Y.P. and P.L., conceptualization; Y.P., Y.H., and Y.L., investigation, data curation, formal analysis, methodology, and visualization; S.A.N., methodology; Z.G., software for AI instructor; P.L. and S.A.N., supervision; P.L., funding acquisition; Y.P. and P.L., writing – original draft; all authors, writing – review & editing.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**
 - Participants
- **METHOD DETAILS**
 - Design
 - Data Acquisition
 - fMRI data preprocessing
 - Neural Time Series Extraction
- **QUANTIFICATIONS AND STATISTICAL ANALYSIS**
 - Behavioral data processing and analysis
 - Eye-tracking data processing and analysis
 - Student–student & student–instructor ISC
 - Classification of group
 - Correlational and mediation analysis
 - ISFC-RSA analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.neuron.2026.04.005>.

Received: October 28, 2025

Revised: March 3, 2026

Accepted: April 3, 2026

Published: April 30, 2026

REFERENCES

1. Paulsen, J., and McCormick, A.C. (2020). Reassessing Disparities in Online Learner Student Engagement in Higher Education. *Educ. Res.* 49, 20–29. <https://doi.org/10.3102/0013189X19898690>.
2. Guo, P.J., Kim, J., and Rubin, R. (2014). How video production affects student engagement. In *Proceedings of the First ACM Conference on*

- Learning @ Scale Conference (ACM), pp. 41–50. <https://doi.org/10.1145/2556325.2566239>.
3. Hollister, B., Nair, P., Hill-Lindsay, S., and Chukoskie, L. (2022). Engagement in Online Learning: Student Attitudes and Behavior During COVID-19. *Front. Educ.* 7, 851019. <https://doi.org/10.3389/educ.2022.851019>.
4. Patricia Aguilera-Hermida, A. (2020). College students' use and acceptance of emergency online learning due to COVID-19. *Int. J. Educ. Res. Open* 1, 100011. <https://doi.org/10.1016/j.ijedro.2020.100011>.
5. Balconi, M., Angioletti, L., and Cassioli, F. (2023). Hyperscanning EEG Paradigm Applied to Remote vs. Face-To-Face Learning in Managerial Contexts: Which Is Better? *Brain Sci.* 13, 356. <https://doi.org/10.3390/brainsci13020356>.
6. De Felice, S., Hatilova, A., Trojan, F., Tsui, I., and Hamilton, A.F. de C. (2023). Autistic adults benefit from and enjoy learning via social interaction as much as neurotypical adults do. *Mol. Autism* 14, 33. <https://doi.org/10.1186/s13229-023-00561-6>.
7. De Felice, S., Vigliocco, G., and Hamilton, A.F. de C. (2021). Social interaction is a catalyst for adult human learning in online contexts. *Curr. Biol.* 31, 4853–4859.e3. <https://doi.org/10.1016/j.cub.2021.08.045>.
8. Kim, J., Merrill, K., Xu, K., and Kelly, S. (2022). Perceived credibility of an AI instructor in online education: The role of social presence and voice features. *Comput. Hum. Behav.* 136, 107383. <https://doi.org/10.1016/j.chb.2022.107383>.
9. Raza, M., Jahangir, Z., Riaz, M.B., Saeed, M.J., and Sattar, M.A. (2025). Industrial applications of large language models. *Sci. Rep.* 15, 13755. <https://doi.org/10.1038/s41598-025-98483-1>.
10. Jeon, J., and Lee, S. (2023). Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. *Educ. Inf. Technol.* 28, 15873–15892. <https://doi.org/10.1007/s10639-023-11834-1>.
11. Wu, R., and Yu, Z. (2024). Do AI chatbots improve students learning outcomes? Evidence from a meta-analysis. *Br. J. Educ. Technol.* 55, 10–33. <https://doi.org/10.1111/bjet.13334>.
12. Tlili, A., Saqr, K., Salha, S., and Huang, R. (2025). Investigating the effect of artificial intelligence in education (AIED) on learning achievement: A meta-analysis and research synthesis. *Inf. Dev.* 41, 825–842. <https://doi.org/10.1177/026666669241304407>.
13. Becker, B. (2025). Will our social brain inherently shape and be shaped by interactions with AI? *Neuron* 113, 2037–2041. <https://doi.org/10.1016/j.neuron.2025.04.034>.
14. Fiorella, L., and Mayer, R.E. (2021). Principles Based on Social Cues in Multimedia Learning: Personalization, Voice, Image, and Embodiment Principles. In *The Cambridge Handbook of Multimedia Learning*, R.E. Mayer and L. Fiorella, eds. (Cambridge University Press), pp. 277–285. <https://doi.org/10.1017/9781108894333.029>.
15. Mayer, R.E. (2021). *Multimedia Learning Third Edit* (Cambridge University Press).
16. Nastase, S.A., Gazzola, V., Hasson, U., and Keysers, C. (2019). Measuring shared responses across subjects using intersubject correlation. *Soc. Cogn. Affect. Neurosci.* 14, 667–685. <https://doi.org/10.1093/scan/nsz037>.
17. Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., and Malach, R. (2004). Intersubject Synchronization of Cortical Activity During Natural Vision. *Science* 303, 1634–1640. <https://doi.org/10.1126/science.1089506>.
18. Meshulam, M., Hasenfratz, L., Hillman, H., Liu, Y.F., Nguyen, M., Norman, K.A., and Hasson, U. (2021). Neural alignment predicts learning outcomes in students taking an introduction to computer science course. *Nat. Commun.* 12, 1922. <https://doi.org/10.1038/s41467-021-22202-3>.
19. Gu, C., Peng, Y., Nastase, S.A., Mayer, R.E., and Li, P. (2024). Onscreen presence of instructors in video lectures affects learners' neural synchrony and visual attention during multimedia learning. *Proc. Natl. Acad. Sci. USA* 121, e2309054121. <https://doi.org/10.1073/pnas.2309054121>.
20. Nguyen, M., Chang, A., Micciche, E., Meshulam, M., Nastase, S.A., and Hasson, U. (2022). Teacher-student neural coupling during teaching and learning. *Soc. Cogn. Affect. Neurosci.* 17, 367–376. <https://doi.org/10.1093/scan/nsab103>.
21. Shen, Y.L., Hyon, R., Wheatley, T., Kleinbaum, A.M., Welker, C.L., and Parkinson, C. (2025). Neural similarity predicts whether strangers become friends. *Nat. Hum. Behav.* 9, 2285–2298. <https://doi.org/10.1038/s41562-025-02266-7>.
22. Parkinson, C., Kleinbaum, A.M., and Wheatley, T. (2018). Similar neural responses predict friendship. *Nat. Commun.* 9, 332. <https://doi.org/10.1038/s41467-017-02722-7>.
23. Dikker, S., Wan, L., Davidesco, I., Kaggen, L., Oostrik, M., McClintock, J., Rowland, J., Michalareas, G., Van Bavel, J.J., Ding, M., et al. (2017). Brain-to-Brain Synchrony Tracks Real-World Dynamic Group Interactions in the Classroom. *Curr. Biol.* 27, 1375–1380. <https://doi.org/10.1016/j.cub.2017.04.002>.
24. Bevilacqua, D., Davidesco, I., Wan, L., Chaloner, K., Rowland, J., Ding, M., Poeppel, D., and Dikker, S. (2018). Brain-to-brain synchrony and learning outcomes vary by student-teacher dynamics: Evidence from a real-world classroom electroencephalography study. *J. Cogn. Neurosci.* 31, 401–411. https://doi.org/10.1162/jocn_a_01274.
25. Sievers, B., Welker, C., Hasson, U., Kleinbaum, A.M., and Wheatley, T. (2024). Consensus-building conversation leads to neural alignment. *Nat. Commun.* 15, 3936. <https://doi.org/10.1038/s41467-023-43253-8>.
26. De Felice, S., Hakim, U., Gunasekara, N., Pinti, P., Tachtsidis, I., and Hamilton, A. (2024). Having a chat and then watching a movie: how social interaction synchronises our brains during co-watching. *Open Neurosci.* 3, kvae006. <https://doi.org/10.1093/oons/kvae006>.
27. Wheatley, T., Thornton, M.A., Stolk, A., and Chang, L.J. (2024). The Emerging Science of Interacting Minds. *Perspect. Psychol. Sci.* 19, 355–373. <https://doi.org/10.1177/17456916231200177>.
28. Iyer, S., Collier, E., Broom, T.W., Finn, E.S., and Meyer, M.L. (2024). Individuals who see the good in the bad engage distinctive default network coordination during post-encoding rest. *Proc. Natl. Acad. Sci. USA* 121, e2306295121. <https://doi.org/10.1073/pnas.2306295121>.
29. Fedorenko, E., Duncan, J., and Kanwisher, N. (2013). Broad domain generality in focal regions of frontal and parietal cortex. *Proc. Natl. Acad. Sci. USA* 110, 16616–16621. <https://doi.org/10.1073/pnas.1315235110>.
30. Wen, T., Duncan, J., and Mitchell, D.J. (2020). Hierarchical Representation of Multistep Tasks in Multiple-Demand and Default Mode Networks. *J. Neurosci.* 40, 7724–7738. <https://doi.org/10.1523/JNEUROSCI.0594-20.2020>.
31. De La Vega, A., Chang, L.J., Banich, M.T., Wager, T.D., and Yarkoni, T. (2016). Large-scale meta-analysis of human medial frontal cortex reveals tripartite functional organization. *J. Neurosci.* 36, 6553–6562. <https://doi.org/10.1523/JNEUROSCI.4402-15.2016>.
32. Stephens, G.J., Silbert, L.J., and Hasson, U. (2010). Speaker-listener neural coupling underlies successful communication. *Proc. Natl. Acad. Sci. USA* 107, 14425–14430. <https://doi.org/10.1073/pnas.1008662107>.
33. Madsen, J., and Parra, L.C. (2022). Cognitive processing of a common stimulus synchronizes brains, hearts, and eyes. *PNAS Nexus* 1, pgac020. <https://doi.org/10.1093/pnasnexus/pgac020>.
34. Hayes, A.F. (2018). Partial, conditional, and moderated moderated mediation: Quantification, inference, and interpretation. *Commun. Monogr.* 85, 4–40. <https://doi.org/10.1080/03637751.2017.1352100>.
35. Allison, T., Puce, A., and McCarthy, G. (2000). Social perception from visual cues: role of the STS region. *Trends Cogn. Sci.* 4, 267–278. [https://doi.org/10.1016/S1364-6613\(00\)01501-1](https://doi.org/10.1016/S1364-6613(00)01501-1).
36. Pearson, J.M., Heilbronner, S.R., Barack, D.L., Hayden, B.Y., and Platt, M.L. (2011). Posterior cingulate cortex: adapting behavior to a changing

