Do People Appreciate the Value of Listening to a Variety of Different Opinions?

Andrew J. Healy*
Loyola Marymount University

July 25, 2006†

Abstract

This paper presents experimental evidence about how individuals learn from information that comes from inside versus outside their ethnic group. In the experiment, Thai subjects observed information that came from Americans and other Thais that they could use to help them answer a series of questions. Consistent with previous research, the subjects display overconfidence in their own opinions and place too low a value on all the information that they observe. Subjects achieve optimality, however, in how they weigh observed American information relative to observed Thai information. The data indicates that subjects understand that outside information has extra value because people from different groups know different things and so have an opportunity to learn from each other. The results demonstrate the importance of forming diverse groups to solve problems.

*E-mail: ahealy@lmu.edu. All data and programs are available for replication purposes upon request.

†I thank Victor Chernozhukov, Esther Duflo, and Sendhil Mullainathan for their advice throughout this project. Special thanks go to Adis Israngkura, Chayun Tantisidivakarn, and Somchai Jitsuchon for helping me arrange the experiment reported in this paper. Thanks also to Ken Chetnakarnkul and Phasook Mengkred for dedicated and superb research assistance. Gary Charness and Dorothea Herreiner provided particularly helpful advice on earlier drafts of this paper. For helpful discussions, I acknowledge and thank Dan Ariely, Ernst Fehr, Jerry Hausman, Dean Karlan, John List, Whitney Newey, Al Roth, Shawn Cole, Brian Healy, Peter Hinrichs, Daniel Paravisini, Chinda Wongngamnit and participants in the MIT development and econometrics lunches. I thank the Russell Sage Foundation and the MIT Shultz Fund for providing financial support for this project and the National Science Foundation for a graduate fellowship.

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 advice she received from neighbors who have experience using the seeds. Or she may decide that she also wants to talk to someone from outside her village, since she knows that 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 use a variety of different opinions to make decisions.

The empirical evidence from the field and the lab on how well agents use information to make decisions is mixed. 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. Information sharing generally improves decision-making, but the lack of sharing across groups sometimes leads groups to choose the wrong decision. In Kenya, entire villages tend to think the same way about whether to give potentially impure water to a sick child (Dearden, Pritchett, and Brown, 2004); even though there is a clear correct answer (to give water), many villages choose the incorrect option. In the lab, subjects sometimes herd on the incorrect

[1] Foster and Rosenzweig (1995), Duflo, Kremer, and Robinson (2004), and Munshi (2004) explore how information sharing affects technology adoption in India and Kenya. Dearden, Pritchett, and Brown (2004), Miguel and Kremer (2004), and Munshi and Myaux (2002) describe how information sharing affects health decisions in developing countries. Duflo and Saez (2002) investigate how communication within social groups influences participation in retirement plans at an American university.

choice when deciding which of two urns produced a randomly drawn ball (Anderson and Holt, 1997). Both in the field and in the lab, agents are not acting irrationally. They choose actions under the constraint that they know a limited amount of information. Increased information sharing could improve decisions. Take the case of a family deciding how to treat its sick child. By talking more with others from outside the village, the family might be able to break the misinformation cycle within the village that leads them to deny water to the child.

Even when outside information is available, though, it is sometimes ignored. When outside information is ignored, the costs are often severe. One stark example involves the Bay of Pigs invasion. The Kennedy administration sought no outside advice from the CIA or State Department, and devised a plan that ignored Castro's popularity, the strength of the Cuban army, and even the size of the island (Janis, 1972; Surowiecki, 2004). Another case involves the Challenger space shuttle disaster. Decision-makers at NASA did not communicate well enough with both internal and external engineers to call off the launch, even though indisputable evidence showed the launch to be extremely risky (Tufte, 1990). Similarly, management at Apple Computer isolated itself from outside information in the mid 1990s, years in which its market share declined significantly (Landry, 1997; Burrows, 2000). Each of these examples involve deeply flawed decisions that arise despite the availability of valuable outside information. The mistakes occur when decision-makers incorrectly believe that they have little to learn from others.

In other environments, recent research demonstrates that when outside opinions are heard, agents often place high value on them. Menon and Pfeffer (2003) showed that corporate managers usually listen closely to the advice given by independent consultants. To explain this high valuation, the authors argue that managers increase their status more by learning from outside knowledge compared to inside knowledge and that the flaws in outside information are less visible due to its scarcity. Another possibility, tested in this paper, is that managers and other economic agents who listen to outside advice appreciate the value of listening to an independent voice. The experimental results provide support for this hypothesis.

In the experiment, Thai subjects consider opinions that come from sources with different cultural backgrounds. The subjects first answer a series of general-knowledge questions that have correct numerical answers. There are three types of questions: 1) Questions about Thailand, 2) Questions about the US, and 3) Questions about both Thailand and the US. The subjects then observe randomly selected answers given by Americans and by other Thais, who had answered the same questions at an earlier date, that they can use to help revise their answers. By looking at how subjects change their answers, the experimental design provides estimates of the weights that subjects apply to observed American answers, to observed Thai answers, and to their own initial answers.

I compare these weights to how a subject would behave to maximize her payments. The data shows that, even when any one Thai answer is equally good as any one American answer, an optimizing Thai should assign significantly higher weight to American answers than to other Thai answers. For example, Americans and Thais are about equally good at answering the questions about both Thailand and the US. The data shows that it is still optimal for a Thai subject to put twice as much weight on observed American answers as on observed Thai answers. This extra value comes from the fact that, in the data, Americans and Thais tend to make different kinds of mistakes. 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 appear to understand this idea, behaving optimally in how they weigh American information relative to Thai information. Subjects achieve optimality despite listening too little to either group. They assign too much weight to their own initial answers, and the extent of overconfidence that I find is consistent with previous studies.[2] Subjects suffer significant costs from this overconfidence, as have other subjects in a variety of field and lab situations.[3] Still, they significantly improve their performance by correctly weighing observed American answers relative to observed Thai answers. Moreover, the experimental design makes it possible to show that subjects appreciate not only how good each group is at answering the questions, but also 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 choose not to seek outside advice. The results are particularly relevant to decision-making groups. When groups are formed to make decisions, any one individual's overconfidence is likely to be of less importance than the group's collective ability to use the available information. Failures to use information effectively, such as the Bay of Pigs invasion, show that groups make mistakes when members all think the same way and outside sources are not consulted. For groups to make the best decisions, the

[2]The experimental design also makes it possible to test whether anchoring causes subjects to overweigh their initial answers (Tversky and Kahneman, 1974). The data shows that anchoring has only a small effect on subject behavior.

[3]For evidence on overconfidence from the lab, see Gigerenzer, Hoffrage, and Kleinbolting (1991), Griffin and Tversky (1992), Camerer and Lovallo (1999), and Hoelzl and Rustichini (2005). Much of the field evidence comes from finance, including Barber and Odean (2001), Scheinkman and Xiong (2003), and Daniel et al. (1998). Other evidence indicates that overconfidence can be significantly reduced by making the target of comparison clearer (Perloff and Fetzer, 1986; Hoorens and Buunk, 1993). Also, for general knowledge tests, paying subjects has been shown to markedly reduce overconfidence (Hoelzl and Rustichini, 2005), and some research suggests that past findings reflect ambiguity aversion more than overconfidence (Greico and Hogarth, 2004).

experiment confirms what the anecdotal evidence suggests. Groups need to access a diverse set of information.

The results underscore the importance of forming diverse groups to solve problems. 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 term, my results show that "the strength of weak ties" carries over to problem-solving. People have more to learn from members of other groups than they do from members of their own group, and their behavior shows that they understand this idea. The experimental results thus show another consequence of the decline in social capital that Putnam (2000) describes. People without access to a diverse group of opinions will make poor decisions.

Section 2 describes the experimental design. In Section 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 Section 5, I estimate the weights subjects give to the information they observe and test a variety of hypotheses that explain that behavior. Section 6 tests the hypothesis that subjects both understand that Americans and Thais make different kinds of mistakes and apply this knowledge to their decisions. Section 7 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. At a later date, 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 revise their answers.

2.1 The Questionnaire

The questionnaire consists of fifteen questions covering a range of topics. Thirteen of the questions contain three distinct parts. For example, one question asks about the January temperature in Bangkok, the January temperature in Boston, and the sum of those temperatures. Another question asks 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 asks about the number of Thai and American 25-29 year-olds with some university education, as well as the sum of those two numbers. Figure 1 shows the format of the questions.

Figure 1: Format of the questions

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

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.[4] 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 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 the individual questions received 200 baht. 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 at answering the questions.

Subjects also completed a demographic survey before completing the questionnaire. The American students were asked to indicate their country of citizenship and the country where they attended high school, since some MIT undergraduates are citizens of other countries. Excluding this group of students from the sample has a negligible effect on the distribution of answers. Therefore, random selection of answers in Stage 2 was based on the entire set of MIT students.

2.3 Stage 2: Subjects observe American and Thai information

In Stage 2, 300 economics undergraduates at Thammasat's Bangkok campus and master's economics students at the National Institute for Development Administration (NIDA) first received

[4] Multiple translations and re-translations by individuals who are fluent in both Thai and English, as well as two pilots, verified the accuracy of the translation.

instructions in Thai (both read aloud and given in a packet) and then answered the questionnaire. 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. The data shows that a subject who maximized her payments would behave in basically the same way as a MSE minimizer. In other words, the experimental design succeeded in giving subjects the objective function of minimizing MSE.

In Stage 2, subjects first 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 from MIT and Thammasat-Rangsit students who answered the same set of questions. Subjects could use this information to help them revise their answers and their payments would be based on their final answers. 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. Payments were based on subjects’ final answers. Figure 2 shows what one group of the Thai subjects saw for the questions about political leaders.

