



This article is part of the topic “Computational Approaches to Social Cognition,” Samuel Gershman and Fiery Cushman (Topic Editors). For a full listing of topic papers, see [http://onlinelibrary.wiley.com/journal/10.1111/\(ISSN\)1756-8765/earlyview](http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1756-8765/earlyview)

Integrating Incomplete Information With Imperfect Advice

Natalia Vélez, Hyowon Gweon

Department of Psychology, Stanford University

Received 16 September 2017; received in revised form 2 September 2018; accepted 14 September 2018

Abstract

When our own knowledge is limited, we often turn to others for information. However, social learning does not guarantee accurate learning or better decisions: Other people’s knowledge can be as limited as our own, and their advice is not always helpful. This study examines how human learners put two “imperfect” heads together to make utility-maximizing decisions. Participants played a card game where they chose to “stay” with a card of known value or “switch” to an unknown card, given an advisor’s advice to stay or switch. Participants used advice strategically based on which cards the advisor could see (Experiment 1), how helpful the advisor was (Experiment 2), and what strategy the advisor used to select advice (Experiment 3). Overall, participants benefited even from imperfect advice based on incomplete information. Participants’ responses were consistent with a Bayesian model that jointly infers how the advisor selects advice and the value of the advisor’s card, compared to an alternative model that weights advice based on the advisor’s accuracy. By reasoning about others’ minds, human learners can make the best of even noisy, impoverished social information.

Keywords: Social learning; Theory of mind; Decision making; Bayesian inference; Computational models

1. Introduction

Our knowledge of the world is neither perfect nor complete. When our own knowledge is insufficient to make a decision, we often turn to others’ advice to guide our choices

Correspondence should be sent to Natalia Vélez and Hyowon Gweon, Department of Psychology, Stanford University, 450 Serra Mall, Jordan Hall, Stanford, CA 94305. E-mails: nvelez@stanford.edu, gweon@stanford.edu

(Boyd, Richerson, & Henrich, 2011; Tomasello, Kruger, & Ratner, 1993). *Social learning*, or learning from others, can help us learn or make better decisions than if we had relied solely on our own knowledge. However, information from others is not perfect; other people's knowledge is often as limited as our own, and the information they provide is sometimes misleading. The current work examines how human learners make utility-maximizing decisions by "putting two heads together," combining their own incomplete knowledge with imperfect advice from others.

Imagine you are standing in line at a busy grocery store, and your friend sends a text message suggesting that you switch to her line. Should you switch or stay put? The difficulty with this decision is that neither you nor your friend has full knowledge of the world; she can't see your line, and you can't see hers. However, you can benefit even from this imperfect advice if you know something about your friend. If she wants to help you find a shorter line, she would only recommend you to switch if her line wasn't too long. Thus, based on your friend's intention and her advice to switch, you can reasonably guess the length of her line and compare it with your current line to make a decision.

This example suggests that, by inferring aspects of the world that are only accessible to others' minds, learners can make more informed decisions than if they relied solely on their own knowledge. Recent computational models have formalized reasoning about others' minds as "inverse planning" (Baker, Jara-Ettinger, Saxe, & Tenenbaum, 2017; Baker, Saxe, & Tenenbaum, 2009). In this framework, agents have a generative model of how others' mental states give rise to their observable actions, and they work backwards from observed actions to infer the unobservable mental states that gave rise to those actions. Here, we propose that this ability to infer other people's mental states—*Theory of Mind*—enables humans to learn even from imperfect social information.

Prior developmental and computational work has emphasized the role of mental state inference in social learning. Even young children actively interpret social information in light of others' knowledge and intention. When a "teacher" pedagogically demonstrates one function of a novel toy, children not only learn the demonstrated function, but also infer it is the toy's *only* function; if the toy had additional functions, a knowledgeable, helpful teacher would have also demonstrated them (Bonawitz et al., 2011). These inferences are consistent with a computational model that assumes that teachers and learners reason about each other's mental states; the teacher selects evidence that maximizes the learner's belief in the correct hypothesis, and the learner draws inferences based on the assumption that the evidence has been selected by a knowledgeable, helpful teacher (Shafto, Goodman, & Griffiths, 2014).

Now, suppose the child in the example above discovers that the toy actually has four functions (rather than one). There are many reasons why the teacher only showed one of its functions; perhaps she was ignorant of these functions, or perhaps she deliberately omitted them to mislead the learner. More recent work has extended models of pedagogical reasoning to formalize how learners use their own knowledge to evaluate informants' knowledge, confidence, and reliability (Bass, Hawthorne, Goodman, & Gweon, 2015; Hawthorne-Madell & Goodman, 2017). Consistent with the predictions of these models, adults and even preschool-aged children appropriately penalized the informant when she

knowingly omitted information (Gweon & Asaba, 2017; Gweon, Pelton, Konopka, & Schulz, 2014) but exonerated her when she was unaware of the additional functions (Bass, Bonawitz, & Gweon, 2017; Bass et al., 2015).

Taken together, prior work suggests that learners make rich inferences about both the world and the informant's mental states. In particular, when learners expect the informant to be knowledgeable and helpful, they rely on social information to draw further inferences about the world; when learners themselves are knowledgeable, they capitalize on their knowledge to infer the informant's intentions or trustworthiness. Critically, however, social learning often unfolds in contexts where neither the learner nor the informant has full access to the true state of the world, as in the grocery store example above. In this case, learners cannot simply trust informants to be knowledgeable and helpful, nor do they have enough world knowledge to directly evaluate their quality. How do learners leverage imperfect advice from others to make better decisions?

The present work explores how learners put two "imperfect" heads together to make utility-maximizing decisions. Participants played a simple card game analogous to the grocery line example above. In the next section, we describe the task and present two computational models that formalize competing hypotheses about how participants integrate their partial knowledge of the world with advice from an advisor. Then, across three experiments, we compare the predictions of the two models with human behavior.¹

2. Computational models

2.1. Task: Switch or stay?

In all experiments, the task had the same basic structure (Fig. 1a). On each trial, two cards were randomly drawn, with different integer values between +1 and +8; one card was visible to participants (*visible card*), the other hidden (*hidden card*). Participants could choose to *stay* and win the points on the visible card, or *switch* and win the points on the hidden card. The goal was to win the most points possible by the end of the task.

