nature human behaviour

Article

https://doi.org/10.1038/s41562-026-02540-2

Mentalizing and heuristics as distinct cognitive strategies in human teaching

Received: 19 March 2025

Accepted: 7 July 2026

Published online: 26 August 2026

Check for updates

Sevan K. Harootonian [1] ✉, Thomas L. Griffiths [1,2], Yael Niv [1,3,5] & Mark K. Ho [4,5]

Teaching is a foundational social behaviour that can result from mentally effortful reasoning or cognitively frugal heuristics. When do people use these strategies to teach? Here we investigated this using behavioural experiments and computational modelling with adult participants recruited via Prolific. Experiment 1 ($N = 100$) revealed robust individual differences: some participants taught by reasoning about a learner's knowledge, consistent with an optimal Bayesian pedagogy model, while others relied on simple heuristics that do not require mentalizing. In two preregistered follow-up experiments, we found that people persist in using heuristics even when they are no longer effective (experiment 2, $N = 253$, $P < 0.001$, rank-biserial $r = 0.287$, 95% confidence interval 0.149–0.419) but that this tendency is pre-empted when inference about a learner's knowledge is scaffolded using an auxiliary task (experiment 3, $N = 759$, $P < 0.001$, partial $\eta_p^2 = 0.107$, 95% confidence interval 0.068–0.148). These results demonstrate sophisticated arbitration between planning and heuristics during teaching and elucidate the more general mechanisms involved in adapting mental effort during social interactions.

Teaching and learning from teachers are familiar elements of everyday social life that underpin distinctively human forms of communication, cultural transmission and cooperation[1,2]. But people can teach in different ways. For example, people can be more or less thoughtful when teaching. Imagine a basketball coach helping a new team member score points (Fig. 1). If the coach is being more thoughtful, she might observe the player's incorrect hand positioning, infer that their knowledge of shooting is wrong, and suggest a better approach. By contrast, if the coach is being less thoughtful, she might provide more generic advice, such as 'Just try shooting closer to the basket'. Thoughtful, tailored teaching is generally more effective than generic feedback. However, thoughtfulness comes at a cost: observing a learner, inferring what they know and selecting an effective pedagogical intervention is cognitively effortful; giving generic feedback is easier. (This is not to say generic feedback is never useful: in professional basketball, centres are typically considered less skilled shooters but have high field goal percentages precisely because they follow the simple rule of shooting near the hoop[3].)

How are teaching strategies shaped by cognitive costs? The thoughtful coaching in our example exemplifies a strategy for teaching that hinges on theory of mind—the ability to reason about others' mental states such as their beliefs, desires and intentions. This kind of teaching requires that the teacher infer the learner's current knowledge state and anticipate how the learner will interpret and respond to the information provided[4–6]. In recent years, frameworks such as Bayesian Theory of Mind[7–10] and Rational Speech Act Theory[11–13] have been proposed as accounts of such mechanisms. These frameworks formalize teaching as expected utility maximization given a model of how a learner will update their beliefs in response to teaching actions[13]. Notably, teaching by mentalizing can be understood as a form of model-based decision-making[14] as it relies on inference and action evaluation using a causal model of another person's mind[15]. Like other forms of model-based decision-making, teaching by mentalizing is flexible and adaptable, allowing teachers to tailor teaching based on an estimate of a learner's knowledge. However, this comes at the price of high cognitive effort[16,17].

[1]Department of Psychology, Princeton University, Princeton, NJ, USA. [2]Department of Computer Science, Princeton University, Princeton, NJ, USA. [3]Princeton Neuroscience, Princeton University, Princeton, NJ, USA. [4]Department of Psychology, New York University, New York, NY, USA. [5]These authors jointly supervised this work: Yael Niv and Mark K. Ho. ✉e-mail: skh@princeton.edu

https://doi.org/10.1038/s41562-026-02540-2

Alternatively, teaching can be guided by heuristic decision-making, as in the less thoughtful coaching strategy in our example. Such an approach relies on adopting simple rules that have been effective previously and can be performed with minimal mental effort or adaptation to the specific situation. Although less emphasized in formal models of communication, heuristic strategies capture how teachers might default to well-practiced patterns of instruction without actively considering the learner's current knowledge state[18–21]. Heuristics offer a lower cognitive cost alternative by relying on shortcuts that require minimal mental effort, but at the cost of reduced flexibility. Like other forms of habitual, automatic or model-free behaviour, heuristics characteristically persist even in situations in which they are not the most effective approach, particularly in novel environments[22].

Mental-state inference has been studied in cognitive science for decades, from spontaneous attributions of moving shapes to controlled laboratory tasks in which participants reason about imagined agents given only a brief cover story[8,11,13,23–27]. Such tasks elicit genuine mentalizing: Vélez et al.[28] showed that teaching an imagined learner in an abstract task recruited the bilateral temporo-parietal junction and medial prefrontal cortex.

Our approach to investigating mentalizing versus heuristic teaching strategies builds on this previous work by combining computational theory with behavioural experiments. First, we begin by formally characterizing the relationship between these two broad strategies by explicitly connecting models of Bayesian teaching[8,27,29,30] with model-based/model-free decision-making[14,25,31]. We then report the results of three experiments that use a graph teaching task (Fig. 2) to study people's teaching behaviours at the level of individual decisions.

Experiment 1 reveals that people spontaneously adopt different teaching strategies, with some favouring heuristics and others engaging in mentalizing. In experiment 2, participants trained in environments where heuristics were effective persisted in using those heuristics even when they became suboptimal. In experiment 3, a scaffolding task that prompted teachers to consider what the learner may not know mitigated heuristic use and encouraged mentalizing, even in heuristic-congruent environments. Together, these results support the view that people adopt cognitive strategies to balance goals and cognitive resources[32].

Computational accounts of teaching strategies

In this section, we present our theoretical framework for characterizing mentalizing versus heuristic teaching. Our approach builds on formal theories from computational social cognition[8,9,25,33,34] and reinforcement learning[35–39].

Teaching by mentalizing

Teaching by mentalizing involves two steps: inferring the learner's knowledge from their behaviour and choosing a teaching intervention that best modifies it. We model teachers as reasoning about a learner who is assumed to act optimally given their knowledge of a task. The teacher and learner share knowledge about the reward structure (for example, that getting the ball in the hoop wins points), but the learner may be missing information about causal structure (for example, what actions cause the ball to end up in the hoop). The challenge is to infer the learner's knowledge from their behaviour and decide what information to provide.

Following previous work, we model teaching by mentalizing as probabilistic reasoning over a generative model of a learner's planning, followed by maximizing expected teaching utility[25,27]. We formalize the learner's knowledge about a task's transition structure as a directed graph, where vertices are states $s \in \mathcal{S}$ and edges represent known deterministic transitions between states $\langle s, s' \rangle$, such that the knowledge about the task is a set of transitions $\mathcal{K} \subseteq \mathcal{S} \times \mathcal{S}$. For simplicity, like in the basketball example, henceforward we assume that the reward function $\mathcal{K} \subseteq \mathcal{S} \times \mathcal{S}$ is known to both learner and teacher from the outset of the task and all knowledge gaps are in the transition function.

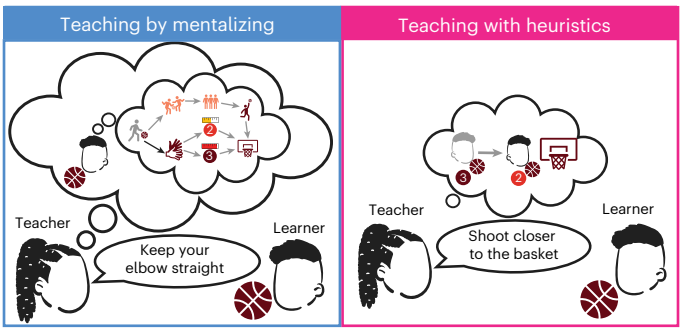


Fig. 1 | Mentalizing and heuristic teaching strategies. Consider a basketball coach watching a player miss a basket and providing advice to help him improve. A heuristic teaching strategy might lead her to simply advise 'shoot closer to the basket', whereas a teaching strategy that involves reasoning about what the learner seems to know (that is, mentalizing) might lead her to provide more tailored advice.

The teacher assumes the learner acts to maximize rewards, given what she knows. That is, the learner maximizes a value function $V_{\mathcal{K}}(s) = R(s) + \gamma \max_{s':\langle s,s' \rangle \in \mathcal{K}} V_{\mathcal{K}}(s')$ by taking actions according to an optimal policy that selects maximal actions (with ties broken uniformly), denoted $\pi_{\mathcal{K}}(s'|s)$.

Because the teacher does not have direct access to the learner's knowledge, she can only infer his knowledge from his behaviour through a process of inverse planning[25]. We model this as Bayesian inference about the learner's task knowledge, $\mathcal{K}$, from a sequence of states, $s_{0:T} = \langle s_0, s_1, \ldots, s_T \rangle$, visited by the learner:

$$P(\mathcal{K}|s_{0:T}) \propto P(\mathcal{K}) \prod_t \pi_{\mathcal{K}}(s_{t+1}|s_t) \quad (1)$$

where $P(\mathcal{K})$ is a prior over knowledge the learner might have and $\prod_t \pi_{\mathcal{K}}(s_{t+1}|s_t)$ is the likelihood of the behaviour observed given knowledge $\mathcal{K}$.

We consider teaching that is purely aligned with the learner's goal to maximize rewards: teaching is valuable to the extent that it provides new information relevant to the learner's task performance. Because the teacher cannot directly observe the learner's knowledge, she must give the best advice she can under uncertainty.

These intuitions can be formalized as a teacher selecting an edge in the graph, $x = \langle s, s' \rangle$, to teach the learner. The learner then updates his representation of the task's transition structure from $\mathcal{K}$ to $\mathcal{K} \cup \{x\}$ and will act optimally given this new knowledge. The teaching utility of an edge is then the difference between the optimal value of the initial task state s_0 given this new transition structure and the original value of s_0 before this knowledge was gained: $V_{\mathcal{K} \cup \{x\}}(s_0) - V_{\mathcal{K}}(s_0)$. The expected teaching utility is the expectation of the teaching utility taken with respect to the teacher's posterior over the learner's knowledge:

$$U(x; s_{0:T}) = \sum_{\mathcal{K}} P(\mathcal{K}|s_{0:T})(V_{\mathcal{K} \cup \{x\}}(s_0) - V_{\mathcal{K}}(s_0)) \quad (2)$$

We assume that this Bayes Optimal Teacher selects an edge to give as advice by (soft) maximizing equation (2) using a Luce choice rule[35,40]. The Bayes Optimal Teacher thereby formalizes teaching by mentalizing by combining inverse planning (equation (1)) with teaching utility maximization (equation (2)).

We consider two additional models that adhere to Bayesian principles but incorporate weaker assumptions about how the learner is represented. The first model replaces the planning likelihood in the posterior inference described by equation (1) with a purely behavioural likelihood, $\mathbf{1}[\langle s_t, s_{t+1} \rangle \in \mathcal{K}]$, which evaluates to 1 if the learner moves

https://doi.org/10.1038/s41562-026-02540-2

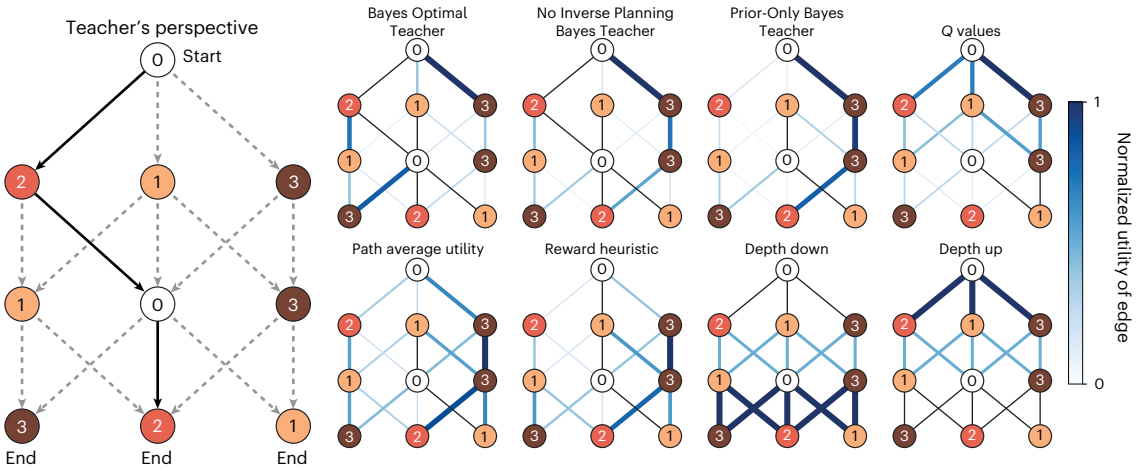


Fig. 2 | Graph teaching task trial and model predictions. Large graph: an example trial graph, in which the nodes represent reward values, grey edges are possible paths, and dark edges indicate a learner's previous path attempting to collect the most points when traversing from start to end. This is a directed deterministic graph where the learner can only move downwards. Each trial involved a graph with different reward points on the nodes, with the same edge structure. Participants took the role of a teacher and were instructed that the learner knows only a subset of edges in the graph and to teach a single edge. Small graphs: model predictions for the trial. Dark-blue edges indicate higher normalized utility (log-probability) predicted by the model.

along an edge in their knowledge structure and 0 otherwise. As such, rather then assuming that the learner is doing the best they can given their knowledge, this teacher gives equal likelihood to any knowledge structure where the learner's behaviour would have been possible. Because this captures a teacher that is not ascribing goal-maximizing planning to the learner during inference, we call this the No Inverse Planning Bayesian Teacher. The second variant we consider ignores the likelihood altogether and maximizes teaching utility with respect to a uniform prior, $P(\mathcal{K})$. As such, rather than giving tailored advice, the teacher would give general advice that would be helpful for the average learner. We refer to this model as the Prior-Only Bayesian Teacher.

To summarize, following previous work in computational social cognition[8,9,13,26,27,29], we model the computations involved in teaching with mentalizing as optimal Bayesian inference over a learner's knowledge and maximization of a learner's expected future performance.

Heuristic teaching

Teaching by mentalizing is flexible, effective and context sensitive, but it is also computationally costly as it relies on inference over possible learner knowledge structures as well as mentally evaluating the effects of different teaching interventions[15,41]. Rather than engaging in such 'model-based' inference and choice, a teacher could rely on a 'model-free' rule or heuristic to teach (for example, the coach simply telling the player to shoot the ball closer to the hoop). We formalize cognitive strategies of this kind as action selection based on heuristic value functions over edges in the transition graph, $H : \mathcal{S} \times \mathcal{S} \to \mathbb{R}$, that correspond to environment features but are not specific to the learner. For example, the heuristic 'teach connections between high-reward states' would correspond to a function $H_{\text{Reward}}(s, s') = R(s) + R(s')$. We then use a softmax choice rule over the heuristic value function to derive a heuristic teaching policy.

As with heuristic decision-making in general, heuristic teaching is computationally cheap because it relies on superficial task features, yet it can be surprisingly effective if it approximates optimal behaviour[18,42,43]. This overlap makes it difficult to discern whether a teacher is genuinely mentalizing or simply following a salient heuristic. Our experimental approach addresses this challenge by designing teaching scenarios in which heuristic teaching and mentalizing diverge (incongruent trials) or converge (congruent trials).