- world. *Trends Cogn. Sci.* 15, 143–151. <https://doi.org/10.1016/j.tics.2011.02.002>.
37. Silbert, L.J., Honey, C.J., Simony, E., Poeppel, D., and Hasson, U. (2014). Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proc. Natl. Acad. Sci. USA* 111, E4687–E4696. <https://doi.org/10.1073/pnas.1323812111>.
38. Simony, E., Honey, C.J., Chen, J., Lositsky, O., Yeshurun, Y., Wiesel, A., and Hasson, U. (2016). Dynamic reconfiguration of the default mode network during narrative comprehension. *Nat. Commun.* 7, 12141. <https://doi.org/10.1038/ncomms12141>.
39. Finn, E.S., Glerean, E., Khojandi, A.Y., Nielson, D., Molfese, P.J., Handwerker, D.A., and Bandettini, P.A. (2020). Idiosyncrony: From shared responses to individual differences during naturalistic neuroimaging. *Neuroimage* 215, 116828. <https://doi.org/10.1016/j.neuroimage.2020.116828>.
40. Bigler, E.D., Mortensen, S., Neeley, E.S., Ozonoff, S., Krasny, L., Johnson, M., Lu, J., Provencal, S.L., McMahon, W., and Lainhart, J.E. (2007). Superior temporal gyrus, language function, and autism. *Dev. Neuropsychol.* 31, 217–238. <https://doi.org/10.1080/87565640701190841>.
41. Correia, J., Formisano, E., Valente, G., Hausfeld, L., Jansma, B., and Bonte, M. (2014). Brain-based translation: fMRI decoding of spoken words in bilinguals reveals language-independent semantic representations in anterior temporal lobe. *J. Neurosci.* 34, 332–338. <https://doi.org/10.1523/JNEUROSCI.1302-13.2014>.
42. Pobric, G., Lambon Ralph, M.A., and Zahn, R. (2016). Hemispheric Specialization within the Superior Anterior Temporal Cortex for Social and Nonsocial Concepts. *J. Cogn. Neurosci.* 28, 351–360. https://doi.org/10.1162/jocn_a_00902.
43. De Felice, S., Di Cio, F., Tompkins, D., Hakim, U., Pinti, P., Vigliocco, G., and Hamilton, A.F. de C. (2025). Learning with others: teacher–learner brain synchrony depends on mutual gaze and joint attention. *Cereb. Cortex* 35, bhaf323. <https://doi.org/10.1093/cercor/bhaf323>.
44. Mayer, R.E., and Fiorella, L. (2022). *The Cambridge Handbook of Multimedia Learning, Third Edition* (Cambridge University Press).
45. Tomasello, Michael, and ebrary, Inc. (2014). *A Natural History of Human Thinking* (Harvard University Press).
46. Wolf, W., and Tomasello, M. (2025). A Shared Intentionality Account of Uniquely Human Social Bonding. *Perspect. Psychol. Sci.* 20, 264–275. <https://doi.org/10.1177/17456916231201795>.
47. Song, H., Kim, J., and Luo, W. (2016). Teacher–student relationship in on-line classes: A role of teacher self-disclosure. *Comput. Hum. Behav.* 54, 436–443. <https://doi.org/10.1016/j.chb.2015.07.037>.
48. Federici, R.A., and Skaalvik, E.M. (2013). Students’ perceptions of emotional and instrumental teacher support: Relations with motivational and emotional responses. *Int. Educ. Stud.* 7, 21–36. <https://doi.org/10.5539/ies.v7n1p21>.
49. Yin, J., Xu, H., Pan, Y., and Hu, Y. (2025). Effects of different AI-driven Chatbot feedback on learning outcomes and brain activity. *NPJ Sci. Learn.* 10, 17. <https://doi.org/10.1038/s41539-025-00311-8>.
50. Schurz, M., Radua, J., Aichhorn, M., Richlan, F., and Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neurosci. Biobehav. Rev.* 42, 9–34. <https://doi.org/10.1016/j.neubiorev.2014.01.009>.
51. Cavanna, A.E., and Trimble, M.R. (2006). The precuneus: a review of its functional anatomy and behavioural correlates. *Brain* 129, 564–583. <https://doi.org/10.1093/brain/awl004>.
52. Rasenberger, M., Özyürek, A., and Dingemanse, M. (2020). Alignment in Multimodal Interaction: An Integrative Framework. *Cogn. Sci.* 44, e12911. <https://doi.org/10.1111/cogs.12911>.
53. Chen, J., Leong, Y.C., Honey, C.J., Yong, C.H., Norman, K.A., and Hasson, U. (2017). Shared memories reveal shared structure in neural activity across individuals. *Nat. Neurosci.* 20, 115–125. <https://doi.org/10.1038/nn.4450>.
54. Zadbood, A., Chen, J., Leong, Y.C., Norman, K.A., and Hasson, U. (2017). How we transmit memories to other brains: constructing shared neural representations via communication. *Cereb. Cortex* 27, 4988–5000. <https://doi.org/10.1093/cercor/bhx202>.
55. Zada, Z., Goldstein, A., Michelmann, S., Simony, E., Price, A., Hasenfratz, L., Barham, E., Zadbood, A., Doyle, W., Friedman, D., et al. (2024). A shared model-based linguistic space for transmitting our thoughts from brain to brain in natural conversations. *Neuron* 112, 3211–3222.e5. <https://doi.org/10.1016/j.neuron.2024.06.025>.
56. Zada, Z., Nastase, S.A., Speer, S., Mwilambwe-Tshilobo, L., Tsoi, L., Burns, S.M., Falk, E., Hasson, U., and Tamir, D.I. (2025). Linguistic coupling between neural systems for speech production and comprehension during real-time dyadic conversations. *Neuron* 114, 774–787.e5. <https://doi.org/10.1016/j.neuron.2025.11.004>.
57. Hasson, U., Ghazanfar, A.A., Galantucci, B., Garrod, S., and Keysers, C. (2012). Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Trends Cogn. Sci.* 16, 114–121. <https://doi.org/10.1016/j.tics.2011.12.007>.
58. Hamilton, A.F. de C. (2021). Hyperscanning: Beyond the Hype. *Neuron* 109, 404–407. <https://doi.org/10.1016/j.neuron.2020.11.008>.
59. Holroyd, C.B. (2022). Interbrain synchrony: on wavy ground. *Trends Neurosci.* 45, 346–357. <https://doi.org/10.1016/j.tins.2022.02.002>.
60. Moreau, Q., and Dumas, G. (2021). Beyond Correlation versus Causation: Multi-brain Neuroscience Needs Explanation. *Trends Cogn. Sci.* 25, 542–543. <https://doi.org/10.1016/j.tics.2021.02.011>.
61. Novembre, G., and Iannetti, G.D. (2021). Hyperscanning Alone Cannot Prove Causality. *Multibrain Stimulation Can. Trends Cogn. Sci.* 25, 96–99. <https://doi.org/10.1016/j.tics.2020.11.003>.
62. Jiang, J., Zheng, L., and Lu, C. (2021). A hierarchical model for interpersonal verbal communication. *Soc. Cogn. Affect. Neurosci.* 16, 246–255. <https://doi.org/10.1093/scan/nsaa151>.
63. Feng, X., and Lu, C. (2026). Interpersonal Neural Synchronization (INS) in Educational Settings: Understanding the Role of INS in Effective Teaching. In *Handbook of Asian Educational Innovation towards the Futures of Education*, J. Chi-Kin Lee, ed. (Springer Nature Singapore), pp. 1–20. https://doi.org/10.1007/978-981-96-3901-4_50-1.
64. Lu, C. (2025). From lab to classroom: the rise of interpersonal neural synchronization in educational neuroscience. *AI Brain Child* 1, 18. <https://doi.org/10.1007/s44436-025-00019-x>.
65. Li, Q., Wang, D., Xiao, W., Tang, Y., Sun, Q., Sun, B., and Hu, Z. (2024). Structured interaction between teacher and student in the flipped classroom enhances learning and interbrain synchrony. *NPJ Sci. Learn.* 9, 73. <https://doi.org/10.1038/s41539-024-00286-y>.
66. Bransford, J.D., Brown, A.L., and Cocking, R.R. (1999). *How People Learn: Brain, Mind, Experience, and School* (National Academies Press). <https://doi.org/10.17226/6160>.
67. Strachan, J.W.A., Albergo, D., Borghini, G., Pansardi, O., Scaliti, E., Gupta, S., Saxena, K., Rufo, A., Panzeri, S., Manzi, G., et al. (2024). Testing theory of mind in large language models and humans. *Nat. Hum. Behav.* 8, 1285–1295. <https://doi.org/10.1038/s41562-024-01882-z>.
68. Sharma, A., Lin, I.W., Miner, A.S., Atkins, D.C., and Althoff, T. (2023). Human–AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support. *Nat. Mach. Intell.* 5, 46–57. <https://doi.org/10.1038/s42256-022-00593-2>.
69. Lee, Y.K., Suh, J., Zhan, H., Li, J.J., and Ong, D.C. (2024). Large Language Models Produce Responses Perceived to be Empathic. In *12th International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 63–71. <https://doi.org/10.1109/ACII61334.2024.00012>.
70. Ovsyannikova, D., de Mello, V.O., and Inzlicht, M. (2025). Third-party evaluators perceive AI as more compassionate than expert humans. *Commun. Psychol.* 3, 4. <https://doi.org/10.1038/s44271-024-00182-6>.