Before making this decision, participants received advice from an advisor who saw some subset of the cards and recommended to "stay" or "switch." In most conditions, the advisor could only see the hidden card (except in Experiment 1, Both and None conditions; see Procedure). The advisor recommended to stay or switch by sampling from the following choice function:

$$P(A = \text{Switch}|C_H) = \frac{1}{1 + e^{-(m_A \times (C_H - C_{med}) + b_A)}} \quad (1)$$

where A is the advice ("switch" or "stay"), C_H is the value of the hidden card (which is visible to the advisor), and m_A (steepness) and b_A (horizontal shift) determine the shape of the advisor's choice function. We derived the choice function from a pilot experiment where human participants played the same card game without an advisor; we assumed

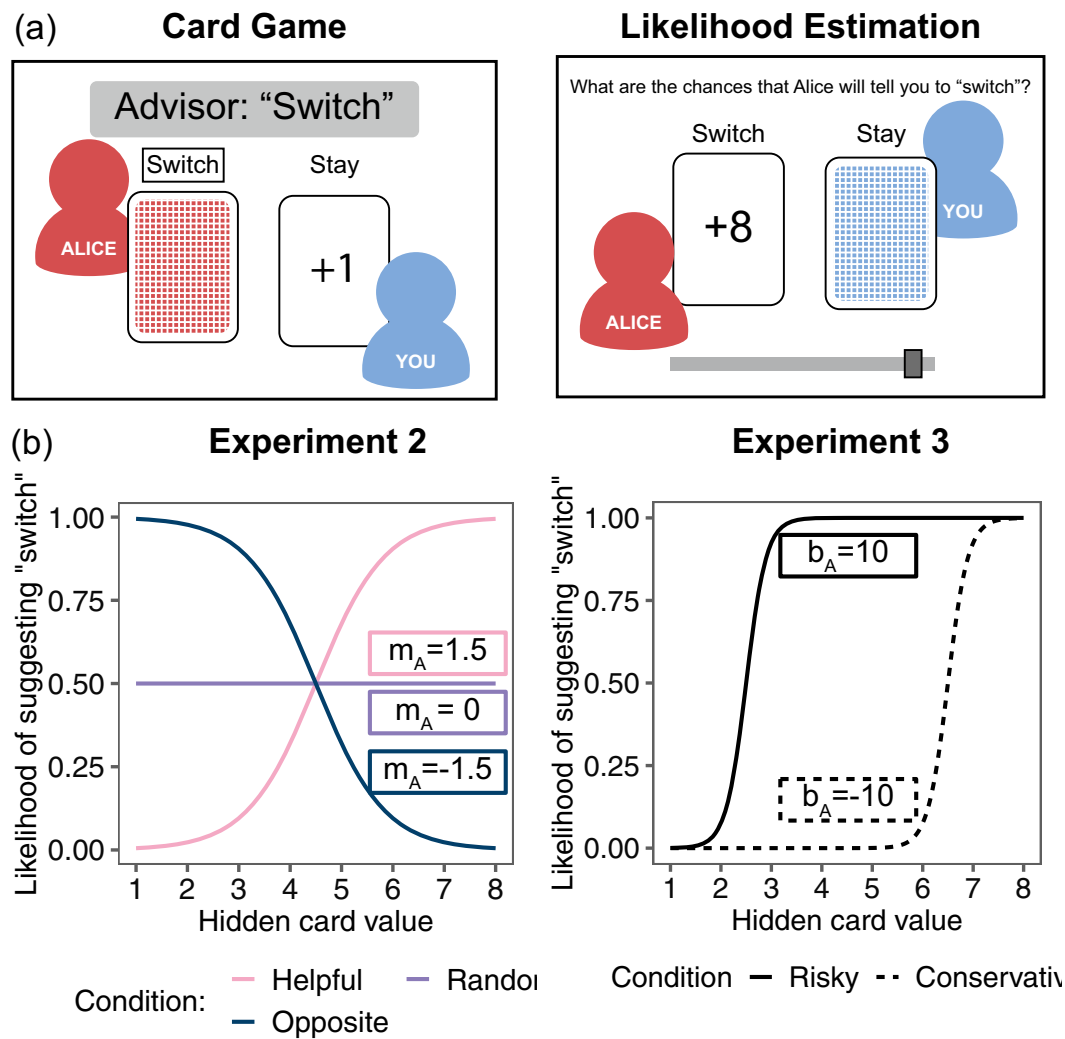


Fig. 1. Procedure. (a) Participants' view of the card game (left) and of the likelihood estimation task (right). (b) Ground-truth advisor choice functions. *Left*: In Experiment 2, advisors varied in their helpfulness (Helpful (pink): $m_A = 1.5$, Random (purple): $m_A = 0$, Opposite (blue): $m_A = -1.5$; all $b_A = 0$). *Right*: In Experiment 3, both advisors were helpful (all $m_A = 5$) but varied in their strategy (Conservative (dashed): $b_A = -10$, Risky (solid) $b_A = 10$).

that a partially knowledgeable yet helpful advisor would make the same choices that a human participant would make to maximize their own payoff. Participants' decisions were well described by a logistic function centered around the median card value ($C_{med} = 4.5$; Supporting Information, *Pilot experiment*). Across experiments, we manipulated the true values of m_A and b_A to change how the advisor selects advice (Fig. 1b).

We compared participants' responses in this task to the predictions of two computational models. Both models make decisions based on the value of the visible card, the

advisor's access to information, and the advisor's advice. When the advisor can see no or both cards (Experiment 1, None and Both), both models assume that the advice is random or always accurate, respectively. Critically, each model uses advice differently when the advisor can only see the hidden card (Experiments 1–3, all other conditions). The *Mental-state Reasoning* model instantiates inverse inference about the advisor's mental states; it infers likely values of the hidden card and compares them to the visible card. We contrast this model with the *Accuracy Heuristic* model, which instantiates a simpler social learning strategy; it weights advice based on its observations of the advisor's accuracy.

