Exp Econ (2009) 12: 386–416
DOI 10.1007/s10683-009-9220-1

How effectively do people learn from a variety of different opinions?

Andrew Healy

Received: 18 July 2008 / Accepted: 17 July 2009 / Published online: 29 July 2009

Abstract This paper presents experimental evidence about how effectively individuals learn from information coming from heterogeneous sources. In the experiment, Thai subjects observed information that came from Americans and from other Thais that they could use to help them answer a series of questions. Despite listening too little to either group, subjects demonstrated a significant amount of statistical sophistication in how they weighed observed American information relative to observed Thai information. The data indicate that subjects understood that outside information has extra value because people from the same group tend to make the same kinds of mistakes. The results illustrate the importance of forming diverse groups to solve problems.

Keywords Information aggregation · Learning · Experimental economics

JEL Classification C91 · D83

1 Introduction

Consider the situation faced by an economic agent who has to make a difficult decision, such as the one faced by a farmer who has to decide whether to start using a new variety of seeds. When the farmer makes her choice, she may feel confident enough to make the decision without any advice. Alternatively, she may consider the

Electronic supplementary material The online version of this article (http://dx.doi.org/10.1007/s10683-009-9220-1) contains supplementary material, which is available to authorized users.

A. Healy (✉)
Department of Economics, Loyola Marymount University, 1 LMU Drive, University Hall 4229, Los Angeles, CA 90045, USA
e-mail: ahealy@lmu.edu

advice she received from neighbors who have experience using the seeds. Or she may decide that she wants to talk to someone from outside her group, since an outsider may have a different experience. Her ability to make the best decision will depend crucially on how much she listens to others, and on the diversity of the opinions that she draws upon. This paper uses an experiment to explore how effectively agents utilize a variety of different opinions to make decisions.

The model and experiment in this paper relate to, and expand upon, previous research into information aggregation. It is distinct from much of that literature in that I am concerned with the ability of individuals to aggregate information from a variety of different sources, while the majority of the existing literature is concerned with the institutions that serve to accumulate individual knowledge. For example, many theoretical and experimental papers have considered the capabilities of auctions and other market institutions to aggregate private information through the price mechanism (Hellwig 1980; Plott and Sunder 1988; Forsythe and Lundholm 1990; Pesendorfer and Swinkels 1997, 2000). An additional set of papers has examined the potential for voting mechanisms to aggregate the information possessed by individual voters (Lohmann 1994; Austen-Smith and Banks 1996; Feddersen and Pesendorfer 1997; Piketty 1999).

While some papers have looked at the process by which individuals use information to make decisions, most of this literature has been concerned with issues relating to sequential decision-making or herding, and has used simple stimuli like the urn-ball design (Anderson and Holt 1997; Goeree et al. 2007; Kraemer et al. 2006). Compared to previous research, my experiment uses a richer informational structure, leading to results that are more easily generalizable. Subjects use their own private information, observed information from other members of their own group, and observed information from members of a different group that has different expertise to answer a series of general-knowledge questions. The design of the experiment makes it possible to test a variety of hypotheses relating to how effectively subjects use information to solve problems.

The experimental design relates to a variety of real-world examples in which individuals can use information to improve decisions. Information sharing within social networks has been shown to influence agricultural technology adoption, health decisions, and savings behavior, but not always for the better.[1] The research shows that individuals put high weight on information learned from others within their own group and that information sharing between groups often does not occur. Would decision-making be improved if individuals had access to a variety of different information sources? Are people generally able to aggregate a variety of different opinions in an intelligent way? The controlled environment of the lab makes it possible to answer yes to both of these questions. In fact, the experimental subjects show an implicit appreciation for subtle statistical ideas in how they weigh information from a variety of different sources.

[1] Foster and Rosenzweig (1995) and Munshi (2004) explore how information sharing affects technology adoption in India and Kenya. Miguel and Kremer (2004) and Munshi and Myaux (2006) describe how information sharing affects health decisions in developing countries. Duflo and Saez (2003) investigate how communication within social groups influences participation in retirement plans at an American university.

In the experiment, Thai subjects first answered a series of general-knowledge questions that had correct numerical answers. There were three types of questions: 1) questions about Thailand, 2) questions about the US, and 3) questions about both Thailand and the US. After answering the questions on their own, subjects observed randomly selected answers given by Americans and by other Thais, who had answered the same questions at an earlier date, that they could use to help them revise their answers. By looking at how subjects changed their answers, it is possible to estimate the weights that they applied to observed American answers, to observed Thai answers, and to their own initial answers.

The questions about both Thailand and the US are a crucial aspect of the experimental design for two reasons. First, the data show that Americans and Thais are equally good at answering these questions. Second, despite the fact that any one Thai answer is equally good as any one American answer, it is possible to show that an optimizing Thai should assign about twice as much weight to American answers as to other Thai answers. This extra value in American information to a Thai subject comes from the fact that Americans tended to make one kind of mistake and Thais tended to make a different kind of mistake. When members of the same group tend to make the same kind of mistake, an agent has more to learn from members of a different group than from other members of her own group.

In general, the subjects appeared to understand this idea, behaving optimally in how they weighed American information relative to Thai information for two of the three types of questions. Although they would have benefited by listening more to both groups, subjects significantly improved their performances by correctly weighing observed American answers relative to observed Thai answers. Moreover, the way that subjects behave when they observe information about the questions about Thailand and the US strongly suggests that subjects appreciated the extra value of an American's independent perspective to a Thai decision-maker.

Subject behavior in the experiment indicates that when agents listen to a diverse group of opinions, they can be expected to carefully consider the available information. The issue of concern is that those independent voices may not be heard at all, either because agents lack access to outside information or because they put too much weight on their personal knowledge and thus choose not to seek outside advice. Failures to use information effectively show that groups make mistakes when all members think the same way and outside sources are not consulted.[2] As one example, the Bay of Pigs fiasco occurred in large part because an insulated group of decision-makers failed to consult independent advisors in the CIA and State Department (Janis 1972; Surowiecki 2004). For decision-makers to avoid these sorts of mistakes, the experiment suggests that ensuring that people have access to a diverse information set may indeed be the crucial issue. The experiment shows that, conditional on those voices

[2]Other research has shown that people benefit in a variety of ways from having a wide range of social contacts. For example, knowing a diverse group of people helps with finding jobs and with psychological wellbeing (Granovetter 1973; Putnam 2000). To borrow Granovetter's phrase, my results show that "the strength of weak ties" carries over to problem-solving. The experimental results thus show another consequence of the decline in social capital that Putnam (2000) describes. People without access to a diverse information set will make poor decisions.

being heard, decision-makers can effectively learn from a variety of different opinions.

The experimental results reported here also help to explain previous experimental results on advice. In a variety of experimental settings, subjects who receive advice from non-experts (often subjects with previous experience in the experiment) come significantly closer to following rational economic models than do unadvised subjects (e.g., Schotter and Sopher 2003, 2006; Schotter 2003; Iyengar and Schotter 2008). These authors generally claim that advice increases efficiency because the process of listening to advice causes decision-makers to think about the problem in a different way (Schotter 2003). This finding from the advice literature provides context for the results reported here which suggest subjects apply subtle statistical ideas to the advice they receive. In addition, the experimental results in this paper provide insight into the reasons why decision-makers learn effectively from advice. Subjects in the experiment appear to appreciate the implications of bias in the advice they receive and they adjust the weight that they give to advice accordingly.

Section 2 describes the experimental design. In Sect. 3, I model the process of using information to make decisions. Section 4 contains the summary statistics that describe the distributions of American and Thai answers to the questions. In Sect. 5, I estimate the weights subjects give to the information they observe and test a variety of hypotheses that explain their behavior. Section 6 describes the results of tests that compare how subjects actually behave to how they optimally would. Section 7 describes how the experimental design makes it possible to test the hypothesis that subjects appreciate the independence in outside information. Section 8 concludes.

2 Experimental design

In the experiment, American students from the Massachusetts Institute of Technology (MIT) and Thai students from Thammasat University's Rangsit campus answered a series of general knowledge questions in December 2003 and January 2004. Soon after that, in January and February 2004, separate groups of Thai students from Thammasat's Bangkok campus and from the National Institute of Development Administration (NIDA) answered the same questions. These students then observed randomly selected answers, given by the MIT and Rangsit students, which they could use to help them revise their answers. Subjects observed between zero and three American answers for each question and between zero and three Thai answers for each question.

Since the aim of the experiment was to estimate the extent to which subjects appreciated the relative accuracies of Americans and Thais and the importance of accounting for the common mistakes made by members of the same group, subjects observed no additional information about the overall American and Thai answer distributions. Given that subjects were not informed about the accuracy of the answers they observed, it was not possible for them to precisely optimize in their usage of the information they observed for each individual question. By looking at subject behavior across questions, however, the experimental design makes it possible to determine the extent to which subjects either systematically underweigh or overweigh American information relative to Thai information. The design also allows for testing of

From 1961-1990, average daily high temperature in January in Bangkok	From 1961-1990, average daily high temperature in January in Boston	Sum
______ °C	+ ______ °C	= ______ °C

Fig. 1 Sample question from the experiment

the hypothesis that subjects appreciate the importance of accounting for the fact that members of the same group make the same kind of mistake when they decide how to weigh the observed information.

2.1 The questionnaire

The questionnaire consisted of fifteen questions covering a range of topics. For example, one question asked about the January temperature in Bangkok, the January temperature in Boston, and the sum of those temperatures. Another question asked for the number of Thai prime ministers since 1960, the number of American presidents since 1960, and the sum of those two numbers. A third question asked about the percentages of Thai and American 25–29 year-olds with some university education, as well as the sum of those two numbers. Figure 1 shows one of the questions. The Appendix contains all of the questions.