71. Reeves, B., and Nass, C. (1996). *The Media Equation: How People Treat Computers, Television, and New Media Like Real People* (Cambridge University Press).
72. Jastrzab, L.E., Chaudhury, B., Ashley, S.A., Koldewyn, K., and Cross, E.S. (2024). Beyond human-likeness: Socialness is more influential when attributing mental states to robots. *iScience* 27, 110070. <https://doi.org/10.1016/j.isci.2024.110070>.
73. Garcia, M.B. (2025). Teachers in the metaverse: the influence of avatar appearance and behavioral realism on perceptions of instructor credibility and teaching effectiveness. *Interact. Learn. Environ.* 33, 4334–4350. <https://doi.org/10.1080/10494820.2025.2462144>.
74. Jin, Y., and Sercu, L. (2025). ChatGPT Interventions in Higher Education: A Systematic Review of Experimental Studies. *J. Comput. Assist. Learn.* 41, e70072. <https://doi.org/10.1111/jcal.70072>.
75. Wang, J., and Fan, W. (2025). The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: insights from a meta-analysis. *Humanit. Soc. Sci. Commun.* 12, 621. <https://doi.org/10.1057/s41599-025-04787-y>.
76. Cheng, M., Lee, A.Y., Rapuano, K., Niederhoffer, K., Liebscher, A., and Hancock, J. (2026). Metaphors of AI indicate that people increasingly perceive AI as warm and human-like. *Commun. Psychol.* 4, 8. <https://doi.org/10.1038/s44271-025-00376-6>.
77. Yin, Y., Jia, N., and Waksak, C.J. (2024). AI can help people feel heard, but an AI label diminishes this impact. *Proc. Natl. Acad. Sci. USA* 121, e2319112121. <https://doi.org/10.1073/pnas.2319112121>.
78. Rubin, M., Li, J.Z., Zimmerman, F., Ong, D.C., Goldenberg, A., and Perry, A. (2025). Comparing the value of perceived human versus AI-generated empathy. *Nat. Hum. Behav.* 9, 2345–2359. <https://doi.org/10.1038/s41562-025-02247-w>.
79. Esteban, O., Markiewicz, C.J., Blair, R.W., Moodie, C.A., Isik, A.I., Erramuzpe, A., Kent, J.D., Goncalves, M., DuPre, E., Snyder, M., et al. (2019). fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat. Methods* 16, 111–116. <https://doi.org/10.1038/s41592-018-0235-4>.
80. Gorgolewski, K., Burns, C.D., Madison, C., Clark, D., Halchenko, Y.O., Waskom, M.L., and Ghosh, S.S. (2011). Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python. *Front. Neuroinform.* 5, 13. <https://doi.org/10.3389/fninf.2011.00013>.
81. Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., et al. (2020). Array programming with NumPy. *Nature* 585, 357–362. <https://doi.org/10.1038/s41586-020-2649-2>.
82. McKinney, W. (2010). Data Structures for Statistical Computing in Python. *SciPy* 2010. <https://doi.org/10.25080/Majora-92bf1922-00a>.
83. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* 17, 261–272. <https://doi.org/10.1038/s41592-019-0686-2>.
84. Hunter, J.D. (2007). Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* 9, 90–95. <https://doi.org/10.1109/MCSE.2007.55>.
85. Waskom, M. (2021). seaborn: statistical data visualization. *J. Open Source Softw.* 6, 3021. <https://doi.org/10.21105/joss.03021>.
86. Chang, L., Jolly, E., Cheong, J.H., Burnashev, A., and Chen, A. (2018). cosanlab/nitools: 0.3.11. Preprint at Zenodo. <https://doi.org/10.5281/zenodo.2229813>.
87. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830. <https://doi.org/10.5555/1953048.2078195>.
88. Vallat, R. (2018). Pingouin: statistics in Python. *J. Open Source Softw.* 3, 1026. <https://doi.org/10.21105/joss.01026>.
89. Seabold S., Perktold J. Statsmodels: Econometric and statistical modeling with python. Proceedings of the 9th Python in Science Conference. 2010. <https://doi.org/10.25080/Majora-92bf1922-011>
90. Preprint at Chamma, A., Frau-Pascual, A., Rothberg, A., Abadie, A., Abraham, A., Gramfort, A., Savio, A., Cionca, A., Sayal, A., Thual, A., et al. (2026). Nilearn Zenodo. <https://doi.org/10.5281/zenodo.18600337>.
91. Brett, M., Markiewicz, C.J., Hanke, M., Côté, M.-A., Cipollini, B., Papadopoulos Orfanos, D., McCarthy, P., Jarecka, D., Cheng, C.P., Larson, E., et al. (2025). nipy/nibabel: 5.3.3. Preprint at Zenodo. <https://doi.org/10.5281/zenodo.17833605>.
92. Pajula, J., and Tohka, J. (2016). How Many Is Enough? Effect of Sample Size in Inter-Subject Correlation Analysis of fMRI. *Comput. Intell. Neurosci.* 2016, 2094601. <https://doi.org/10.1155/2016/2094601>.
93. Kemmerer, D. (2022). *Cognitive Neuroscience of Language* (Routledge). <https://doi.org/10.4324/9781138318427>.
94. Peirce, J., Gray, J.R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., and Lindeløv, J.K. (2019). PsychoPy2: Experiments in behavior made easy. *Behav. Res. Methods* 51, 195–203. <https://doi.org/10.3758/s13428-018-01193-y>.
95. Hsu, C.T., Clariana, R., Schloss, B., and Li, P. (2019). Neurocognitive Signatures of Naturalistic Reading of Scientific Texts: A Fixation-Related fMRI Study. *Sci. Rep.* 9, 10678. <https://doi.org/10.1038/s41598-019-47176-7>.
96. Richlan, F., Gagl, B., Hawelka, S., Braun, M., Schurz, M., Kronbichler, M., and Hutzler, F. (2014). Fixation-related fMRI analysis in the domain of reading research: using self-paced eye movements as markers for hemodynamic brain responses during visual letter string processing. *Cereb. Cortex* 24, 2647–2656. <https://doi.org/10.1093/cercor/bht117>.
97. Psychology Software Tools, Inc. (2016) E-Prime 3.0. Preprint.
98. Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., and Gee, J.C. (2010). N4ITK: Improved N3 Bias Correction. *IEEE Trans. Med. Imaging* 29, 1310–1320. <https://doi.org/10.1109/TMI.2010.2046908>.
99. Avants, B.B., Epstein, C.L., Grossman, M., and Gee, J.C. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.* 12, 26–41. <https://doi.org/10.1016/j.media.2007.06.004>.
100. Zhang, Y., Brady, M., and Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans. Med. Imaging* 20, 45–57. <https://doi.org/10.1109/42.906424>.
101. Fonov, V.S., Evans, A.C., McKinstry, R.C., Almlí, C.R., and Collins, D.L. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *Neuroimage* 47, S102. [https://doi.org/10.1016/S1053-8119\(09\)70884-5](https://doi.org/10.1016/S1053-8119(09)70884-5).
102. Ciric, R., Thompson, W.H., Lorenz, R., Goncalves, M., MacNicol, E.E., Markiewicz, C.J., Halchenko, Y.O., Ghosh, S.S., Gorgolewski, K.J., Poldrack, R.A., et al. (2022). TemplateFlow: FAIR-sharing of multi-scale, multi-species brain models. *Nat. Methods* 19, 1568–1571. <https://doi.org/10.1038/s41592-022-01681-2>.
103. Jenkinson, M., Bannister, P., Brady, M., and Smith, S. (2002). Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *Neuroimage* 17, 825–841. [https://doi.org/10.1016/S1053-8119\(02\)91132-8](https://doi.org/10.1016/S1053-8119(02)91132-8).
104. Greve, D.N., and Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *Neuroimage* 48, 63–72. <https://doi.org/10.1016/j.neuroimage.2009.06.060>.
105. Power, J.D., Mitra, A., Laumann, T.O., Snyder, A.Z., Schlaggar, B.L., and Petersen, S.E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *Neuroimage* 84, 320–341. <https://doi.org/10.1016/j.neuroimage.2013.08.048>.
106. Madsen, J., Júlio, S.U., Gucik, P.J., Steinberg, R., and Parra, L.C. (2021). Synchronized eye movements predict test scores in online video