2.2. Mental-state reasoning model

Consistent with prior Bayesian Theory of Mind models (e.g., Baker et al., 2017; Hawthorne-Madell & Goodman, 2017), the Mental-state Reasoning model works backwards from the advisor's observable advice to infer the unobservable states that gave rise to the advice—namely, how the advisor selects advice (i.e., the advisor's choice function) and what the advisor can see (i.e., her beliefs about the value of the hidden card). The Mental-state Reasoning model assumes that the advisor recommends to stay or switch based on a choice function that depends on the advisor's access to the card values. When the advisor can only see the hidden card, the model infers the value of the hidden card on each trial (C_H) given the advice (A) and the visible card value (C_V):

$$P(C_H|A, C_V) \propto P(C_H|C_V)P(A|C_H) \quad (2)$$

where $P(C_H|C_V)$, a discrete distribution that is 0 for $C_H = C_V$ and uniform everywhere else, represents the learner's prior belief about the value of the hidden card, and $P(A|C_H)$ (Eq. 1) represents the likelihood of the observed advice based on the advisor's belief about the hidden card value. The model jointly infers the parameters of the advisor's choice function (m_A and b_A) and the value of the hidden card (C_H).

Based on its beliefs about the value of the hidden card ($P(C_H|A, C_V)$), the model then computes an expected value ($E[C_H|A, C_V]$). Finally, the model selects an action (L) by comparing the value of the visible card (C_V) to the expected value of the hidden card using a softmax choice rule:

$$P(L = \text{Switch}|A, C_V) = \frac{1}{1 + e^{-m_L \times (E[C_H|A, C_V] - C_V)}} \quad (3)$$

where m_L is a free parameter that modulates how much the learner's choices are influenced by expected rewards.

2.3. Alternative model: Accuracy heuristic

Our alternative model is inspired by a body of work suggesting that learners tend to follow advice from reliable people (Boyd et al., 2011; Harris, 2012; Henrich, 2015; see

Sobel & Kushnir, 2013 for a review of related developmental studies). While the Mental-state Reasoning model represents the full distribution of possible values of the hidden card based on the inferred choice function of the advisor, the Accuracy Heuristic model uses a simpler approximation. This model assumes that the value of the hidden card is high when the advisor suggests to switch ($R_A = 8$) and low when the advisor suggests to stay ($R_A = 1$; for a similar approach, see Biele, Rieskamp, & Gonzalez, 2009; Toelch, Bach, & Dolan, 2013). This value is weighted by an estimate of the advisor's accuracy ($\text{acc} \in (0, 1)$, where 1 indicates that the advisor is perfectly accurate). The weight parameter $w_{\text{acc}} = (\text{acc} - 0.5) \times 2$ transforms acc into a weight between -1 and 1 ; for example, when the model believes the advice is random ($\text{acc} = 0.5$), it is not influenced by the advice ($w_{\text{acc}} = 0$). Finally, the model selects an action by combining the weighted influence of the advice to the value of the visible card (C_V):

$$P(L = \text{Switch} | C_V, A) = \frac{1}{1 + e^{-m_L \times (w_{\text{acc}} \times (R_A - C_{\text{med}}) - (C_V - C_{\text{med}}))}}$$

2.4. Model fitting

We inferred our model parameters using a hierarchical Bayesian model implemented in PyMC3 (Salvatier, Wiecki, & Fonnesebeck, 2016). All models sampled parameter values for individual participants from a group-level distribution (see Fig. S5 for the inferred posterior distribution of these parameters). For the Mental-state Reasoning model, m_A and b_A were each sampled from a Gaussian distribution ($m_A \sim N(\mu_m, \sigma_m)$, $b_A \sim N(\mu_b, \sigma_b)$); for the Accuracy Heuristic model, acc was sampled from a Beta distribution ($\text{acc} \sim \text{Beta}(\alpha_{\text{acc}}, \beta_{\text{acc}})$). Both models sampled the free parameter m_L from a Gamma distribution ($L \sim \text{Gamma}(k, \theta)$; see Gershman, 2016).² We put uninformative (i.e., uniform) priors over the group-level parameters, with a range consistent with the true values of m_A and b_A within our task ($\mu_m, \mu_b \sim \text{Uniform}(-20, 20)$; $\sigma_m, \sigma_b, k, \theta \sim \text{Uniform}(0.5, 40)$). Parameter values for both models were conditioned on participants' choices, and the value of the hidden card on trials where they chose to switch; for the Accuracy Heuristic model, we transformed the observed value of the hidden card into a binary value indicating whether the advisor's advice was correct.

We estimated the posterior distribution over model parameters by collecting 750,000 samples over 3 MCMC chains using the Metropolis-Hastings algorithm. We generated predicted data (posterior predictives) by iterating over 10,000 samples from the posterior distribution over the group-level parameters. In each iteration, we sampled as many new values of the model parameters from the group-level distributions (e.g., m_A , b_A , and m_L for Mental-state Reasoning) as there were participants in our actual sample. These parameters were used to generate predicted responses (i.e., by entering these new values into Eqs. 3 and 4) which were then aggregated into an average predicted rate of switching. In the results below, we report model fit as the correlation between the highest posterior density estimate from the posterior predictive distribution and average participant

responses. The better our models describe the data, the more our predicted data should resemble the data we observed.

3. Experiment 1

In Experiment 1, we asked whether participants use advice differently, based on the advisor's access to information about the card values. Across conditions, we varied whether the advisor could see both cards, no cards, or only the hidden card.

3.1. Methods

3.1.1. Participants

Fifty participants participated in an online study for pay on Amazon Mechanical Turk. Participants were paid \$2 for their time, plus a performance bonus of up to \$2 proportional to their final score. All participants in this and subsequent experiments had US IP addresses and provided informed consent in accordance with the requirements of the institutional review board.

3.1.2. Procedure

Participants played an online card game presented using jsPsych (De Leeuw, 2015). In Experiment 1, participants were explicitly told that the advisor is always trying to be helpful (Supporting Information). Across three within-participant conditions, we manipulated the advisor's access to information about the card values. In the None condition, the advisor could not see any cards and thus gave random advice; in the Both condition, the advisor saw both cards and deterministically recommended the best action. In our critical condition, the Hidden condition, the advisor saw only the hidden card and gave advice generated using Eq. 1 ($m_A = 1.5$, $b_A = 0$). This advisor is more likely to recommend to switch as the hidden card value increases; however, because their recommendation is based on just the hidden card, blindly following the advice does not guarantee a higher payoff.