In addition to heuristic teaching models based on superficial features, we created two non-mentalizing utility models that treat teaching as a pure puzzle-solving problem and ignore any learner representation. The first is a tabular Q-value model that assigns each edge $x = \langle s, s' \rangle$ utility $U_Q(x) = Q(s, s')$, where Q values were learned via Q-learning[44] The second is a path-averaged utility model that assigns higher utility to edges that tend to appear on paths with higher total rewards. Model outputs for an example trial are shown in Fig. 2.

In experiment 1, we focus on identifying individual differences in teaching strategies using our graph teaching task, including determining the kinds of heuristics people adopt on our task. Experiment 2 was designed to test the influence of training on teaching performance, specifically, how environmental features impact participants' use of heuristics for teaching. Experiment 3 examined how scaffolding mental state inference with an auxiliary task could mitigate the tendency to use heuristics and lead to an overall improvement in teaching.

Results

Individual differences in teaching strategies (experiment 1)

Our first experiment aimed to validate a teaching task and assess whether people spontaneously adopt different teaching strategies. Participants ($N = 100$) played the role of teachers in the experiment. They were shown 40 graphs with possible edges connecting nodes with rewards to be collected, and a single path from top to bottom already taken by the learner (Fig. 2). Each participant was instructed that each of the 40 trials involved a distinct learner who may not know all the possible edges (paths) between nodes, and that they had taken a path that collected as many points as they could, given their knowledge. The participant's task was to teach each learner a single edge to help them improve their performance as much as possible. Critically, there was no feedback presented, meaning that the teachers did not know whether their teaching in fact improved each learner's performance. All participants were asked to teach in the same 40 graphs. See Supplementary Fig. 1 for a step-by-step schematic that walks through a single trial from the learner's and teacher's perspectives.

The goal of experiment 1 was to assess the range of teaching strategies people spontaneously adopt. For our first analysis, we focused on task performance. Performance was quantified as the expected teaching utility of chosen edges (equation (2)) normalized by the best edge on a graph (that is, choosing the best edge yields a teaching score of 1 and any other choices a score of <1).

As shown in Fig. 3a, the performance distribution is roughly bimodal, with one cluster of participants centred at a normalized

https://doi.org/10.1038/s41562-026-02540-2

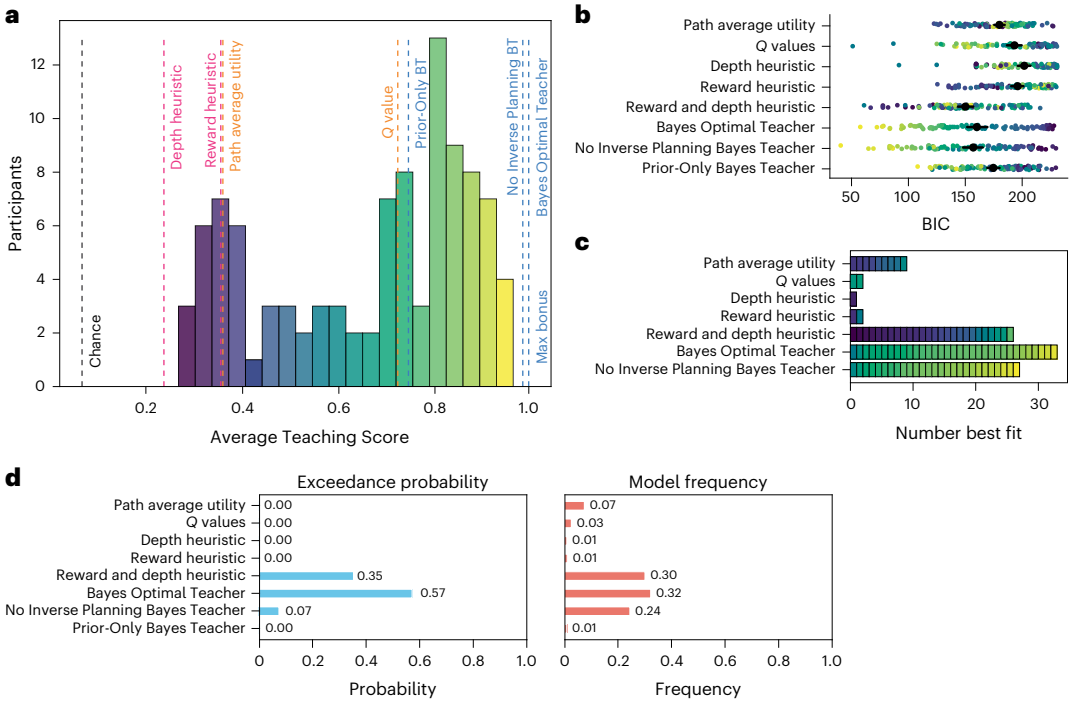


Fig. 3 | Bimodal distribution of teaching scores reveals two primary teaching strategies. a, Histogram of participants' average teaching score across 40 trials. Bar colours reflecting each bin's mean teaching score on a continuous blue-to-yellow scale (blue, low; yellow, high). The dotted vertical lines represent the argmax performance of each model as a pure teaching strategy, with pink labels denoting heuristic models and blue labels denoting Bayesian mentalizing models. Participants show a bimodal distribution of performance, with mentalizing models corresponding to high-performing participants and heuristic models corresponding to low-performing participants. **b**, BIC scores for different models across participants ($n = 100$ participants), with black circles showing the mean and horizontal whiskers showing bootstrapped 95% CI (1,000 resamples). Coloured dots are BIC scores for individual participants, with the colour reflecting the participant's Teaching Score using the colour scheme from **a**. A lower BIC score indicates a better fit. Heuristic models show better fit for lower-performing participants (blue hues for lower individual BIC scores) whereas Bayesian models show better fit for high performing participants (yellow hues). **c**, The number of participants best fit by each model, where the 'best-fitting' model is defined as the one with the lowest BIC among all models for that participant. Colours in **b**and **c** represent individuals' average performance as per **a**. **d**, Random-effects model comparison. Left: exceedance probability, that is, the probability that each model is more prevalent than all others. Right: estimated population frequency, that is, the fraction of participants best explained by each model, estimated via variational Bayes. BT, Bayes teacher.

teaching utility of about 0.80 and another cluster centred around 0.35. To interpret these results, we calculated the ceiling performance of several different mentalizing and heuristic models on the task under an argmax policy (Methods). As shown in the figure, the performance of the (simplified) mentalizing models overlaps with the higher-performing cluster of participants, while heuristic models—particularly the reward heuristic—aligns with the low-performing cluster.

Comparing participant performance with that of idealized mentalizing and heuristic models provides a coarse-grained view into the range of teaching strategies people employ. To gain a more fine-grained understanding of people's teaching strategies, we fit each of the mentalizing and heuristic models to individual participant choices and calculated Bayesian information criterion (BIC) scores per model per participant. Figure 3b shows the mean BIC scores with 95% confidence intervals (CIs) for the different models (black circles with whiskers; lower scores reflect a better model fit), together with individual BIC scores for each participant represented by coloured points, where the colour reflects each participant's teaching score. The best-fitting model combined two heuristics (reward and depth heuristic model), with the other best-fitting models being the Bayes Optimal Teacher and the No Inverse Planning Bayes Teacher. Consistent with mentalizing being more effective than heuristics, the multiple-heuristic model fit the lower-performing individuals best (blue hues for participants with especially low BIC scores) while the mentalizing models fit the higher-performing individuals best (yellow hues for low BIC-score participants). Figure 3c shows the number of participants best fit by each model, indicating that no single model provides the best fit for all participants. To obtain a stricter comparison, we conducted a random-effects model comparison following[45]. For each model, we computed the exceedance probability, the probability that the model is more prevalent than any alternative given the data, and we estimated population frequencies, the fraction of participants best explained by each model. Exceedance favoured the Bayes Optimal Teacher ($\phi = 0.57$, estimated population frequency $r = 0.32$) and the reward and depth heuristic model ($\phi = 0.36$, $r = 0.30$), with the No Inverse Planning Bayesian Teacher also accounting for a substantial fraction ($\phi = 0.07$, $r = 0.24$) (Fig. 3d). These frequency estimates indicate that no single model dominates; rather, the leading models jointly account for most participants.

Experiment 1 therefore validated our Graph Teaching Task and modelling framework, while demonstrating that people spontaneously adopt distinct teaching strategies reflecting mentalizing and heuristic algorithms. But how malleable are these strategies and are they sensitive to task contingencies? For example, heuristics are more cognitively frugal compared with mentalizing, so a desire to minimize cognitive effort might shape people's choice of one versus the other kind of strategy. In experiments 2 and 3, we aimed to investigate whether people would adapt their cognitive strategy in a manner reflecting the costs associated with mentalizing.

Strategic deployment of teaching strategies (experiment 2)

Experiment 2 examined whether people adaptively adjust their teaching strategy based on the structure of teaching trials encountered.

https://doi.org/10.1038/s41562-026-02540-2

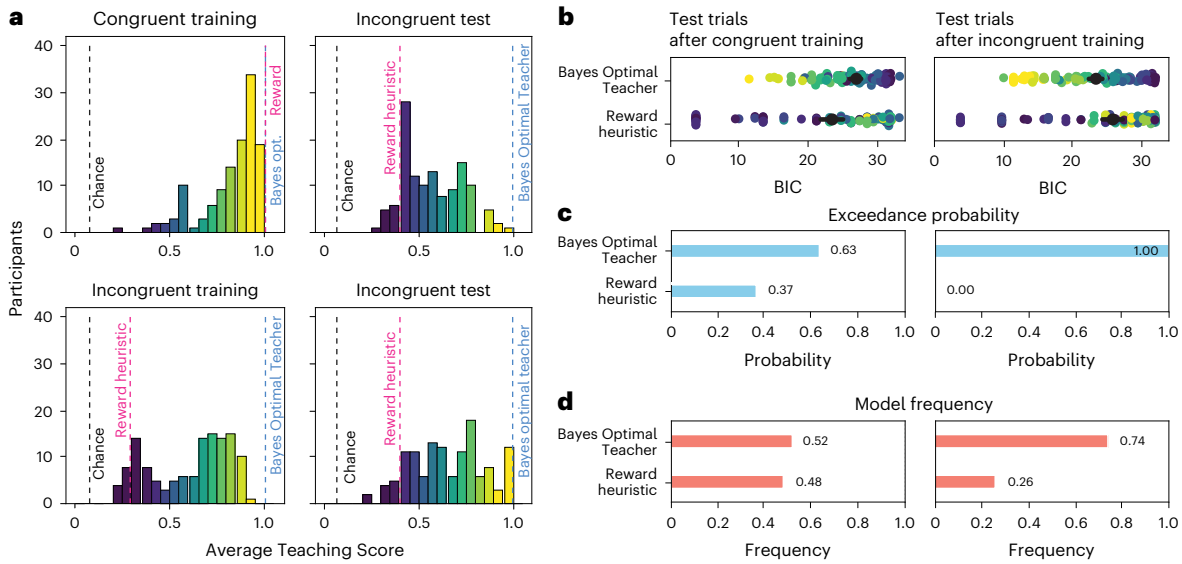


Fig. 4 | Teaching performance is affected by training conditions. a, Histogram of participants' average Teaching Score across 20 training trials (left) and 5 test trials (right), with bar colours reflecting each bin's mean Teaching Score on a continuous blue-to-yellow scale (blue/purple, low; yellow, high). Dotted vertical lines represent the performance of each model as a pure teaching strategy (pink labels, heuristic; blue labels, Bayesian). **b**, BIC scores for the Bayes Optimal Teacher model and reward heuristic model for each group on test trials alone (heuristic congruent: $n = 125$; heuristic incongruent: $n = 128$). Black circle and whiskers: mean BIC and bootstrapped 95% CI (1,000 resamples). Coloured dots are BIC scores of individual participants, with the colour reflecting their Teaching Score using the scale from **a** (yellow colours indicate higher scores). Lower BIC values indicate a better model fit. **c**, Exceedance probability for each model on test trials, computed from the Dirichlet posterior over model frequencies, shown separately for the heuristic-congruent (left) and heuristic-incongruent (right) group. **d**, Estimated population frequency on test trials (random-effects variational Bayes), that is, the fraction of participants best explained by each model, shown separately for heuristic-congruent (left) and heuristic-incongruent (right) groups.

Specifically, we test the hypothesis that people adopt mentalizing versus heuristic strategies depending on the usefulness of a simple, cognitively frugal heuristic. Motivated by the observation in experiment 1 that participants spontaneously use a reward heuristic on our Graph Teaching Task, we ran a preregistered[46] experiment in which we systematically manipulated the usefulness of this heuristic on an initial set of trials.

The experiment consisted of 20 training trials followed by five test trials. Each of $N = 253$ participants was assigned to either the heuristic-congruent training condition ($N = 125$), in which the reward heuristic values were aligned with the Bayes optimal teaching values in the training trials, or the heuristic-incongruent training condition ($N = 128$), in which the heuristic and optimal values diverged during training. Both groups then received the same set of five test trials, all of which had heuristic values that were incongruent with optimal values. As in experiment 1, there was no feedback presented on any trial. Training and test phases were not indicated to participants, who simply completed 25 trials of the task without additional instructions.

This design enabled us to compare the effectiveness of different teaching strategies in environments where a readily available feature can serve as a heuristic for optimal teaching versus environments where this feature is not useful as a heuristic. Figure 4a shows that heuristic-congruent participants performed well on training, while heuristic-incongruent participant performance had a similar bimodal pattern as in experiment 1, where some individuals' performance aligns with use of the (here, specifically unhelpful) heuristic while others performed more optimally.

Our main analysis compared normalized performance and model fits on the test trials between the two conditions. Here, heuristic-congruent participants performed worse than heuristic-incongruent participants on the test trials (Fig. 4a; Mann–Whitney U test: $U = 5704$, $n_1 = 125$, $n_2 = 128$, $P < 0.001$, rank-biserial $r = 0.287$).

Similarly, choice models fit to the test trials showed that the Bayes optimal model accounted for the behaviour of heuristic-incongruent participants better than the reward heuristic model (Fig. 4b; Mann–Whitney U test: $U = 5461$, $n_1 = 128$, $n_2 = 128$, $P < 0.001$, rank-biserial $r = 0.333$), but this was not the case for participants in the heuristic-congruent group (Mann–Whitney U test: $U = 8059$, $n_1 = 125$, $n_2 = 125$, $P = 0.667$, rank-biserial $r = -0.032$). Consistent with these findings, the random-effects model comparison showed the same pattern: in the heuristic-incongruent group, the Bayes Optimal Teacher had an exceedance probability of $\phi = 1.00$ and estimated population frequency of $r = 0.74$, whereas in the heuristic-congruent group, the reward heuristic was comparatively more prevalent ($r = 0.49$) than in the incongruent group ($r = 0.26$) (Fig. 4c,d). This indicates that some participants appear more susceptible to environmental features and shift towards heuristic strategy when those features are congruent, whereas others maintain mentalizing-based strategies.