- education. *Proc. Natl. Acad. Sci. USA* 118, e2016980118. <https://doi.org/10.1073/pnas.2016980118>.
107. Kriegeskorte, N., Mur, M., and Bandettini, P. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* 2, 4. <https://doi.org/10.3389/neuro.06.004.2008>.
108. Finn, E.S., Corlett, P.R., Chen, G., Bandettini, P.A., and Constable, R.T. (2018). Trait paranoia shapes inter-subject synchrony in brain activity during an ambiguous social narrative. *Nat. Commun.* 9, 2043. <https://doi.org/10.1038/s41467-018-04387-2>.
109. Chen, P.A., Jolly, E., Cheong, J.H., and Chang, L.J. (2020). Intersubject representational similarity analysis reveals individual variations in affective experience when watching erotic movies. *Neuroimage* 216, 116851. <https://doi.org/10.1016/j.neuroimage.2020.116851>.
110. Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Res.* 27, 209–220.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
fMRI and behavioral data	This paper	OSF: https://doi.org/10.17605/OSF.IO/FU5H6 ; DOI: 10.17605/OSF.IO/FU5H6
Original code	This paper	OSF: https://doi.org/10.17605/OSF.IO/FU5H6 ; DOI: 10.17605/OSF.IO/FU5H6
Software and algorithms		
fMRIPrep 23.2.3	Esteban et al. ⁷⁹	10.1038/s41592-018-0235-4 ; RRID: SCR_016216
Nipype 1.8.6	Gorgolewski et al. ⁸⁰	10.3389/fninf.2011.00013 ; RRID: SCR_002502
Numpy 1.26.4	Harris et al. ⁸¹	10.1038/s41592-020-2649-2 ; RRID: SCR_008633
Pandas 2.2.3	McKinney ⁸²	10.25080/Majora-92bf1922-00a ; RRID: SCR_018214
SciPy 1.13.1	Virtanen et al. ⁸³	10.1038/s41592-019-0686-2 ; RRID: SCR_008058
matplotlib 3.9.2	Hunter ⁸⁴	10.1109/MCSE.2007.55 ; RRID: SCR_008624
seaborn 0.13.2	Waskom ⁸⁵	10.21105/joss.03021 ; RRID: SCR_018132
nltools 0.5.1	Chang et al. ⁸⁶	10.5281/zenodo.2229813
scikit-learn 1.5.2	Pedregosa et al. ⁸⁷	10.5555/1953048.2078195 ; RRID: SCR_002577
pingouin 0.5.5	Vallat ⁸⁸	10.21105/joss.01026 ; RRID: SCR_022261
statsmodels 0.14.4	Seabold et al. ⁸⁹	10.25080/Majora-92bf1922-011 ; RRID: SCR_016074
nilearn 0.10.4	Chamma et al. ⁹⁰	10.5281/zenodo.18600337 ; RRID: SCR_001362
nibabel 5.3.2	Brett et al. ⁹¹	10.5281/zenodo.17833605 ; RRID: SCR_002498