Figure 2: Sample of information that subjects observe

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

Answers given by Thai students

1. 15	+	1. 20	=	1. 35

Answers given by American students

1. 1	+	1. 9	=	1. 10
2. 5		2. 7		2. 12
3. 7		3. 10		3. 17

The subjects in this group observed that one randomly selected Thai student thought the number of Thai prime ministers was 15, the number of American presidents was 20, and the sum was 35. They also observed that a randomly selected American student thought thought the answers were 1, 9, and 10, respectively. A second American student thought the answers were 5, 7, and 12; a third American student thought the answers were 7, 10, and 17.

The selection of answers that subjects observed proceeded as follows: First, I randomly selected whether subjects observed information about the Thailand questions, the US questions, the sum questions, or all three. In the example in Figure 2, subjects observed information about all three types of questions. Second, I randomly chose how many Thai answers that subjects observed (up to three). Third, I randomly chose how many Thai answers that subjects observed. In the example in Figure 2, subjects observed 1 Thai answer and 3 American answers. Finally, I randomly selected which Thai and American answers that subjects saw. Proceeding in this way yielded twenty randomly selected sets of information that subjects could observe.

In total, each experimental session took approximately 50 minutes. We conducted a total of 25 sessions: 17 at Thammasat and 8 at NIDA. Total payments averaged approximately 280 baht ($7) per subject.

2.4 Controlling for anchoring

Tversky and Kahneman (1974) show 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 their example, an experimenter spins a wheel in front of a group of students. Then students answer a question about the number of African countries in the UN. When the wheel gives a higher number, students give much larger answers. In my experiment, subjects first answer the questions and then update their answers based on what they observe. Thus, anchoring presents a serious concern in this experiment; a subject provides her own answer that she can anchor to and that number contains meaning relevant to the task, unlike the random number which affects students' answers about countries in the UN.

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 treatment group. Subjects in the main group may anchor to their initial answers because they express them.[5] In the experimental data, a subject who anchors would fail to sufficiently change the answer that she gives after observing new information. If anchoring is present in the main group, the 42 subjects who do not provide their private beliefs will choose final answers closer to the answers they observe. The data will

[5] Another possibility is that subjects anchor to their private beliefs because they think about them before they process the information they observe. The fact that subjects observe the other students' answers at the same time they see the question should minimize this type of anchoring.

show that anchoring has a small and statistically insignificant effect on subject behavior in this experiment.

3 A Model of Information Aggregation

3.1 Summary statistics

Here, I describe a model of information aggregation that shows how the Thai subjects should weigh the information they see under the assumption that they are mean-squared error (MSE) minimizers. The model applies to each question type (Bangkok/Thailand, Boston/US, or sum) separately. In later sections, I show that the primary results for how subjects should behave remain unchanged under weaker assumptions than those implied by the model. Moreover, the data confirms the validity of the assumptions.

For a question q, take individual i in group j, where the group is American or Thai, to have a private signal x_{ijq} about the correct answer for the question. 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 and the MSE for Thais for each question q.

The group MSE can be broken down into consistent estimators for 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* $s_{jq}^2 = \frac{1}{N_j}\sum_{i=1}^{N_j}(x_{ijq} - \overline{x}_{jq})^2$ *and* $\alpha_{jq}^2 = \frac{1}{N_j}\sum_{i=1}^{N_j}(\overline{x}_{jq} - Truth_q)^2$*, the MSE for group* j *for question* q *can be expressed as*

$$\Delta_{jq}^2 = s_{jq}^2 + \alpha_{jq}^2$$

Proof. *See Appendix A.1.1* ■

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

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

$$\rho_{jq} = \frac{\alpha_{jq}^2}{\Delta_{jq}^2}$$

For group j for question q, ρ_{jq} captures what share of total MSE comes from the common group bias. If individuals in group j tend to make different kinds of mistakes from other individuals in their group, ρ_{jq} will be low, as MSE will primarily be caused by individual-level variation. If people in a group make the same kind of mistake, ρ_{jq} will be high, as group bias will cause 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, $\frac{\Delta_A^2}{\Delta_T^2}$, 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, $\rho_{jq} = \frac{\alpha_{jq}^2}{\Delta_{jq}^2}$, and $\widehat{\alpha}_{jq} = \overline{x}_{jq} - Truth_q$, the maximum-likelihood estimators (MLE) for these parameters are:*

$$\frac{\Delta_A^2}{\Delta_T^2} = \frac{1}{Q}\sum_{q=1}^{Q}\frac{\Delta_{Aq}^2}{\Delta_{Tq}^2},\ \rho_A = \frac{1}{Q}\sum_{q=1}^{Q}\rho_{Aq},\ \ \textit{and}\ \ \rho_T = \frac{1}{Q}\sum_{q=1}^{Q}\rho_{Tq}.$$

Proof. *See Appendix A.1.2* ■

These three parameters determine an optimal rule for how the Thai subjects should weigh the information they observe if they apply the same rule across questions. If the group bias share for Americans is low, for example, then multiple American guesses would provide significantly more information about the correct answer than a single American opinion. A high American group bias share, however, implies that Americans tend to make the same kind of mistake; therefore a large group of American answers contains only slightly more information than a single American opinion.

When the American (Thai) group bias share is greater than zero, optimal behavior implies that a subject puts a lower weight on any one piece of American (Thai) information when she observes more American (Thai) answers. Note that an optimizing Thai subject puts lower weight on an observed Thai answer even when she observes only one Thai answer. As a Thai individual, she shares the same group bias with the observed Thai.

3.2 A model of subject behavior

Described here is a model of how the Thai subjects treat the information they observe. Previous research on overconfidence suggests that a Thai subject may not treat an observed Thai answer in the same way as her own answer. To account for overconfidence, a subject's own perceived

MSE is modeled as a fraction c of another Thai's.[6] Where Δ_{sq}^2 is a subject's perceived MSE for question q,

$$\Delta_{sq}^2 = c\Delta_{Tq}^2 \ , \tag{1}$$

and overconfidence implies $c < 1$.

A standard way to measure overconfidence involves comparing a subject's confidence interval for a given quantity to what it should be (Cesarini, Sandewall, and Johanneson (2003)). Consider as an example a weather forecaster who has to choose a 95% confidence interval for tomorrow's temperature. An overconfident forecaster will choose a confidence interval for tomorrow's temperature that is smaller than it should be. The actual temperature will be outside her confidence interval more than 5% of the time. The modeling of overconfidence in equation (1) corresponds to this idea. A subject perceives her confidence interval to be c times the width of another Thai's, for any given significance level. It is also assumed that a subject perceives her squared bias to be the same fraction c of another Thai's squared bias.

A subject who minimizes the expected MSE of her final answer will weigh information according to her perceptions of how good Americans are relative to Thais at answering the questions and how much of each group's MSE comes from group bias. Her perceptions of $\frac{\Delta_T^2}{\Delta_A^2}$, ρ_T, and ρ_A determine her behavior. The actual values of the parameters determine what she would optimally do.[7]

[6] Although overconfidence will lead subjects to perform less well at answering the questions, it should not be assumed that overconfidence is irrational. Since self-esteem may lead to higher effort or increased utility, overconfidence may well be rational (Benabou and Tirole, 2002; Koszegi, 2005).

[7] One possible reason that subject perceptions of different group's abilities at answering the questions could deviate from reality is that individuals could be overconfident not just in themselves, but in their cultural group. Recent research suggests, that in some environments, agents are overconfident in members of their peer groups (Healy, 2006).

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:

$$
\begin{aligned}
E(MSE) &= E(y_{iq} - Truth_q)^2 \\
& E\left(\lambda_T(x_{Tq,1} + ... + x_{Tq,n_T}) + \lambda_A(x_{Aq,1} + ... + x_{Aq,n_A}) + \lambda_s x_{iq} - Truth_q\right)^2
\end{aligned}
$$

where

$$
\begin{aligned}
\lambda_s &= \text{weight for own information} \\
\lambda_T &= \text{weight for any one piece of Thai information} \\
\lambda_A &= \text{weight for any one piece of American information}
\end{aligned}
$$

Assuming independence between the American and Thai group biases and that the weights given to all information sum to one, the expressions in Proposition 2 below capture the weights that subjects would optimally use. The expressions below define optimal behavior conditional on any level of overconfidence. To fully minimize MSE, subjects should also choose $c = 1$, avoiding both underconfidence and overconfidence.

Proposition 3 *The following expressions define the MSE-minimizing weights that subjects should use to evaluate information:*

$$
\begin{aligned}
\textit{Weighing self relative to other Thais:} \quad & \frac{\lambda_s}{\lambda_T} = \frac{1}{c} + \frac{1-c}{c}\left(\frac{\rho_T}{1-\rho_T}\right)n_T & (2) \\
\textit{Weighing self relative to Americans:} \quad & \frac{\lambda_s}{\lambda_A} = \left(\frac{\Delta_A^2}{\Delta_T^2}\right)\left(\frac{1-\rho_A}{c} + \frac{\rho_A}{c}n_A\right) & (3) \\
\textit{Weighing Americans relative to other Thais:} \quad & \frac{\lambda_A}{\lambda_T} = \left(\frac{\Delta_T^2}{\Delta_A^2}\right)\left(\frac{1+(n_T-1)\rho_T - c\rho_T^2 n_T}{(1+(n_A-1)\rho_A)(1-\rho_T)}\right) & (4)
\end{aligned}
$$

Proof. *See Appendix A.1.3.* ■

Under the model, the same equations describe subject behavior, with the actual parameter values substituted by subjects' perceptions of these parameters. For example, consider how a subject weighs her initial answer relative to the Thai answers she observes. As described by equation (2), she will choose a weight ratio that depends on her perception of her own accuracy relative to other Thais (captured by c) and how much group bias she perceives there to be among Thais (captured by ρ_T).