The design of the questions had two purposes. The first goal was to generate questions that Thais were likely to know better than Americans, questions that Americans were likely to know better than Thais, and questions they might know equally well (and the data confirm that this is the case for the sum questions). Second, the case in which subjects see information about the sum question makes it possible to test the hypothesis that they appreciate the independence in outside information.

2.2 Stage 1: Creating a pool of American and Thai answers

In Stage 1 of the experiment, 116 introductory economics students at MIT and 130 introductory economics students at Thammasat University's Rangsit campus answered the series of questions. Students had 15 minutes to answer the survey. In both countries, students answered the questionnaire at the end of introductory economics classes. In Thailand, the questionnaire and instructions were given in Thai. Each group answered the questions in the standard units prevailing in their respective countries. For example, Americans answered temperature questions in degrees Fahrenheit and Thais answered temperature questions in degrees Celsius.

Subjects received monetary rewards for answering accurately. For the American students, the top three performers on the entire set of questions received $50 each and the top fifteen performers on some of the individual questions received $10 each. Among the Thai students, the top five performers on the overall questionnaire received 1000 baht (approximately $25) and the top twenty on some of the individual questions received 200 baht.[3] To determine the top performers for each question, students were ranked according to the distance of their answers from the correct answer. The additional rewards for the Thai students reflected the larger sample size. The rewards for the individual questions were included to ensure that students who felt they had little chance of winning the overall awards still had sufficient incentive to try hard to answer the questions well.

2.3 Stage 2: Showing American and Thai information to subjects

In Stage 2, 300 economics undergraduates at Thammasat's Bangkok campus and master's economics students at the National Institute for Development Administration (NIDA) received instructions in Thai. The instructions were read aloud by an assistant at the same time that the subjects read the written instructions. Subjects were informed that they would receive 100 baht for participating and 20 baht for each question that they answered within a range of the correct answer. The incentives were intended to provide subjects with the objective of minimizing the mean-squared error (MSE) of their answers, while keeping the instructions as simple as possible. In the initial instruction packet, subjects were not told that they would be receiving additional information to help them choose their final answers to the questions.

After the first set of instructions had been read, subjects answered all of the questions using Microsoft Excel in computer labs at NIDA and Thammasat. They directly answered the Bangkok/Thailand and Boston/US questions, and the sum was calculated from those answers. After all subjects answered the questions, they received a second set of instructions.

Subjects were told that they would observe randomly selected answers given by Thammasat-Rangsit students, which subjects knew to be a different campus of Thammasat on the far outskirts of the city, and MIT students who answered the same set of questions.[4] The randomly selected answers from other students were provided in a separate packet. For each question, subjects saw the heading "Answers from Thai students" followed by the Thai information, and then "Answers from American students" followed by the American information. Figure 2 shows what one group of the Thai subjects saw for the question about political leaders. The subjects were told that their payments would be based on the final answers that they gave after observing the information.

In addition to the answers themselves, I randomly selected three features of the data that subjects observed: 1) the type of question (Bangkok/Thailand, Boston/US,

[3]Thai per-capita GDP is about one-fifth of American GDP in purchasing power parity terms, so that the Thai rewards were somewhat higher in relative income terms. At the same time, this difference somewhat overstates the disparity among the participants, since the mean income for Thai college graduates income was about one-third of mean income for American college graduates in 2001 (author's calculation using the 2001 Thai Labor Force Survey).

[4]While it was a concern that the subjects might perceive MIT as a brand name that could cause them to overestimate the relative quality of American answers, the data suggest that subjects in fact do not overestimate the relative accuracy of observed American answers compared to observed Thai answers. Moreover, the test for whether subjects appreciate the value of outside information was designed to be valid for any perceptions of American accuracy relative to Thai accuracy.

Since January 1, 1960, number of Thai prime ministers		Since January 1, 1960, number of American presidents		Sum
______	+	______	=	______

Answers given by Thai students

1. 18

Answers given by American students

1. 7
2. 9
3. 10

Fig. 2 Sample of information that subjects observed

just the sum, or all three questions)[5] for which subjects observed information, 2) how many Thai students' answers subjects observed (up to three), and 3) how many American students' answers subjects observed (up to three). The subjects who saw the information in Fig. 2, for example, observed one Thai opinion and three American opinions for the American part of the question about political leaders. Finally, I randomly selected which Thai and American answers that subjects saw. Proceeding in this way, I produced 20 different sets of information that subjects could observe.

In total, each experimental session took approximately 50 minutes. I conducted 25 sessions, 17 at Thammasat and 8 at NIDA. Payments averaged approximately 280 baht ($7) per subject, so that subjects earned about $8.40 per hour, or about three times the median wage in Bangkok.[6] The experimental instructions are in the Appendix (see electronic supplementary material).

2.4 Controlling for anchoring

Tversky and Kahneman (1974) showed that individuals will tend to stick to a number that is given to them, even when that number is irrelevant to the question at hand, a phenomenon they called anchoring. In my experiment, subjects first answered the questions and then updated their answers based on what they observed. Thus, anchoring presents a serious concern in this experiment; a subject provided her own answer to which she can anchor and that number contains meaning relevant to the task, unlike the random number which affected students' answers in Tversky and Kahneman (1974).

[5] Seeing information for all three questions is functionally equivalent to just seeing information about the Bangkok/Thailand and Boston/US questions.

[6] Subject payments did not significantly change across sessions. The point estimate obtained by regressing subject payments on session number is slightly negative, but very close to zero, suggesting that no significant sharing of answers occurred between students in different sessions.

Due to these concerns, an additional 42 students observed information and answered the questions without first providing their private beliefs. To test for anchoring, I compare these students to the students in the main group. The data show that anchoring had a small and statistically insignificant effect on subject behavior in the experiment. Details on the test for anchoring can be found in the Appendix.

3 A model of information aggregation

Here, I describe a model of information aggregation that shows how the Thai subjects should weigh the information they observe to minimize the mean-squared error (MSE) of their answers. The model applies to each question type (Bangkok/Thailand, Boston/US, or sum) separately. For a question q, take individual i in group j (either A for American or T for Thai), to have a private belief x_{ijq} about the correct answer for the question.[7] The MSE, Δ^2_{jq}, for a group j for question q is then

$$\Delta^2_{jq} = \frac{1}{N_j}\sum_{i=1}^{N_j}(x_{ijq} - Truth_q)^2,$$

where N_j is the number of group j members in the sample and $Truth_q$ is the correct answer for question q. A group that is comparatively better at answering a question will have a lower MSE for that question. The distributions of American and Thai answers give the MSE for Americans, Δ^2_{Aq}, and the MSE for Thais, Δ^2_{Tq}, for each question q.

The group MSE can be broken down into estimators of the population variance for the group (s^2_{jq}) and the squared group bias (α^2_{jq}), where $\overline{x}_{jq}$ is the mean answer given by group j for question q.

Proposition 1 *Where* $s^2_{jq} = \frac{1}{N_j}\sum_{i=1}^{N_j}(x_{ijq} - \overline{x}_{jq})^2$ *and* $\alpha^2_{jq} = \frac{1}{N_j}\sum_{i=1}^{N_j}(\overline{x}_{jq} - Truth_q)^2$, *the MSE for group j for question q can be expressed as*

$$\Delta^2_{jq} = s^2_{jq} + \alpha^2_{jq}.$$

Proof See the Appendix. □

This decomposition reflects the fact that the total error made by the group consists of individual and group components. The individual component, s^2_{jq}, comes from the variation in answers given by members of the same group. The group component, α^2_{jq}, comes from the distance between the group mean and the correct answer.

[7] The empirical results indicate that there are no significant differences between individuals with different personal characteristics (i.e. gender) or socioeconomic status in how they treat the information available to them (results available upon request).

Define the fraction of a group's MSE that comes from group bias by ρ_{jq}:

$$\rho_{jq} = \frac{\alpha^2_{jq}}{\Delta^2_{jq}}$$

The group bias share, ρ_{jq}, will be high when group bias causes most of the group's MSE.

To analyze subject behavior, I focus on three parameters, averaged across questions: 1) the American-to-Thai MSE ratio, $\frac{\Delta^2_A}{\Delta^2_T}$, which captures how accurately the Americans answer the questions relative to Thais, 2) the American group bias share, ρ_A, which captures the share of American MSE for which group bias is responsible, and 3) the Thai group bias share, ρ_T, which captures the share of Thai MSE for which group bias is responsible.

Proposition 2 *Where Q is the number of questions, the maximum-likelihood estimators (MLE) for these three parameters are:*

$$\frac{\Delta^2_A}{\Delta^2_T} = \frac{1}{Q}\sum_{q=1}^{Q}\frac{\Delta^2_{Aq}}{\Delta^2_{Tq}}, \qquad \rho_A = \frac{1}{Q}\sum_{q=1}^{Q}\rho_{Aq}, \quad \textit{and} \quad \rho_T = \frac{1}{Q}\sum_{q=1}^{Q}\rho_{Tq}.$$

Proof See the Appendix. □

A subject's implicit perceptions of $\frac{\Delta^2_T}{\Delta^2_A}$, ρ_T, and ρ_A determine the average weights that she will apply to the information that she observes. The actual values of the parameters determine what she would optimally do. For the sake of simplicity, the model assumes that subjects treat each American answer that they see in the same way and each Thai answer that they see in the same way, for any given question. As will be discussed, the regression used to estimate subject behavior can relax this assumption, and doing so has no significant effect on the estimates for the average weights that subjects apply to the information that they observe.

To account for overconfidence and individual heterogeneity in aptitude at answering the questions, I model a subject to perceive her own MSE to be a fraction c of her actual MSE, where her actual MSE is taken to be $\phi_i \Delta^2_{Tq}$. The ϕ_i parameter accounts for heterogeneity in aptitude, with $\phi_i < 1$ (> 1) corresponding to a subject actually being better (worse) at answering the question than the mean accuracy for all Thais. The mean of ϕ_i is one by definition, so that the mean squared error for all Thais is Δ^2_{Tq}.