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Participants

One instructor (author Y.P.) was scanned using fMRI while giving a verbal lecture with accompanying PowerPoint slides. The author served as the instructor to facilitate data collection and ensure full control over the experiment as one condition required one-on-one pre-lecture social interaction with each student. The instructor had three consecutive years of college-level teaching or teaching assistant experience and consistently received high student evaluations for her teaching quality.

Seventy college students from the Hong Kong Polytechnic University participated in this study. All participants were right-handed and native speakers of Mandarin Chinese. They all reported having normal hearing and normal or corrected-to-normal vision. Participants had no history of mental or neurological disorders. The study was approved by the Institutional Review Board of The Hong Kong Polytechnic University. All participants provided written informed consent before taking part in the experiment and received monetary compensation for their participation.

Data from thirteen participants were excluded from the statistical analyses reported below due to nausea (2), excessive sleepiness during the video lecture (3), excessive head motion (head movements > 3 mm; 3), eye tracker malfunctioning during scanning (3), technical issues during AI instructor interaction (1), and failure to complete human instructor interaction (1). The final sample reported below included 57 participants (33 women and 24 men; mean age 22.5 (SD = 2.18), range 18 to 27 years of age). Participants reported their race or ethnicity as Asian = 57. Sex and gender were not considered in any analyses because we did not expect any differences in gaze alignment, neural alignment or learning based on these variables.

Our final sample included 57 participants (No interaction: $n = 20$; Human interaction: $n = 19$; AI interaction: $n = 18$). The choice of the sample size in our study was based on consideration of other naturalistic neuroimaging studies that used similar designs and ISC analyses as ours (e.g., $n = 17^{53}$; $n = 18^{18}$); effect sizes for ISC-based designs are difficult to estimate from traditional task-based fMRI studies and there is no established framework for a priori power estimation for multi-level ISC analyses. Empirical validation studies also demonstrate that ISC statistics with ~ 20 participants show high convergence ($r = 0.92$) with estimates of large-sample participants of up to 130.⁹²

METHOD DETAILS

Design

This study adopted a between-subjects design. Participants were randomly assigned to three experimental conditions: no social interaction condition (No interaction group; $n = 20$), human instructor social interaction condition (Human interaction group; $n = 19$) and AI instructor social interaction condition (AI interaction group; $n = 18$). A summary of the demographic information of the three groups of participants can be found in [Table S1](#).

Social interaction session

In this study, both the human and AI instructors followed a structured pre-lecture prompt (see [Table S2](#) for full interaction template). The interaction began with the instructor's self-introduction highlighting her teaching experience in cognitive neuroscience and a warm invitation for the student to share their name, major, and year of study. After the student introduced themselves, the teacher responded using the student's name directly, a subtle but important move to foster interpersonal connection and social presence.⁴⁹ The instructor next asked how the student's field of study might relate to brain or cognitive science, followed by a discussion of the student's hobbies or leisure activities and whether they might connect to neuroscience. Each response from the student was met with an affirming and curiosity-sparking elaboration. For instance, if a student expressed an interest in playing badminton and noted that it benefited both the body and the brain, the instructor would first affirm the accuracy of this observation with an encouraging tone. The instructor might then explain that physical activities activate the motor cortex and cerebellum, regions involved in coordination, precision, and motor planning. Regular engagement in physical activity not only strengthens these motor-related areas but also enhances attention, reaction speed, and executive functions, while promoting mental well-being through the release of dopamine and endorphins. The structured interaction concluded with an encouraging preview of the upcoming lesson, highlighting the relevance of neuroscience to daily life and motivating students to approach the content with curiosity and engagement.

The human instructor was introduced by the experimenters (authors Y.H. and Y.L.) with the following statement: "Before you start video learning in the scanner, the lecturer would like to have a chat with you." Following this, the human instructor conducted the pre-lecture social interaction with the student in a preparation room adjacent to the MRI scanner. To maintain a clearly defined instructional role and avoid conflating the instructor with the experimental facilitators, the remainder of the experiment, including all procedural instructions, technical operations, and participant management, was handled solely by the experimenters. This arrangement ensured that the human instructor was perceived exclusively as the lecturer within the context of the learning task.

The AI instructor was introduced by the same experimenters with the following statement: "Before you view the video lecture delivered by an AI instructor in the scanner, we would like you to have a chat with the AI instructor." Following this introduction, the AI instructor–student interaction commenced with the assistance of the experimenter. The AI instructor appeared on a computer screen and followed a scripted prompt to complete the interaction session, resembling the human instructor. To maintain continuity with the human instructor condition, the AI instructor featured a customized avatar based on the human instructor's voice and appearance, including matching outfit, hairstyle, and facial features. The technical implementation of the AI instructor was elaborated in [Table S3](#) and [Figure S1](#). The in-scanner video learning session commenced immediately after the interaction with the human/AI instructors to ensure the efficiency and continuity of the procedure.

In the No interaction condition, participants proceeded directly to the video lecture without any pre-lecture interaction activity with an instructor. Critically, however, all participants, regardless of condition, first completed an identical preparation phase that included a 7-item self-rating questionnaire assessing prior knowledge and a 2-minute practice video with 2 comprehension questions to familiarize participants with the in-scanner learning procedures. This ensured equivalent task familiarization and baseline orientation across all groups. Following the preparation, experimenters provided only necessary procedural instructions, informing participants: "You will now watch a video lecture in the scanner." All participants then watched the same 14-minute video lecture, ensuring identical learning duration, content exposure, and instructor presentation across conditions. The absence of pre-lecture interaction resulted in a shorter total session duration (8–10 minutes less than the Human and AI interaction groups). However, we interpret this duration difference to weigh against a confounding factor: namely, if session length were driving results, the No interaction group should show better learning due to reduced fatigue. Further, this condition also mirrors ecologically valid online learning contexts (e.g., MOOCs and YouTube) where learners typically access content directly without preceding social activities, while duration-matched controls would introduce their own confounds; e.g., artificial performance depression from filler activities or priming effects from any social interaction (see De Felice et al.²⁶ where irrelevant daily conversations may also induce enhanced neural synchrony measured by fNIRS during subsequent movie co-watching).

Stimuli

The instructor developed a 14.25-minute Lesson and a 2.5-minute Review (the 2.5-minute review was excluded from the main analysis and reasons for exclusion were elaborated in SI Figure S2), both accompanied by PowerPoint slides (see Figure 1A). The instructional material was extracted and adapted from a college-level course on Neurolinguistics taught at The Hong Kong Polytechnic University. The slides were based on the first chapter of the textbook *Cognitive Neuroscience of Language*⁹³ and covered foundational topics of cognitive neuroscience, including the basic structure of a neuron, information transmission between neurons, and hierarchical coding by neurons.