Equation (2) is the ratio of the self-weight to the weight for other Thais and thus the accuracy of Thais does not matter. But group bias does matter. When ρ_T is high and subjects are overconfident, the weight that subjects should put on other Thais becomes small. When group bias accounts for more of total MSE, an overconfident subject puts more trust in her reading of the joint Thai information than in another Thai's. If $c = 1$, the subject is not overconfident and she puts equal weight on herself and any other Thai.

Equation (3) is the ratio of the self-weight to the weight subjects should put on observed Americans when $n_T = 0$. Subjects should put more weight on Americans when Americans are more accurate, a situation that corresponds to a low $\frac{\Delta_A^2}{\Delta_T^2}$ value. When c is low, subjects put less weight on American answers, since subjects perceive themselves to be better at answering the question. When ρ_A is high, subjects should treat each additional American answer after the first as providing little added value. Dividing each side by n_A gives the optimal ratio of own-weight to the total weight given to all American information.

$$\frac{\lambda_s}{n_A \lambda_A} = \left(\frac{\Delta_A^2}{\Delta_T^2}\right)\left(\frac{1-\rho_A}{c n_A} + \frac{\rho_A}{c}\right)$$

This equation shows more clearly that when ρ_A is high, subjects should put less weight on each American answer when they observe more of them.

Equation (4) gives the weight ratio that subjects should assign to an 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 $c\rho_T^2 n_T$ term. For reasonable values of ρ_T, overconfidence has a small, but noticeable, effect on the relative weight ratio that subjects should use for American versus Thai information. When overconfidence is high (c is lower), subjects put more relative weight on Americans 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 (4) when $n_T = 1$, but the American group bias term only enters when $n_A > 1$.

The experimental data on how subjects update their answers provide estimates of the actual weights that subjects use. While the experiment does not directly observe subjects' perceptions of $\frac{\Delta_T^2}{\Delta_A^2}$, ρ_T, and ρ_A, the following sections show how subject behavior makes it possible to test a variety of hypotheses relating to subjects' implicit perceptions.

3.3 A check on the optimal weights

To supplement the model in the previous section, I also estimate the optimal weights that subjects should use by regressing the correct answer to the questions on a subject's initial answer, the average of the American answers she observes, and the average of the Thai answers she observes. The regression by definition minimizes the MSE. Checking how closely these estimates of the optimal weights match the model's estimates provides a test of the model's applicability. The model's usefulness derives from making it possible to separately analyze the effects of overconfidence on subject behavior.

Included in the regression are question dummies for the three categories of questions: meteorology, economics/politics, and social/cultural questions. I run the regressions separately for the cases where subjects observe information for the Bangkok/Thailand questions, the Boston/US questions, and the sum questions. To run the regression that estimates optimal behavior, I first standardize the data to make it possible to compare answers across questions. The data appendix, Appendix A.2, describes how I standardize the data.

Where

$$\begin{aligned} Truth_q &= \text{(standardized) correct answer for question } q, \\ x_{iq} &= \text{(standardized) initial answer given by subject } i \text{ for question } q, \\ \overline{x}_{iAq} &= \text{(standardized) average observed American answer for question } q, \\ \overline{x}_{iTq} &= \text{(standardized) average observed Thai answer for question } q, \\ C_q &= \text{vector of dummy variables for question category for question } q, \end{aligned}$$

the regression equation is

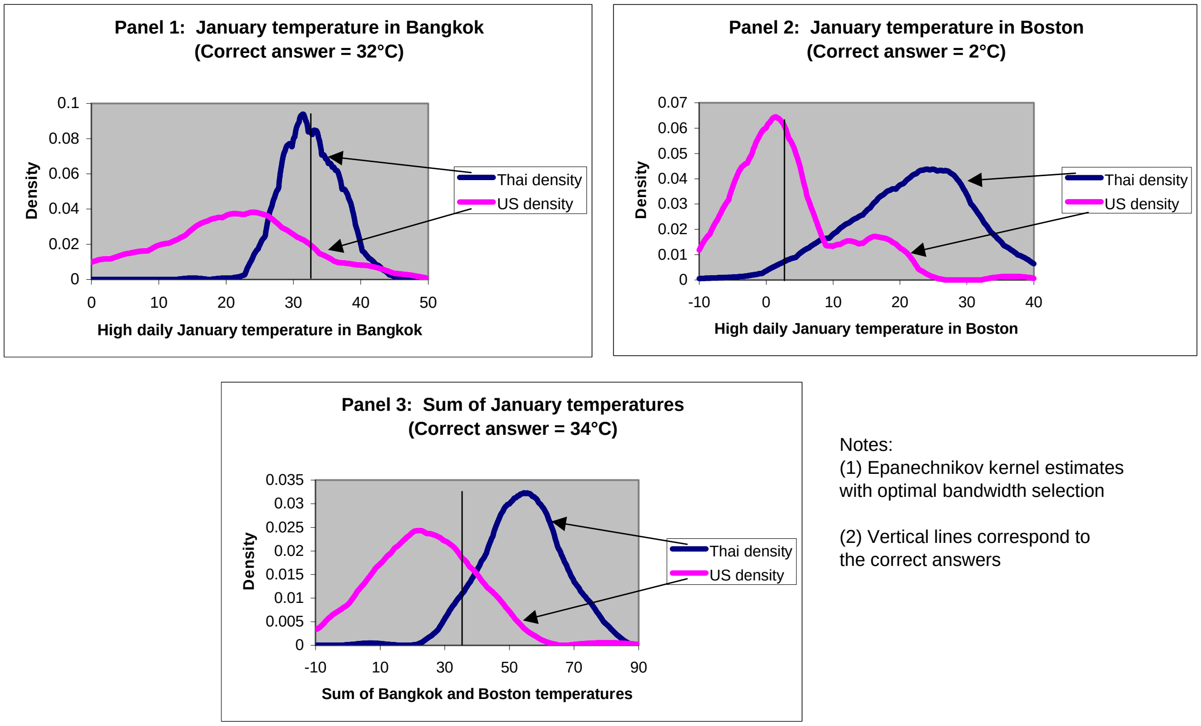
$$Truth_q = \alpha_s x_{iq} + \alpha_A \overline{x}_{iAq} + \alpha_T \overline{x}_{iTq} + C_q' \alpha_q + \varepsilon_{iq} \tag{5}$$

Relying only on a linearity assumption, this regression estimates the average weights that subjects should use to minimize their mean-squared error. I discuss the results from estimating (5) in Section 5. As described there, I also include terms that account for the number of American and Thai answers that a subject observes and different relative American-to-Thai accuracy for different questions. To estimate how subjects actually weigh information, I replace $Truth_q$ in (5) with the subjects' final answers to the questions.

4 Summary Statistics

The data shows that, across questions, Thais tend to make one kind of mistake and Americans make their own kind of mistake. This group bias means that American information contains extra value for a Thai subject. As an example of what the data looks like, Figure 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 (truth=$32°C$) and Panel 2 shows that Thais have a mean answer of $20°C$ for the Boston January temperature (truth=$2°C$).

Figure 3: Kernel density estimates for American and Thai answers about January temperature



Notes:
(1) Epanechnikov kernel estimates with optimal bandwidth selection

(2) Vertical lines correspond to the correct answers

It is important to note that other questions show a different pattern than the question about January temperature. For other questions, the Americans average is not as 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 for the sum question 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 an opportunity to learn from the other group.

Across questions, the data provides 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. The estimates in Table 1 come from the 116 Americans and the 430 Thais who either never observed anyone else's answers or who answered the questions before observing other subjects' opinions.[8]

Table 1: Relative accuracy of Americans and Thais

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

Note: Bootstrapped standard errors in parentheses

Consider the second column in Table 1. Thais have about one-half the MSE of Americans for the Thai questions. 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

[8] 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 estimated variances and correlations.

she puts on any Thai answer. The data will show that group bias means that a subject should actually put about 2.2 times more 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. The group bias share expresses what share of the mistakes that subjects make can be attributed to the common group error. Table 2 displays the estimated group bias shares for each question type. Column (1) contains estimates of ρ_T and column (2) contains estimates of ρ_A for each question type.

Table 2: Group effects for Americans and for Thais

Question type	Estimated Thai group bias share	Estimated American group bias share
	(1)	(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

Table 2 shows that each group's bias share is higher for the question types 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, the group effect 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 the group effect 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 is $20°C$. As I show in Section 6, the fact that Thai group bias is biggest for the Boston/US questions and American group bias is biggest for the Bangkok/Thailand questions has important implications for how a subject would optimally behave when she sees information for the sum question only.

In summary, the answer distributions show significant group biases for both Americans and Thais. The presence of Thai group bias means that American answers have extra value to a Thai subjects. An optimizing Thai subject needs to account for group bias when deciding how to weigh the American opinions she observes compared to the Thai opinions she observes.

5 Estimating subject behavior

5.1 Regression estimates

A simple regression provides the weights that subjects give to the information they observe and to their own private beliefs. This involves regressing subjects' final answers after observing information on the initial answers they gave before observing information, the American answers they observe, and the Thai answers they observe. If subjects increase (decrease) their answers more in response to high (low) observed American answers than to high (low) observed Thai answers, the regression will estimate that subjects assign a higher weight to the American answers.