Scaffolding teaching by mentalizing (experiment 3)

Experiment 3 tested whether persistence observed in experiment 2 could be explained by the cognitive cost of mentalizing. We reasoned that if people use heuristics to avoid the cost of mentalizing, then reducing the cost of mentalizing should decrease the use of heuristics and increase mentalizing when appropriate. Thus, in the preregistered experiment 3[47], we introduced an auxiliary scaffolding task designed to reduce the cost of mentalizing. Our main prediction was that people who received this scaffolding would engage in more mentalizing, teaching in a way that is more consistent with the Bayes Optimal Teacher.

Participants were placed into one of six conditions ($N = 759$) organized into a 2×3 design defined by two training conditions (heuristic congruent and heuristic incongruent) and three scaffolding conditions (no scaffolding, inference scaffolding and reward scaffolding). Sample sizes per condition were approximately equal (ranging from 120 to 134; Table 1); see the Methods for details. Participants in the no-scaffolding condition performed the same task as in experiment 2 (with some minor differences in the training trials; Methods). Participants in the inference scaffolding condition were shown the same trials but were

https://doi.org/10.1038/s41562-026-02540-2

Table 1 | Experiment 3 per condition participant count

	Heuristic congruent	Heuristic incongruent
Inference scaffolding	120	134
No scaffolding	130	130
Reward scaffolding	122	123

asked to complete an auxiliary scaffolding task at the start of each training trial. Specifically, before teaching an edge, they were asked to mark three edges they believed the learner was unaware of. They received no feedback. This task was intended to scaffold mentalizing and highlight differences between the teacher's knowledge and the learner's knowledge. Finally, participants in the reward scaffolding condition performed a different auxiliary task where they were asked to mark three edges connected to the nodes with the highest values. Here, the auxiliary task encouraged use of a reward heuristic rather than mentalizing. In addition, the reward scaffolding condition allowed us to control for the general effect of having an auxiliary task. In both the inference and reward scaffolding conditions, after making their selections, participants then taught one edge to the learner before moving on to the next learner and graph. Participants in all three conditions performed the same set of test trials without any auxiliary tasks.

Consistent with our predictions, we found that a scaffolding intervention that engaged participants in inferring what a learner does not know increases the use of mentalizing and reduces reliance on heuristics. These conclusions are supported by both behavioural and model-based analyses, and Fig. 5a shows participant performance for all conditions.

For our behavioural analyses, we used a linear mixed-effects model to examine the effects of the two training conditions (heuristic congruent and heuristic incongruent) and three scaffolding conditions (no scaffolding, inference scaffolding and reward scaffolding), and their interaction, on the dependent variable (Teaching Score), and included random intercepts for each participant to account for repeated measures. The analysis revealed a significant main effect of scaffolding conditions ($F(2, 753.64) = 45.2939$, $P < 0.001$, partial $\eta_p^2 = 0.107$); the main effect of training condition was not significant ($F(1, 753.65) = 3.79$, $P = 0.052$, partial $\eta_p^2 = 0.005$). There was a significant interaction between training conditions and scaffolding conditions ($F(2, 753.65) = 4.23$, $P = 0.015$, partial $\eta_p^2 = 0.107$). Post-hoc pairwise comparisons with Bonferroni adjustment indicated that in the no-scaffolding condition, the heuristic-incongruent training condition had higher Teaching Scores compared with the heuristic-congruent training condition (Mann–Whitney U test: $U(130, 130) = 6{,}377.0$, $P < 0.001$, rank-biserial $r = 0.245$), replicating our findings from experiment 2. By contrast, no significant differences between training conditions were found in the reward scaffolding ($U(122, 123) = 7{,}228.5$, $P = 0.436$, rank-biserial $r = 0.037$) or inference scaffolding conditions ($U(120, 134) = 8074.5$, $P = 0.954$, rank-biserial $r = -0.004$). These comparisons indicate that the main effect of training condition was primarily driven by the observed difference between the two no-scaffolding training conditions.

Next, we focused on whether the scaffolding interventions affected participants' performance at test, regardless of training condition. In the Inference Scaffolding condition, where participants were asked to mark three edges the learner may not know, we observed higher performance at test compared with the no-scaffolding condition, significantly so for the heuristic-congruent training conditions (Mann–Whitney U test: $U(120, 130) = 10{,}575$, $P < 0.001$, rank-biserial $r = -0.356$) but not for the heuristic-incongruent training conditions ($U(134, 130) = 9{,}828$, $P = 0.071$, rank-biserial $r = -0.128$). This suggests that the intervention influenced participants to use a mentalizing teaching strategy on test trials even when they were no longer prompted to infer which edges the learner may not know. This finding underscores the impact of the scaffolding intervention to improve teaching even when environmental factors promote a heuristic strategy (Fig. 5b).

Conversely, participants in the reward scaffolding condition with heuristic-incongruent training performed worse at test than those in the no-scaffolding condition (Mann–Whitney U test between reward scaffolding and no scaffolding: $U(123, 130) = 5{,}252$, $P < 0.001$, rank-biserial $r = 0.343$), although this effect was not significant for the heuristic-congruent training conditions where participants in the no-scaffolding condition performed as poorly as those in the reward scaffolding condition (Mann–Whitney U test: $U(122, 130) = 6971$, $P = 0.097$, rank-biserial $r = 0.121$). Together, these results highlight the power of scaffolding, with the latter results demonstrating that if implemented incorrectly, scaffolding can also lead to deleterious consequences (Fig. 5b).

Finally, we also performed cognitive modelling analyses by fitting all models to the test trials as in experiment 2 (Fig. 6a). In the inference scaffolding condition, the Bayes Optimal Teacher model was the better-fitting model across both training groups (congruent: $\Delta BIC_{OBM-Reward}$ −3.73, 95% CI −5.91 to −1.55, Wilcoxon $P < 0.001$; incongruent: ΔBIC −4.59, 95% CI −6.39 to −2.78, $P < 0.001$). By contrast, in the reward scaffolding condition, the Reward Heuristic model was the better-fitting model across both training conditions (congruent: ΔBIC +5.59, 95% CI 3.10 to 8.07, $P = 0.006$; incongruent: ΔBIC +6.26, 95% CI 3.70 to 8.82, $P = 0.001$). These results suggest that our scaffolding conditions were effective in inducing behaviour consistent with each of the two alternative strategies. In other words, participants' strategies aligned with the type of scaffolding they received: mentalizing when inference was scaffolded, heuristic when reward was scaffolded. Replicating the findings from experiment 2, in the no-scaffolding condition, the Bayes Optimal Teacher model provided a better fit following incongruent training (ΔBIC −1.93, 95% CI −3.71 to −0.14, Wilcoxon $P = 0.003$), while the two models did not differ following congruent training (ΔBIC +0.99, 95% CI −0.84 to +2.81, $P = 0.415$). While no significant differences across congruency were observed within either the reward scaffolding or inference scaffolding groups, we collapsed across congruency for the random-effects model comparison shown in Fig. 6b (full random-effects comparisons by scaffolding × congruency are reported in Supplementary Fig. 9). Consistent with the BIC results, the exceedance probabilities and estimated population frequencies favoured the Bayes Optimal Teacher in the inference scaffolding condition ($\phi = 1.00$, $r = 0.81$) and the reward heuristic in the reward scaffolding condition ($\phi = 0.95$, $r = 0.55$) (Fig. 6b). Overall, these results support the intuition that the scaffolding interventions are influencing participants' teaching strategies and that this influence is strong enough to overcome the environmental effects seen in experiment 2.

Discussion

We investigated the strategies that people use to teach, focusing on costly mentalizing strategies that tailor advice to specific learners and frugal heuristic strategies that provide generically useful information. Our experimental results and analyses demonstrate that participants adopt both kinds of strategies to teach (experiment 1), that teaching experiences can systematically promote the use and persistence of heuristic strategies (experiment 2) and that scaffolding the computations involved in teaching by mentalizing can counteract a tendency towards heuristic teaching while scaffolding directed at task features can instead reinforce it (experiment 3). As we discuss next, these findings contribute to developing an integrated, theoretically grounded understanding of the psychology of teaching that spans computational modelling, education research and social cognition more broadly.

First, our analyses shed new light on teaching by bridging two influential theoretical frameworks from computational cognitive science: Bayesian social cognition[8,9,13,27,29] and reinforcement learning[14,38]. We proposed that, because the mentalizing strategies captured by Bayesian inference and action selection constitute model-based

https://doi.org/10.1038/s41562-026-02540-2

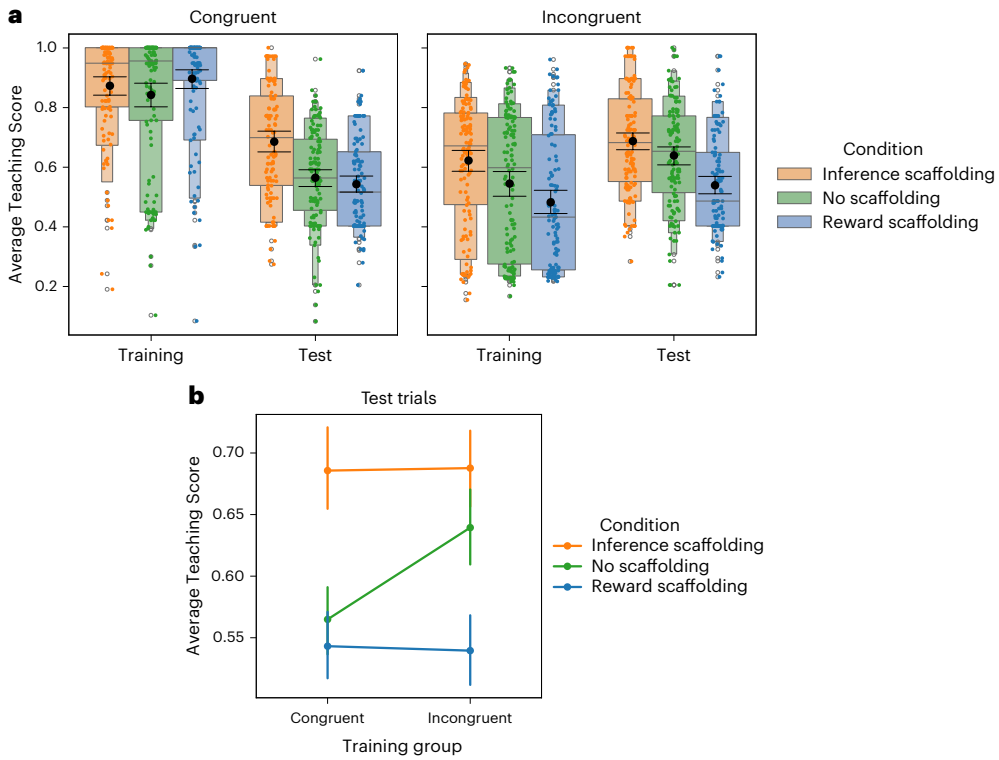


Fig. 5 | Inference scaffolding enhances teaching performance. a, Average Teaching Score for each training condition and scaffolding condition for training trials (left) and test trials (right). Note that, while training trials were different across the congruent and incongruent conditions, test trials were the same for all training condition. The plot shows the distribution of teaching scores using letter-value plots (boxenplots), which display nested boxes representing progressively smaller quantile ranges from the median outwards: the innermost box spans the interquartile range (25th–75th percentiles) and successive boxes extend to further quantile boundaries, providing a detailed view of the distribution shape including tails. Central grey line: median Teaching Score. The black circle and whiskers represent the mean Teaching Score across individuals with bootstrapped 95% CI (1,000 resamples). Sample sizes per condition: inference scaffolding (Congruent $n = 120$, Incongruent $n = 134$), no scaffolding ($n = 130$, $n = 130$) and reward scaffolding ($n = 122$, $n = 123$). **b**, Replotting of test trial performance for all six groups for ease of comparison. Data are presented as mean values ± bootstrapped 95% CI. The Inference Scaffolding groups performed best, and similarly regardless of congruent or incongruent training, whereas the reward scaffolding groups performed worse (again, seemingly uninfluenced by training condition). Replicating experiment 2, the no-scaffolding groups performed better (that is, more consistent with mentalizing) in the incongruent training condition.

decision-making[15], people should intelligently arbitrate between these more effortful strategies and easier heuristic, rule-based, or model-free strategies[32,48–50]. Empirically, this predicts we should see artefacts of strategy arbitration that are typically found in other domains, such as strategy persistence[22,51,52], cognitive rigidity[53] and insensitivity to contextual shifts (for example, hysteresis effects)[54,55]. Our results demonstrating the adoption and persistence of heuristic strategies (experiment 2) and the effects of scaffolding computations with auxiliary tasks (experiment 3) are consistent with these predictions.

Beyond strategy arbitration, our findings also reveal the possibility of more complex interactions between mentalizing and heuristics. For instance, in our experiments, participants never receive explicit feedback on the quality of their teaching yet systematically adopt different strategies as a function of their teaching experiences. This suggests that mentalizing can serve not only as a teaching strategy but also as a mechanism for evaluating heuristics: By simulating, with one's theory of mind, how a learner would respond to different teaching interventions, teachers can obtain internal feedback about each edge's value. We speculate that some participants in the congruent conditions initially used mentalizing to discover and adopt the heuristic (a form of inference amortization[56]), but once they stopped mentalizing, they could no longer reassess whether the heuristic remained effective, a form of planning fallibility in which individuals default to habitual responses[57,58].

Second, by relaxing the assumption that teachers infer and act optimally to account for suboptimal heuristic teaching, our work provides new ways to connect formal theories of teaching to informal constructs in the educational literature. While recognizing that our abstract task is a highly simplified version of the rich social complexity present in interactive classroom instruction, our finding that people spontaneously teach by mentalizing and with heuristics mirrors the distinction between routine and adaptive expertise developed in the educational literature[59]. For example, Stigler and Miller[60–62] draw on this distinction when they note that classroom instruction in the USA is dominated by routine strategies: teachers solve an example problem and then supervise students who apply the same procedure. Japanese classrooms, by contrast, foster adaptive strategies; students begin with a challenging problem and teachers provide personalized feedback as the students struggle to solve it independently. The latter approach demands greater adaptive expertise because teachers must continuously model each learner's evolving understanding.