Each trial consisted of three parts (Fig. S2): participants first saw a reminder of the advisor's information access and her advice (1.5 s); next, participants saw the value of the visible card and decided whether to stay or switch (self-paced); finally, participants received feedback on how many points they had earned (1 s).

Participants played two blocks of 28 trials in each condition (6 blocks, 168 trials total); the values of the visible and hidden cards were counterbalanced so that participants were shown all possible combinations of visible and hidden cards in each condition.

To ensure that participants accurately tracked the advisor's information access, at the beginning of each block, participants were asked to indicate which cards the advisor could see ($M(SD)$ accuracy = 98(0.95)%). Because Experiment 1 addressed *how* participants use the advice given the advisor's information access, rather than *whether* they

can track her information access, this blocked design allowed us to make the advisor’s information access as obvious as possible and minimize unnecessary task demands.

3.2. Results

Fig. 2 shows participants’ responses and model predictions. In the None condition, participants almost completely ignored the advisor’s advice. In the Both condition, there was a strong tendency to follow the advice.³ Critically, participants’ responses were influenced by the advice even in the Hidden condition, where the advisor had partial, non-overlapping information about the cards. Furthermore, participants benefited from this imperfect advice, compared to random advice (total score: Hidden $M(SE) = 312.16(1.31)$; None $M(SE) = 304.54(1.58)$; $W = 637.5$, $p < .001$, paired Wilcoxon test). Both models showed strong correlations with participants’ choices (Mental-state Reasoning: $r(44) = .981$, $R^2 = .96$; Accuracy Heuristic: $r(44) = .980$, $R^2 = .96$; difference between correlations: $t = 0.25$, $p < .80$).

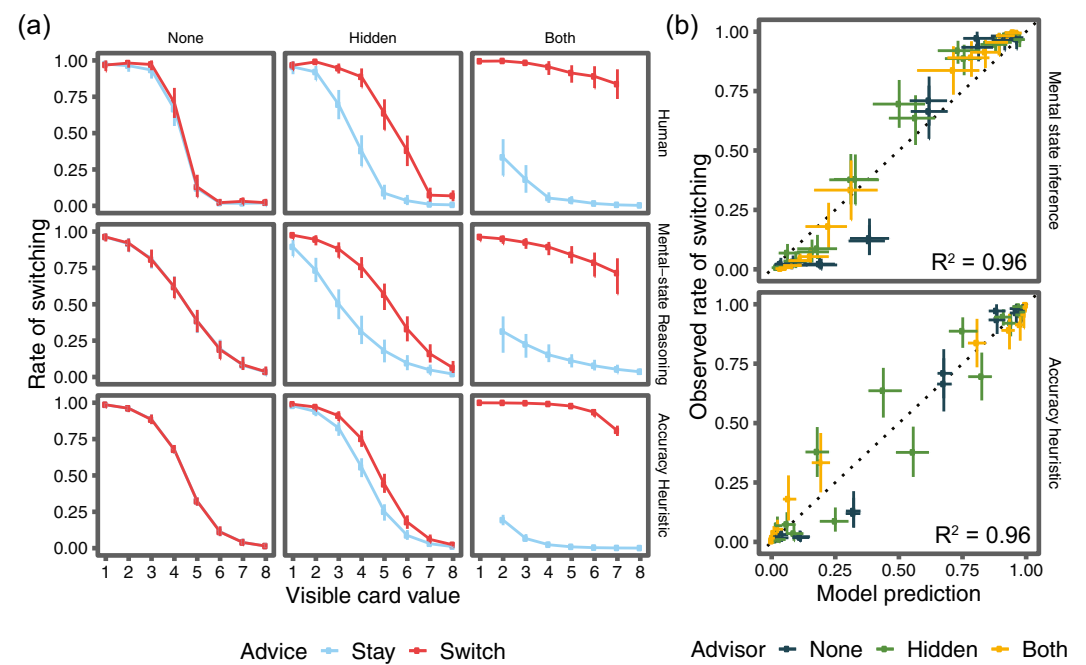


Fig. 2. (a) Experiment 1 results: behavioral data (top) and model predictions (middle & bottom). Plots show how often participants switched, based on the advisor’s access to information (columns), advice, and the value of the visible card. (b) Comparing human response curves (Observed rate of switching) to model posterior predictive (Model predictions). Error bars denote 95% bootstrapped confidence intervals (human) and credible intervals (model).

Overall, we found that participants adjusted their use of advice based on the advisor's access to information, and that they benefited even from advice based on incomplete and non-overlapping information about the cards. However, we were unable to distinguish whether participants made these decisions by inferring the advisor's mental states (Mental-state Reasoning) or by relying on her accuracy (Accuracy Heuristic); both models captured the pattern of human behavior.

4. Experiment 2

In Experiment 2, participants played the game with an advisor who could only see the hidden card (as in Experiment 1 "Hidden" condition); we asked whether participants adjust their use of the advice based on the advisor's helpfulness.

4.1. Methods

4.1.1. Participants

One hundred and fifty participants were recruited on Amazon Mechanical Turk and randomly assigned to one of three between-subject conditions: Helpful ($N = 56$), Opposite ($N = 50$), and Random ($N = 44$).

4.1.2. Procedure

The task was similar to Experiment 1; the key difference was that the advisor (introduced as "Alice") always saw just the hidden card. We simplified the trial structure by removing the reminder of the advisor's information access and presenting the value of the visible card and the advisor's advice on a single screen (Fig. 1a). Participants played 8 blocks of 14 trials (112 trials total).

Across three between-participant conditions, we manipulated how the advisor selected advice (Fig. 1b). The advisor in the *Helpful* condition was identical to the Hidden condition of Experiment 1; this advisor was more likely to suggest to switch as the value of the hidden card increased ($m_A = 1.5$, $b_A = 0$). In the remaining conditions, the advisor gave the *Opposite* of the helpful advisor's advice ($m_A = -1.5$, $b_A = 0$) or selected advice at *Random* ($m_A = 0$, $b_A = 0$). Unlike in Experiment 1, where participants were told that the advisor is always helpful, in Experiment 2 participants had to learn how she selects advice through their interactions with her in the card game.