The following regression estimates the average weights that a subject puts on American answers (β_A), other Thai answers (β_T), and her own initial answer (β_s). Where

$$y_{iq} = \text{(standardized) final answer given by subject } i \text{ for question } q,$$

$$x_{iq} = \text{(standardized) initial answer given by subject } i \text{ for question } q,$$

$$\overline{x}_{iAq} = \text{(standardized) average observed American answer for question } q,$$

$$\overline{x}_{iTq} = \text{(standardized) average observed Thai answer for question } q,$$

$$C_q = \text{vector of dummy variables for question category for question } q,$$

I estimate the following regression equation:

$$y_{iq} = \beta_s x_{iq} + \beta_A \overline{x}_{iAq} + \beta_T \overline{x}_{iTq} + C_q' \beta_q + \varepsilon_{iq} \tag{6}$$

Optimal behavior implies that a subject chooses different weights for the American and Thai average when she observes different amounts, n_{iA} and n_{iT}, of observed American and Thai information. When a subject observes more American answers, a higher weight should be assigned to the American average since it contains more information. A high American group bias share, though, means that there is less new information in each additional American answer and that a subject should put less weight on each individual American answer when she sees more of them. To see how subjects actually do change the weights they give to information depending on how much they observe, I include terms to account for this in the regression.

In addition, if subjects have some knowledge of the group MSEs for individual questions, they will apply higher weight to American information for those questions that Americans answer better relative to Thais. To see if subjects behave this way, the regression can also be expanded to include a term that captures relative group accuracy across questions. Details and results for all the specifications that I consider are contained in Appendix A.2.[9]

The basic results for the three types of questions are summarized in Table 3.

[9] I include the cases where individuals see answers for all three types of questions in the regressions for the

Table 3: Summary of estimated weights

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

Note:
(1) Regression standard errors are in parentheses.
(2) These estimates come from the regressions in columns 1, 5, and 9 of Table A1.

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 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 than to other Thai answers, choosing 0.464 as the weight for their initial answers, 0.090 for the weight given to observed Thai answers, and 0.463 for the weight given to observed American 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, giving

Thailand and US questions. One possibility is that there are spillovers across question types when subjects observe all answers for all three types of questions. For example, a subject who sees an American who answers well for the Thailand question may put more weight on that American for the US question. Testing for this possibility generally produces insignificant results. Details are available upon request.

estimates of 0.731 for the own-weight, 0.068 for the Thai weight, and 0.165 for the American weight.

The results (shown in Table A1 in the Appendix) also indicate that subjects account for different group accuracies across questions. For all three types of questions, subjects put significantly more weight on American information for those questions on which Americans perform relatively better. For the Bangkok/Thailand questions, an increase of 0.1 in the accuracy index (Thai MSE divided by the sum of American and Thai MSE) causes subjects to increase the weight given to American answers by 0.02. Given that subjects assign a weight of 0.058 to American answers, this represents a substantial increase.

It may seem surprising that subjects are able to appreciate the accuracy of Americans relative to Thais for individual questions. That they do is perhaps less surprising when the individual questions are considered. Relative to American answers, Thai answers are much more accurate for the question about temperature in Bangkok than for the question about female-labor force participation in Thailand. It seems reasonable to expect that the subjects would understand that it takes individual experience in Bangkok to know the weather there, but there may be more general knowledge involved with making an educated guess about labor-force demographics. Therefore, the Thai subjects may and apparently do put higher relative weight on Thai information for the question about Bangkok weather.

5.2 Anchoring

As discussed earlier, anchoring presents a major concern in this experiment. Subjects may adhere more closely to their initial answers because they express them. To test for this, a treatment group of 42 students observed randomly selected answers from Americans and Thais without first providing their private beliefs about the answers to the questions. These subjects observe one of the same twenty sets of information that the subjects in the main group observe. By comparing subjects in the anchoring group to subjects in the main group who see the same information, I can test for anchoring. If subjects anchor to their private signals when they reveal them, the subjects who do not reveal their private signals will choose final answers closer to the answers they observe.[10]

To test for anchoring, I consider the following regressions:

$$(y_{iq} - \overline{x}_{iTq})^2 = \theta_{1T} Anchor_i + Version_{iv}' \theta_{2T} + \varepsilon_{iq} \quad (7)$$

$$(y_{iq} - \overline{x}_{iAq})^2 = \theta_{1A} Anchor_i + Version_{iv}' \theta_{2A} + \varepsilon_{iq}, \quad (8)$$

where $Anchor_i$ is a dummy that is one if the subject belongs to the anchoring treatment group and zero otherwise. $Version_{iv}$ is a dummy that is one if subject i observed version v out of the 20 possible sets of randomly selected information. Including the version dummies creates a direct comparison between subjects in the anchoring group and subjects in the main group who observed the same information.

The first (second) regression looks at the distance between subjects' final answers and the Thai (American) answers they observe. If subjects in the main group anchor, then the following two conditions should hold: $\theta_{1T} < 0$ and $\theta_{1A} < 0$. These conditions state that subjects who do not provide their private beliefs before observing answers end up closer to those answers than subjects who answered the questions on their own first. Table 5 shows that the coefficients are always close to zero. The estimates are insignificant except for the distance between subjects'

[10] Another way to test for anchoring involves looking at whether or not subjects in the anchoring group do better at answering the questions than the students in the main group because these 42 students do not suffer the losses associated with anchoring. On average, subjects in the anchoring treatment group appear to do slightly better, but the difference is not significant.

answers and the American average they see for the Boston/US questions.

The first three entries in column (1) of Table 4 come from running regression (7) for each of the three types of questions and the last three entries in column (1) come from running (8) for each of the three types of questions. The results indicate, for the questions about Thailand, that the squared distance between subjects' final answers and the Thai average they observe is .013 units lower for subjects in the anchoring group than for subjects in the main group.

Table 4: Effect of anchoring on distance from observed information

Independent variable: Dummy for anchoring treatment group	Effect of not having a chance to anchor	Average squared distance for subjects not in the anchoring group
	(1)	(2)
A. Dependent variable: Squared distance from average of observed Thai answers		
Questions about Thailand	-.013	.827
	(.021)	(1.773)
Questions about the US	-.051	1.798
	(.065)	(3.23)
Questions about sum	-.026	1.75
	(.049)	(3.585)
B. Dependent variable: Squared distance from average of observed American answers		
Questions about Thailand	.042	1.716
	(.078)	(2.571)
Questions about the US	-.046	1.378
	(.022)	(2.945)
Questions about sum	-.049	1.781
	(.07)	(3.314)

Note:
For column (1), the robust regression standard errors are reported in parentheses.
For column (2), the standard deviation is in parentheses.

To put the regression coefficients into perspective, column (2) in Table 4 shows the average mean-squared distance from of subjects' final answers from the observed Thai or American average for the 300 subjects who were not in the anchoring group. Since the data has been standardized, the standard deviation for each question is one. The table indicates that, on average, subjects in the main group choose final answers that are 1.835 standard deviations from the observed Thai average for the US question. The regression estimates that the students in the anchoring group choose final answers that are .051 standard deviations closer to the observed Thai average than the main group. In other words, compared to subjects in the main group, subjects who do not have the chance to anchor choose final answers that are about $\frac{.051}{1.835} = 3\%$ closer in squared distance to the observed Thai average. In general, the results in Table 5 follow this pattern. Subjects who are not given the chance to anchor end up insignificantly closer to the answers they observe than other subjects who answer the questions first and observe the same information. Anchoring has a small and usually insignificant effect on subject behavior.

6 Tests for optimal behavior

6.1 Estimated optimal weights

As discussed in Section 3.3, regressing the correct answers to the questions on subjects' initial answers, the average observed American answer, and the average observed Thai answer gives estimates of the optimal weights subjects should use. Table A2 in the Appendix shows in detail the optimal weights obtained in this way for all three types of questions. Panel A of Table 5 summarizes the main results.

Table 5: Estimating the optimal weights

A. Estimates from regression with the correct answers as the dependent variable

Optimal weight	Questions about Thailand	Questions about US	Questions about sum
	(1)	(2)	(3)
Own initial answer	.292	.063	.179
	(.021)	(.017)	(.026)
Thai average	.458	.115	.250
	(.023)	(.02)	(.024)
American average	.250	.822	.571
	(.018)	(.016)	(.026)

Notes:
(1) Regression standard errors are in parentheses.
(2) These estimates come from the regressions in column 1, 5, and 9 in Table A2.

B. Estimates from the econometric model

Optimal weight	Questions about Thailand	Questions about US	Questions about sum
	(1)	(2)	(3)
Own initial answer	.258	.100	.154
	(.007)	(.008)	(.009)
Thai average	.458	.174	.332
	(.014)	(.015)	(.02)
American average	.284	.726	.514
	(.021)	(.024)	(.028)

Notes:
(1) Bootstrapped standard errors are in parentheses.
(2) These estimates come from the parameter estimates described in Tables 1 and 2.

For the Bangkok/Thailand questions, an optimally-behaving subject would put a weight of 0.292 on her private belief, a weight of 0.458 on the observed Thai average, and a weight of 0.250 on the observed American average. For the Boston/US questions, subjects would optimally choose a weight of 0.063 for their initial answers, 0.115 for the observed Thai average, and 0.822 for the observed American average. For the sum questions, the regression estimates that a subject should choose 0.179 for the self weight, 0.250 for the weight given to observed

Thai answers, and 0.571 for the weight given to observed American answers. Notice that these regression estimates of optimal behavior do not account for overconfidence. These are the weights that a subject should use assuming overconfidence is not present.

The model described in Sections 3.1 and 3.2 provides a different way to estimate 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 bias share into equations (2), (3), and (4). These estimates were reported in Tables 1 and 2. Panel B of Table 5 displays the econometric model's estimates of the optimal weights subjects should apply given that subjects do not show overconfidence ($c = 1$).