The educational literature also suggests that teaching by mentalizing not only enables better information transfer, but also makes teachers more effective at navigating the motivational dimensions of teaching. Teachers who adopt a mentalizing stance can more effectively lower students' stress[63] and anxiety[64], and are better at reducing violence[65] and disruptive behaviour[66]. Mentalizing interventions, usually involving perspective taking similar to our inference scaffolding condition, cut teachers' disciplinary actions in half[67], increase academic motivation[68], increase student empathy and emotion regulation[69], and improve teachers' own mental health[70]. Future work could investigate

https://doi.org/10.1038/s41562-026-02540-2

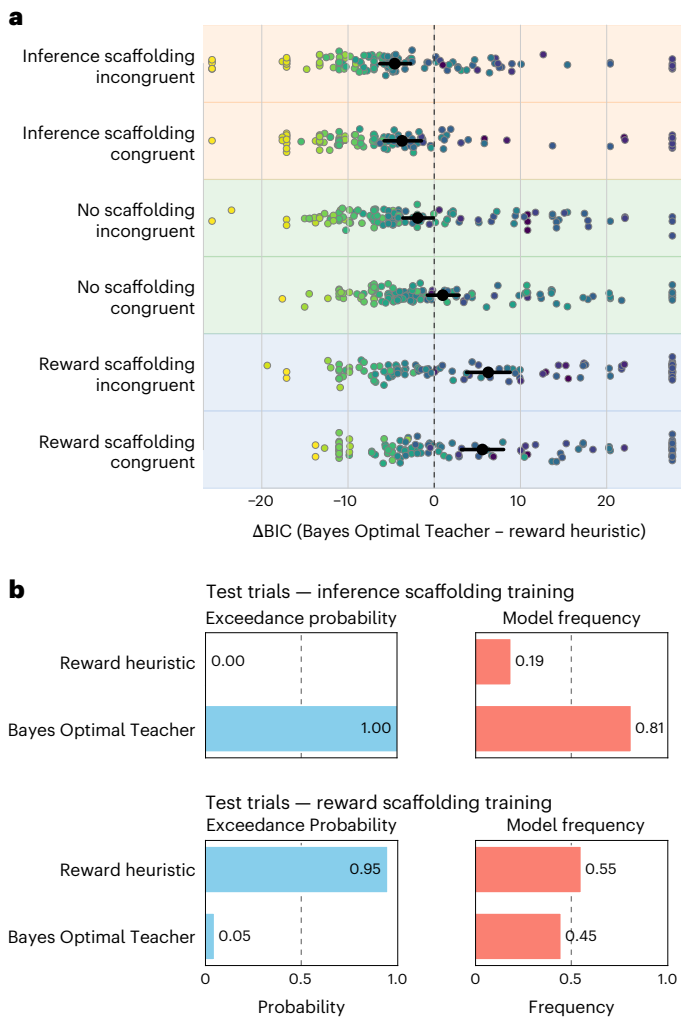


Fig. 6 | Model fits across training conditions and scaffolding conditions in experiment 3. a, Difference in BIC scores between the Bayes Optimal Teacher model and reward heuristic model across the two training and three scaffolding conditions (sample sizes per condition as in Table 1). Negative values indicate a better fit for the optimal Bayesian teacher model, while positive values indicate a better fit for the reward heuristic model. Black circles show the mean difference; horizontal whiskers show the bootstrapped 95% CI of the mean difference in BIC (1,000 resamples). Coloured dots represent individual BIC differences, with the colour reflecting the participant's Teaching Score (similar to Fig. 3, yellow: higher score, blue: lower score). **b**, Random-effects model comparison (collapsed across congruency). Inference scaffolding (top) favours the Bayes Optimal Teacher, while reward scaffolding (bottom) favours the reward heuristic.

the mechanistic basis for these findings by, for example, combining our framework with models of interpersonal emotion regulation and motivation[71,72].

Finally, our findings contribute to the growing literature on how cognitive costs shape human social cognition. Although previous work in psychology has long recognized that people avoid full theory of mind reasoning when simpler strategies suffice[73–76], this insight has been difficult to square with accounts of mentalizing based on inverse planning and Bayesian decision-making[8,27]. Our work complements several recent lines of research that have begun to tackle these issues, including work by Hawkins et al.[77] that examines how communicators adjust theory of mind use based on division of labour, work by Pöppel and Kopp[78] that introduces a satisficing framework for limiting mental effort, and work by Bass et al.[79] that studies how learners' perceptions of teaching effort shape their receptiveness to a teacher. Furthermore, while we have focused on the domain of pedagogy, our empirical findings are relevant to recent theoretical proposals that highlight people's use of more or less cognitively effortful strategies in other social domains, including social perception (for example, the institutional stance)[80] and moral deliberation (for example, resource-rational contractualism)[81]. A fruitful direction for future work will be to investigate how the arbitration of mentalizing and heuristic strategies in these other domains compares to their arbitration in the communicative setting studied here.

Our study has limitations. First, participants may have relied on hybrid strategies that blur the line between heuristic and inference-based approaches or reflect Bayesian processes under extreme priors, effectively ignoring new evidence[82]. Second, the absence of feedback may have prevented participants from recognizing and adjusting suboptimal strategies, highlighting the importance of feedback in real-world teaching. Finally, although our scaffolding intervention appeared to help participants adjust their teaching strategies, further research is needed to assess how enduring these changes might be outside controlled settings.

In short, the current work investigates the factors that lead people to adopt mentalizing versus heuristic teaching strategies. This approach provides new opportunities to bridge computational cognitive science with education research and to deepen our understanding of the cognitive processes that underpin human social interactions.

Methods

This research complies with all relevant ethical regulations. This study was approved by the Institutional Review Board of Princeton University (protocol 11968), and all participants provided informed consent.

Task

In all experiments, participants acted as teachers. They were shown a graph where a learner had already taken a path from start to finish, collecting points along the way. They were informed that the learner might not have known all possible edges (connections) between nodes, but took (one of) the best path(s) they could have given their knowledge. Participants' task was to select one edge to teach, which the learner would then add to its available paths to recalculate and execute the highest-scoring trajectory. No feedback was provided during the trials, so participants had to make decisions without knowing how effective their teaching was. Before starting the main task, participants underwent training, which included seeing two example trials from both the learner's and teacher's perspectives. This was followed by three practice trials in which participants received feedback on the learner's outcome and the quality of their teaching choice. After the practice phase, participants took a comprehension quiz. Those who failed the quiz twice were excluded from the study. After completing the main task, participants completed a post-task debriefing questionnaire to provide feedback on their strategies and the task.

Experiment 1. Participants completed 40 trials, with 20 unique graphs and their 20 horizontally flipped versions presented in random order. Participants were reminded that each trial involved a new learner, and no feedback was provided during the task, keeping the focus on participants' teaching strategies based on the observed trajectory.

Experiment 2. Participants were assigned to one of two teaching conditions: congruent or incongruent. In the congruent condition, the optimal teaching strategy in each of the first 20 'training' trials aligned with a simple reward heuristic, where the highest-value edge (that is, the edge whose nodes had the largest total reward) was the best choice. In the incongruent condition, in each of the first 20 trials the reward heuristic led to suboptimal performance, thus selecting the best edge to teach required more complex inference of the learner's knowledge. All participants then continued to complete an identical set of five

incongruent 'test' trials, with no feedback provided. The transition from teaching to test trials was not signalled. In addition, each trial had a 30-s time limit. If a participant did not respond within this limit, the trial was considered missed, and no bonus pay was awarded.

Experiment 3. Participants were again placed into congruent and incongruent groups as in experiment 2 and assigned to one of three scaffolding conditions: reward scaffolding, inference scaffolding or no scaffolding. Participants in the scaffolding conditions completed an intermediate task before teaching during the initial 15 training trials. In the reward scaffolding condition, they were asked to identify three edges connected to high-value nodes before selecting one edge to teach. In the inference scaffolding condition, participants identified three edges they believed the learner was unaware of before choosing an edge to teach. The no-scaffolding condition replicated the set-up from experiment 2 without the intermediate tasks. No feedback was provided during the main task. All participants completed the same three practice trials as in previous experiments, but those in the scaffolding conditions also completed two additional practice trials related to their specific scaffolding task. Participants completed 15 teaching trials followed by an identical set of 5 incongruent test trials for all conditions, for a total of 20 trials. Similar to experiment 2, all trials had a 30 second time limit. Unlike experiment 2, the transition from the teaching to the test phase was signalled for those in the scaffolding conditions, as they no longer had to complete the intermediate tasks once they reached the test phase.

Participants

Participants ($N = 100$) were recruited from Prolific using jsPsych[83], NivTurk[84] and AnyChart[85] for experiment 1 (61 women, 39 men; mean age 34.00 ± 10.38 years); $N = 265$ for experiment 2 (143 women, 122 men; mean age 36.46 ± 10.94 years); $N = 788$ for experiment 3 (379 women, 409 men; mean age 38.99 ± 12.31 years). In all experiments, participants were paid a base rate of US$3.25 and a performance-contingent bonus of up to US$6 in experiment 1 and US$2 in experiments 2 and 3. Participants earned an average of US$4.02 ± US$1.26 and spent 19.16 ± 10.44 min on experiment 1, earned an average of US$1.36 ± US$0.40 and spent 13.07 ± 5.15 min on experiment 2, and earned an average of US$1.41 ± US$0.47 and spent 16.82 ± 8.18 min on experiment 3. We followed the preregistered data exclusion criteria, removing ($N = 14$ total across all experiments) participants who had half of their test trials missing due to exceeding the 30-s time limit and participants that expressed confusion or misunderstanding about the task in the post-task debriefing form ($N = 27$ total across all experiments). The final number of participants was $N = 100$ for experiment 1, $N = 253$ for experiment 2 and $N = 759$ for experiment 3 (Table 1).

Experiment 2 was preregistered on 14 June 2023 (https://osf.io/ckg98), and experiment 3 was preregistered on 18 April 2024 (https://osf.io/x9h2v). We report the hypotheses, outcome variables and analyses as preregistered, with the following exceptions: (1) Mann–Whitney U tests were used in place of the preregistered independent two-sample t-tests because teaching score distributions were non-normal and bounded on [0, 1]; (2) the random-effects variational Bayes model comparison[45] was added in response to a reviewer request during peer review; and (3) two additional non-mentalizing utility models (tabular Q-value and path-averaged utility) were added during revision to provide a more comprehensive model comparison. Sample sizes for experiment 2 were determined by a power analysis based on a pilot study ($N = 35$, observed $d = 0.531$, $\alpha = 0.001$, power 0.80), yielding a target of $N = 124$ per condition. The final sample ($N = 253$) fell slightly below the planned $N = 275$ due to recruitment; however, both conditions exceeded the minimum required by the power analysis. Sample sizes for experiment 3 were determined by two power analyses: a simulation-based analysis using the `simr` package in R, which indicated approximately 85 participants per condition were needed for the scaffolding main effect ($\alpha = 0.001$, power 0.80), and a pilot-based analysis ($d = 0.445$, $\alpha = 0.01$, power 0.80) requiring $N = 120$ per condition for the replication of the training effect. We recruited above these minimums ($N = 759$ across 6 conditions, all cells ≥120) to allow for exclusions. Participants were assigned to experimental conditions via the Prolific recruitment platform; within each condition, training trial order was randomized client side. Data collection and analysis were not performed blind to the conditions of the experiments. All statistical tests were two-tailed. Mann–Whitney U tests were used rather than t-tests because the teaching score data were bounded on [0, 1] and non-normally distributed; visual inspection of Q–Q plots confirmed departures from normality (Supplementary Fig. 10), supporting the use of non-parametric tests.

Modelling

We used two types of models to explain participant behaviour: Bayesian inference models and heuristic-based models. The Bayesian models aim to infer the learner's knowledge and maximize the utility of teaching an edge, while the heuristic models rely on task features rather than learner features to choose an edge to teach.

Bayesian teacher models

The Bayesian teacher models consist of three variants: the Bayes Optimal Teacher, the No Inverse Planning Bayesian Teacher and the Prior-Only Bayesian Teacher. Each model follows the same general structure: (1) modelling the learner's behaviour, (2) inferring the learner's knowledge and (3) calculating the utility of teaching each edge.

Learner model. We model the learner as an agent navigating a directed deterministic graph, represented as a Markov decision process $M_L = \langle S, \mathcal{T}_L, R \rangle$, where S is the set of nodes (states), $\mathcal{T}_L$ is the learner's transition model (a subset of the true transitions $\mathcal{T}$) $\mathcal{T}_L \subseteq \mathcal{T}$ where $\mathcal{T} \subset S \times S$, and $R : S \to \mathbb{R}$ is the known reward function, which assigns rewards to each node. We assume that the learner follows the optimal policy, given their knowledge of the task, $\pi_L(s'|s, \mathcal{T}_L)$, resulting in a trajectory $\zeta = \{s_0, s_1, \ldots, s_T\}$ (as illustrated by the solid black lines in Fig. 2, right).

Utility calculation. For each Bayesian model, the utility of teaching an edge $e = (s, s')$ is the improvement in the learner's performance, defined as the difference in cumulative reward between the learner's current trajectory ζ and the new trajectory ζ_e:

$$U(e|\zeta) = \sum_{\mathcal{T}_L} P(\mathcal{T}_L|\zeta)(G(\zeta_e) - G(\zeta)), \tag{3}$$

where $G(\zeta)$ is the cumulative reward when following ζ and $P(\mathcal{T}_L|\zeta)$ is the probability distribution over the learner's transition model (see below).

Inference variants. The main difference between the Bayesian models lies in how $P(\mathcal{T}_L|\zeta)$ is computed, which depends on the likelihood function used in the inference.

Bayes Optimal Teacher. The Bayes Optimal Teacher uses inverse planning to infer the learner's transition model by considering how likely the observed trajectory ζ is under the learner's optimal policy, given their transition model $\mathcal{T}_L$:

$$P_{\mathrm{BOT}}(\mathcal{T}_L|\zeta) \propto P_{\mathrm{IRL}}(\zeta|\mathcal{T}_L)P(\mathcal{T}_L), \tag{4}$$

where $P(\mathcal{T}_L)$ is uniform over all graphs with at least one possible trajectory from the start state to a terminal state, and $P_{\mathrm{IRL}}(\zeta|\mathcal{T}_L)$ uses inverse reinforcement learning (IRL) to compute a distribution of possible graphs that produce ζ under the assumption that the learner is attempting to act optimally according to policy π^* (ref. 25):

$$P_{IRL}(\zeta|\mathcal{T}_L) = \prod_t \pi(s_{t+1}|s_t; \mathcal{T}_L) \tag{5}$$

where t is each time step beginning from S_0. In our task specifically, the likelihood $P_{\mathrm{IRL}}(\zeta|\mathcal{T}_{\mathrm{L}})$ can take on one of three values: it equals 1 if ζ uniquely represents the optimal trajectories for $\mathcal{T}_L$, equals $\frac{1}{n}$ if there are n equally optimal trajectories (including ζ) and equals 0 if ζ is not an optimal trajectories for $\mathcal{T}_{\mathrm{L}}$.

No Inverse Planning Bayesian Teacher. The No Inverse Planning Bayesian Teacher can be viewed as a version of the Bayes Optimal Teacher with a simplified likelihood function that assigns a value of 1 if the learner's transition model $\mathcal{T}_{\mathrm{L}}$ includes the observed trajectory ζ as a possible trajectory, and 0 otherwise:

$$P_{\mathrm{noIRLBT}}(\mathcal{T}_{\mathrm{L}}|\zeta) \propto P_{\zeta\subset\mathcal{T}_{\mathrm{L}}}(\zeta|\mathcal{T}_{\mathrm{L}})P(\mathcal{T}_{\mathrm{L}}), \tag{6}$$

where $P(\mathcal{T}_{\mathrm{L}})$ is uniform over all graphs with at least one possible trajectory from the start state to a terminal state, and $P_{\zeta\subset\mathcal{T}_{\mathrm{L}}}(\zeta|\mathcal{T}_{\mathrm{L}})$ is zero for learner graphs that do not include ζ and 1 if they do:

$$P_{\zeta\subset\mathcal{T}_L}(\zeta|\mathcal{T}_L) \propto \begin{cases} 1 & \text{if}\ \zeta \subset \mathcal{T}_L \\ 0 & \text{if}\ \zeta \not\subset \mathcal{T}_L \end{cases} \tag{7}$$

This means the No Inverse Planning Bayesian Teacher considers all $\mathcal{T}_{\mathrm{L}}$ that include ζ, without evaluating the optimality of the trajectory.