After the instructional slides were finalized, the instructor was scanned using fMRI while delivering the lecture. The instructor repeated the same lecture three times across two separate scanning sessions. During the recordings, slides were presented using PsychoPy2⁹⁴ and displayed via a projector onto a rear-projection screen mounted at the back of the scanner bore. The instructor viewed the slides through a mirror attached to the head coil and advanced them manually using a button box. Speech was captured using an MRI-compatible microphone with real-time active noise cancellation (FOMRI III; Optoacoustics Ltd), allowing for high-quality audio recording in the scanner environment.

Following the instructor scans, the audio recording with the highest sound and presentation⁸⁹ quality was selected and further denoised using Adobe Audition. The cleaned audio was then transcribed, timestamped, and aligned with the corresponding slide presentation. To simulate the instructor's social presence, consistent with standard practices on mainstream video learning platforms, the instructor later lip-synced to the denoised audio recording for the full duration of the lecture. The recorded talking-head video was subsequently integrated into the lecture slides to create a cohesive video lecture. To validate the naturalness of the lip syncing, eleven independent participants watched the full lecture video and were asked whether they noticed any unnatural synchronization; none of the participants reported detecting that the audio and visual speech were not originally recorded together.

Behavioral Measurements

Following multimedia learning studies,¹⁵ seven self-rating questions were used as a pre-test to assess whether all learners had comparable levels of prior knowledge in the subject domain prior to the video lecture. A one-way ANOVA showed no significant group differences in self-rated prior knowledge ($F(2, 54) = 2.27$, $p = 0.112$, $\eta^2 = 0.078$, 95% CI [0.008, 0.277]). After the learning phase, participants completed three standardized post-tests designed to measure comprehension (in-scanner; pre-defined as easy and hard comprehension questions), recall, and transfer (outside of the scanner; see specific questions in Table S4). Additionally, a learning experience questionnaire was administered to assess participants' perception of the instructor, the instructor's delivery quality and their own effort levels.

Procedure

The experiment involved two separate visits to the laboratory. During the first visit, participants signed the consent form and completed a background questionnaire to collect demographic information.

During the second visit, participants first completed the prior knowledge test, followed by a practice session outside the MRI scanner. For participants assigned to the Human interaction group and the AI interaction group, a social interaction session with the instructor (Human or AI) was conducted as described above, typically lasting 8–10 minutes (duration varied slightly depending on participants' responses and follow-up questions). Participants in the No interaction group did not undergo the social interaction session. Afterward, all participants proceeded with the MRI scanning for the video learning session.

Inside the MRI scanner, participants underwent a structural MRI scan and a functional MRI scan. During the main multimedia learning session, simultaneous fMRI and eye-tracking data were collected using the fixation-related fMRI scanning paradigm.^{95,96} This session included a preparation phase for the eye-tracking, a structural run and a functional run during which participants watched a lecture video (lecture duration = 14.25 minutes) and answered comprehension questions (30 multiple-choice questions, with a maximum of 30 seconds to respond to each). Students responded to the comprehension questions using an MRI-compatible button box. The video lecture and comprehension test were presented using E-Prime 3.0.⁹⁷ The total in-scanner session lasted approximately 40 minutes.

After exiting the scanner, participants completed the post-learning recall and transfer tests, as well as the learning experience questionnaire (see Table S4 for example questions).

Data Acquisition

Eye-movement data

In parallel with fMRI acquisition, both the instructor's eye movements during in-scanner lecture delivery and the participants' eye movements during in-scanner video lecture learning were recorded. Eye-tracking was implemented using a long-range mount, MRI-compatible eye-tracker (EyeLink 1000 Plus, SR Research), integrated with PsychoPy2 (for the instructor) and E-prime3 (for the students) for synchronized data collection. The camera was positioned at the rear of the scanner bore and captured eye movements via a reflective mirror mounted above the head coil. The distance from the eye-tracker camera to the reflected eye position was approximately 120 cm. Monocular recording was performed on the right eye. A standard 13-point calibration and validation procedure was conducted for the right eye, followed by a drift check to ensure data quality.

MRI data

MRI data was acquired using a 3T Siemens Magnetom Prisma Fit scanner with a 64-channel head coil. Functional data were acquired first to minimize the delay from the social interaction session to the video learning phase and to maintain participants' alertness. Because initial calibration and validation for eye-tracking could prolong the scanning time, this order reduced the risk of fatigue and attentional decline.

After an initial localizing scan, one run of functional images were acquired using a multiband T2*-weighted echo-planar imaging (EPI) sequence (34 interleaved axial slices, TR = 1 s, TE = 30 ms, slice thickness = 3.5 mm, FOV = 240 × 240 mm², matrix size = 80 × 80, in-plane resolution = 3 × 3 mm², inter-slice gap = 10% of slice thickness, flip angle = 58°, multiband acceleration factor = 2, acquisition time depending on self-paced question answering speed ranging from 1239 s to 1499 s). The functional sequence covered the whole brain except for the lower extent of the cerebellum. Immediately afterward, an anatomical T1-weighted MPRAGE image was acquired (176 ascending sagittal slices, FOV = 256 mm, TR = 1570 ms, TE = 2.15 ms, flip angle = 9°, GRAPPA in-plane acceleration factor = 2, acquisition time = 220 s).

fMRI data preprocessing

Anatomical and functional MRI data were preprocessed using fMRIPrep 23.2.3⁷⁹ based on Nipype 1.8.6⁸⁰ and Nilearn 0.10.2.⁹⁰

T1-weighted images were corrected for intensity non-uniformity (INU) with N4BiasFieldCorrection,⁹⁸ distributed with ANTs 2.5.0,⁹⁹ and used as a reference throughout the workflow. The T1w reference was then skull-stripped with a Nipype 1.8.6 implementation of the ants-BrainExtraction.sh workflow (from ANTs), using OASIS30ANTs as the target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white matter (WM), and gray matter (GM) was performed on the brain-extracted T1w image using FAST (FSL¹⁰⁰). Volume-based spatial normalization to MNI152 standard space (MNI152NLin2009cAsym) was performed through nonlinear registration with antsRegistration (ANTs 2.5.0⁹⁹), using the ICBM 152 Nonlinear Asymmetrical template version 2009c,¹⁰¹ accessed via TemplateFlow 23.1.0.¹⁰²

For each BOLD run per subject, the following preprocessing was performed using. First, a reference volume was generated using a custom fMRIPrep methodology for use in head-motion correction. Head-motion parameters with respect to the BOLD reference (transformation matrices and six rotation and translation parameters) were estimated prior to any spatiotemporal filtering using MCFLIRT (FSL¹⁰³). The BOLD reference was then co-registered to the T1w reference using boundary-based registration implemented in bbrregister (FreeSurfer¹⁰⁴), configured with six degrees of freedom. Several confound time series were calculated from the preprocessed BOLD signal, including framewise displacement (FD¹⁰⁵), DVARS, and region-wise global signals within CSF, WM, and whole-brain masks. Volumes exceeding an FD threshold of 0.5 mm or a standardized DVARS threshold of 1.5 were flagged as motion outliers. All spatial transformations including head-motion correction, co-registration to the anatomical reference, and normalization to MNI152 space were composed and applied in a single interpolation step using cubic B-spline interpolation via *ni-transforms*, minimizing interpolation-induced smoothing.