For the Bangkok/Thailand questions, a subject should apply a weight of 0.258 to her initial answer, 0.459 to the Thai average she observes, and 0.284 to the American average she observes. For the Boston/US questions, the corresponding weights are 0.100, 0.174, and 0.726. For the sum questions, the model estimates that a subject would optimally choose 0.154 for the self-weight, 0.332 for the weight given to observed Thai answers, and 0.514 for the weight given to observed American answers. Bootstrapping gives the standard errors for these estimates.

For all three types of questions, the regression's estimates of the optimal weights (Panel A) and the econometric model's estimates (Panel B) in Table 5 match closely and equality cannot be rejected. The close correspondence between the estimates provides additional evidence supporting the importance of accounting for group bias in determining optimal behavior. If, in the model, the group bias shares ρ_A and ρ_T are assumed to be zero, the hypothesis of equality between the two sets of estimates of optimal behavior is rejected for all three types of questions. In other words, the regression estimates of optimal behavior in Panel A account for the implications of group bias for how the Thai subjects should behave.

6.2 Construction of confidence intervals

Simulations using the regression coefficients from (6) 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 ways in which subjects could be behaving. I focus on two such confidence intervals: 1) the simple weight ratio, which expresses the relative weight a subject would use if she understood each group's accuracy but ignored group bias, and 2) the optimal weight ratio described in equation (4) of Section 3, which expresses how a subject would behave if she correctly perceived each group's accuracy and accounted for group bias.

To summarize, I focus on the following three weight ratios:

$$\begin{aligned}
\text{Actual} &= \frac{\beta_A}{\beta_T} \\
\text{Simple} &= \left(\frac{\Delta_T^2}{\Delta_A^2}\right) \\
\text{Optimal} &= \left(\frac{\Delta_T^2}{\Delta_A^2}\right)\left(\frac{1+(n_T-1)\rho_T - c\rho_T^2 n_T}{(1+(n_A-1)\rho_A)(1-\rho_T)}\right)
\end{aligned}$$

The regression estimates of how subjects behave give a confidence interval for the actual weight ratio that subjects use for all three types of questions. 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. Table 6 reports these confidence intervals for each of the three question types. The p-values in the table correspond to tests that I describe in the next subsection.

Table 6: Comparing how subjects relatively weigh American and Thai information

	Thailand questions (1)	US questions (2)	Sum questions (3)
Actual weight ratio	.231	5.143	2.49
95% confidence interval	(.131,.352)	(3.144,15.17)	(1.45,5.366)
Simple weight ratio	.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 ratios for different overconfidence levels			
No overconfidence	.571	4.423	1.843
	(.475,.694)	(3.467,5.63)	(1.466,2.332)
	p=0.000	*p=0.622*	*p=0.346*
Estimated overconfidence	.603	5.043	2.168
	(.476,.773)	(3.635,7.345)	(1.597,3.136)
	p=0.000	*p=0.979*	*p=0.75*

Notes: (1) Bootstrapped 95% confidence intervals are in parentheses.
(2) The bootstrap for the actual weights accounts for correlation in the coefficient estimates.
(3) *p* -values compare the given weight ratio to the actual weight ratio.

The optimal weight ratio increases as overconfidence increases. As discussed earlier, an overconfident subject already puts high weight on Thai information by assigning a high weight to her own beliefs. To compensate for this extra weight on Thai information, an overconfident but otherwise optimizing subject will put more weight on observed Americans relative to observed Thais. Even for very large overconfidence, though, the direct effect of group bias on the optimal weight ratio is larger than the effect of overconfidence. 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 to the level seen in the data causes the optimal weight ratio to further rise to 5.04.[11]

[11] Estimating the overconfidence parameter, c, involves estimating a non-linear model. Specifically, for each of the question types, if it is assumed that $\rho_A = \rho_T$ and c is allowed to be a function of $n_{iA} + n_{iT}$, then the

To summarize, the data on how subjects use the information they observe to update their answers give confidence intervals for the ratio that subjects use to weigh observed American information relative to observed Thai information. The Thai and American answer distributions for the series of questions gives confidence intervals for the optimal weight ratio that subjects should apply. The data provides these estimates of optimal behavior, conditional on any amount of overconfidence.

6.3 Testing behavioral hypotheses

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

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

Under H_0, subjects will choose

$$\frac{\beta_A}{\beta_T} = \left(\frac{\Delta_T^2}{\Delta_A^2}\right) \ . \tag{9}$$

Rejection of the prediction (9) implies rejection of H_0.

The second row of Table 6 displays the results of this test for all three types of questions. For the Bangkok/Thailand questions and the sum questions, we can reject this hypothesis at the 5% level ($p = 0$ and $p = 0.017$, respectively). For the Boston/US questions, we can reject

overconfidence parameters $c(n_{iA} + n_{iT})$ can be estimated. The standard error bounds for the overconfidence estimates are large. The estimated value of the overconfidence parameter c falls in the range between 0.25 and 0.5, on average. Details and regression results are available upon request.

it at the 10% level ($p = 0.065$). We reject the hypothesis for the Bangkok 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, subjects do not correctly perceive how accurate Americans are relative to Thais while at the same time failing to recognize the presence of group bias.

Now consider the hypothesis of optimal behavior, H_1:

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

This hypothesis states that subjects correctly perceive the MSE of Thais relative to Americans and also correctly account for group bias. Under H_1, subjects understand each group's accuracy and correctly value the independence in American information. 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.

Under H_1, subjects will choose the optimal weight ratio

$$\frac{\beta_A}{\beta_T} = \left(\frac{\Delta_T^2}{\Delta_A^2}\right)\left(\frac{1 + (n_T - 1)\rho_T - c\rho_T^2 n_T}{(1 + (n_A - 1)\rho_A)(1 - \rho_T)}\right) \tag{10}$$

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, the one in which they need the least help at answering the questions. On the other hand, for the Boston/US and sum questions, we cannot reject H_1 for any level of overconfidence.[12]

[12] It is interesting that the Thai subjects fail to use the optimal relative weights only for the questions about Bangkok or Thailand. They 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 that the Thailand heading causes

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 the value of c implied by subject behavior, the actual and optimal weight ratios match up remarkably closely in these two cases. For the Boston/US questions, the optimal weight ratio estimate is 5.04 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 ($p = 0.979$). For the sum questions, the actual weight ratio estimate is 2.49, compared to the optimal estimate of 2.17. The test for equality gives a p-value of 0.750. In summary, individuals who display the same amount of overconfidence as the subjects in the experiment would optimally weigh American answers relative to Thai answers in a very similar way to how the subjects actually behave.

Biased behavior, however, could still explain subject behavior for the Boston/US and sum questions. Under this hypothesis, H_2, subjects perceive Americans to be better than they actually are compared to Thais and they ignore group bias:

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

Under H_1, subjects put extra weight on American information because they understand the extra value in American information that comes from the fact that Americans and Thais make different kinds of mistakes. Under H_2, subjects put extra weight on American information because subjects incorrectly perceive American answers to be better than they really are.

The experimental design can distinguish between H_1 and H_2. Subject behavior on the sum question gives a direct test of the hypothesis that subjects ignore group bias for any perception

subjects to underweigh American information. In other words, only when the question is clearly about a Thai's area of presumed expertise do the subjects listen too little to Americans.

of $\left(\frac{\Delta_T^2}{\Delta_A^2}\right)$, how accurate Americans are relative to Thais. By looking at how subjects who observe answers only for the sum question update their answers for both the Bangkok/Thailand and Boston/US questions, it is possible to test the hypothesis that subjects ignore group bias when choosing their final answers.

To explain this test, I expand the earlier notation that applied when each question type was considered separately. Define:

$$\begin{aligned}\rho_{jk} &= \text{group bias share in total MSE for group } j \text{ for question type } k \\ \Delta_{jk}^2 &= \text{mean-squared error for group } j \text{ for question type } k\ .\end{aligned}$$

For example, ρ_{Ta} is the group bias share in total MSE for Thais answering the Boston/US questions.

Consider the case when a subject uses observed answers for the sum question to update her answer for the Bangkok/Thailand questions. A subject updates her answer for a Bangkok/Thailand question based on her initial answer for a Bangkok/Thailand question and the distance between the answers she observes for the sum question and her initial answer for the sum question. When a subject observes answers above her own for the sum question and uses them to update her answer, she is likely to revise upwards her answer for the Bangkok/Thailand question.

Define $\left(\frac{\phi_A}{\phi_T}\right)_{Thai}$ to be the weight ratio that subjects assign to American answers relative to Thai answers to the sum question when they update for the Bangkok/Thailand question. Analogously, define $\left(\frac{\phi_A}{\phi_T}\right)_{US}$ 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. Consider the following proposition:

Proposition 4 *If* $(\rho_{Tt})_{perceived} = (\rho_{Ta})_{perceived} = 0$, *then a subject will choose* $\left(\frac{\phi_A}{\phi_T}\right)_{Thai} =$

$\left(\frac{\phi_A}{\phi_T}\right)_{US}$

Proof. See Appendix A.1.4 ■

Proposition 4 refers to a subject who implicitly perceives there to be no group bias for both the Bangkok/Thailand and Boston/US questions. When observing information about the sum question, this subject will use the same weight ratio to update for the Bangkok/Thailand questions as she uses to update for the Boston/US questions. Consider the hypothesis, G_0, that the perceived group bias shares are zero.

$$G_0 : (\rho_{Tt})_{perceived} = (\rho_{Ta})_{perceived} = 0$$

This hypothesis states that subjects ignore Thai group bias for Thais for both the Thai and US questions. Notice that G_0 encompasses H_0, the hypothesis that subjects perceive total MSE correctly and ignore group bias, and H_2, the hypothesis that subjects perceive Americans to have lower MSE relative to Thais than they actually do and ignore group bias. Rejection of G_0 means we must also reject H_2.