Where Δ^2_{Sq} is a subject's perceived MSE for question q,

$$\Delta^2_{Sq} = c\phi_i \Delta^2_{Tq}. \tag{1}$$

If subjects are overconfident, they perceive themselves to be better than they actually are, so that $c < 1$.[8]

Define y_{iq} to be the final answer that individual i gives after observing information about question q. For the case where subjects see n_A American answers ($x_{Aq,1}$ through x_{Aq,n_A}) and n_T Thai answers ($x_{Tq,1}$ through x_{Tq,n_T}), a subject's objective function is:

$$\begin{aligned} E(MSE) &= E(y_{iq} - Truth_q)^2 \\ &= E\big(\lambda_{Ti}(x_{Tq,1} + \cdots + x_{Tq,n_T}) + \lambda_{Ai}(x_{Aq,1} + \cdots + x_{Aq,n_A}) \\ &\quad + \lambda_{si} x_{iq} - Truth_q\big)^2 \end{aligned}$$

where

λ_{si} = weight for initial answer,
λ_{Ti} = weight for any one piece of Thai information,
λ_{Ai} = weight for any one piece of American information.

Assuming independence between the American and Thai group biases and that the weights given to all information sum to one, the expression in Proposition 3 below captures the weights that subjects would optimally use to weigh the American information they observe relative to the Thai information they observe, for any level of overconfidence. Similar expressions can be derived for how subjects would weigh their initial answers both relative to observed Thais and observed Americans. These expressions are left to the Appendix, since all the hypothesis tests of interest in the paper concern the weight ratio below.

Proposition 3 *The following expression defines the MSE-minimizing weights that subjects should use to evaluate the American information they observe relative to the Thai information they observe*:

$$\frac{\lambda_{Ai}}{\lambda_{Ti}} = \left(\frac{\Delta_T^2}{\Delta_A^2}\right)\left(\frac{1 + (n_T - 1)\rho_T - c\phi_i \rho_T^2 n_T}{(1 + (n_A - 1)\rho_A)(1 - \rho_T)}\right). \tag{2}$$

Proof See the Appendix. □

Equation (2) gives the weight ratio that subjects should assign to an observed American answer relative to an observed Thai answer. Not surprisingly, subjects should put higher weight on American information when Δ_A^2 is low relative to Δ_T^2. Also when ρ_A is low and ρ_T is high, subjects should put higher relative weight on American information.

The overconfidence parameter enters the expression in a second-order way through the $c\phi_i \rho_T^2 n_T$ term. When overconfidence is high (c is lower), subjects should

[8]This modeling of overconfidence corresponds to the idea that a subject perceives her confidence interval to be c times the width of another Thai's, for any given significance level. Examples of experimental evidence on overconfidence where subjects answer questions about trivia or other topics include Gigerenzer et al. (1991), Camerer and Lovallo (1999), and Hoelzl and Rustichini (2005).

put more weight on Americans relative to observed Thais because an overconfident subject trusts her perception of the common Thai information for a given question more than another Thai's perception. Another way to think of this idea is that overconfident subjects already put high weight on Thai information through the high weight they give to themselves. A Thai who is overconfident but otherwise rational will then put higher weight on observed Americans than on observed Thais.

Notice also that increases in ρ_A only cause subjects to put less weight on individual American answers when n_A is greater than one, but increases in ρ_T cause subjects to put less weight on observed Thais even when only one Thai is observed. When one Thai is observed, there are two Thai answers to consider: a subject's own answer and the one she observes. As a result, the Thai group bias term enters (2) when $n_T = 1$, but the American group bias term only enters when $n_A > 1$.

The analysis to follow estimates the average weights used across subjects to weight information. To determine the average weight ratios that would be used under the model simply requires averaging the ratio in (2) across subjects. The equation is linear in the individual-specific parameter, ϕ_i, so that averaging across subjects leads to a simple result.

Proposition 4 *The following expression defines the average weights that will be used across subjects under the model to evaluate observed American information relative to observed Thai information*:

$$\frac{\lambda_A}{\lambda_T} = \left(\frac{\Delta_T^2}{\Delta_A^2}\right)\left(\frac{1 + (n_T - 1)\rho_T - c\rho_T^2 n_T}{(1 + (n_A - 1)\rho_A)(1 - \rho_T)}\right). \tag{3}$$

Proof The result follows immediately from Proposition 3 and the fact that the mean of ϕ_i is one by definition. □

The model's prediction of how subjects should weigh observed Americans relative to observed Thais can be obtained through estimating the relevant parameters in (3). Those predictions can then be compared to the actual weights that subjects use in the experiment to test hypotheses relating to subjects' implicit perceptions of $\frac{\Delta_T^2}{\Delta_A^2}$, ρ_T, and ρ_A.

4 Summary statistics

The data show that, across questions, Thais tend to make one kind of mistake and Americans tend to make their own kind of mistake. As an example of what the data look like, Fig. 3 shows kernel density estimates for the Thai and American answers for the questions about January temperature in Bangkok and Boston. Panel 1 shows that Americans have a mean of 20°C for the Bangkok temperature (correct answer = 32°C) and Panel 2 shows that Thais have a mean answer of 20°C for the Boston January temperature (correct answer = 2°C).

It is important to note that other questions show a different pattern than the question about January temperature. For other questions, the American average is not as

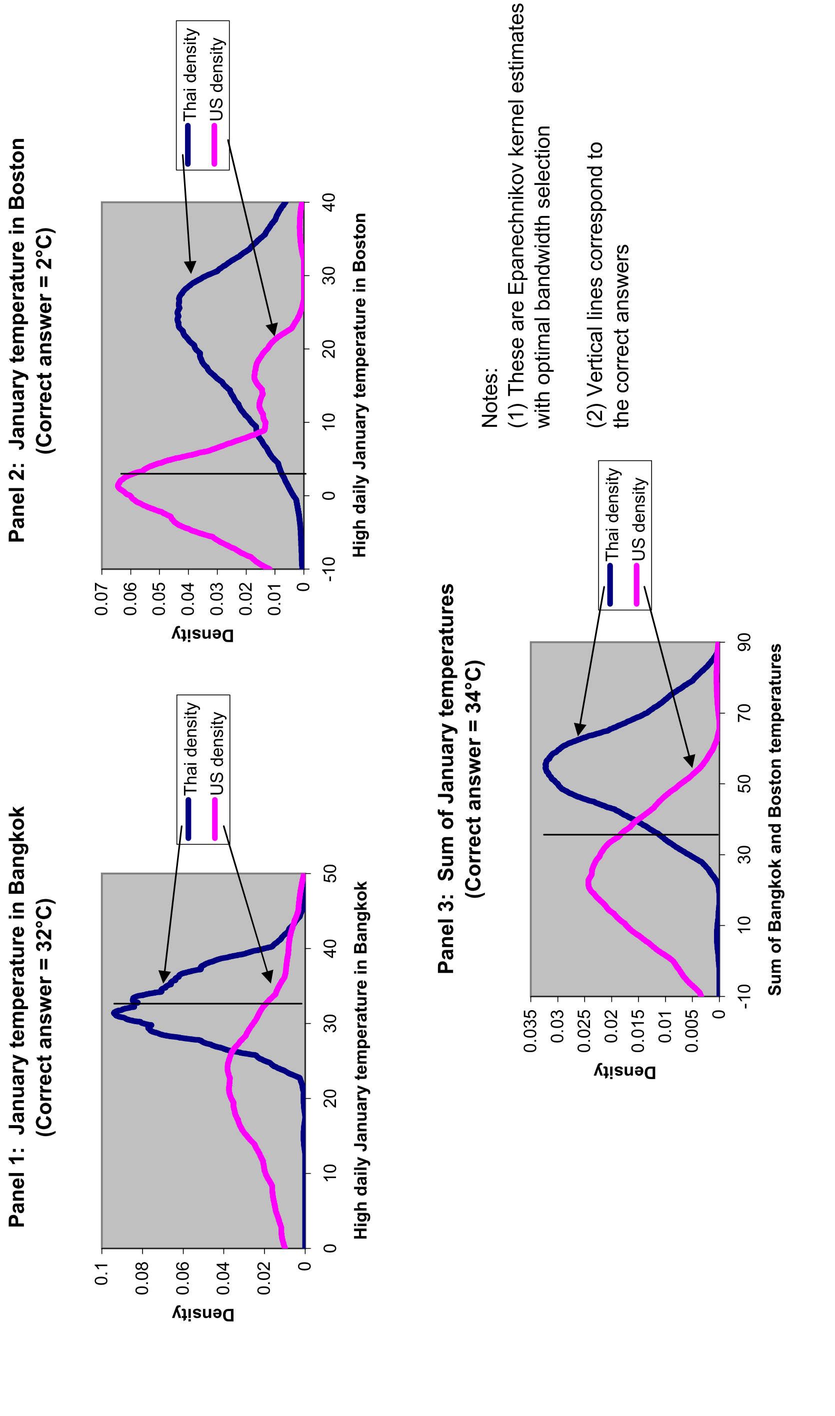

Fig. 3 Kernel density estimates for American and Thai answers about January temperature

Table 1 Relative accuracy of Americans and Thais

Question type	$\frac{\text{Thai MSE}}{\text{Thai MSE+US MSE}} = \frac{\Delta_T^2}{\Delta_T^2+\Delta_A^2}$ (1)	$\frac{\text{Thai MSE}}{\text{US MSE}} = \frac{\Delta_T^2}{\Delta_A^2}$ (2)
Type 1 (Questions about Thailand)	.341	.517
	(.008)	(.018)
Type 2 (Questions about US)	.755	3.086
	(.013)	(.216)
Type 3 (Questions about the sum)	.565	1.299
	(.01)	(.052)

Note: Bootstrapped standard errors in parentheses

close to the correct answer to the US question and the Thai average is not as close to the correct answer to the Thailand question. Also, for some questions, the American and Thai averages are either both above or both below the correct answer. The extra value of information from the other group is thus not an artifact of the experimental design. Thais have more to learn from Americans than from other Thais because, across questions, the American and Thai answer distributions are in different places. In other words, Americans and Thais make different kinds of mistakes and this fact creates a greater opportunity to learn from the other group than from one's own group.