After the fMRIPrep preprocessing, spatial smoothing was applied to the functional data using a Gaussian kernel with a full-width at half maximum (FWHM) of 6 mm. We then performed voxelwise denoising using a general linear model (GLM), which included the following nuisance regressors: the six head motion parameters estimated during realignment, their first derivatives, the squares of each term (i.e., 24 motion parameters in total), dummy regressors for volumes identified as outliers based on global signal deviation and framewise displacement (temporal derivative), and average CSF signal. We did not apply a separate high-pass filter; instead, linear and quadratic trends were modeled explicitly in the GLM to account for slow drifts and physiological artifacts. These preprocessing steps were implemented uniformly for both the human instructor's and the students' MRI data. These further steps were performed mainly using niftools⁸⁶ on Python 3.10 (<http://www.python.org>).

Neural Time Series Extraction

We selected ROIs of three networks implicated in learning and comprehension (the default mode network; DMN^{18,20}), cognitive resource allocation (the multiple demand network; MD²⁹) and socio-emotional processing (Limbic ROIs²⁸). DMN includes the ventromedial prefrontal cortex (VMPFC), dorsomedial prefrontal cortex (DMPFC), temporo-parietal junction (TPJ), precuneus (PC), and posterior cingulate cortex functionally combined with posterior TPJ (pcc/pTPJ), the superior temporal sulcus extending into temporal poles (STS/TP), and hippocampus (DMN regions depicted in Figure 3B^{28,31}). MD covers the middle frontal gyrus, inferior frontal sulcus, anterior insula, intraparietal sulcus, supplementary motor area, and anterior cingulate cortex (see MD network regions in Figure 4A). Limbic ROIs comprise: the dorsal anterior cingulate cortex (dACC), anterior insula (AI), amygdala, and nucleus accumbens (NAc), consistent with Iyer et al.²⁸ (see limbic regions in Figure 3C).

For each participant, we extracted the parcel-wise averaged BOLD time series by computing the mean time series across all voxels within each labeled parcel within the DMN, limbic, and MD networks during the video learning. For the instructor, we conducted similar operations for the neural responses during the lecture delivery for the DMN. Following prior recommendations,¹⁶ we excluded the pre-video fixation period as well as the first 10 seconds of lecture delivery (for the instructor) and video watching (for the learners) to minimize transient, non-selective responses to stimulus onset. After these exclusions, a total of 845 functional volumes were retained for subsequent analyses. We used Python libraries Nilearn⁹⁰ and nibabel⁹¹ for neural time series extraction.

QUANTIFICATIONS AND STATISTICAL ANALYSIS

Behavioral data processing and analysis

Learning performance was assessed using comprehension, recall and transfer scores. Two independent groups of raters (each consisting of two raters) scored the recall and transfer tests separately, yielding inter-rater reliability of 0.88 (95% CI: [0.7682, 0.9300], $p < 0.001$) and 0.89 (95% CI: [0.7982, 0.9392], $p < 0.001$), respectively. SciPy⁸³ and Pingouin⁸⁸ libraries were used for behavioral data analysis.

Eye-tracking data processing and analysis

Eye-tracking data were analyzed primarily using EyeLink Data Viewer (Version 3.2.1; SR Research Ltd.) and Python 3.10 (<http://www.python.org>). Following recent eye intersubject correlation (ISC) studies,^{33,106} we computed student–student and student–instructor gaze ISC during video lecture learning. For both analyses, we first extracted moment-to-moment eye gaze positions along the vertical and horizontal axes from the instructor during lecture delivery and from the students during video watching.

For student–student eye ISC ($n = 57$), we applied a leave-one-out approach: for each participant, we computed the Pearson correlation between their gaze position and the average gaze position of all other participants separately for the vertical and horizontal directions. For student–instructor eye ISC (only for the No and Human interaction groups where the instructor’s eye gaze data was available; $n = 39$), we used a pairwise approach, computing the Pearson correlation between each student’s gaze position and the human instructor’s gaze position, again separately for vertical and horizontal axes. The vertical and horizontal ISC values were then averaged to derive a student–student and a student–instructor gaze ISC score for subsequent analyses.

Prior to ISC calculation, missing gaze data due to eye blinks were linearly interpolated.^{33,106} We then performed two ANCOVAs, each with the same covariate participants’ prior knowledge scores, to separately examine (1) the effect of social interaction (No interaction vs. Human interaction) and (2) the effect of instructor type (Human interaction vs. AI interaction). We mainly used python libraries Numpy,⁸¹ Pandas⁸² and SciPy⁸³ (e.g., interpolate and SciPy.stats) for gaze ISC computation and statistical analysis.

Student–student & student–instructor ISC

Neural alignment among subjects obtained via separate fMRI scanning were quantified using intersubject correlation (ISC^{16,17}). Two ISC analyses were conducted: (i) student–student ISC ($n = 57$), calculated using a leave-one-out approach where each student’s z-scored response time course was correlated using Pearson correlation to the average z-scored response time course of the remaining ($N - 1$) students at each ROI. (ii) student–instructor ISC (for the Human and No interaction groups only, available neural data was leveraged to examine student–instructor DMN neural alignment; $n = 39$), calculated using a pairwise approach where ISC was calculated by correlating each student’s z-scored response time series to the instructor’s z-scored response time series at each ROI. ISC values were computed separately for each ROI within the three networks of interest (the DMN, limbic network, and MD network) and Fisher’s z-transformation was applied to normalize the correlation coefficients. We also tested for a lag between student and instructor neural responses during learning in the DMN as suggested by prior literature^{37,54} by shifting the students’ time series by 1–6s³² relative to the instructor’s time series (instructor preceding the students) and calculating ISC at each lag.

We assessed the impact of social interaction (No interaction vs. Human interaction) and interaction instructor type (Human interaction vs. AI interaction) on neural alignment by evaluating group differences using a non-parametric permutation test.¹⁶ Specifically, we computed the observed difference in mean ISC between the No interaction and Human interaction groups as well as the Human interaction and AI interaction groups, respectively. Then, we generated a null distribution by randomly shuffling group labels 1,000 times and recalculating the difference in mean ISC for each permutation. The empirical two-tailed p-value was calculated as the proportion of permuted differences whose absolute value was greater than or equal to the absolute value of the observed difference. This approach provides a non-parametric estimate of the statistical significance of group differences in ISC without relying on distributional assumptions. We corrected for multiple comparisons across ROIs in each network by controlling the false discovery rate (FDR) at $q = 0.05$. We also computed 95% confidence intervals via non-parametric bootstrapping. ISC analysis was mainly conducted with libraries including nlttools,⁸⁶ Nilearn,⁹⁰ nibabel,⁹¹ Numpy,⁸¹ Pandas⁸² and SciPy.⁸³

Classification of group

To evaluate whether the Human interaction, AI interaction, and No interaction groups exhibited comparable neural distinctiveness, we conducted a multi-class classification analysis based on ISC features. We extracted ISC values from the DMN and MD network and used them as input features.

We implemented a machine learning pipeline consisting of three sequential steps: 1) feature scaling via z-scoring, 2) univariate feature selection using ANOVA F-tests, and 3) classification using a support vector machine (SVM) with a linear kernel (SVC with class balancing enabled). To maximize statistical power while minimizing overfitting given the modest sample size, model evaluation was performed using leave-one-subject-out cross-validation.

Classification performance was assessed based on overall binary accuracy, classification metrics (precision, recall, and F1 score), and the confusion matrix. The binary accuracy was defined by whether the predicted group label matched the actual group label for each participant. The theoretical chance-level accuracy for three groups was 0.33. To statistically assess whether the observed classification accuracy exceeded chance, we conducted a permutation test by randomly shuffling group labels and repeating the entire classification procedure 10,000 times to generate a null distribution of classification accuracies. The empirical p-value was calculated as the proportion of permutations in which the shuffled-label accuracy exceeded or equaled the observed accuracy. Here, python library scikit-learn⁸⁷ was used to implement SVM.