Prior-Only Bayesian Teacher. The Prior-Only Bayesian Teacher does not perform inference based on the observed trajectory and assumes a uniform prior over all possible transition models (that is, all graphs with at least one possible trajectory from the start state to a terminal state):

$$P_{\mathrm{Prior-onlyBT}}(\mathcal{T}_L|\zeta) = P(\mathcal{T}_L). \tag{8}$$

Bayesian model fitting. We used a softmax function to assign likelihood to each choice a participant made according to each Bayesian teacher model:

$$P(Y) = \frac{e^{\beta U_{\mathrm{model}}}}{\sum e^{\beta U_{\mathrm{model}}}}, \tag{9}$$

where the inverse temperature β is the free parameter.

Heuristic models

Our heuristic models do not include a learner model. Instead, they rely on task features to identify the most useful teaching edge. These models employ a set of heuristics based on task structure, such as the sum of node values for each edge or the position of the edge in the graph, rather than probabilistic inference. These models therefore offer a low-cost alternative to Bayesian models, focusing on general task structure rather than tailoring decisions to the specific learner.

Reward heuristic. The reward heuristic prioritizes edges based on the sum of the rewards at the connected nodes. The utility of an edge $e_{x,y}$ is determined by the following feature:

$$F_{reward}(e_{x,y} = \langle s_i, s_j\rangle) = R(s_i) + R(s_j) \tag{10}$$

where $R(s_i)$ and $R(s_j)$ represent the rewards at nodes s_i and s_j, respectively. Higher sums indicate higher utility and priority for selection.

Depth heuristic. This heuristic evaluates edges on the basis of their vertical position in the graph. Edges closer to the start (top of the graph) can be riskier because a continuing path from them may not be known; however, they may also have more impact as they potentially open up whole swaths of the graph map. By contrast, edges near the bottom tend to be safer as their use does not require additional knowledge, but their effect is local. We therefore used a depth feature $F_{\mathrm{Depth}}(e_{x,y} = [s_i, s_j]) = y$ (where y represents the vertical height of state s_j) together with logistic regression (see below) to parameterize an individual preference for either earlier or later edges.

Heuristic model fitting. We used a multinomial logistic regression to fit preferences for reward and depth features to the participant's choices:

$$P(Y) = \frac{e^{\beta_1 F_{\mathrm{reward}} + \beta_2 F_{\mathrm{depth}}}}{\sum e^{\beta_1 F_{\mathrm{reward}} + \beta_2 F_{\mathrm{depth}}}} \tag{11}$$

Non-mentalizing utility models

In addition to our Bayesian inference-based teachers, we implemented two models that compute edge utility without considering a learner. Unlike simple rule-based heuristics, these models compute and compare a utility for every candidate edge, similar to solving a puzzle.

Tabular *Q*-value model. We obtained state-action values via a Q-learning algorithm[44] implemented with the MSDM Python package[86]. In our graph, each edge (e) is a tuple (s, s'), that is, the transition from state s to next state s'. We therefore equate an edge with the action a that (deterministically) realizes $s \to s'$ and write

$$U_Q(e) = Q(s, a) = R(e) \tag{12}$$

Here, $R(e)$ is the expected cumulative reward for traversing that edge under the optimal policy.

Path-averaged utility model. This model enumerates every complete path P from the start state to any terminal state. For each path

$$p = (e_0, e_1, \ldots, e_T), \quad G(p) = \sum_{t=0}^{T} R(e_t),$$

we compute its total return $G(p)$. Then for each edge e, we compute all paths that include $eP(e) = \{p \in P: e \in p\}$, and define the edge's utility as the average return over those paths:

$$U_{\mathrm{path}}(e) = \frac{1}{|P(e)|} \sum_{p\in P(e)} G(p). \tag{13}$$

Intuitively, this model assigns higher utility to edges that tend to appear on paths with higher total rewards.

Choice rule. Both models produce a utility for every edge. We convert these into a decision probability via softmax,

$$P(e) = \frac{\exp(\beta\, U(e))}{\sum_{e'} \exp(\beta\, U(e'))}, \tag{14}$$

where β is an inverse-temperature parameter fit to human data.

Model comparisons

To assess which of our candidate teaching strategy models best accounts for human behaviour, we used both fixed-effects BIC comparison and a random-effects framework. The former approach highlights individual differences between participants that are better explained by each of the models, whereas the latter approach treats each participant's model identity as a latent variable, allowing different participants to follow different strategies and yielding direct estimates of the population-level proportions for each model[45]. Following the model comparison framework of ref. 45, we compute exceedance probabilities that quantify how likely each model is to be the most prevalent in the dataset.

Fixed-effects BIC

For each participant n and model k, we compute the BIC:

$$\mathrm{BIC}_{n,k} = -2\,\log p\left(Y_n|M_k,\theta_{n,k}\right) + K_k\,\log N,$$

where $\theta_{n,k}$ is the maximum-likelihood estimate of the parameters for model k for participant n, K_k is the number of free parameters and N is the number of observations[87]. Lower BIC values indicate stronger support for a given model.

Random-effects variational Bayes

Each participant n draws a latent label $z_n \in \{1, \ldots, K\}$ indicating which model generated their data: $z_n|\mathbf{m} \sim \mathrm{Categorical}(\mathbf{m})$, $\mathbf{m} \sim \mathrm{Dirichlet}(\boldsymbol{\alpha}_0)$, $\alpha_{0,k} = 1$, where $\sim$ denotes 'is distributed as', $z_n = k$ means 'participant n uses model k', and $\mathbf{m} = (m_1, \ldots, m_K)$ are unknown population frequencies of each model, to be inferred from the data. Following ref. 45, we place an uninformative Dirichlet prior $(\alpha_{0,k} = 1)$ on $\mathbf{m}m$.

We approximate the intractable posterior $p(\mathbf{m}, \{z_n\}|\{Y_n\})$ via a factorized variational distribution $q(\mathbf{m})\prod_n q(z_n)$. The variational-Bayes expectation–maximization updates then iterate until convergence:

E-step (participant responsibilities). For each participant n and model k, compute the approximate log-evidence[87]

$$\ell_{n,k} \approx -\frac{1}{2}\mathrm{BIC}_{n,k},$$

and update the responsibility (probability that participant n uses model k)

$$u_{n,k} = q(z_n = k) = \frac{\exp\left(\ell_{n,k} + \mathbb{E}_{q(\mathbf{m})}[\log m_k]\right)}{\sum_{j=1}^{K}\exp\left(\ell_{n,j} + \mathbb{E}_{q(\mathbf{m})}[\log m_j]\right)},$$

where under the Dirichlet posterior, $\mathrm{Dirichlet}(\boldsymbol{\alpha})$,

$$\mathbb{E}_{q(\mathbf{m})}[\log m_k] = \psi(\alpha_k) - \psi\left(\sum_{j=1}^{K}\alpha_j\right),$$

and ψ is the digamma function.

M-step (population update). Update the Dirichlet posterior parameters:

$$\alpha_k = \alpha_{0,k} + \sum_{n=1}^{N} u_{n,k},$$

and extract the new model-frequency estimate:

$$r_k = \frac{\alpha_k}{\sum_{j=1}^{K}\alpha_j}$$

Iterate E and M steps until $\|\mathbf{r}^{\mathrm{new}} - \mathbf{r}^{\mathrm{old}}\|$ is $\leq 10^{-6}$.

Exceedance probability

After fitting the random-effects model, we obtain a Dirichlet posterior $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)$ over the population frequencies $\mathbf{m}$. To quantify how strongly the data support each model k as the most prevalent strategy, we compute the exceedance probability

$$\phi_k = \Pr\left(m_k > m_j\,\forall j \neq k|\boldsymbol{\alpha}\right).$$

Intuitively, ϕ_k (ranges from 0 (model k never wins) to 1 (model k always wins)) is the probability that model k wins against all others in the population. We estimate ϕ_k by Monte Carlo sampling with S samples:

$$\phi_k \approx \frac{1}{S}\sum_{s=1}^{S}\mathbf{1}\left[m_k^{(s)} = \max_j m_j^{(s)}\right], \quad \mathbf{m}^{(s)} \sim \mathrm{Dirichlet}(\boldsymbol{\alpha}), \quad (s = 1, \ldots, S).$$

In practice, we draw S independent samples $\mathbf{m}^{(s)}$ from the posterior Dirichlet distribution. For each sample, we compute $\max_j m_j^{(s)}$, the largest component (that is, the 'winning' model frequency) in that hypothetical population. The indicator $\mathbf{1}[\cdot]$ equals 1 when model k attains this maximum and 0 otherwise. Averaging these indicators over all S samples yields ϕ_k, the proportion of draws in which model k is the most frequent.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Anonymized participant data are available via GitHub at https://github.com/sharootonian/CognitiveStrategiesInTeaching.

Code availability

The code used for all models and analyses is available via GitHub at https://github.com/sharootonian/CognitiveStrategiesInTeaching.

References

1. Csibra, G. & Gergely, G. Natural pedagogy. *Trends Cogn. Sci.* **13**, 148–153 (2009).
2. Tomasello, M., Carpenter, M., Call, J., Behne, T. & Moll, H. Understanding and sharing intentions: the origins of cultural cognition. *Behav. Brain Sci.* **28**, 675–691 (2005).
3. NBA/ABA career leaders and records for field goal percentage. *Basketball Reference* https://www.basketball-reference.com/leaders/fg_pct_career.html?sr&utm_source=direct&utm_medium=Share&utm_campaign=ShareTool#tot (2023, accessed 13 September 2024).
4. Chi, M. T., Siler, S. A., Jeong, H., Yamauchi, T. & Hausmann, R. G. Learning from human tutoring. *Cogn. Sci.* **25**, 471–533 (2001).
5. Frith, C. D. & Frith, U. The social brain: allowing humans to boldly go where no other species has been. *Philos. Trans. R. Soc. B* **361**, 2201–2214 (2006).
6. Saxe, R. Uniquely human social cognition. *Curr. Opin. Neurobiol.* **16**, 235–239 (2006).
7. Jara-Ettinger, J. Theory of mind as inverse reinforcement learning. *Curr. Opin. Behav. Sci.* **29**, 105–110 (2019).
8. Baker, C. L., Jara-Ettinger, J., Saxe, R. & Tenenbaum, J. B. Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nat. Hum. Behav.* **1**, 0064 (2017).
9. Jara-Ettinger, J., Gweon, H., Schulz, L. E. & Tenenbaum, J. B. The naïve utility calculus: computational principles underlying commonsense psychology. *TrendsCogn. Sci.* **20**, 589–604 (2016).
10. Jern, A. & Kemp, C. A decision network account of reasoning about other people's choices. *Cognition* **142**, 12–38 (2015).
11. Shafto, P. & Goodman, N. D. Teaching games: statistical sampling assumptions for learning in pedagogical situations. In *Proc. 30th Annual Conference of the Cognitive Science Society* 1632–1637 (Cognitive Science Society, 2008).
12. Goodman, N. D. & Frank, M. C. Pragmatic language interpretation as probabilistic inference. *TrendsCogn. Sci.* **20**, 818–829 (2016).
13. Chen, A. M., Palacci, A., Vélez, N., Hawkins, R. D. & Gershman, S. J. A hierarchical Bayesian model of adaptive teaching. *Cogn. Sci.* **48**, e13477 (2024).
14. Daw, N. D., Niv, Y. & Dayan, P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* **8**, 1704–1711 (2005).
15. Ho, M. K., Saxe, R. & Cushman, F. Planning with theory of mind. *Trends Cogn. Sci.* **26**, 959–971 (2022).
16. Shenhav, A., Botvinick, M. M. & Cohen, J. D. Expected value of control modulates cognitive control. *Nat. Neurosci.* **16**, 1227–1235 (2013).