Participants also completed a *Likelihood Estimation* task twice, before and after the main card game (Fig. 4a). On each trial (8 trials total, randomized across all card values), participants saw the card game from Alice's perspective, with only the "hidden" card visible, and answered the question: "What are the chances that Alice will say 'switch'?". Participants responded using a slider ranging from 0% ("Definitely 'stay'") to 100% ("Definitely 'switch'"). Because participants had to learn about Alice's helpfulness from experience, this task allowed us to probe participants' initial beliefs about how Alice selects advice, as well as how their beliefs changed with experience.

4.2. Results

Results from the Likelihood Estimation task suggested that participants indeed learned about their advisor. Although the initial responses suggested a prior belief that the advisor is helpful, their responses after the card game appropriately reflected their experience with the advisor (see Figs. 4a, S8). Participants' responses in the card game further suggested that this learning happened early in the task (Fig. S5).

Participants' use of advice in the main task varied depending on the advisor's helpfulness (see Supporting Information, *Experiment 2* for regression analyses testing the effect of advice on participant responses). Descriptively, as in the Hidden condition in Experiment 1, participants in the Helpful condition tended to follow the advisor's advice; by contrast, participants in the Random condition tended to ignore the advice, mirroring the None condition in Experiment 1. Accordingly, participants earned more points in the Helpful condition (total score: $M(SE) = 627.86(5.17)$) than in the Random condition (total score: $M(SE) = 609.52(5.54)$); Helpful vs. Random: $W = 534$, $p < .001$, Wilcoxon test). By contrast, while participants showed a weak but significant tendency to go against the Opposite advisor's advice, they did not earn more in the Opposite condition (total score: $M(SE) = 615.08(5.02)$) than in the Random condition (Opposite vs. Random: $W = 903.5$, $p = .137$). Note that the Opposite advisor's advice was as informative as that of the Helpful advisor; thus, a perfectly rational learner could benefit even from the Opposite advisor's advice by going against it. Our results suggest that participants' use of the Opposite advisor's advice was not optimal. This may reflect people's tendency to be optimistic about others' helpfulness (Leong & Zaki, 2018), though the role of lower level factors such as inhibitory control cannot be ruled out.

Both models captured key differences across conditions in participants' choices (see Fig. 3). However, we found some discrepancies between the behavioral data and model predictions. First, both models overestimated how often participants went against the advice in the Opposite condition. Second, the Accuracy Heuristic model also predicted that participants would go against the advice in the Random condition; this is because the estimated mean of w_{acc} (mode: 0.41, 95% credible intervals: [0.38, 0.44]) was lower than the actual observed accuracy in the Random condition (i.e., how often the advisor was observed to be right when participants switched; mean: 49.9%, 95% CI: [47.28, 52.97]). As a result, the Mental-state Reasoning model captured more variance in human behavior ($r(46) = .961$, $R^2 = .92$) than the Accuracy Heuristic model ($r(46) = .912$, $R^2 = .83$; test between correlations: $t = 3.95$, $p < .001$).

These discrepancies might stem from our models' assumption that participants' beliefs about the advisor are static, whereas the behavioral data suggest that participants started with the assumption that the advisor is helpful regardless of condition and learned about the advisor over time (Figs. 4 and S5). Even though this learning happened early in the task, the models might have overestimated participants' learning by ignoring changes in participants' beliefs over time. Thus, even though the Mental-state Reasoning model showed a stronger correlation with human behavior than the Accuracy Heuristic model, we designed an additional experiment to better tease apart the two models.

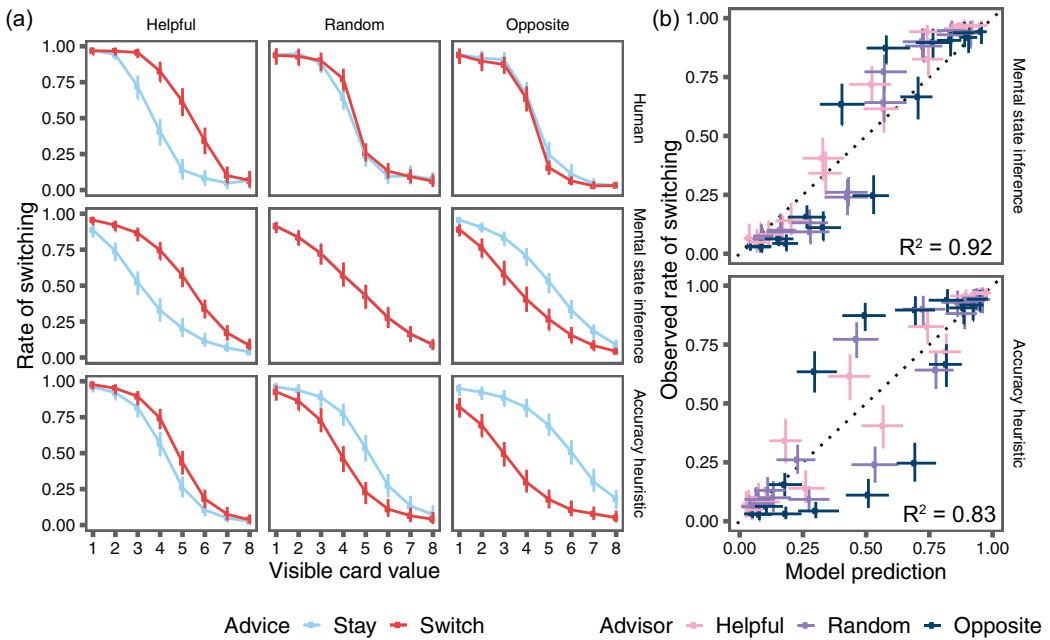


Fig. 3. (a) Experiment 2 results: behavioral data (top) and model predictions (middle and bottom). (b) Comparing human response curves (Observed rate of switching) to model posterior predictive (Model Predictions). Error bars denote 95% bootstrapped confidence intervals (human) and credible intervals (model).

5. Experiment 3

In Experiment 3, the two conditions were matched in advisor accuracy but differed in their strategies. If participants rely solely on the advisor's accuracy, there should be no difference in how they use the advice (as predicted by the Accuracy Heuristic model); however, if participants reason about how the advisor selects advice, their use of advice should differ across conditions (as predicted by the Mental-state Reasoning model).