Correlational and mediation analysis

To examine how neural alignment and visual alignment jointly contribute to multimedia learning, we computed Spearman correlations between neural ISC, gaze ISC, and learning outcomes (see a similar approach in Gu et al.¹⁹). Statistical significance was assessed

using a one-sided permutation test,¹⁸ in which student behavioral measures or eye-movement ISC values were randomly shuffled 1,000 times while keeping the neural ISC scores fixed, generating a null distribution of correlation values.

To assess broader relationships across groups, we concatenated data from all three conditions and computed partial Spearman correlations between gaze ISC, neural ISC, and learning performance, controlling for group membership. Specifically, we regressed out group effects from each variable and then computed Spearman correlations between the residuals. Statistical significance was again assessed via a one-sided permutation test (1,000 iterations), where we randomly permuted the residual neural ISC values, which were held constant across all correlation tests, while keeping the residuals of other variables fixed.

In both sets of correlation analyses, empirical one-tailed p-values were defined as the proportion of permuted correlations greater than or equal to the observed correlation, following the method of Meshulam et al.¹⁸ Multiple comparisons across ROIs within each network were corrected using false discovery rate (FDR) correction at $q = 0.05$. In parallel, we computed 95% confidence intervals using non-parametric bootstrapping.

Furthermore, we conducted mediation analysis to test whether neural alignment served as a mediator between gaze alignment and overall learning performance or whether gaze alignment mediated the relationship between neural alignment and learning performance. In these models, both neural alignment and gaze alignment were tested as potential mediators. The analyses were conducted using nonparametric bootstrap resampling (5,000 iterations) with a significance level of $\alpha = 0.05$. These mediation models were theory-driven and focused on a small number of a priori hypothesized pathways (e.g., STS and PCC). Lastly, for the student-student level mediation analysis, we conducted the analysis with the full sample ($n = 57$, all three groups) and performed a robustness check restricted to Human and No interaction groups only ($n = 39$, matching the subsample for student-instructor analyses; results presented in Table S13). We primarily used Numpy,⁸¹ Pandas,⁸² SciPy,⁸³ Scikit-learn⁸⁷ and Pingouin⁸⁸ libraries for these analyses.

ISFC-RSA analysis

In the present study, we leveraged ROIs from DMN, implicated in comprehension and learning, and the limbic system, involved in socio-emotional processing, to examine whether functional interactions between these networks were associated with learning outcomes.

To assess the relationship between group-level socio-emotional engagement and individual cognitive processing, we computed intersubject functional connectivity (ISFC) between each learner's DMN ROI time series and the group-averaged time series of each limbic ROI (leave-one-out approach by excluding the current participant). ISFC is used here due to its higher signal-to-noise ratio compared to simple functional connectivity analysis (i.e., single out stimulus-driven functional connectivity by excluding individual idiosyncratic patterns). For each subject, we Pearson correlated the time series of each DMN ROI with the corresponding group-level limbic ROI time series separately, generating a vector of ISFC values across DMN-limbic ROI pairs.

Given the number of ROIs (7 from the DMN and 4 limbic regions), resulting in 28 ISFC values per participant, we operationalized the analysis from a holistic perspective by adopting the framework of intersubject representational similarity analysis (ISRSA³⁹). Akin to representational similarity analysis (RSA¹⁰⁷), ISRSA computes the correlation between two subject-by-subject similarity matrices: one derived from neural measures and the other from behavioral measures. Here, instead of constructing the neural matrix from pairwise ISC values, we adopted the ISFC values. Neural similarity matrices were constructed by reshaping each participant's ISFC values into a feature vector, computing pairwise correlation distances between participants, and converting these distances into similarity values ($1 - \text{correlation distance}$). The behavioral similarity matrices were constructed upon the in-scanner comprehension scores. Pairwise similarity between participants was computed such that participants with either higher or more similar scores were treated as more similar, whereas those with lower or more divergent scores were treated as less similar. The ISFC-RSA correlation thus quantifies the extent to which similarity in functional connectivity profiles, thus interaction across these two networks, predicts similarity in learning performance across participants.

Specifically, we constructed two types of pairwise similarity matrices to construct the pairwise learning performance matrices. This decision was made as we did not hold specific hypotheses regarding the interaction between these two networks and learning performance. There could be two possible scenarios: 1) those who exhibit high-performing individuals tend to exhibit more similar network interaction patterns, while lower-performing individuals may show more idiosyncratic patterns (see Finn et al.¹⁰⁸, and Iyer et al.²⁸ for examples); 2) subjects with similar learning outcomes, regardless of performance level, will exhibit similar neural interaction profiles (see Chen et al.¹⁰⁹ for an example).

To entertain the first scenario, we followed 'Anna Karenina' or AnnaK principle.³⁹ Specifically, we constructed pairwise learning performance similarity matrices where higher pairwise similarity emerges among subjects with higher learning performance scores. To do so, an absolute position on the behavioral scale for each subject was estimated by sorting learning performance where a lower rank indicates lower performance. The similarity (i.e., $S_{i,j}$) of each pair of subjects was computed using Equation (1), where P_i and P_j were the absolute positions of subjects i and j respectively and N was the number of subjects. Hence, $S_{i,j}$ calculated from the pairs of subjects with higher behavioral scores would be greater compared to $S_{i,j}$ obtained from the remaining pairs of subjects.

$$S_{i,j} = \frac{(P_i + P_j)}{2 * N} \quad (\text{Equation 1})$$

To accommodate the second situation, we also organized the pairwise similarity matrices in a way that higher pairwise similarity emerges from similar behavioral scores, which is named as the 'nearest-neighbor' or NN principle.³⁹ We first obtained the absolute

position on individual behavioral scale by sorting the learning outcomes measured by comprehension and transfer tasks. The similarity (i.e., $S_{ni,j}$) of each pair of subjects was computed using Equation 2, where P_i and P_j indicated the absolute positions of subjects i and j respectively and N was the number of subjects. Therefore, $S_{ni,j}$ estimated from the pairs of subjects with similar learning outcomes would be greater than $S_{ni,j}$ calculated from the pairs of subjects with less similar learning outcomes.

Note that during actual analysis, we followed the implementation details of Finn et al.³⁹ to calculate the difference between two subjects' ranks using Euclidean distance. These distances were then converted to similarity scores to allow matrix correlation with the ISFC similarity matrix.

$$S_{ni,j} = 1 - \left| \frac{(P_i - P_j)}{N} \right| \quad (\text{Equation 2})$$

Lastly, the statistical significance for the ISFC-learning representational similarity was assessed using a nonparametric Mantel test.¹¹⁰ Specifically, we randomly permuted participant labels of the ISFC similarity matrix (i.e., rows and columns) 10,000 times to generate a null distribution of surrogate Spearman's correlation coefficients. The observed Spearman's correlation coefficient was then compared to this null distribution to derive a p-value for each analysis. To implement ISFC-RSA analysis, we used Python libraries nltools,⁸⁶ nilearn,⁹⁰ nibabel,⁹¹ SciPy⁸³ and scikit-learn.⁸⁷

Graph abstract (Panel 1), Figures 1, 5, and 7 of this manuscript include several icon resources from Flaticon.com (<https://www.flaticon.com/>). These resources are available for both personal and commercial use under the Flaticon License, provided that appropriate attribution is included.

Here we make the attribution statement in each respective figure caption:

Icons from Flaticon.com are used for the Graphical Abstract, Figures 1, 5, and 7 under Free license.