The data provide estimates of the average Thai-to-American MSE ratio for each of the three types of questions. For the questions about Thailand, $\widehat{\frac{\Delta_T^2}{\Delta_A^2}}$ is 0.517, meaning that the expected squared distance between a randomly selected American answer and the correct answer is about twice as large as the expected squared distance between a randomly selected Thai answer and the correct answer. I will describe this kind of result as Thais being twice as accurate as Americans for the questions about Thailand. Table 1 summarizes the relative Thai-to-American accuracy for each of the three question types.[9] Consider the second column in the table. Thais have about one-half the MSE of Americans for the questions about Thailand, while Americans are three times more accurate for the questions about the US and about 1.3 times more accurate for the sum questions. These ratios exactly describe the weights that a subject should use if group bias did not matter. For the sum questions, for example, a subject should put 1.3 times more weight on any observed American answer than she puts on any Thai answer. However, as I show in the next section, the presence of group bias means that a subject should actually put more than twice as much weight on American answers than on Thai answers.

The experimental design also enables me to estimate the share of group bias in total MSE for each question type, both for Americans and for Thais. Table 2 displays

[9]Thai subjects were informed that the answers they observed came from MIT students and Thammasat-Rangsit students. So if subjects had different perceptions about Thammasat-Rangsit than the universe of all Thai subjects in the experiment, it would be appropriate to use only the 130 Thai students from Stage 1 to calculate variances and correlations. Limiting the calculations to the Stage 1 students has almost no effect, and certainly no significant effect, on the parameter estimates.

Table 2 Group effects for Americans and for Thais

Question type	Estimated Thai group bias share (ρ_T) (1)	Estimated American group bias share (ρ_A) (2)
Type 1 (Questions about Thailand)	.234	.307
	(.066)	(.056)
Type 2 (Questions about US)	.362	.227
	(.088)	(.063)
Type 3 (Questions about the sum)	.336	.277
	(.081)	(.062)

Note: Bootstrapped standard errors in parentheses

the estimated group bias shares for each question type. Column (1) contains estimates of the Thai group bias share and column (2) contains estimates of the American group bias share.

Table 2 shows that each group's bias share is higher for the question type that the group knows less well. For Thais, group bias is responsible for the smallest share, 23%, of total MSE for the Bangkok/Thailand questions and the largest share, 36%, for the Boston/US questions. In contrast, group bias is responsible for the smallest share of total American MSE for the Boston/US questions and the largest share for the Bangkok/Thailand questions. For example, Thais make small errors about the average high daily January Bangkok temperature and group bias causes a small share of that error. On the other hand, Thais make much larger errors for the January Boston temperature, and a larger share of their mistakes comes from the fact that the group mean for Thais was 20°C. As I show in Sect. 7, the fact that Thai group bias is biggest for the Boston/US questions has important implications for how a subject would optimally behave when she sees information for the sum question only.

To summarize, the distributions of Thai and American answers show significant group biases for each group for each type of question. The presence of group bias means that American answers have extra value to a Thai subject. An optimizing Thai subject needs to account for group bias when deciding how to weigh the American information she observes compared to the Thai information she observes.

5 Estimating subject behavior

A regression of subjects' final answers after observing information on the initial answers they gave before observing information and the answers that they observed gives estimates of the average weights that subjects put on American answers (β_A), other Thai answers (β_T), and their own initial answers (β_S). To make answers comparable across questions, I standardize each answer by dividing by the mean of the American and Thai standard errors for any given question. I also include dummy variables for the three categories of questions: geography, economics/politics, and demography. Where

Table 3 Summary of estimated weights

Actual weight	Thailand questions (1)	US questions (2)	Sum questions (3)
Own initial answer (β_S)	.653	.464	.731
	(.016)	(.02)	(.019)
Thai average (β_T)	.238	.09	.068
	(.02)	(.03)	(.02)
American average (β_A)	.056	.463	.165
	(.012)	(.019)	(.024)
N	1008	1053	557

Notes:

(a) Regression standard errors are in parentheses

(b) Regressions include dummies for the three question categories

(c) The regressions in columns (1) and (2) each include the cases for which subjects observed information for both Bangkok/Thailand questions and the US questions

y_{iq} = (standardized) final answer given by subject i for question q,
x_{iq} = (standardized) initial answer given by subject i for question q,
$\overline{x}_{iAq}$ = (standardized) average observed American answer for question q,
$\overline{x}_{iTq}$ = (standardized) average observed Thai answer for question q,
C_q = vector of dummy variables for question category for question q.

I estimate the following regression equation:

$$y_{iq} = \beta_S x_{iq} + \beta_A \overline{x}_{iAq} + \beta_T \overline{x}_{iTq} + C_q' \beta_q + \varepsilon_{iq}. \tag{4}$$

The basic results, obtained by estimating (4) for each of the three types of questions, are summarized in Table 3. For the Bangkok/Thailand questions, subjects put a weight of 0.653 on their private beliefs, 0.238 on the observed Thai average, and 0.056 on the observed American average. Thus, the model estimates that subjects assign 4.2 times more weight to the observed Thai answers than to the American answers for the Bangkok/Thailand questions. When subjects observe information about the Boston/US questions, they assign approximately 5.1 times more weight to American answers as to other Thai answers. When subjects observe answers for the sum question, the regression estimates that they assign 2.4 times more weight to American answers than to observed Thai answers.

In addition to these simple regressions, I also consider a variety of robustness checks. To account for heterogeneity at the individual level, I consider regressions that include subject fixed effects. To account for potential heterogeneity across the different sets of information that subjects could observe, I include information set fixed effects. The robustness of the results to the inclusion of these fixed effects indicates that unobserved heterogeneity is not driving the main results.[10] As Table 4

[10] I thank an anonymous reviewer for suggesting this robustness check.

Table 4 Results obtained by including subject and information set fixed effects

Dependent variable: Subjects' final answers						
Regressor	Questions about Thailand		Questions about US		Questions about sum	
	(1)	(2)	(5)	(6)	(9)	(10)
Subject's initial answer	.647	.652	.433	.455	.737	.729
	(.019)	(.016)	(.023)	(.02)	(.027)	(.019)
Thai average	.241	.238	.106	.098	.12	.112
	(.024)	(.024)	(.032)	(.033)	(.028)	(.028)
American average	.046	.046	.489	.489	.175	.18
	(.015)	(.015)	(.021)	(.021)	(.033)	(.033)
Subject fixed effects?	Y	N	Y	N	Y	N
Information set fixed effects?	N	Y	N	Y	N	Y
R-squared	.759	.67	.696	.566	.864	.752
Number of observations	1008	1008	1052	1052	557	557

Notes:

(1) Regression standard errors are in parentheses

(2) Regressions include dummies for the three question categories (geography, economics/politics, and demography)

(3) The regressions in columns (1) and (2) each include the cases for which subjects observed information for both Bangkok/Thailand questions and the US questions

demonstrates, the coefficient estimates do not change significantly when subject and information set fixed effects are included in the regressions.

Then I allow for the weights that subjects use to depend on the number of answers that they observe and to depend on the individual question. Finally, I account for the possibility that subjects may weigh information differently depending on the variance of the information they observe and on the distance of observed information from their initial answers. All of these robustness checks are reported in the Online Appendix (see electronic supplementary material).

It is important to note that subjects improve their earnings considerably by changing their answers after observing information. Compared to subjects who do not observe information for a given question, subjects who observe information earn, on average, 12% more for the Bangkok/Thailand questions, 65% more for the Boston/US questions, and 26% more for the sum questions.

6 Tests for optimal behavior

The model described in Sect. 3 provides estimates of the optimal weights that a subject should use. These estimates come from substituting the estimates of the American-to-Thai MSE ratio, the American group bias share, and the Thai group

Table 5 Estimating the optimal weights

Optimal weight	Questions about Thailand (1)	Questions about US (2)	Questions about sum (3)
Own initial answer (β_S)	.245	.087	.146
	(.009)	(.010)	(.011)
Thai average (β_T)	.480	.168	.300
	(.014)	(.015)	(.02)
American average (β_A)	.274	.745	.554
	(.021)	(.024)	(.028)

Notes:

(a) The estimates come from the parameter estimates in Tables 1 and 2

(b) Bootstrapped standard errors are in parentheses

bias share into the solution for optimal behavior produced by the model.[11] Table 5 displays these estimates of the optimal weights given that $c = 1$, that subjects are not overconfident in their answers.

For the Bangkok/Thailand questions, a subject should apply a weight of 0.245 to her initial answer, 0.480 to the Thai average she observes, and 0.274 to the American average she observes. For the Boston/US questions, the corresponding weights are 0.087, 0.168, and 0.745. For the sum questions, the model estimates that a subject would optimally choose 0.146 for the self-weight, 0.300 for the weight given to observed Thai answers, and 0.554 for the weight given to observed American answers. Bootstrapping gives the standard errors for these estimates.