5.1. Methods

5.1.1. Participants

One hundred fifty participants were recruited from Amazon Mechanical Turk and randomly assigned to the Conservative ($N = 83$) or Risky ($N = 67$) condition.

5.1.2. Procedure

The procedure was similar to Experiment 2, but the conditions differed in how the advisor selected advice (Fig. 1b). In the *Conservative* condition ($m_A = 5$, $b_A = -10$), the advisor mostly recommended to stay, unless the hidden card had a very high value (e.g., +7 or +8). In the *Risky* condition ($m_A = 5$, $b_A = 10$), the advisor mostly recommended to

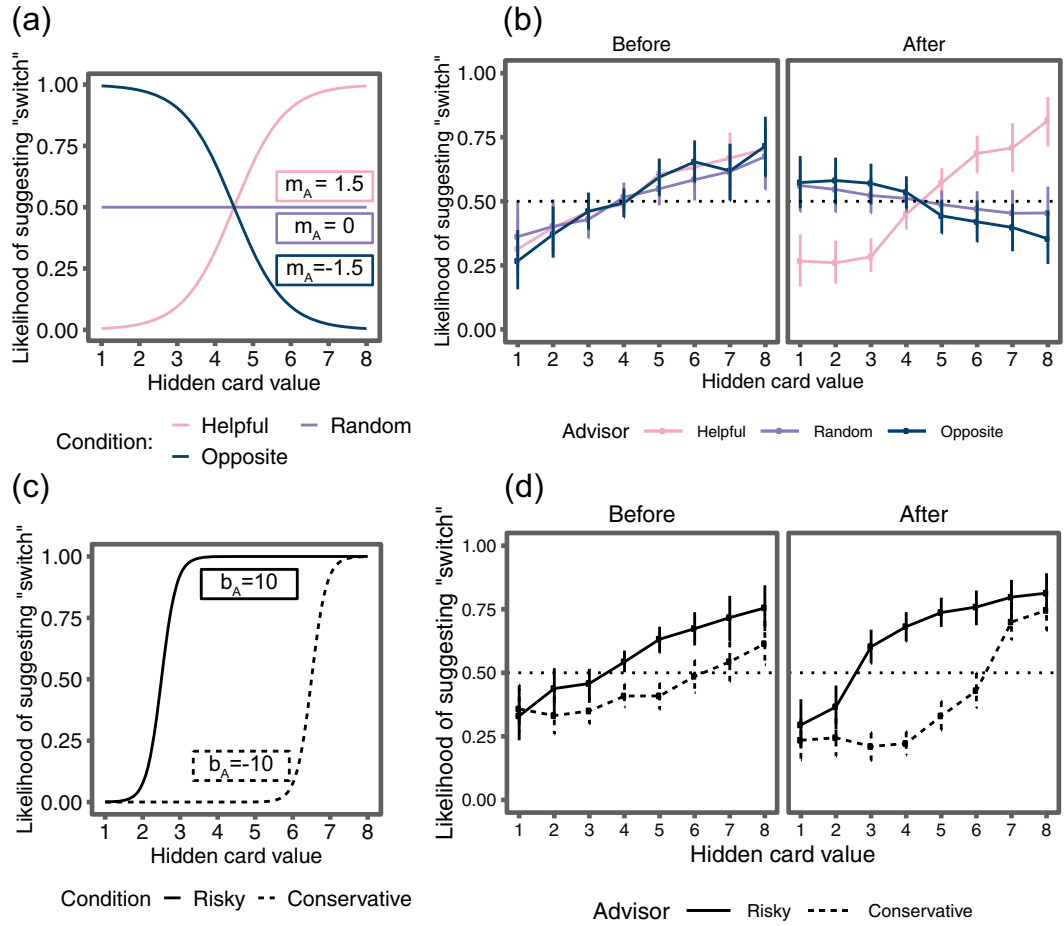


Fig. 4. (a and c) Ground-truth advisor choice functions for Experiments 2 (a) and 3 (c). (b and d) Likelihood Estimation results for Experiment 2 (b) and 3 (d). Before and after playing the card game, participants estimated the likelihood that the advisor would tell them to “switch” for each possible value of the hidden card. Error bars denote bootstrapped 95% CI.

switch, unless the hidden card had a very low value (e.g., +1 or +2). Both advisors were equally helpful; on average, they recommended the correct action in 71.2% of trials. To help participants learn about the advisor’s strategy without giving away information that favors one model over the other, participants were given a short prompt about Alice before the task. In the Risky condition, the prompt read: “Alice tends to tell people to ‘switch’ often”; in the Conservative condition, “switch” was replaced with “stay.”

5.2. Results

Participants’ responses in the Likelihood Estimation task suggested that they learned how the advisor generates advice (Figs. 4d and S8). Unlike in Experiment 2, participants’

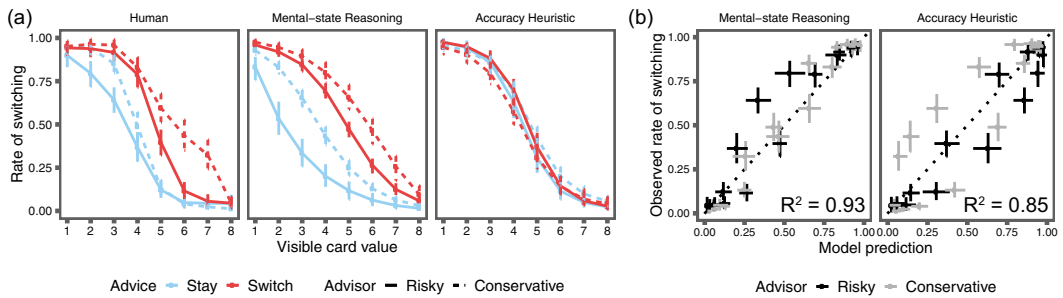


Fig. 5. Experiment 3 behavioral results and model predictions. Error bars denote 95% bootstrapped confidence intervals (human) and credible intervals (model).

responses diverged across conditions before the start of the game, likely due to the prompt they received about the advisor (Supporting Information, *Likelihood Estimation model*).