17. Kool, W. & Botvinick, M. M. The cost of cognitive control: the influence of cognitive load on decision-making. *J. Exp. Psychol. Gen.* **146**, 1438–1447 (2017).
18. Gigerenzer, G., Todd, P. M. & the ABC Research Group. *Simple Heuristics That Make Us Smart* (Oxford Univ. Press, 1999).
19. Todd, P. M., Gigerenzer, G. & the ABC Research Group. *Ecological Rationality: Intelligence in the World* (Oxford Univ. Press, 2012).
20. Kirschner, P. A., Sweller, J. & Clark, R. E. Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educ. Psychol.* **41**, 75–86 (2006).
21. Taplin, J. E., Schwenk, A. & Shavinina, L. V. Heuristic and analytic processes in reasoning. *Think. Reason.* **10**, 433–455 (2004).
22. Wood, W. & Neal, D. T. A new look at habits and the habit–goal interface. *Psychol. Rev.* **114**, 843–863 (2007).
23. Heider, F. & Simmel, M. An experimental study of apparent behavior. *Am. J. Psychol.* **57**, 243–259 (1944).
24. Frank, M. C. & Tenenbaum, J. B. Three ideal observer models for rule learning in simple languages. *Cognition* **120**, 360–371 (2011).
25. Baker, C. L., Saxe, R. & Tenenbaum, J. B. Action understanding as inverse planning. *Cognition* **113**, 329–349 (2009).
26. Ho, M. K., Littman, M., MacGlashan, J., Cushman, F. & Austerweil, J. L. Showing versus doing: teaching by demonstration. In *30th Conference on Neural Information Processing Systems (NIPS 2016)* (eds Lee, D. D., Sugiyama, M., von Luxburg, U., Guyon, I. & Garnett, R.) 3035–3043 (Curran Associates, 2016).
27. Shafto, P., Goodman, N. D. & Griffiths, T. L. A rational account of pedagogical reasoning: teaching by, and learning from, examples. *Cogn. Psychol.* **71**, 55–89 (2014).
28. Vélez, N., Chen, A. M., Burke, T., Cushman, F. A. & Gershman, S. J. Teachers recruit mentalizing regions to represent learners' beliefs. *Proc. Natl Acad. Sci. USA* **120**, e2215015120 (2023).
29. Rafferty, A. N., LaMar, M. M. & Griffiths, T. L. Inferring learners' knowledge from their actions. *Cogn. Sci.* **39**, 584–618 (2015).
30. Rafferty, A. N., Brunskill, E., Griffiths, T. L. & Shafto, P. Faster teaching via pomdp planning. *Cogn. Sci.* **40**, 1290–1332 (2016).
31. Kool, W., Cushman, F. A. & Gershman, S. J. Competition and cooperation between multiple reinforcement learning systems. in *Goal-Directed Decision Making: Computations and Neural Circuits* (eds Morris, R., Bornstein, A. & Shenhav, A.) 153–178 (Academic Press, 2018).
32. Lieder, F. & Griffiths, T. L. Strategy selection as rational metareasoning. *Psychol. Rev.* **124**, 762–794 (2017).
33. Goodman, N. D., Baker, C. L. & Tenenbaum, J. B. in *Cause and Intent: Social Reasoning in Causal Learning* 2759–2764 (Cognitive Science Society Austin, 2009).
34. Lucas, C. G. et al. The child as econometrician: a rational model of preference understanding in children. *PLoS ONE* **9**, e92160 (2014).
35. Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction* (MIT Press, 2018).
36. Botvinick, M. & Weinstein, A. Model-based hierarchical reinforcement learning and human action control. *Philos. Trans. R. Soc. B* **369**, 20130480 (2014).
37. Dayan, P. & Niv, Y. Reinforcement learning: the good, the bad and the ugly. *Curr. Opin. Neurobiol.* **18**, 185–196 (2008).
38. Niv, Y. Reinforcement learning in the brain. *J. Math. Psychol.* **53**, 139–154 (2009).
39. Niv, Y. Learning task-state representations. *Nat. Neurosci.* **22**, 1544–1553 (2019).
40. Luce, R. D. *Individual Choice Behavior* Vol. 4 (Wiley, 1959).
41. Frith, C. & Frith, U. Theory of mind. *Curr. Biol.* **15**, R644–R645 (2005).
42. Gigerenzer, G. & Gaissmaier, W. Heuristic decision making. *Annu. Rev. Psychol.* **62**, 451–482 (2011).
43. Lieder, F. & Griffiths, T. L. Resource-rational analysis: understanding human cognition as the optimal use of limited computational resources. *Behav. Brain Sci.* **43**, e1 (2020).
44. Watkins, C. J. & Dayan, P. Q-learning. *Mach. Learn.* **8**, 279–292 (1992).
45. Piray, P., Dezfouli, A., Heskes, T., Frank, M. J. & Daw, N. D. Hierarchical Bayesian inference for concurrent model fitting and comparison for group studies. *PLoS Comput. Biol.* **15**, e1007043 (2019).
46. Harootonian, S. K., Niv, Y. & Ho, M. K. Mentalizing as a cognitively effortful teaching strategy. *OSF* https://osf.io/n6gq5 (2023).
47. Harootonian, S. K., Niv, Y. & Ho, M. K. Scaffolding strategies: overcoming theory of mindeffort and heuristic pitfalls in teaching. *OSF* https://osf.io/r2vw6 (2024).
48. Lieder, F. & Griffiths, T. L. When to use which heuristic: a rational solution to the strategy selection problem. In *Proc. 37th Annual Conference of the Cognitive Science Society* (eds Noelle, D. C., Dale, R., Warlaumont, A. S., Yoshimi, J., Matlock, T., Jennings, C. D. & Maglio, P. P.) 1362–1367 (Cognitive Science Society, 2015).
49. Griffiths, T. L., Lieder, F. & Goodman, N. D. Rational use of cognitive resources: levels of analysis between the computational and the algorithmic. *Top. Cogn. Sci.* **7**, 217–229 (2015).
50. Ho, M. K. et al. People construct simplified mental representations to plan. *Nature* **606**, 129–136 (2022).
51. Diamond, A. Executive functions. *Annu. Rev. Psychol.* **64**, 135–168 (2013).
52. Holton, E., Niv, Y. & O'Reilly, J. X. The adaptive value of stubborn goals. *Trends Cogn. Sci.* **30**, 379–390 (2026).
53. Ho, M. K., Cohen, J. D. & Griffiths, T. L. Rational simplification and rigidity in human planning. *Psychol. Sci.* **34**, 1281–1292 (2023).
54. Tsetsos, K., Chater, N. & Usher, M. Salience driven value integration explains decision biases and preference reversal. *Proc. Natl Acad. Sci. USA* **109**, 9659–9664 (2012).
55. Busemeyer, J. R. & Townsend, J. T. Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. *Psychol. Rev.* **100**, 432 (1993).
56. Dasgupta, I., Schulz, E., Goodman, N. D. & Gershman, S. J. Remembrance of inferences past: amortization in human hypothesis generation. *Cognition* **178**, 67–81 (2018).
57. Kool, W., McGuire, J. T., Rosen, Z. B. & Botvinick, M. M. Decision making and the avoidance of cognitive demand. *J. Exp. Psychol. Gen.* **139**, 665 (2010).
58. Ho, M. K., Abel, D., Griffiths, T. L. & Littman, M. L. The value of abstraction. *Curr. Opin. Behav. Sci.* **29**, 111–116 (2019).
59. Hatano, G. & Inagaki, K. Two courses of expertise. *Annu. Rep. Res. Center Child Dev.* **6**, 27–36 (1984).
60. Stigler, J. W. *The TIMSS Videotape Classroom Study: Methods and Findings from an Exploratory Research Project on Eighth-Grade Mathematics Instruction in Germany, Japan, and the United States* (US Department of Education, Office of Educational Research and Improvement, 1999).
61. Stigler, J. W. & Hiebert, J. Improving mathematics teaching. *Educ. Leadersh.* **61**, 12–17 (2004).
62. Stigler, J. W. & Miller, K. F. Expertise and expert performance in teaching. in *The Cambridge Handbook of Expertise and Expert Performance* 2nd edn (eds Ericsson, K. A., Hoffman, R. R., Kozbelt, A. & Williams, A. M.) 431–452 (Cambridge Univ. Press, 2018).
63. Sorgenfrei, M. et al. *Classroom Wellbeing Toolkit: Simple Ways to Support Secondary Students' Mental Health* (Anna Freud Centre and Early Intervention Foundation, 2022); https://www.AnnaFreud.org/classroom-wellbeing-toolkit
64. Schäfer, J. Ö, Naumann, E., Holmes, E. A., Tuschen-Caffier, B. & Samson, A. C. Emotion regulation strategies in depressive and anxiety symptoms in youth: a meta-analytic review. *J. Youth Adolesc.* **46**, 261–276 (2017).

65. Twemlow, S. W., Fonagy, P. & Sacco, F. C. A developmental approach to mentalizing communities: II. the peaceful schools experiment. *Bull. Menninger Clin.* **69**, 282–304 (2005).
66. Twemlow, S. W., Fonagy, P., Campbell, C. & Sacco, F. C. Creating a peaceful school learning environment: attachment and mentalization efforts to promote creative learning in kindergarten through fifth-grade elementary school students with broad extension to all grades and some organizations. in *Handbook of Attachment-Based Interventions* (eds Steele, H. & Steele, M.) 431–452 (Guilford Press, 2017).
67. Okonofua, J. A., Paunesku, D. & Walton, G. M. Brief intervention to encourage empathic discipline cuts suspension rates in half among adolescents. *Proc. Natl Acad. Sci. USA* **113**, 5221–5226 (2016).
68. Nišponská, M. Mentalization and its use in schools. *Pedagogika* **72**, 535–552 (2022).
69. Chelouche-Dwek, G. & Fonagy, P. Mentalization-based interventions in schools for enhancing socio-emotional competencies and positive behaviour: a systematic review. *Eur. Child Adolesc. Psychiatry* **34**, 1295–1315 (2025).
70. Schwarzer, N.-H. et al. The relationship between global distress, mentalizing and well-being in a german teacher sample. *Curr. Psychol.* **42**, 1239–1248 (2023).
71. Niven, K. The four key characteristics of interpersonal emotion regulation. *Curr. Opin. Psychol.* **17**, 89–93 (2017).
72. Chen, T., Houlihan, S. D., Chandra, K., Tenenbaum, J. B. & Saxe, R. Intervening on emotions by planning over a theory of mind. In *Proc. 46th Annual Meeting of the Cognitive Science Society* (eds Samuelson, L. K., Frank, S. L., Toneva, M., Mackey, A. & Hazeltine, E.) 5887–5894 (Cognitive Science Society, 2024).
73. Apperly, I. A. & Butterfill, S. A. Do humans have two systems to track beliefs and belief-like states?. *Psychol. Rev.* **116**, 953 (2009).
74. Apperly, I. *Mindreaders: The Cognitive Basis of "Theory of Mind"* (Psychology Press, 2010).
75. Keysar, B., Barr, D. J., Balin, J. A. & Brauner, J. S. Taking perspective in conversation: the role of mutual knowledge in comprehension. *Psychol. Sci.* **11**, 32–38 (2000).
76. Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J. & Bodley Scott, S. E. Seeing it their way: evidence for rapid and involuntary computation of what other people see. *J. Exp. Psychol. Hum. Percept. Perform.* **36**, 1255 (2010).
77. Hawkins, R. D., Gweon, H. & Goodman, N. D. The division of labor in communication: speakers help listeners account for asymmetries in visual perspective. *Cogn. Sci.* **45**, e12926 (2021).
78. Pöppel, J. & Kopp, S. Satisficing models of Bayesian theory of mind for explaining behavior of differently uncertain agents. In *Proc. 17th International Conference on Autonomous Agents and MultiAgent Systems* 470–478 (IFAAMAS, 2018).
79. Bass, I., Espinoza, C., Bonawitz, E. & Ullman, T. D. Teaching without thinking: negative evaluations of rote pedagogy. *Cogn. Sci.* **48**, e13470 (2024).
80. Jara-Ettinger, J. & Dunham, Y. The institutional stance. *Behav. Brain Sci.* https://doi.org/10.1017/S0140525X25103427 (2025).
81. Levine, S., Chater, N., Tenenbaum, J. B. & Cushman, F. Resource-rational contractualism: a triple theory of moral cognition. *Behav. Brain Sci.* **47**, e188 (2024).
82. Parpart, P., Jones, M. & Love, B. C. Heuristics as bayesian inference under extreme priors. *Cogn. Psychol.* **102**, 127–144 (2018).
83. De Leeuw, J. R. jspsych: a javascript library for creating behavioral experiments in a web browser. *Behav. Res. Methods* **47**, 1–12 (2015).
84. Zorowitz, S. & Bennett, D. Nivturk. *Zenodo* https://doi.org/10.5281/zenodo.6609218 (2022).
85. AnyChart javascript charting library. *AnyChart* https://www.anychart.com/ (2023).
86. Ho, M. K., Correa, C. G. & Ritter, D. Models of sequential decision making (msdm) *GitHub* https://github.com/markkho/msdm (2021).
87. Kass, R. E. & Raftery, A. E. Bayes factors. *J. Am. Stat. Assoc.* **90**, 773–795 (1995).

Acknowledgements

We thank N. Klevak for helping coding the pilot experiment. We thank the Niv lab as a whole for helpful discussions.

Author contributions

S.K.H. conceptualized the study, designed the experiments, developed the computational models, collected and analysed data, created visualizations, wrote the original draft and subsequent edits and revisions. T.L.G. provided supervision and reviewed and edited the manuscript. Y.N. provided supervision, acquired funding, contributed to methodology, reviewed and edited the manuscript. M.K.H. provided supervision, contributed to methodology, reviewed and edited the manuscript. All authors reviewed and approved the final manuscript.

Funding

Y.N. discloses support for the research of this work from the Simons Foundation International (record ID: SFI-PD-Pivot Mentor-00009521). S.K.H. discloses support for the research of this work from the National Science Foundation Graduate Research Fellowship Program (grant number DGE-2444107). T.L.G. and M.K.H. declare no relevant funding.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41562-026-02540-2.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-026-02540-2.

Correspondence and requests for materials should be addressed to Sevan K. Harootonian.

Peer review information *Nature Human Behaviour* thanks Ali Mahmoodi and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

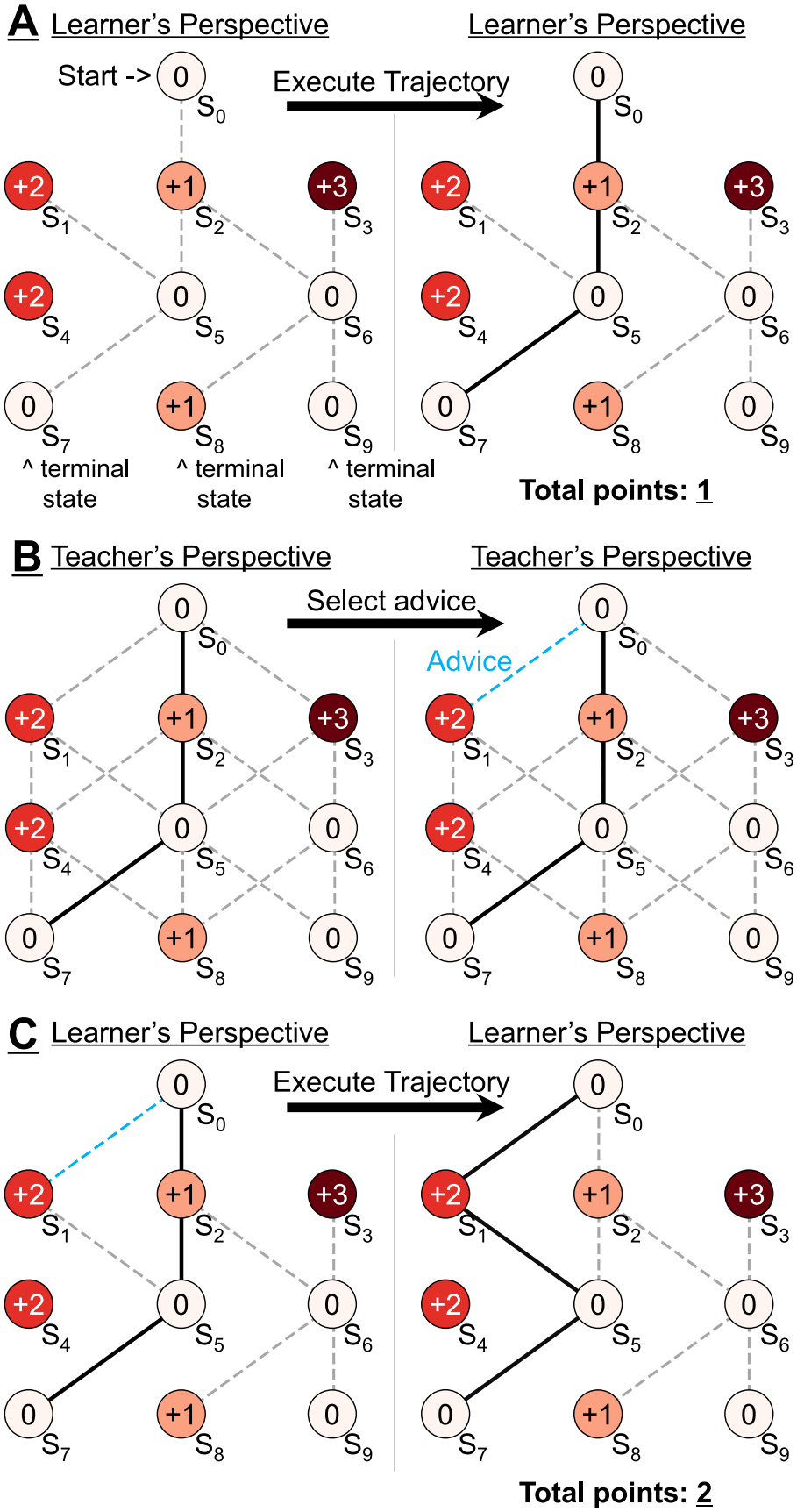
Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2026