6.1 Construction of confidence intervals

Simulations using the regression coefficients and their variance-covariance matrix obtained by estimating (4) give a confidence interval for $\frac{\beta_A}{\beta_T}$, the weight ratio that expresses how subjects actually weigh American compared to Thai information. Also, the distributions of Thai and American answers make it possible to generate confidence intervals for weight ratios that subjects would use under different hypotheses that could explain their behavior. I focus on confidence intervals for two such weight ratios: 1) the simple weight ratio, which expresses how a subject would behave if she understood each group's accuracy but ignored group bias, and 2) the optimal weight ratio described in equation (3), which expresses how a subject would behave if she correctly perceived each group's accuracy and accounted for group bias. The parameter estimates in Tables 1 and 2 give confidence intervals for the simple weight ratio and for the optimal weight ratio for any value of the overconfidence parameter c.

[11] The optimal weights can also be estimated by replacing a subject's final answer in (4) with the correct answer to the question. The weights obtained in this way are nearly the same as those produced by the model.

To summarize, I consider the following three weight ratios:

$$\text{Actual} = \frac{\beta_A}{\beta_T},$$

$$\text{Simple} = \left(\frac{\Delta_T^2}{\Delta_A^2}\right),$$

$$\text{Optimal} = \left(\frac{\Delta_T^2}{\Delta_A^2}\right)\left(\frac{1+(n_T-1)\rho_T - c\rho_T^2 n_T}{(1+(n_A-1)\rho_A)(1-\rho_T)}\right).$$

Table 5 reports confidence intervals for these three ratios for each of the three question types. For the optimal weight ratio, I consider three values of the overconfidence parameter: $c = 0.25$, $c = 0.5$, and $c = 1$.[12] The p-values in the table correspond to tests that I describe in the next subsection.

As the table shows, the optimal weight ratio increases substantially due to group bias. The optimal weight ratio also increases as overconfidence increases, but not as much. Consider the Boston/US questions. With no overconfidence, the presence of group bias causes the optimal weight ratio to increase from 3.09 to 4.42. Increasing overconfidence by dropping c from 1 to 0.25 causes the optimal weight ratio to further rise to 5.04.

For the Bangkok/Thailand questions, the simple and optimal models are both rejected. Subjects put excessive weight on observed Thai information relative to observed American information for these questions that are specifically about a Thai's area of expertise. For the other two types of questions, however, while it is possible to reject the simple model, the optimal model cannot be rejected and corresponds closely to actual subject behavior. This non-rejection of the optimal model for the Boston/US and sum questions is only consistent with the optimal model describing subject behavior. In Table 7 in Sect. 7, I use information about how subjects update their answers when they observe information about the sum question to reject alternative explanations for subject behavior.

6.2 Hypothesis tests

By looking at how subjects weigh American information relative to Thai information, I can test a variety of hypotheses relating to subjects' perceptions about both the accuracy of American answers relative to Thai answers and the extent of group bias for each group. Consider the hypothesis, H_0, that subjects correctly perceive the accuracy of Thais relative to Americans, but ignore group bias. Under this hypothesis, subject behavior will reflect the following perceptions:

$$H_0\colon \quad \left(\frac{\Delta_T^2}{\Delta_A^2}\right)_{perceived} = \left(\frac{\Delta_T^2}{\Delta_A^2}\right), \quad (\rho_T)_{perceived} = 0, \quad (\rho_A)_{perceived} = 0.$$

[12] A nonlinear model can be used to estimate c, but the standard errors on the resulting estimates are large. The estimated value of c falls between 0.20 and 0.45, on average.

Table 6 Comparing how subjects relatively weigh American and Thai information

	Thailand questions (1)	US questions (2)	Sum questions (3)
Actual weight ratio $\left(\frac{\beta_A}{\beta_T}\right)$	.231	5.143	2.49
	(.131, .352)	(3.144, 15.17)	(1.45, 5.366)
Simple weight ratio $\left(\frac{\Delta_A^2}{\Delta_T^2}\right)$	.517	3.086	1.299
	(.477, .554)	(2.737, 3.577)	(1.205, 1.406)
	$p = 0.000$	$p = 0.065$	$p = 0.017$
Optimal weight ratio $\left(\frac{\lambda_A}{\lambda_T}\right)$			
$c = 1$	.571	4.423	1.843
	(.475, .694)	(3.467, 5.63)	(1.466, 2.332)
	$p = 0.000$	$p = 0.622$	$p = 0.346$
$c = 0.5$	.606	4.897	2.048
	(.476, .759)	(3.62, 6.938)	(1.549, 2.877)
	$p = 0.000$	$p = 0.879$	$p = 0.572$
$c = 0.25$	.616	5.125	2.135
	(.468, .798)	(3.662, 7.592)	(1.525, 2.991)
	$p = 0.000$	$p = 0.986$	$p = 0.676$

Notes:

(a) Bootstrapped 95% confidence intervals are in parentheses

(b) The bootstrap for the actual weights accounts for correlation in the coefficient estimates

(c) p-values compare the given weight ratio to the actual weight ratio

Under H_0, as shown in Sect. 3, subjects will choose

$$\frac{\beta_A}{\beta_T} = \left(\frac{\Delta_T^2}{\Delta_A^2}\right). \tag{5}$$

Rejection of the prediction (5) implies rejection of H_0.

The second row of Table 6 displays the results of testing for the equality of the ratios in (5) for all three types of questions. For the Bangkok/Thailand questions and the sum questions, we can reject equality at the 5% level ($p < .001$ and $p = 0.017$, respectively). For the Boston/US questions, we can reject it at the 10% level ($p = 0.065$). We reject the hypothesis for the Bangkok/Thailand questions due to subjects choosing too low a weight for American answers relative to Thai answers. It is rejected for the Boston/US and sum questions due to subjects relatively overweighing American answers. For all three types of questions, the results reject the simple model, in which subjects both correctly perceive the groups' relative accuracies and simultaneously ignore group bias.

Now consider the hypothesis of optimal behavior, H_1:

$$H_1\colon \quad \left(\frac{\Delta_T^2}{\Delta_A^2}\right)_{perceived} = \left(\frac{\Delta_T^2}{\Delta_A^2}\right), \quad (\rho_T)_{perceived} = \rho_T, \quad (\rho_A)_{perceived} = \rho_A.$$

This hypothesis states that subjects correctly perceive the MSE of Thais relative to Americans and also correctly account for group bias. Compared to a subject who behaves according to H_0, a subject who behaves according to H_1 will put more weight on American answers because she appreciates the value of an American's independent perspective to a Thai subject.

Under H_1, subjects will choose the optimal weight ratio

$$\frac{\beta_A}{\beta_T} = \left(\frac{\Delta_T^2}{\Delta_A^2}\right)\left(\frac{1+(n_T-1)\rho_T - c\rho_T^2 n_T}{(1+(n_A-1)\rho_A)(1-\rho_T)}\right). \tag{6}$$

Table 6 displays the results of the above test for a variety of possible values of the overconfidence parameter. For the Bangkok/Thailand questions, H_1 is rejected. For all values of overconfidence, the test gives a p-value of nearly zero. Thais put too little weight on American answers in this case.[13] On the other hand, for the Boston/US and sum questions, we cannot reject H_1 for any level of overconfidence. For $c = 1$, the optimal weight ratio estimates are 4.42 and 1.84, compared to the actual weight ratio estimates of 5.14 and 2.49. Tests of equality give p-values of 0.622 and 0.346, respectively.

Now consider the optimal weight ratios for the Boston/US and sum questions when overconfidence is taken into account. Given $c = 0.25$, the actual and optimal weight ratios match up quite closely. For the Boston/US questions, the optimal weight ratio estimate is 5.13 and the actual weight ratio estimate is 5.14. The test for equality between the two, not surprisingly, gives a p-value of nearly one ($p = 0.986$). For the sum questions, the optimal weight ratio estimate is 2.14, compared to the actual weight ratio of 2.49 ($p = 0.676$).

While the optimal model is consistent with subject behavior for the Boston/US and sum questions, a different hypothesis could still explain subject behavior. Under this hypothesis, H_2, subjects perceive Americans to be better than they actually are compared to Thais and they ignore group bias:

$$H_2\colon \quad \left(\frac{\Delta_T^2}{\Delta_A^2}\right)_{perceived} > \left(\frac{\Delta_T^2}{\Delta_A^2}\right), \quad (\rho_T)_{perceived} = 0, \quad (\rho_A)_{perceived} = 0.$$

Under H_1, subjects put extra weight on American information because they understand the additional value in American information that comes from the fact that

[13] For the questions about Bangkok or Thailand, subjects should put a high weight on observed Thais compared to observed Americans, but they choose an even higher relative weight than they optimally would. Given that subjects use the optimal weights for the sum questions, which include the Thailand questions, it seems likely that the Thailand heading for the questions about Thailand causes subjects to underweigh American information. In other words, only when the question is clearly about a Thai's area of expertise do the subjects make significant mistakes.