Under G_0, subjects will choose

$$\left(\frac{\phi_A}{\phi_T}\right)_{Thai} = \left(\frac{\phi_A}{\phi_T}\right)_{US} \tag{11}$$

The left-hand side of (11) represents the weight ratio subjects use to relatively weigh American and Thai answers for the sum question to revise their answers for the Thailand questions. The right-hand side of (11) represents the analogous ratio that subjects use to update for the US questions. Rejection of the equality in (11) would imply rejection of G_0.

The following two regressions give the parameter estimates needed to conduct the above test. The first equation expresses the change in a subject's answer for the Thai question (subscript

t) as a function of the distance between the average observed American answer for the sum question and her own answer for the sum question (subscript s) 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.

Where

$$\begin{aligned} y_{iqt} &= \text{final answer given by subject } i \text{ to the Thai part of question } q, \\ x_{iqt} &= \text{initial answer given by subject } i \text{ to the Thai part of question } q, \\ y_{iqa} &= \text{final answer given by subject } i \text{ to the US part of question } q, \\ x_{iqa} &= \text{initial answer given by subject } i \text{ to the US part of question } q, \\ x_{iqs} &= \text{initial answer given by subject } i \text{ for the sum part of question } q, \end{aligned}$$

the regression equations are:

$$y_{iqt} - x_{iqt} = \phi_{A,Thai}(\overline{x}_{iAqs} - x_{iqs}) + \phi_{T,Thai}(\overline{x}_{iTqs} - x_{iqs}) + C_q'\phi_1 + \varepsilon_{iqt} \quad (12)$$

$$y_{iqa} - x_{iqa} = \phi_{A,US}(\overline{x}_{iAqs} - x_{iqs}) + \phi_{T,US}(\overline{x}_{iTqs} - x_{iqs}) + C_q'\phi_2 + \varepsilon_{iqu} \quad (13)$$

Table 7 reports the results from estimating equations (12) and (13). Notice that $\left(\frac{\hat{\phi}_A}{\hat{\phi}_T}\right)_{Thai} = \frac{.075}{.056} = 1.34$ and $\left(\frac{\hat{\phi}_A}{\hat{\phi}_T}\right)_{US} = \frac{.226}{.081} = 2.79$. The regression results provide the inputs needed to test (11). We can reject (11) at a 10% level ($p = 0.058$). At a 10% level, we reject G_0, the hypothesis that subjects fail to take group bias into account, regardless of how they perceive American and Thai accuracy.

Table 7: Subject updating for the Thai and US questions when they see 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 and initial answer (*for sum question*)	.056 (.013)	.081 (.019)
ϕ_A = Distance between observed American average and initial answer (*for sum question*)	.075 (.014)	.226 (.022)
p-value for test of : $\left(\frac{\phi_A}{\phi_T}\right)_{Thai} = \left(\frac{\phi_A}{\phi_T}\right)_{US}$	0.058	
N	548	544

Notes:
(1) Regression standard errors are in parentheses.
(2) Regressions include dummies for question categories (meteorology, economic/political, and social/cultural).

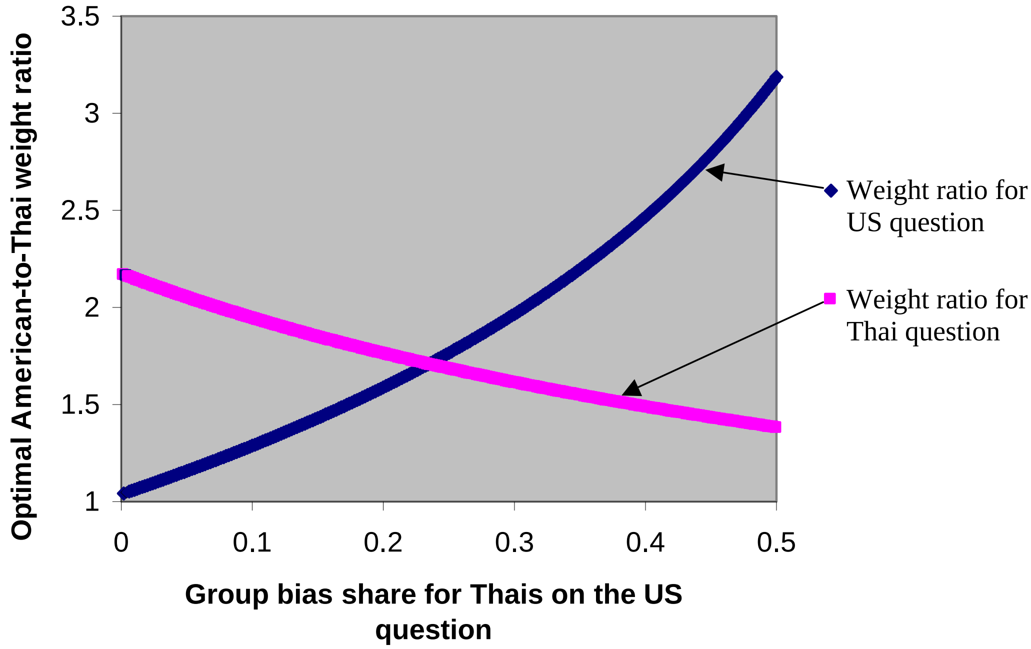
In contrast, correctly accounting for group bias would lead subjects to behave in a way that accords with how they actually behave. If the perceived group bias share for Thais for the Boston/US questions (ρ_{Ta}) is greater than the perceived group bias share for Thais for the Bangkok/Thailand questions (ρ_{Tt}), then subjects will put a higher relative weight on observed Americans for the US questions than for the Thailand questions. The proof of Proposition 4 demonstrates that:

$$\rho_{Ta} > \rho_{Tt} \Rightarrow \left(\frac{\phi_A}{\phi_T}\right)_{Thai} < \left(\frac{\phi_A}{\phi_T}\right)_{US}$$

To understand the intuition, consider a subject updating her answer for the Bangkok/Thailand question after observing answers for the sum question. If Thai group bias for the Bangkok/Thailand questions (ρ_{Tt}) is high, she should put less weight on Thais relative to Americans for the same reasons seen earlier; group bias means each additional Thai answer contains less new information. On the other hand, if Thai group bias for the Boston/US questions (ρ_{Ta}) is high, she should put *higher* weight on observed Thai answers for the sum question. When ρ_{Ta} 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 ρ_{Ta} were equal to one, all Thais would give the same answer to the US questions. Then, a subject could exactly deduce the observed individual's private belief about the Thai question from her answer to the sum question.

The data shows that the effects of group bias can explain how subjects actually weigh the information they observe. If subjects applied the estimated actual variance estimates and estimated actual group bias shares from Tables 1 and 2, they would choose $\left(\frac{\hat{\phi}_A}{\hat{\phi}_T}\right)_{Thai} = 1.51$ and $\left(\frac{\hat{\phi}_A}{\hat{\phi}_T}\right)_{US} = 2.53$. Figure 4 shows how the optimal weight ratios for the two types of questions vary as a function of ρ_{Ta}, holding the other parameters constant at their estimated values. The graph shows that, when updating their answers for the Bangkok/Thailand questions, subjects should put less weight on observed Americans answers to the sum question and more weight on observed Thai answers when ρ_{Ta} is high.

Figure 4: Optimal behavior when information for the sum question is observed



In summary, the earlier results showed that subjects used approximately the optimal weight ratio for the Boston/US and sum questions. We could explain this behavior in two ways. Either subjects appreciate the importance of group bias or subjects overestimate $\frac{\Delta_T^2}{\Delta_A^2}$, the ratio of Thai MSE to American MSE. By looking at how subjects update separately for the Bangkok/Thailand and Boston/US questions when they observe answers for the sum, we can reject the latter possibility. On the other hand, the hypothesis that subjects appreciate each group's accuracy and the extra value in an American's independent perspective to a Thai subject describes the data quite closely.

7 Conclusion

This paper demonstrates that economic agents can learn effectively from a diverse set of opinions. The Thai subjects attain significant improvements in their answers by correctly weighing observed American answers relative to observed Thai answers. Moreover, subject behavior shows that agents can be expected to account for not only the relative quality of different opinions, but also the value of drawing on a variety of independent perspectives.

The model presented in this paper may seem to require a level of statistical sophistication that demands too much of any experimental subject. Almost certainly, subjects do not reason along the lines of Bayesian updating, mulling over their beliefs about group bias. Their actions, however, reflect implicit perceptions. The experimental design made it possible to test a variety of hypotheses regarding these perceptions. The data supports the hypothesis that subjects understand that American information has extra value to a Thai subject due to the fact that members of the same group tend to make the same kind of errors.

Part of the subjects' success in the experiment may derive from Thailand's openness to foreigners and relatively positive view of the US. In contrast, only 15% of Indonesians had a favorable impression of the US around the time that I conducted this experiment (Pew Research Center, 2003).[13] It may be the case that, in general, Indonesian decision-makers would have a more difficult time optimally using information from Americans than Thais due to more negative feelings toward the US. For this and other reasons, it is unclear to what extent the results in this paper apply across cultures. Future research could apply this experimental design to students from other countries to determine the generality of my results.

[13] The US invasion of Iraq is largely responsible for the feelings of Indonesians towards the US. In 2000, 75% of Indonesians had a favorable view of the US. In 2002, 61% of Indonesians still viewed the US favorably.

The results illustrate the importance of forming diverse groups to solve problems. Unfortunately, the desire for cohesiveness often prevents diverse groups from being formed. Psychological evidence suggests that homogeneous groups become close-knit more easily than diverse groups (Janis, 1972). As homogeneous groups become more close-knit, they believe more in the group's invulnerability, creating a cycle in which the group becomes increasingly more close-knit and insulated from outside opinions. The experimental results match with the psychological evidence pointing towards "groupthink" as the reason groups fail to call upon a diverse set of opinions. The individuals in the experiment appreciate the value of listening to a variety of different opinions.