https://doi.org/10.1038/s41562-026-02540-2

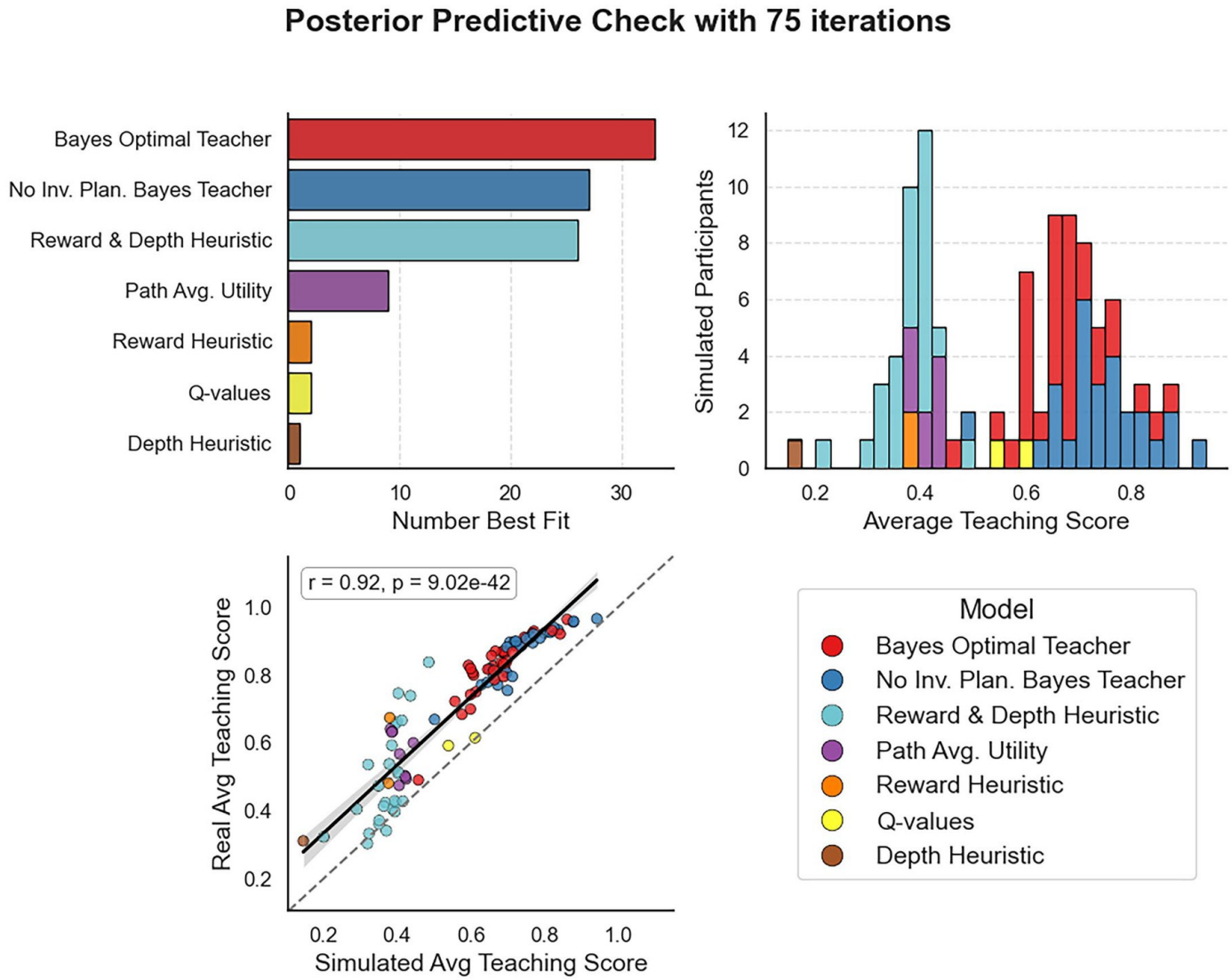


Extended Data Fig. 1 | See next page for caption.

https://doi.org/10.1038/s41562-026-02540-2

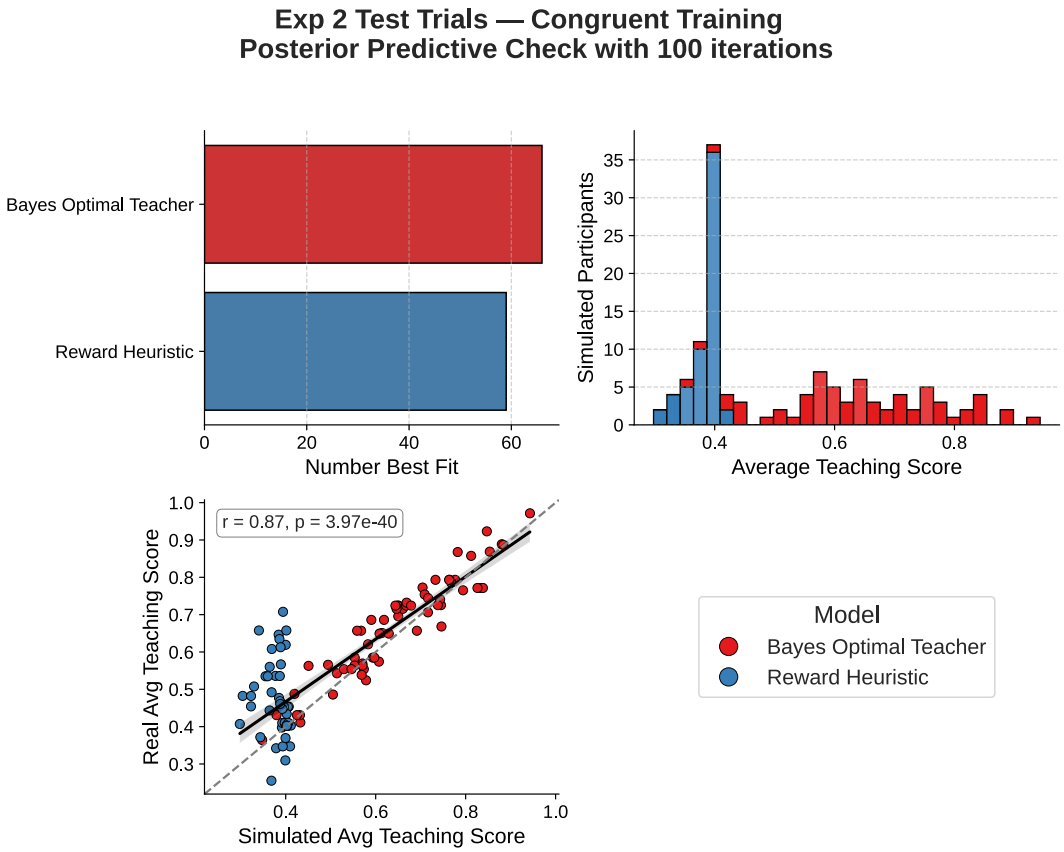
Extended Data Fig. 1 | Mechanics of the Graph Teaching Task. Graph Teaching Task: The circles represent states/locations (example: S5) and the positive value inside the circle represents the reward obtained when that state is entered. The dotted lines represent possible deterministic transitions by the learner and the solid line represents the learner's executed trajectory. The total points are the sum of all the rewards the learner collected along their executed "top-to-bottom" trajectory. **A**) Learner computes and executes the best available trajectory, **B**) Teacher observes the learner's trajectory and selects an edge to teach. **C**) Learner incorporates the taught edge and computes and executes a new trajectory. Note: participants only see part **B** and make assumptions for parts **A** and **C**.

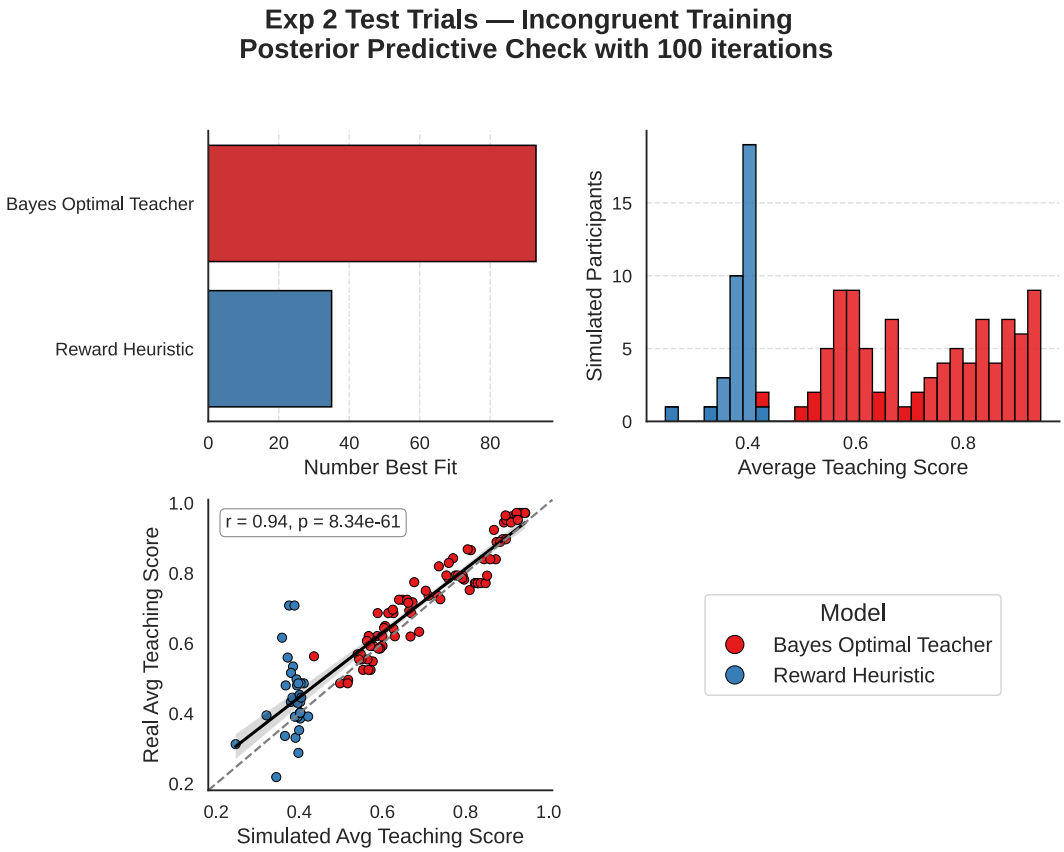
https://doi.org/10.1038/s41562-026-02540-2



Extended Data Fig. 2 | Posterior Predictive Check for Experiment 1.

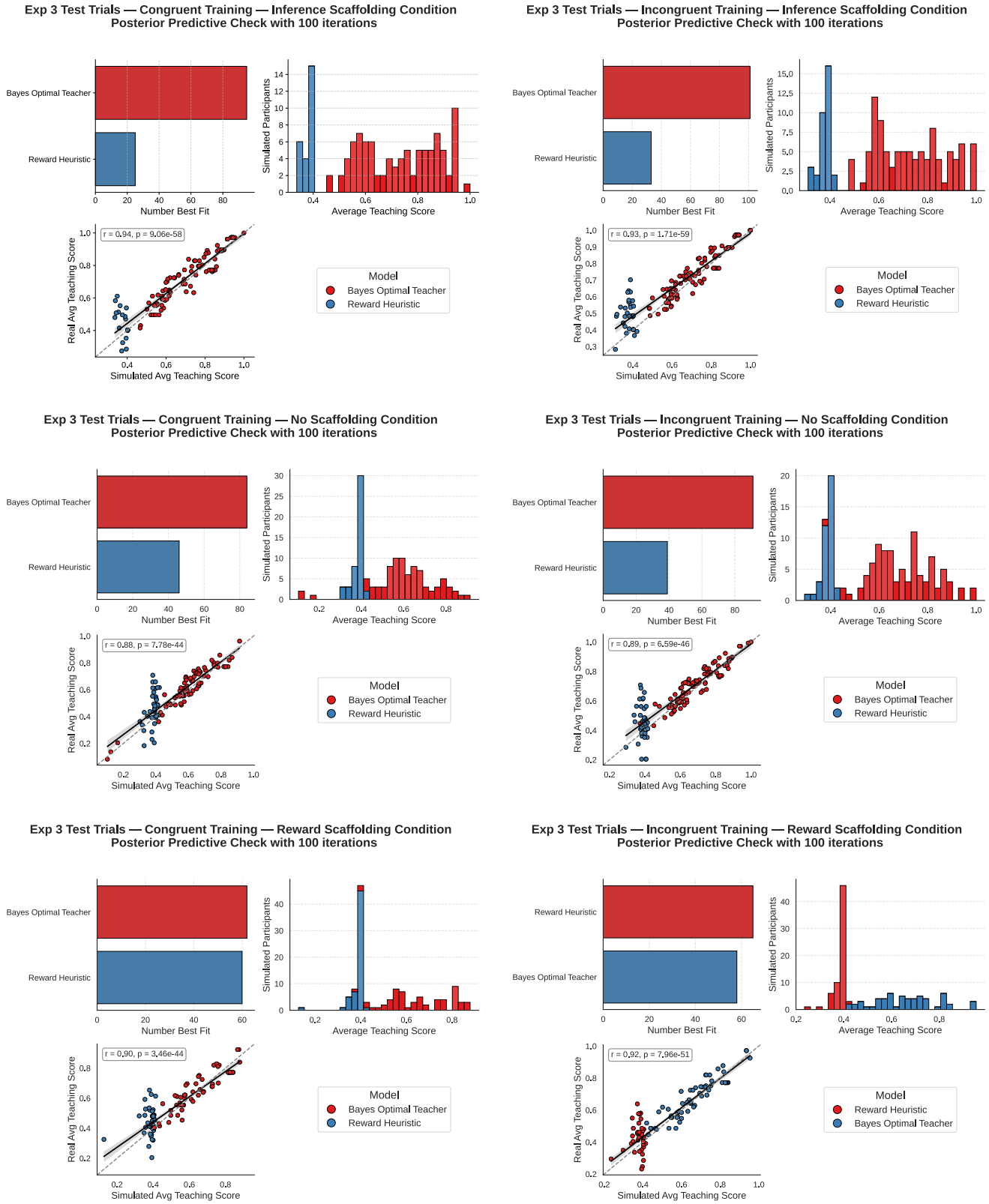
https://doi.org/10.1038/s41562-026-02540-2