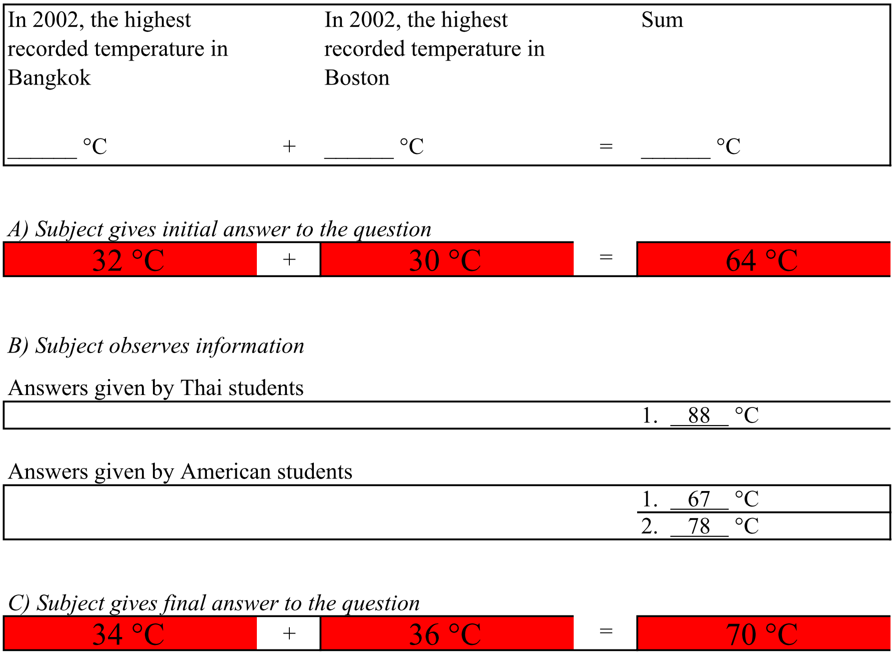

Fig. 4 Updating answers based on information about the sum question

Americans and Thais make different kinds of mistakes. Under H_2, subjects put extra weight on American information because they incorrectly perceive American answers to be better than they really are.

7 Do subjects value the independence in outside information?

The presence of the sum questions makes it possible to distinguish between H_1 and H_2, and thus to reject the hypothesis that subjects fail to value the independence in outside information. Specifically, the test that distinguishes between these hypotheses uses the data generated by the case in which subjects observe information about the sum question only. Consider the questions about the highest temperature in 2002 in Bangkok and Boston, as shown in Fig. 4. In part A of the figure, the subject gives her initial answers to the questions. In part B, she observes information about others' opinions for the sum question only. In part C, she uses the observed information to update her answers for both the question about Bangkok temperature and the question about Boston temperature. In general, by looking at how subjects *separately* update their answers for the Thailand questions and the US questions, it is possible to test the hypothesis that subjects ignore group bias when accumulating the knowledge that they observe.

To explain this test, it is necessary to expand the earlier notation that applied when each question type was considered separately.[14] Define:

$\rho_{j,k}$ = group bias share in total MSE for group j for question type k,
$\Delta^2_{j,k}$ = mean-squared error for group j for question type k.

For example, $\rho_{T,US}$ is the group bias share for Thais answering the Boston/US questions.

Define $(\frac{\phi_A}{\phi_T})_{Thai}$ to be the weight ratio that subjects assign to American answers to the sum question relative to Thai answers to the sum question when they update for the Bangkok/Thailand question. In Fig. 4, this weight ratio captures the weight that subjects put on the observed American answers (67°C and 78°C) relative to the observed Thai answer (88°C) when they revise their answers for the Bangkok/Thailand question (here, the subject updates from 32°C to 34°C). Analogously, define $(\frac{\phi_A}{\phi_T})_{US}$ to be the weight ratio that subjects assign to American answers relative to Thai answers to the sum question when they update for the Boston/US question. In Fig. 4, this ratio describes the weight subjects use for the observed American answers (67°C and 78°C) relative to the observed Thai answer (88°C) when they decide how to update for the Boston/US question (here, the subject updates from 30°C to 36°C).

Consider the following proposition:

Proposition 5 *If* $(\rho_{T,Thai})_{perceived} = (\rho_{T,US})_{perceived} = 0$, *a subject who observes information for the sum question will choose* $(\frac{\phi_A}{\phi_T})_{Thai} = (\frac{\phi_A}{\phi_T})_{US}$.

Proof See the Appendix. □

Proposition 5 refers to subjects who implicitly perceive there to be no Thai group bias for both the Bangkok/Thailand and Boston/US questions. When observing information about the sum question, these subjects will use the same weight ratio to update for the Bangkok/Thailand questions as they use to update for the Boston/US questions.

Consider the hypothesis, G_0, that the perceived group bias shares are zero.

$$G_0\colon \quad (\rho_{T,Thai})_{perceived} = (\rho_{T,US})_{perceived} = 0.$$

This hypothesis states that subjects ignore Thai group bias for the Bangkok/Thailand and Boston/US questions. Notice that rejection of G_0 would imply rejection of H_2, and that G_0 puts no restrictions on subjects' perceptions of the groups' relative accuracies.

Under G_0, as stated in Proposition 5, subjects will choose

$$\left(\frac{\phi_A}{\phi_T}\right)_{Thai} = \left(\frac{\phi_A}{\phi_T}\right)_{US}. \tag{7}$$

[14] For the sake of simplicity, I do not consider individual-level heterogeneity in answering the question in this extension of the model, although doing so leads to the same predictions for average behavior across subjects.

Table 7 Updating for the Thai and US questions after observing answers for the sum

Dependent variables: (1) Final answer for Thai question (2) Final answer for US question		
Regression weights	Thailand questions (1)	US questions (2)
ϕ_T = Distance between observed Thai average	.056	.081
and initial answer (*for sum question*)	(.013)	(.019)
ϕ_A = Distance between observed American average	.075	.226
and initial answer (*for sum question*)	(.014)	(.022)
p-value for test of $(\frac{\phi_A}{\phi_T})_{Thai} = (\frac{\phi_A}{\phi_T})_{US}$	0.058	
N	548	544

Notes:

(a) Regression standard errors are in parentheses

(b) Regressions include dummies for question categories (meteorology, economic/political, and demography)

Rejection of the equality in (7) would imply rejection of G_0. The following two regressions give the parameter estimates needed to conduct a test of the equality of the ratios in (7). The first equation expresses the change in a subject's answer for the Thai question as a function of the distance between the average observed American answer for the sum question and her own answer for the sum question and the distance between the average observed Thai answer for the sum question and her own answer for the sum question. The second equation expresses how subjects update their answers for the US question after observing information relating to the sum question.

The regression equations are:

$$y_{iq,Thai} - x_{iq,Thai} = \phi_{A,Thai}(\overline{x}_{iAq,Sum} - x_{iq,Sum}) + \phi_{T,Thai}(\overline{x}_{iTq,Sum} - x_{iq,Sum}) + C_q'\phi_1 + \varepsilon_{iq1}, \tag{8}$$

$$y_{iq,US} - x_{iq,US} = \phi_{A,US}(\overline{x}_{iAq,Sum} - x_{iqs,Sum}) + \phi_{T,US}(\overline{x}_{iTq,Sum} - x_{iq,Sum}) + C_q'\phi_2 + \varepsilon_{iq2}. \tag{9}$$

Table 7 reports the estimated weight ratios obtained by estimating equations (8) and (9). Notice that $(\frac{\hat{\phi}_A}{\hat{\phi}_T})_{Thai} = \frac{.075}{.056} = 1.34$ and $(\frac{\hat{\phi}_A}{\hat{\phi}_T})_{US} = \frac{.226}{.081} = 2.79$. The regression results provide the inputs needed to test the equality in (7). The p-value for that test is .058.[15] At a 10% level, we thus reject G_0, the hypothesis that subjects fail to take group bias into account, regardless of how they perceive American and Thai accuracy.

In contrast, correctly accounting for group bias would lead subjects to behave in a similar way to how they actually do behave. If the perceived Thai group bias share for

[15]The variance–covariance matrix of all the regression coefficients is used to determine the p-value for the test of equality.

the Boston/US questions ($\rho_{T,US}$) is greater than the perceived Thai group bias share for the Bangkok/Thailand questions ($\rho_{T,Thai}$), then subjects will put a higher relative weight on observed Americans for the US questions than for the Thailand questions. The proof of Proposition 5 demonstrates that:

$$\rho_{T,US} > \rho_{T,Thai} \Rightarrow \left(\frac{\phi_A}{\phi_T}\right)_{Thai} < \left(\frac{\phi_A}{\phi_T}\right)_{US}.$$

To understand the intuition, consider a subject updating her answer for the question about Thailand after observing information about the sum question. If Thai group bias for the questions about Thailand ($\rho_{T,Thai}$) is high, she should put less weight on Thais relative to Americans because each additional Thai answer contains little new information. On the other hand, if Thai group bias for the Boston/US questions ($\rho_{T,US}$) is high, she should put *higher* weight on observed Thai answers for the sum question. When $\rho_{T,US}$ is high, Thai subjects have a better idea of what other Thai answers about the sum mean for what those observed students believe about the Thai question. For example, if $\rho_{T,US}$ was equal to one, all Thais would give the same answer to the US question. Then, a subject could exactly deduce an observed Thai's private belief about the Bangkok/Thailand question from her answer to the sum question.

The data show that the effects of group bias can explain how subjects actually weigh the information they observe. If subjects applied the actual variance estimates and actual group bias shares from Tables 1 and 2, they would choose $(\frac{\hat{\phi}_A}{\hat{\phi}_T})_{Thai} = 1.51$ and $(\frac{\hat{\phi}_A}{\hat{\phi}_T})_{US} = 2.53$, similar ratios to the estimates of optimal behavior of 1.34 and 2.79.

In summary, the results expressed in Tables 5 and 6 provide tests of the three main hypotheses (H_0, H_1, and H_2) for the Boston/US and sum questions. H_0, the hypothesis that subjects correctly estimate the accuracy of Americans relative to Thais and ignore group bias, is rejected in Table 6 because subjects put too much weight on observed Americans relative to its predictions. In Table 7, I reject the general hypothesis (G_0) that subjects ignore group bias for *any* perceptions about how accurate Americans are relative to Thais. Rejection of this hypothesis also implies rejection of H_2, the hypothesis that subjects overestimate American accuracy and ignore group bias. Left standing is H_1, the hypothesis that subjects weigh information correctly because they appreciate the extra value in an American's independent perspective to a Thai subject.

8 Conclusion

This paper demonstrates that economic agents can learn effectively from a variety of information sources. The experimental design made it possible to demonstrate that subjects achieved optimality due to their abilities to appreciate subtle statistical ideas. Specifically, subjects appear to appreciate the implications of the fact that members of the same group tend to make similar kinds of mistakes. This fact means that subjects have more to learn from out-group members than from other members of their own group, even when each group is equally good at answering the question.