Note that the two advisors were equally accurate but differed in their tendency to suggest to “switch” or “stay.” Participants followed both advisors’ advice equally often ($W = 2201.5$, $p = .052$, Wilcoxon test; Fig. S5) and earned as many points in both conditions ($W = 2524$, $p = .484$; Fig. S3). Critically, however, as predicted by the Mental-state Reasoning model, participants used each advisor’s advice differently: They were more likely to follow advice when the Risky advisor suggested to stay and when the Conservative advisor suggested to switch (Figs. 5 and S6). This tendency was present even within the first two blocks and strengthened over time (Fig. S5d). The Mental-state Reasoning model ($r(46) = .966$, $R^2 = .93$) explained more variance in behavior than the Accuracy Heuristic model ($r(44) = .924$, $R^2 = .85$; difference between correlations: $t = 2.22$, $p < .035$).

6. General discussion

Humans are uniquely successful social learners. Theory of Mind has generally been considered a key ingredient of that success (Boyd et al., 2011; Henrich, 2015; Heyes, 2012; Tomasello et al., 1993), particularly when learning from knowledgeable, helpful others (Shafto et al., 2014). However, people are rarely perfectly knowledgeable or helpful; in the present work, we provide evidence that people engage in mental-state reasoning to learn even from imperfect advice based on incomplete knowledge.

We found that participants flexibly adjusted their use of advice based on the advisor’s access to information (Experiment 1), helpfulness (Experiment 2), and strategy (Experiment 3). As a result, they earned more points than they would have without the advisor, even when the advice was often inaccurate. Collectively, our behavioral results suggest that participants “put two heads together,” integrating their own knowledge and the advisor’s advice to make better decisions.

Our work also begins to shed light on *how* such integration occurs. Theories of social learning differ in how they characterize the role of social cognition in learning effectively from others; the computational models presented here instantiate two competing hypotheses. One prominent idea is that humans have evolved a specialized mechanism for tracking others' accuracy and selectively learn from reliable informants (Harris, 2012; Henrich, 2015; Heyes, 2012). A large body of literature suggests that even preschool-aged children preferentially learn from people who were previously accurate over those who were inaccurate (see Sobel & Kushnir, 2013 for a review). Consistent with this idea, some computational models of social learning or advice-taking assign an additional reward to options that are recommended by others (Biele et al., 2009), weighted by their accuracy or confidence (Boorman, O'Doherty, Adolphs, & Rangel, 2013; Campbell-Meiklejohn, Simonsen, Frith, & Daw, 2017; Leong & Zaki, 2018; Toelch et al., 2013). These models are largely agnostic about *how* informants generate advice and instead treat advice as a reward-predictive cue. More broadly, these accounts rely on learners' ability to form representations of informant reliability based on observable cues, assumptions, or prior experience. While Theory of Mind might be a prerequisite for forming such representations (Harris, 2012; Henrich, 2015), the value of learning comes from *accepting* or *rejecting* others' testimony or advice, rather than inferring the contents of others' minds. Our Accuracy Heuristic model instantiates this idea.

Another proposal is that Theory of Mind reasoning allows learners to infer the informant's mental states (Shafra et al., 2014); given observed data (e.g., advice), learners work backwards to infer the hypothesis from which the informant sampled her advice (i.e., her belief about the world). Consistent with this idea, adults, and even young children, use pedagogically sampled information to draw inferences that go beyond the face value of the data, recovering the informant's belief about the correct hypothesis (Bonawitz et al., 2011; Shafra et al., 2014). However, prior work has largely focused on cases where informants are assumed to be knowledgeable and helpful, or where learners themselves have enough evidence to evaluate informants' knowledge and helpfulness (Bass et al., 2015; Hawthorne-Madell & Goodman, 2017). More recent computational and empirical work on inductive inferences in social contexts suggests that even infants can use statistical information to jointly infer the sampling process (i.e., how an adult samples evidence) and hidden properties of the world (i.e., the adult's belief about hidden properties of objects; Gweon, Tenenbaum, & Schulz, 2010). Consistent with this work, our Mental-state Reasoning model shows how such joint inference can be useful even when the advisor's knowledge is imperfect; by jointly inferring the process by which the advisor selects advice and her beliefs about the hidden card, learners can make better decisions even when both they and the advisor have partial, non-overlapping knowledge.

Our results suggest that there are cases where simple sensitivity to accuracy fails to explain human behavior. When advisors were matched in their overall accuracy but differed in their strategy (Experiment 3), the Accuracy Heuristic model failed to predict people's choices. Importantly, our results do not necessarily mean that learners don't use accuracy in social learning, nor that learners always engage in rich inferences about others' minds. The fact that both models capture human behavior in Experiments 1 and 2

suggests that learners could use, and benefit from, either strategy. The key implication of our work is that such heuristics can only go so far; there are cases where learners must (and do) reason about other minds. Using a simple yet novel social decision-making task, we demonstrated how mental-state reasoning enables learners to benefit even from social information from impoverished sources. Our work connects previous models of social decision-making that describe how learners' decisions are influenced by social information (yet are agnostic about how such information is generated; Bahrami et al., 2010; Biele et al., 2009; Leong & Zaki, 2018), with models that formalize how learners use observed actions to recover others' mental states (yet do not directly address how such inferences impact decision making under uncertainty; Baker et al., 2009, 2017; Shafto et al., 2014; see also Hawthorne-Madell & Goodman, 2017).

One limitation of our current work is that we used a simulated agent with static strategies as the "advisor" and did not consider learning. While the parameters of the advisor's strategies were rooted in average human behavior, real human advisors may show richer, more complex behaviors. For example, advisors might change their strategies over time or even tailor their advice based on the learner's behavior; participants, in turn, might also adjust continually to the advice. Because our task did not implement such dynamics, one might wonder whether inferring the value of the hidden card "qualifies" as mental-state reasoning, and raise the possibility that participants would behave similarly if the advice were replaced with non-social, probabilistic cues. More broadly, we believe that the nature of "advice" is inherently social; it is information selected and provided by an agent to another agent, with an underlying intent and knowledge. Thus, even if we implemented a non-social version of the task (e.g., replacing the advisor with a machine), people might attribute various goals or strategies to the machine, making it difficult to distinguish it from a "social" task. On the other hand, note that participants failed to capitalize on the Opposite advisor's advice (Experiment 2); if this failure is due to people's tendency to overestimate other humans' accuracy Leong and Zaki (2018), receiving the same advice from a mechanical device may actually help people better assess its accuracy.