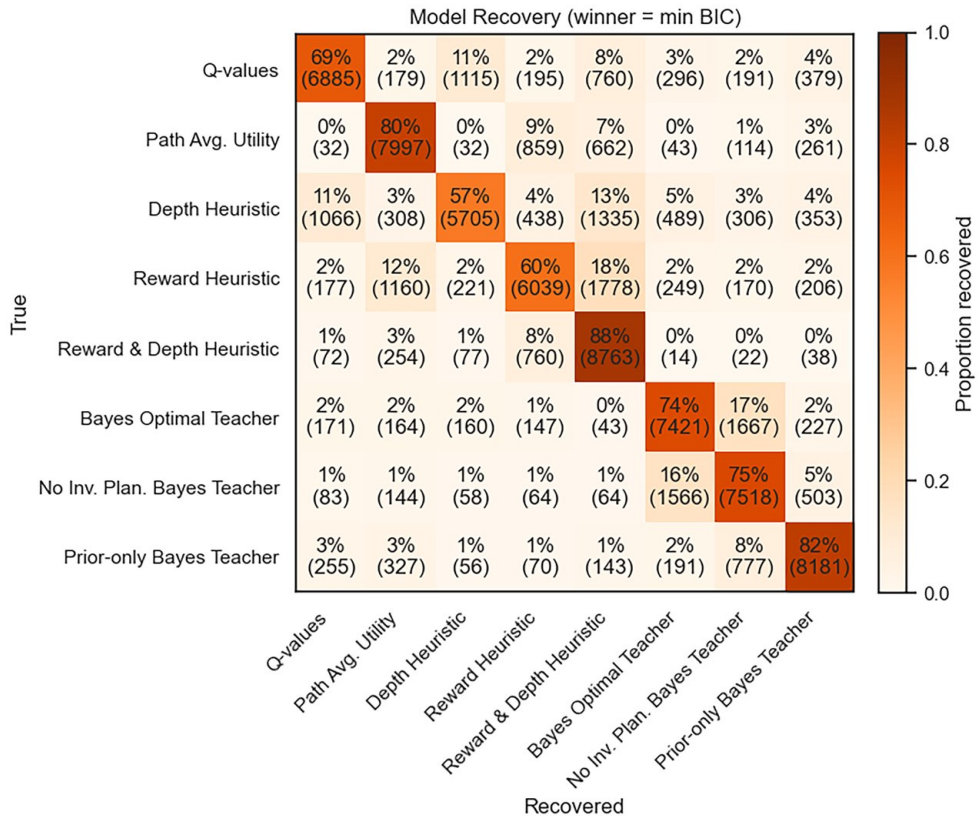
Extended Data Fig. 3 | Posterior Predictive Check for Experiment 2.

https://doi.org/10.1038/s41562-026-02540-2



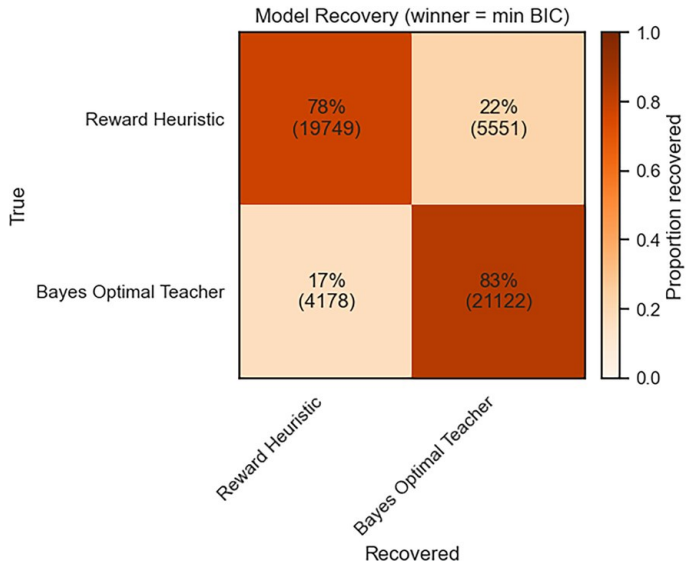
Extended Data Fig. 4 | Posterior Predictive Check for Experiment 3.

https://doi.org/10.1038/s41562-026-02540-2



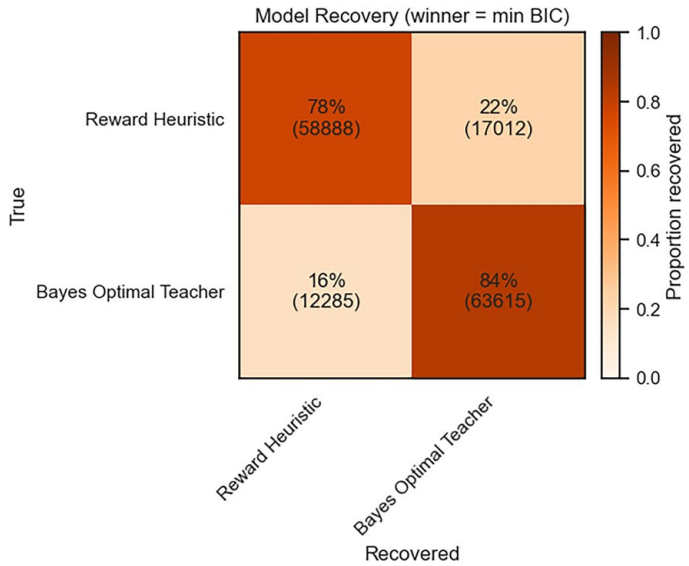
Extended Data Fig. 5 | Model Recovery for Experiment 1.

https://doi.org/10.1038/s41562-026-02540-2



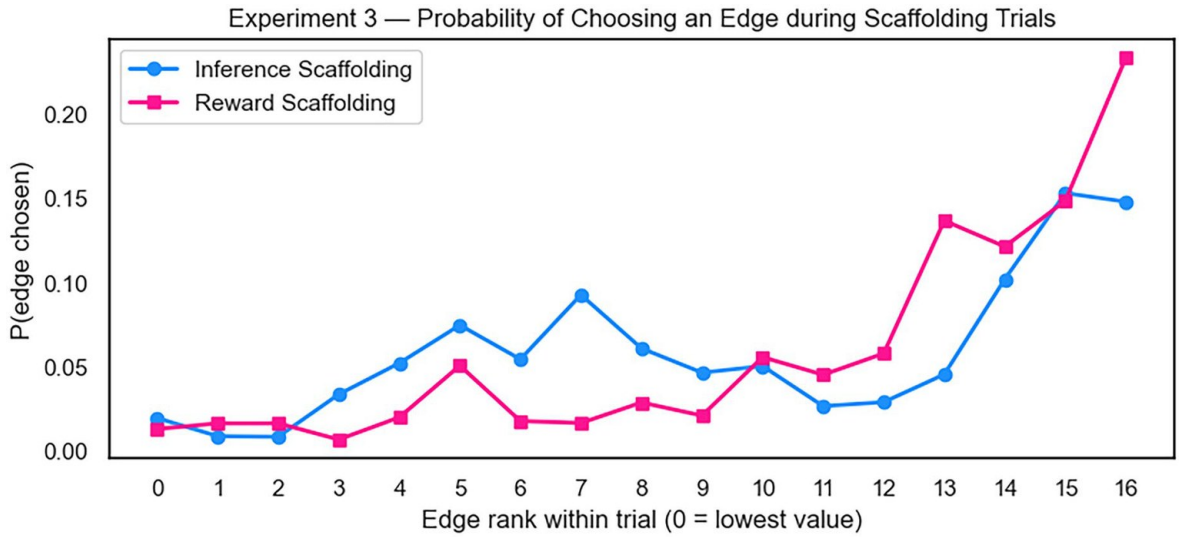
Extended Data Fig. 6 | Model Recovery for Experiment 2.

https://doi.org/10.1038/s41562-026-02540-2



Extended Data Fig. 7 | Model Recovery for Experiment 3.

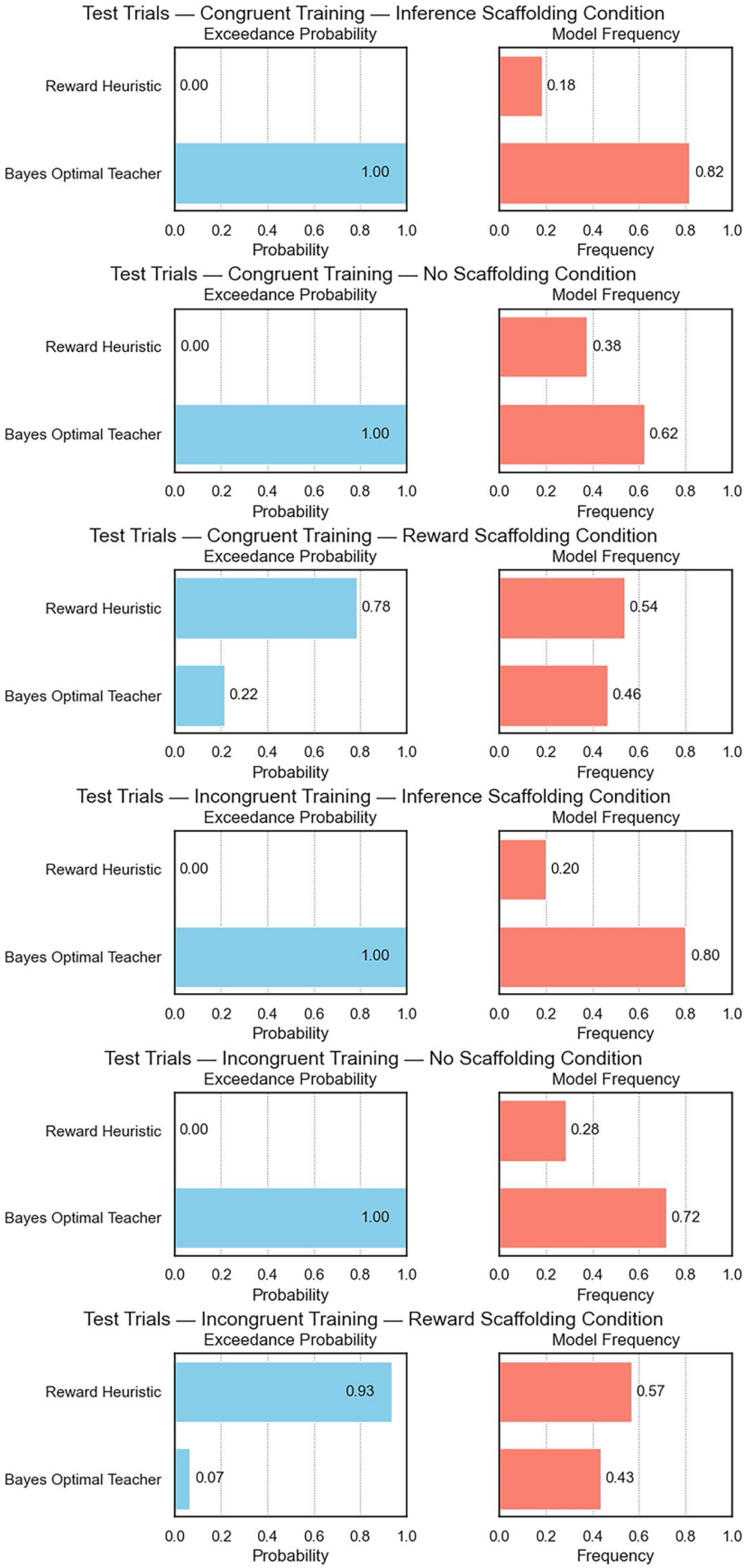
https://doi.org/10.1038/s41562-026-02540-2



Ranks are defined **within each condition**:
• Inference Scaffolding: edges ranked by increasing probability the learner doesn't know them (Bayes Optimal Teacher).
• Reward Scaffolding: edges ranked by increasing reward feature value.
Numeric ranks therefore have different meanings across the two lines.

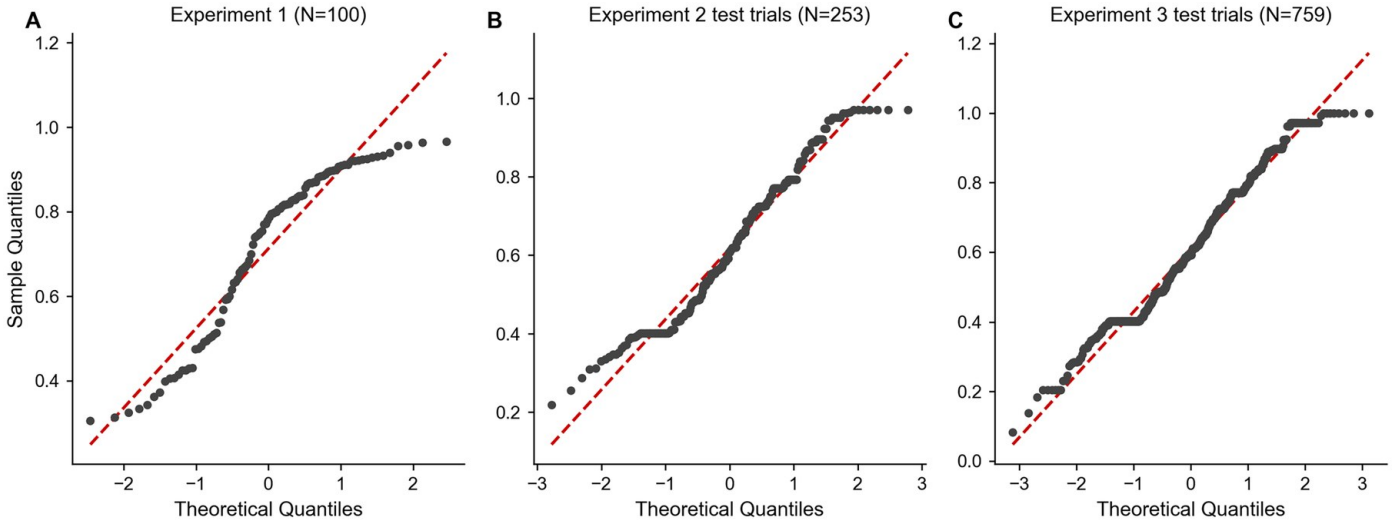
Extended Data Fig. 8 | Participants' Responses During Inference and Reward-Scaffolding Trials. Correlation between participants' responses and model predictions in the scaffolding task. Blue shows probabilities of choosing edges in the inference scaffolding condition, with edges ordered according to the predicted probability that an edge is unknown to the learner according to the Bayes-Optimal Teacher (0 - most likely known, 16 - most likely unknown). Pink shows probabilities of choosing edges in the reward scaffolding condition, with edges ordered according to the value of each edge as a function of the sum of the two nodes it connects, as defined by the Reward Heuristic (0 - lowest value, 16 - highest value; note that some edges are tied, with different ties for each graph shown). In both cases, a higher x-axis value corresponds to edges more consistent with the model's prediction for this condition. The order differs between the Bayes-Optimal Teacher and the Reward Heuristic. Because there were three choices on each trial, the y-axis shows the probability that an edge was one of the three selected (with a maximum value of 33.3%). This was computed by normalizing the proportion of responses, averaged over all trials and subjects. The figure shows that most responses aligned with the model predictions, indicating that participants performed the scaffolding task as intended.

https://doi.org/10.1038/s41562-026-02540-2



Extended Data Fig. 9 | Random-Effect Model Comparison for each Congruency group and Scaffolding condition.

https://doi.org/10.1038/s41562-026-02540-2



Extended Data Fig. 10 | Normal Q-Q plots of per-participant mean teaching scores confirm non-normality across all three experiments. Each point represents one participant's mean teaching score (**A**: Experiment 1, mean across all 40 trials, N = 100; **B**: Experiment 2, mean across 5 test trials, N = 253; **C**: Experiment 3, mean across 5 test trials, N = 759). The red dashed line shows the expected relationship under normality. Departures from this line are evident in all panels, reflecting the bounded [0, 1] nature of teaching scores and their non-normal distributions. Shapiro-Wilk tests con- firmed significant departures from normality in all three experiments (Experiment 1: $W = 0.906$, $p = 2.62 \times 10\text{-}6$; Experiment 2: $W = 0.965$, $p = 7.39 \times 10\text{-}6$; Experiment 3: $W = 0.980$, $p = 9.31 \times 10\text{-}9$), supporting the use of Mann-Whitney U tests in place of independent-samples t-tests.