The results suggest the importance of forming diverse groups to solve problems, since subjects appear to appreciate the value of outside advice when they are aware of it. Unfortunately, the desire for cohesiveness often prevents diverse groups from being formed, since homogeneous groups become close-knit more easily (Janis 1972). Since the experimental results reported here illustrate that when agents listen to a diverse group of opinions, they can be expected to carefully consider the available information, additional experimental results would be useful to investigate the issue of why people fail to consult outside opinions, and what can be done to encourage people to obtain a variety of different opinions before making decisions.

Finally, the experimental results reported here also help to explain previous experimental results on advice. Subjects in the experiment not only appear to learn from the valuable information they observe, they also recognize the nature of the mistakes that advisors make and adjust their weights accordingly. The results of this experiment thus suggest that advised subjects outperform unadvised ones in part due to an ability to recognize the bias in advisors' decisions.

Acknowledgements I thank Victor Chernozhukov, Esther Duflo, and Sendhil Mullainathan for their advice and support. Special thanks go to Adis Israngkura, Chayun Tantisidivakarn, and Somchai Jitsuchon for helping to arrange the experiment. I thank Ken Chetnakarnkul and Phasook Mengkred for dedicated research assistance. For helpful discussions, I acknowledge and thank Dan Ariely, Gary Charness, Shawn Cole, Ernst Fehr, Jerry Hausman, Dorothea Herreiner, Peter Hinrichs, Dean Karlan, John List, Whitney Newey, Daniel Paravisini, Al Roth, Chinda Wongngamnit, and seminar participants at the University of Essex, MIT, the 2005 ESA meetings, and the 2007 AEA meetings. I am grateful for financial support from a Russell Sage Small Grant in Behavioral Economics and from an MIT Shultz Fund grant. I also thank the National Science Foundation for a graduate fellowship. All errors are my own.

Appendix

A.1 Proofs

A.1.1 Proof of Proposition 1

From the definition of the MSE for group j for question q:

$$\Delta_{jq}^2 = \frac{1}{N_j}\sum_{i=1}^{N_j}(x_{ijq} - \overline{x}_{jq} + \overline{x}_{jq} - Truth_q)^2.$$

Expanding the expression gives:

$$\Delta_{jq}^2 = \frac{1}{N_j}\sum_{i=1}^{N_j}\left[(x_{ijq} - \overline{x}_{jq})^2 + 2(x_{ijq} - \overline{x}_{jq})(\overline{x}_{jq} - Truth_q) + (\overline{x}_{jq} - Truth_q)^2\right].$$

Since $\overline{x}_{jq} = \frac{1}{N_j}\sum_{i=1}^{N_j} x_{ijq}$, the middle term drops out, giving the result:

$$\Delta^2_{jq} = \frac{1}{N_j}\sum_{i=1}^{N_j}(x_{ijq} - \overline{x}_{jq})^2 + (\overline{x}_{jq} - Truth_q)^2 = s^2_{jq} + \alpha^2_{jq}.$$

A.1.2 Proof of Proposition 2

Consider group j (either A or T). For a given question q, the MLE for the true mean-squared error (MSE) of group j answers is:

$$\widehat{\sigma^2_{jq}} = \frac{1}{N_j}\sum_{i=1}^{N_j}(x_{ijq} - Truth_q)^2, \tag{10}$$

where $\overline{x}_{jq}$ is the mean answer for group j members for question q, the MLE for the population variance is

$$s^2_{jq} = \frac{1}{N_j}\sum_{i=1}^{N_j}(x_{ijq} - \overline{x}_{jq})^2. \tag{11}$$

Since the sample variance, s^2_{jq}, is the part of total MSE that does not come from group bias and the MLE of a function is the function of the MLEs, the sample variance can be expressed as:

$$s^2_{jq} = (1 - \widehat{\rho}_{jq})\widehat{\sigma^2_{jq}}.$$

Substitution of (10) into (11) then gives

$$1 - \widehat{\rho}_{jq} = \frac{\sum_{i=1}^{N_j}(x_{ijq} - \overline{x}_{jq})^2}{\sum_{i=1}^{N_j}(x_{ijq} - Truth_q)^2} \Rightarrow \widehat{\rho}_{jq} = \frac{\sum_{i=1}^{N_j}(\overline{x}_{jq} - Truth_q)^2}{\sum_{i=1}^{N_j}(x_{ijq} - Truth_q)^2} = \frac{\widehat{\alpha^2_{jq}}}{\widehat{\sigma^2_{jq}}}.$$

Since the questions are assumed to be independent, the MLE for ρ_j is the mean of $\widehat{\rho}_{jq}$ across questions. Likewise, the MLE for $\frac{\Delta^2_A}{\Delta^2_T}$ is the mean of $\widehat{\frac{\Delta^2_{Aq}}{\Delta^2_{Tq}}}$ across questions.

A.1.3 Proof of Proposition 3

Assuming that $n_T\lambda_{Ti} + n_A\lambda_{Ai} + \lambda_{si} = 1$ so that the sum of the weights put on all pieces of information is one, the expected value of a subject's MSE can be expressed as:

$$E(MSE) = E\left(\lambda_{Ti}\sum_{k=1}^{n_T}(x_{Tk} - Truth) + \lambda_{Ai}\sum_{k=1}^{n_A}(x_{Ak} - Truth) + \lambda_{si}(x_{si} - Truth)\right)^2.$$

Also assume the group biases are uncorrelated, so that

$$E((x_{Tk_T} - Truth)(x_{Ak_A} - Truth)) = 0,$$

as holds true in the experimental data. Then expanding the expression for MSE gives

$$E(MSE) = n_T\lambda_{Ti}^2\Delta_T^2 + n_T(n_T-1)\lambda_{Ti}^2\rho_T\Delta_T^2 + n_A\lambda_{Ai}^2\Delta_A^2 + n_A(n_A-1)\lambda_{Ai}^2\rho_A\Delta_A^2 + 2cn_T\lambda_{Ti}\lambda_{si}\rho_T\Delta_T^2 + c\lambda_{Ti}^2\Delta_T^2.$$

Taking derivatives with respect to the weights gives the expression in Proposition 3.

A.1.4 Proof of Proposition 5

Consider the case when a subject uses observed answers for the sum question to update her answer for the Bangkok/Thailand question. To minimize notation, the subscripts for the group bias terms have been shortened so that, for example, $\rho_{T,Thai}$ is written as $\rho_{T,t}$. When she observes n_A American answers and n_T Thai answers, her expected mean-squared error is

$$\begin{aligned}
E(MSE) &= E\Big(\phi_T\left((x_{Ts1}-x_{is})+\cdots+(x_{Tsn_T}-x_{is})\right)+\phi_A\big((x_{As1}-x_{is})+\cdots \\
&\quad + (x_{Asn_A}-x_{is})\big)+x_{it}-\mu_t\Big)^2 \\
&= E\left(\phi_T\sum_{k=1}^{n_T}(\varepsilon_{Ttk}+\varepsilon_{Tak})+\phi_A\sum_{k=1}^{n_A}(\varepsilon_{Atk}+\varepsilon_{Aak})\right. \\
&\quad \left. -(n_T\phi_T+n_A\phi_A)(\varepsilon_{it}+\varepsilon_{ia})+\varepsilon_{it}\right)^2 \\
&= n_T\phi_T^2\left[(\Delta_{Tt}^2+\Delta_{Ta}^2)+(n_T-1)(\rho_{Tt}\Delta_{Tt}^2+\rho_{Ta}\Delta_{Ta}^2)\right] \\
&\quad + n_A\phi_A^2[\Delta_{At}^2+\Delta_{Aa}^2+(n_A-1)(\rho_{At}\Delta_{At}^2+\rho_{Aa}\Delta_{Aa}^2)] \\
&\quad + (1-n_T\phi_T-n_A\phi_A)^2\alpha\Delta_{Tt}^2+(n_T\phi_T+n_A\phi_A)^2\alpha\Delta_{Ta}^2 \\
&\quad + 2(1-n_T\phi_T-n_A\phi_A)n_T\phi_T\alpha\rho_{Tt}\Delta_{Tt}^2 \\
&\quad - 2(n_T\phi_T+n_A\phi_A)n_T\phi_T\alpha\rho_{Ta}\Delta_{Ta}^2,
\end{aligned}$$

where ϕ_A is the weight given to American answers for the sum question and ϕ_T is the weight given to other Thai answers for the sum question.