The problem thus appears to come from poorly functioning groups as opposed to poorly functioning individuals. 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 key, demonstrated by fiascoes like the Bay of Pigs invasion and the Challenger disaster, is to make sure that decision-makers hear those independent voices.

References

[1] Anderson, Lisa and Charles Holt (1997). "Information Cascades in the Laboratory," *American Economic Review, 87,* 847-862.

[2] Barber, Brad and Terrence Odean (2001). "Boys Will Be Boys: Gender, Overconfidence, and Common Stock Investment," *Quarterly Journal of Economics, 116*, 261-292.

[3] Benabou, Roland and Jean Tirole (2002). "Self-confidence and Personal Motivation," *Quarterly Journal of Economics*, *117*, 871-915.

[4] Burrows, Peter (2000). "Apple; Yes, Steve, You Fixed It. Congrats! Now What's Act Two?" *Business Week*, *3692*, 102-112.

[5] Camerer, Colin and Dan Lovallo (1999). "Overconfidence and Excess Entry: An Experimental Approach," *American Economic Review, 89,* 306-318.

[6] Cesarini, David, Orjan Sandewall, and Magnus Johanneson (2003). "Confidence Interval Estimation Tasks and the Economics of Overconfidence," Stockholm School of Economics working paper.

[7] Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam (1998). "Investor Psychology and Security Market Over- and Under-reactions," *Journal of Finance, 53*, 1839-1886.

[8] Dearden, Kirk, Lant Pritchett and Jeff Brown (2004). "Learning from Neighbors: Social Learning about Child Feeding During Dysentery Episodes," working paper.

[9] Duflo, Esther, and Emanuel Saez (2003). "The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment," *Quarterly Journal of Economics, 118,* 815-842.

[10] Duflo, Esther, Michael Kremer and Jonathan Robinson (2003). "Understanding Technology Adoption: Fertilizer in Western Kenya," working paper.

[11] Foster, Andrew and Mark Rosenzweig (1995). "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture," *Journal of Political Economy, 103*, 1176-1209.

[12] Gigerenzer, Gerd, U. Hoffrage, and H. Kleinbolting (1991). "Probabilistic Mental Models: A Brunswikian Theory of Confidence," *Psychological Review*, *98*, 506-528.

[13] Granovetter, Mark (1973). "The Strength of Weak Ties," *American Journal of Sociology*, *78*, 1360-1380.

[14] Greico, Daniela and Robin Hogarth (2004). "Excess Entry, Ambiguity Seeking, and Competence," working paper.

[15] Griffin, Dale and Amos Tversky (1992). "The Weighting of Evidence and the Determinants of Confidence," *Cognitive Psychology, 24,* 411-435.

[16] Healy, Andrew (2006). "Is My Group Better Than Yours? Overconfidence and Peer Networks," working paper.

[17] Hoorens, V. and B.P. Buunk (1993). "Social Comparison of Health Risks: Locus of Control, the Person-Positivity Bias, and Unrealistic Optimism," *Journal of Applied Social Psychology, 23,* 291-302.

[18] Janis, Irving L. (1972). *Victims of Groupthink: A Psychological Study of Foreign Policy Decisions and Fiascoes.* Boston, MA: Houghton-Mifflin.

[19] Klar, Yechiel and Eilath Giladi (1997). "No One in My Group Can Be Below the Group's Average: A Robust Positivity Bias in Favor of Anonymous Peers," *Journal of Personality and Social Psychology, 73,* 885-901.

[20] Koszegi, Botond (2005). "Ego Utility, Information Acquisition, and Task Choice," forthcoming, *Journal of the European Economic Association.*

[21] Landry, John (1997). "If an Organization Has a Memory, Will it Need a Therapist?" *Proceedings of the 13th Annual Hawaii International Conference on System Sciences*, 1-9.

[22] Larrick, Richard and Jack Soll (2003). "Intuitions about Combining Opinions: Misappreciation of the Averaging Principle," INSEAD working paper.

[23] Menon, Tanya and Jeffrey Pfeffer (2003), "Valuing Internal vs. External Knowledge: Explaining the Preference for Outsiders," *Management Science.*

[24] Miguel, Edward and Michael Kremer (2004). "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities," *Econometrica, 72,* 159-217.

[25] Munshi, Kaivan and Jacques Myaux (2005), "Social Norms and the Fertility Transition," *Journal of Development Economics*, forthcoming.

[26] Munshi, Kaivan (2004). "Social Learning in a Heterogeneous Population: Technology Diffusion in the Indian Green Revolution," *Journal of Development Economics, 73/1,* 185-213.

[27] Perloff, L. and B. Fetzer (1986). "Self-Other Judgments and Perceived Vulnerability to Victimization," *Journal of Personality and Social Psychology, 50,* 502-510.

[28] Pew Research Center (2003). *The Pew Global Attitudes Project.* http://people-press.org/presentations/PewGlobalAttitudesViewsofaChangingWorld.pdf.

[29] Putnam, Robert (2000). *Bowling Alone.* New York: Simon & Schuster.

[30] Scheinkman, Jose and Wei Xiong (2003). "Overconfidence and Speculative Bubbles," *Journal of Political Economy*, 111, 1183-1220.

[31] Surowiecki, James (2004). *The Wisdom of Crowds.* New York, NY: Doubleday.

[32] Tversky, Amos and Daniel Kahneman (1974). "Judgment Under Uncertainty: Heuristics and Biases," *Science*, *185,* 1124-1131.

[33] Wright, G.N., et al. (1978). "Cultural Differences in Probabilistic Thinking," *Journal of Cross-Cultural Psychology, 9,* 285-299.

[34] Yates, J. Frank, et al. (2002). "Probability Judgment Across Cultures," in *Heuristics and Biases: The Psychology of Intuitive Judgment,* Gilovich, Griffin, and Kahneman, eds., 271-291.

A 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:

$$\begin{aligned} \Delta_{jq}^2 &= \frac{1}{N_j} \sum_{i=1}^{N_j} (x_{ijq} - \overline{x}_{jq})^2 \; + \; (\overline{x}_{jq} - Truth_q)^2 \\ &= s_{jq}^2 + \alpha_{jq}^2 \end{aligned}$$

QED.

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 distance of group j answers from the truth, σ^2_{Tq}, is:

$$\widehat{\sigma^2_{jq}} = \frac{1}{N_j}\sum_{i=1}^{N_j}(x_{ijq} - Truth_q)^2 \tag{14}$$

The total MSE between a group's answers and the correct answer consists of the sample variance and the squared group bias. Where $\overline{x}_{jq}$ is the average answers for group j members for question q, the MLE for the sample variance, s^2_{jq}, is

$$s^2_{jq} = \frac{1}{N_j}\sum_{i=1}^{N_j}(x_{ijq} - \overline{x}_{jq})^2 \tag{15}$$

Since the sample variance 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 (14) into (15) then gives

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

which leads to

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

Since the questions are assumed to be independent, the MLE for ρ_j is just the average of the estimates provided by all the Q questions:

$$\widehat{\rho}_j = \frac{1}{Q}\sum_{q=1}^{Q}\widehat{\rho}_{jq} = \frac{1}{Q}\sum_{q=1}^{Q}\frac{\widehat{\alpha^2_{jq}}}{\widehat{\sigma^2_{jq}}}$$

Similarly, question independence implies that averaging across questions gives the MLE for the ratio of American and Thai MSEs:

$$\frac{\widehat{\Delta_A^2}}{\Delta_T^2} = \frac{1}{Q}\sum_{q=1}^{Q} \frac{\widehat{\Delta_{Aq}^2}}{\widehat{\Delta_{Tq}^2}}$$

QED.

A.1.3 Proof of Proposition 3

Assuming that $n_T\lambda_T + n_A\lambda_A + \lambda_s = 1$ 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:

$$E(MSE) = E\left(\lambda_T \sum_{k=1}^{n_T}(x_{Tk} - Truth) + \lambda_A \sum_{k=1}^{n_A}(x_{Ak} - Truth) + \lambda_s(x_s - Truth)\right)^2.$$

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

$$\begin{aligned} E(MSE) &= n_T\lambda_T^2\Delta_T^2 + n_T(n_T-1)\lambda_T^2\rho_T\Delta_T^2 + n_A\lambda_A^2\Delta_A^2 + n_A(n_A-1)\lambda_A^2\rho_A\Delta_A^2 \\ &\quad +2cn_T\lambda_T\lambda_s\rho_T\Delta_T^2 + c\lambda_T^2\Delta_T^2 \end{aligned}$$

The first two terms in the above expression describe the expected MSE of all the observed Thai answers and the error that comes from the shared group bias among the n_T Thais (a total of $n_T(n_T-1)$ interactions). The next two terms capture the analogous errors for the observed American answers. The next-to-last term describes the shared group bias between the subject herself and the other Thais she observes. The final term describes the subject's own perceived MSE.

Taking the derivatives with respect to the weights gives the expressions in Proposition 2. QED.

A.1.4 Proof of Proposition 4

Consider the case when a subject uses observed answers for the sum question to update her answer for the Bangkok/Thailand question. A subject updates her answer for the Bangkok/Thailand question based on her initial answer for the Bangkok/Thailand question and the distance between the answers she observes for the sum question and her initial answer for the sum question.