In the current work, communication between the learner and the advisor was unidirectional (i.e., the learner received "advice" before the decision was made), and their access to information was fixed throughout. An important open question is how a learner and an advisor might communicate bidirectionally when the goal is to cooperate to accumulate the most points together or to compete against each other, and how such interactions might change based on the relative information access or skill of the two players. In addition, recent work suggests that learners not only benefit from advice before a decision, but also from evaluative feedback after a decision or action has taken place (Ho, MacGlashan, Littman, & Cushman, 2017); such learning may be especially useful when learners do not receive immediate feedback about their decisions. Our task opens up possibilities for further investigating the dynamics of dyadic social interactions by manipulating various qualities of agents (e.g., knowledge, competence, intention) and the nature of social information (e.g., advice, feedback, demonstration).

Decisions are rarely made with certainty, because our knowledge of the world is incomplete and imperfect. However, social information is selected by people with rich mental lives; people's goals, knowledge, and intentions structure the information they choose to communicate. The richness and sophistication of human cognition allows us to capitalize on this nature of social information and leverage the knowledge that resides in others' minds. By "putting two heads together," we can make the best of even imperfect advice.

Notes

1. Materials, data, and analysis scripts for this project are available at <http://osf.io/w37up>.
2. The Gamma distribution constrains m_L to positive values. In effect, m_L acts as a temperature parameter; in the extremes, the model can make completely random or perfectly greedy choices, but it cannot systematically favor the option with the lowest value. In contrast, we use a Gaussian distribution for m_A because m_A can take on negative values when the advisor gives adversarial advice (Experiment 2, Opposite condition).
3. While participants were explicitly told that the advisor can see both cards, they still showed a slight bias to "stay" when visible card had a very high value (+7) and to "switch" when it was very low (+2). This may reflect a small degree of uncertainty about the advisor's reliability, which was captured by both models by assuming that participants used a softmax choice rule.

References

- Bahrami, B., Olsen, K., Latham, P. E., Roepstorff, A., Rees, G., & Frith, C. D. (2010). Optimally interacting minds. *Science*, 329(5995), 1081–1085.
- Baker, C. L., Jara-Ettinger, J., Saxe, R., & Tenenbaum, J. B. (2017). Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behavior*, 1, 1–10.
- Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*, 1–21.
- Bass, I., Bonawitz, L. B., & Gweon, H. (2017). Didn't know, or didn't show? Preschoolers consider epistemic state and degree of omission when evaluating teachers. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (pp. 105–110). Austin, TX: Cognitive Science Society.
- Bass, I., Hawthorne, D., Goodman, N. D., & Gweon, H. (2015). Not by number alone: The effect of teacher's knowledge and its value in evaluating "sins of omission." In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 166–171). Austin, TX: Cognitive Science Society.
- Biele, G., Rieskamp, J., & Gonzalez, R. (2009). Computational models for the combination of advice and individual learning. *Cognitive Science*, 33(2), 206–242.

- Bonawitz, E., Shafto, P., Gweon, H., Goodman, N. D., Spelke, E., & Schulz, L. (2011). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition*, 120(3), 322–330.
- Boorman, E. D., ODoherty, J. P., Adolphs, R., & Rangel, A. (2013). The behavioral and neural mechanisms underlying the tracking of expertise. *Neuron*, 80 (6), 1558–1571.
- Boyd, R., Richerson, P. J., & Henrich, J. (2011). The cultural niche: Why social learning is essential for human adaptation. *Proceedings of the National Academy of Sciences*, 108, 10918–10925.
- Campbell-Meiklejohn, D., Simonsen, A., Frith, C. D., & Daw, N. D. (2017). Independent neural computation of value from other people's confidence. *Journal of Neuroscience*, 37(3), 673–684.
- De Leeuw, J. R. (2015). jspsych: A javascript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, 47(1), 1–12.
- Gershman, S. J. (2016). Empirical priors for reinforcement learning models. *Journal of Mathematical Psychology*, 71, 1–6.
- Gweon, H., & Asaba, M. (2017). Order matters: Competence and limitations in young children's evaluation of under-informative teachers. *Child Development*, 89(3), e278–e292.
- Gweon, H., Pelton, H., Konopka, J. A., & Schulz, L. E. (2014). Sins of omission: Children selectively explore when teachers are under-informative. *Cognition*, 132(3), 335–341.
- Gweon, H., Tenenbaum, J. B., & Schulz, L. E. (2010). Infants consider both the sample and the sampling process in inductive generalization. *Proceedings of the National Academy of Sciences*, 107(20), 9066–9071.
- Harris, P. L. (2012). *Trusting what you're told: How children learn from others*. Cambridge, MA: Harvard University Press.
- Hawthorne-Madell, D., & Goodman, N. D. (2017). Reasoning about social sources to learn from actions and outcomes. *Decision*, Advance online publication.
- Henrich, J. (2015). *The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter*. Princeton, NJ: Princeton University Press.
- Heyes, C. (2012). New thinking: The evolution of human cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1599), 2091–2096.
- Ho, M. K., MacGlashan, J., Littman, M. L., & Cushman, F. (2017). Social is special: A normative framework for teaching with and learning from evaluative feedback. *Cognition*, 1–16.
- Leong, Y. C., & Zaki, J. (2018). Unrealistic optimism in advice taking: A computational account. *Journal of Experimental Psychology: General*, 147(2), 170–189.
- Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. (2016). Probabilistic programming in python using pymc3. *PeerJ Computer Science*, 2, e55.
- Shafto, P., Goodman, N. D., & Griffiths, T. L. (2014). A rational account of pedagogical reasoning: Teaching by, and learning from, examples. *Cognitive Psychology*, 71 (C), 55–89.
- Sobel, D. M., & Kushnir, T. (2013). Knowledge matters: How children evaluate the reliability of testimony as a process of rational inference. *Psychological Review*, 120(4), 779–797.
- Toelch, U., Bach, D. R., & Dolan, R. J. (2013). The neural underpinnings of an optimal exploitation of social information under uncertainty. *Social Cognitive and Affective Neuroscience*, 9(11), 1746–1753.
- Tomasello, M., Kruger, A., & Ratner, H. (1993). Cultural learning. *Behavioral and Brain Sciences*, 16, 495–552.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix S1. Supplementary methods and results.