Taking the derivatives with respect to ϕ_A and ϕ_T gives the optimal weights. The ratio $\frac{\phi_A}{\phi_T}$ can be expressed as a function of the optimal weight ratios derived in Proposition 4 for how subjects should weigh American information relative to Thai information for the Bangkok/Thailand questions and the Boston/US questions: $(\frac{\lambda_A}{\lambda_T})_{Thai}$ and $(\frac{\lambda_A}{\lambda_T})_{US}$. The optimal weight ratio is

$$\left(\frac{\phi_A}{\phi_T}\right)_{Thai} = \frac{\left(\frac{\lambda_A}{\lambda_T}\right)_{Thai}+\left(\frac{\lambda_A}{\lambda_T}\right)_{US}\frac{y_A}{y_T}+\frac{cn_T\Delta_{Ta}^2(\rho_{Tt}-\rho_{Ta})(1-\rho_{Ta})}{y_T}}{1+\frac{y_A}{y_T}\frac{1-\rho_{Tt}}{1-\rho_{Ta}}+\frac{cn_T\Delta_{Ta}^2(\rho_{Ta}-\rho_{Tt})}{y_T}},$$

where $y_A = (1 + (n_A - 1)\rho_{Aa})(1 - \rho_{Ta})$ and $y_T = (1 + (n_A - 1)\rho_{At})(1 - \rho_{Tt})$. If the perceived group bias shares for Thais for the Thailand and US questions, ρ_{Tt} and ρ_{Ta}, are zero, this reduces to the following expression:

$$\left(\frac{\phi_A}{\phi_T}\right)_{Thai} = \frac{\Delta_{Tt}^2 + \Delta_{Ta}^2}{\Delta_{At}^2 + \Delta_{Aa}^2}.$$

The same line of reasoning implies that, if ρ_{Tt} and ρ_{Ta} are zero,

$$\left(\frac{\phi_A}{\phi_T}\right)_{US} = \frac{\Delta_{Tt}^2 + \Delta_{Ta}^2}{\Delta_{At}^2 + \Delta_{Aa}^2}.$$

This equation gives the desired result:

$$(\rho_{Tt})_{perceived} = (\rho_{Ta})_{perceived} = 0 \Rightarrow \left(\frac{\phi_A}{\phi_T}\right)_{Thai} = \left(\frac{\phi_A}{\phi_T}\right)_{US}.$$

A.2 Derivation of additional weight ratios

In the paper, I describe the weight ratio predicted by the model for how subjects will use observed American information relative to observed Thai information. Here, I describe the analogous weight ratios for how subjects will weigh their initial answers relative to the Thai answers they observe and for how they will weigh their initial answers relative to the American answers they observe.

Proposition 6 *The following expressions define the MSE-minimizing weights that subjects should use to evaluate their initial answers relative to the opinions they observe*:

Weighing Americans relative to self: $$\frac{\lambda_{Ai}}{\lambda_{si}} = \left(\frac{\Delta_T^2}{\Delta_A^2}\right)\left(\frac{c\phi_i}{1 - \rho_A - \rho_A n_A}\right),$$

Weighing other Thais relative to self: $$\frac{\lambda_{Ti}}{\lambda_{si}} = \frac{(1 - \rho_T)c\phi_i}{1 - \rho_T + \rho_T n_T - \rho_T n_T c\phi_i}.$$

Proof The following expression defines the expected mean squared error for a subject as a function of the weights she uses:

$$E(MSE) = n_T\lambda_{Ti}^2\Delta_T^2 + n_T(n_T - 1)\lambda_{Ti}^2\rho_T\Delta_T^2 + n_A\lambda_{Ai}^2\Delta_A^2 + n_A(n_A - 1)\lambda_{Ai}^2\rho_A\Delta_A^2 + 2cn_T\lambda_{Ti}\lambda_{si}\rho_T\Delta_T^2 + c\lambda_{Ti}^2\Delta_T^2.$$

Taking derivatives with respect to the weights, setting equal to zero, and solving, gives the above expressions. □

To determine the average weight ratios that would be used under the model requires averaging the above equations across subjects. The first equation is linear in the individual specific parameter, so that averaging across subjects leads to a simple

Table 8 Testing for anchoring

Dependent variables: (1) Squared distance from observed Thai answers
(2) Squared distance from observed American answers
Independent variable: Dummy for anchoring treatment group

	Distance from Thai answers (1)	Distance from American answers (2)
Questions about Thailand	−.013	.042
	(.021)	(.078)
Questions about the US	−.051	−.046
	(.065)	(.022)
Questions about sum	−.026	−.049
	(.049)	(.07)

Note: Regression standard errors are reported in parentheses

result, given that the mean of the individual accuracy parameter is one by definition. While the second equation is not linear in ϕ_i, the earlier equation derived for $\frac{\lambda_A}{\lambda_T}$ and the expression below for $\frac{\lambda_A}{\lambda_s}$, combined with the condition that the weights must sum to one leads to an expression for $\frac{\lambda_T}{\lambda_s}$ that is also independent of ϕ_i.

Proposition 7 *The following expressions define the average weights that will be used across subjects under the model to evaluate their initial answers relative to the opinions they observe*:

$$\textit{Weighing Americans relative to self:} \quad \frac{\lambda_A}{\lambda_s} = \left(\frac{\Delta_T^2}{\Delta_A^2}\right)\left(\frac{c}{1-\rho_A-\rho_A n_A}\right),$$

$$\textit{Weighing other Thais relative to self:} \quad \frac{\lambda_T}{\lambda_s} = \frac{(1-\rho_T)c}{1-\rho_T+\rho_T n_T-\rho_T n_T c}.$$

Proof The first result is straightforward, since the expression is linear in ϕ_i and the mean of ϕ_i is one by definition. Combining the first result with the result in (3), along with the condition that $\lambda_s + n_T\lambda_T + n_A\lambda_A = 1$ gives the second result. □

The above expressions, along with (3) are used to generate the optimal weights in Table 5.

A.3 Testing for anchoring

To test for anchoring, I consider the following regressions:

$$(y_{iq} - \overline{x}_{iTq})^2 = \theta_{1T} Anchor_i + Version'_{iv}\theta_{2T} + \varepsilon_{iq}, \tag{12}$$

$$(y_{iq} - \overline{x}_{iAq})^2 = \theta_{1A} Anchor_i + Version'_{iv}\theta_{2A} + \varepsilon_{iq}. \tag{13}$$

The first (second) regression looks at the distance between subjects' final answers and the Thai (American) answers they observe. $Anchor_i$ is a dummy that is one for the 42 subjects in the anchoring group, and $Version'_{iv}$ is a vector of information set fixed effects. Including the information set fixed effects creates a comparison between the anchoring group subjects and main group subjects who observed the same information.

If subjects anchor, then $\theta_{1T} < 0$ and $\theta_{1A} < 0$, so that subjects who do not provide their private beliefs before observing answers end up closer to those observed answers than subjects who answer the questions on their own first. As seen in Table 8, the coefficients are always close to zero and insignificant except for the distance from the American average for the Boston/US questions.

References

Anderson, L., & Holt, C. (1997). Information cascades in the laboratory. *American Economic Review*, *87*, 847–862.
Austen-Smith, D., & Banks, J. (1996). Information aggregation, rationality, and the Condorcet Jury Theorem. *American Political Science Review*, *90*, 34–45.
Camerer, C., & Lovallo, D. (1999). Overconfidence and excess entry: an experimental approach. *American Economic Review*, *89*, 306–318.
Duflo, E., & Saez, E. (2003). The role of information and social interactions in retirement plan decisions: evidence from a randomized experiment. *Quarterly Journal of Economics*, *118*, 815–842.
Feddersen, T., & Pesendorfer, W. (1997). Voting behavior and information aggregation in elections with private information. *Econometrica*, *65*, 1029–1058.
Forsythe, R., & Lundholm, R. (1990). Information aggregation in an experimental market. *Econometrica*, *58*, 309–347.
Foster, A., & Rosenzweig, M. (1995). Learning by doing and learning from others: human capital and technical change in agriculture. *Journal of Political Economy*, *103*, 1176–1209.
Gigerenzer, G., Hoffrage, U., & Kleinbolting, H. (1991). Probabilistic mental models: a Brunswikian theory of confidence. *Psychological Review*, *98*, 506–528.
Goeree, J., Palfrey, T., Rogers, B., & McKelvey, R. (2007). Self-correcting information cascades. *Review of Economic Studies*, *74*, 733–762.
Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology*, *78*, 1360–1380.
Hellwig, M. F. (1980). On the aggregation of information in competitive markets. *Journal of Economic Theory*, *22*, 477–498.
Hoelzl, E., & Rustichini, A. (2005). Overconfident: do you put your money on it? *Economic Journal*, *115*, 305–318.
Iyengar, R., & Schotter, A. (2008). Learning under supervision: an experimental study. *Experimental Economics*, *11*, 154–173.
Janis, I. L. (1972). *Victims of groupthink: A psychological study of foreign policy decisions and fiascoes*. Boston: Houghton-Mifflin.
Kraemer, C., Noth, M., & Weber, M. (2006). Information aggregation with costly information and random ordering: experimental evidence. *Journal of Economic Behavior and Organization*, *59*, 423–432.
Lohmann, S. (1994). Information aggregation through costly political action. *American Economic Review*, *84*, 518–530.
Miguel, E., & Kremer, M. (2004). Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica*, *72*, 159–217.
Munshi, K. (2004). Social learning in a heterogeneous population: technology diffusion in the Indian Green Revolution. *Journal of Development Economics*, *73*, 185–213.
Munshi, K., & Myaux, J. (2006). Social norms and the fertility transition. *Journal of Development Economics*, *80*, 1–38.
Pesendorfer, W., & Swinkels, J. (1997). The loser's curse and information aggregation in common value auctions. *Econometrica*, *65*, 1247–1281.

Pesendorfer, W., & Swinkels, J. (2000). Efficiency and information aggregation in auctions. *American Economic Review*, *90*, 499–525.
Piketty, T. (1999). The information-aggregation approach to political institutions. *European Economic Review*, *43*, 791–800.
Plott, C., & Sunder, S. (1988). Rational expectations and the aggregation of diverse information in laboratory security markets. *Journal of Economic Behavior and Organization*, *56*, 1085–1118.
Putnam, R. (2000). *Bowling alone*. New York: Simon & Schuster.
Schotter, A. (2003). Decision making with naive advice. *American Economic Review*, *93*, 196–201.
Schotter, A., & Sopher, B. (2003). Social learning and coordination conventions in intergenerational games: An experimental study. *Journal of Political Economy*, *111*, 498–529.
Schotter, A., & Sopher, B. (2006). Advice and behavior in intergenerational ultimatum games: An experimental approach. *Games and Economic Behavior*, *58*, 365–393.
Surowiecki, J. (2004). *The wisdom of crowds*. New York: Doubleday.
Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, *185*, 1124–1131.