When she observes n_A American answers and n_T Thai answers, her expected mean-squared prediction error is

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

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 optimal weight ratio $\frac{\phi_A}{\phi_T}$ can be expressed as a function of the optimal weight ratios derived in Proposition 2 for how subjects should weigh American information relative to Thai information for the Bangkok/Thailand questions and the Boston/US questions: $\left(\frac{\lambda_A}{\lambda_T}\right)_{Thai}$ and $\left(\frac{\lambda_A}{\lambda_T}\right)_{US}$. The optimal weight ratio is

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

where

$$\begin{aligned} y_A &= (1+(n_A-1)\rho_{Aa})(1-\rho_{Ta}) \\ y_T &= (1+(n_A-1)\rho_{At})(1-\rho_{Tt}) \end{aligned}$$

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 simple expression:

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

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

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

This gives the desired result:

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

QED.

A.2 Data Appendix

A.2.1 Standardization

Where the subscript i denotes the individual, j denotes the group (either A or T), and q denotes the question, any variable y_{ijq} is standardized in the following way:

$$z_{ijq} = \frac{y_{ijq} - \overline{y}_{Tq}}{\frac{1}{2}(s_{Tq} + s_{Aq})} .$$

In the above equation, s_{Tq} is the standard deviation of the Thai answers, s_{Aq} is the standard deviation of the American answers, and $\overline{y}_{Tq}$ is the mean of the Thai answers, for question q.

Another possibility is to just use the standard deviation of the Thai answers to standardize the data. For questions where Thais have a small standard deviation, this standardization would cause the standardized American mean to be far from zero if the American and Thai means are different. These points can then have excessive influence on the regression reported in the text. Still, the results remain essentially the same if the data are standardized in this way.

A.2.2 Additional Tables

The regression can be expanded to include terms that account for the number of answers that subjects observe and the relative accuracy of Americans and Thais across questions. To account for accuracy, I employ the following measure:

$$Acc_q = \frac{\widehat{\Delta}^2_{Tq}}{\widehat{\Delta}^2_{Aq} + \widehat{\Delta}^2_{Tq}} . \tag{16}$$

This measure is the ratio of Thai MSE for question q to the sum of American and Thai MSE. When Acc_q is high, Americans are better relative to Thais for the question q.[14]

The full regression allows the weights that a subject uses to vary with how much American and Thai information she sees and also on the question that the subject is answering, so that:

$$\beta_s = \beta_{s,1} + \beta_{s,2} n_{iA} + \beta_{s,3} n_{iT} + \beta_{s,4} Acc_q$$

$$\beta_A = \beta_{A,1} + \beta_{A,2} n_{iA} + \beta_{A,3} n_{iT} + \beta_{A,4} Acc_q$$

$$\beta_T = \beta_{T,1} + \beta_{T,2} n_{iA} + \beta_{T,3} n_{iT} + \beta_{T,4} Acc_q$$

For example, the regression that includes terms involving the number of observed American answers is:

$$y_{iq} = \beta_{s,1} x_{iq} + \beta_{s,2} n_{iA} x_{iq} + \beta_{A,1} \overline{x}_{iAq} + \beta_{A,2} n_{iA} \overline{x}_{iAq} + \beta_{T,1} \overline{x}_{iTq} + \beta_{T,2} n_{iA} \overline{x}_{iTq} + Q' \beta_q + \varepsilon_{iq} \quad (17)$$

Estimating these regressions gives the average weights that subjects put on Thai answers, American answers, and their own initial answers for each question type. Table A1 describes subject behavior for the all three types of questions.

[14] Subjects may also want to account for the spread in the answers they observe. For example, when a subject sees two answers that are near each other and one answer that is extreme, she may ignore the extreme answer, deeming it irrelevant. To look at this, I expand the regression to include a term that captures the spread in the observed answers. The spread is measured simply as the standard deviation in the observed information. Inclusion or exclusion of terms involving the standard deviation of observed information does not affect the results.

Table A1: Estimated weights for the US and sum questions

Dependent variable: Subjects' final answers

Regressor	Questions about Thailand				Questions about US				Questions about sum			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Subject's initial answer ($\beta_{s,1}$)	.653	.654	.65	.662	.464	.474	.48	.492	.731	.731	.731	.728
	(.016)	(.016)	(.016)	(.018)	(.02)	(.02)	(.021)	(.021)	(.019)	(.02)	(.02)	(.02)
Initial answer • number of observed American answers ($\beta_{s,2}$)		.002	.008	.004		.003	.001	.013		.001	-.001	-.001
		(.019)	(.019)	(.019)		(.024)	(.024)	(.024)		(.024)	(.024)	(.024)
Initial answer • number of observed Thai answers ($\beta_{s,3}$)			-.038	-.04			.041	.045			-.002	.002
			(.019)	(.019)			(.024)	(.024)			(.026)	(.026)
Initial answer • accuracy index ($\beta_{s,4}$)				-.142				-.273				.266
				(.084)				(.127)				(.13)
Thai average ($\beta_{T,1}$)	.238	.242	.242	.24	.09	.086	.054	.064	.068	.07	.064	.076
	(.02)	(.021)	(.025)	(.026)	(.03)	(.03)	(.036)	(.036)	(.02)	(.025)	(.028)	(.032)
Thai average • number of observed American answers ($\beta_{T,2}$)		.004	.002	.012		-.015	-.004	.003		.006	.011	.008
		(.024)	(.024)	(.025)		(.034)	(.035)	(.038)		(.03)	(.031)	(.03)
Thai average • number of observed Thai answers ($\beta_{T,3}$)			.001	-.002			-.053	-.038			-.024	-.017
			(.025)	(.026)			(.038)	(.041)			(.037)	(.037)
Thai average • accuracy index ($\beta_{T,4}$)				.04				-.123				-.151
				(.098)				(.221)				(.204)
American average ($\beta_{A,1}$)	.056	.049	.039	.063	.463	.481	.476	.455	.165	.092	.088	.081
	(.012)	(.014)	(.015)	(.016)	(.019)	(.02)	(.021)	(.021)	(.024)	(.037)	(.037)	(.038)
American average • number of observed American answers ($\beta_{A,2}$)		-.016	0	-.003		.075	.09	.07		-.07	-.069	-.061
		(.017)	(.019)	(.019)		(.022)	(.023)	(.024)		(.032)	(.032)	(.032)
American average • number of observed Thai answers ($\beta_{A,3}$)			-.036	-.027			-.046	-.053			-.023	-.012
			(.018)	(.018)			(.024)	(.024)			(.025)	(.025)
American average • accuracy index ($\beta_{A,4}$)				.232				.398				.351
				(.055)				(.124)				(.139)
Number of observations	1008	986	986	986	1052	1032	1032	1032	557	548	548	548

Notes:
(1) Regression standard errors are in parentheses.
(2) The regression in column (1) includes the data for questions 4 and 8 where subjects saw either 0, 5, 10, or 20 American and Thai answers.
(3) Regressions include dummies for the three question categories (meteorology, economic/political, and social/cultural).

Table A2 reports the results of estimating regressions using the correct answers to the questions as the dependent variables with the same regressors as the specifications in Table A1. The estimates in Table A2 describe how subjects would optimally behave.

Table A2: MSE minimizing behavior for the all three types of questions

Dependent variable: Correct answers to the questions

Regressor	Questions about Thailand				Questions about US				Questions about sum			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Subject's initial answer ($\beta_{s,1}$)	.292	.289	.266	.285	.063	.057	.042	.086	.179	.148	.152	.142
	(.021)	(.021)	(.021)	(.019)	(.017)	(.017)	(.017)	(.014)	(.026)	(.028)	(.029)	(.022)
Initial answer • number of observed American answers ($\beta_{s,2}$)		.034	.052	.032		-.055	-.048	-.029		-.117	-.118	-.121
		(.024)	(.024)	(.021)		(.019)	(.019)	(.015)		(.034)	(.034)	(.025)
Initial answer • number of observed Thai answers ($\beta_{s,3}$)			-.114	-.115			-.079	-.095			.018	-.044
			(.023)	(.02)			(.02)	(.016)			(.036)	(.027)
Initial answer • accuracy index ($\beta_{s,4}$)				-.897				-.747				-.517
				(.086)				(.083)				(.132)
Thai average ($\beta_{T,1}$)	.458	.437	.497	.403	.115	.079	.092	.157	.25	.249	.247	.405
	(.022)	(.023)	(.025)	(.023)	(.02)	(.02)	(.021)	(.016)	(.024)	(.032)	(.034)	(.027)
Thai average • number of observed American answers ($\beta_{T,2}$)		-.145	-.187	-.151		-.137	-.135	-.056		.038	.038	.022
		(.029)	(.029)	(.026)		(.023)	(.023)	(.019)		(.035)	(.035)	(.026)
Thai average • number of observed Thai answers ($\beta_{T,3}$)			.182	.144			.036	.118			-.006	-.001
			(.028)	(.025)			(.022)	(.018)			(.036)	(.027)
Thai average • accuracy index ($\beta_{T,4}$)				-.231				-1.406				-2.316
				(.091)				(.1)				(.154)
American average ($\beta_{A,1}$)	.25	.274	.237	.312	.822	.865	.866	.757	.571	.603	.601	.453
	(.018)	(.019)	(.02)	(.018)	(.016)	(.016)	(.017)	(.014)	(.026)	(.036)	(.038)	(.029)
American average • number of observed American answers ($\beta_{A,2}$)		.112	.135	.119		.192	.183	.084		.08	.08	.099
		(.024)	(.026)	(.022)		(.019)	(.019)	(.015)		(.033)	(.033)	(.025)
American average • number of observed Thai answers ($\beta_{A,3}$)			-.068	-.028			.043	-.023			-.011	.045
			(.025)	(.022)			(.019)	(.015)			(.031)	(.023)
American average • accuracy index ($\beta_{A,4}$)				1.128				2.153				2.833
				(.064)				(.082)				(.14)
Number of observations	1008	986	986	986	1052	1032	1032	1032	557	548	548	548

Notes:
(1) Regression standard errors are in parentheses.
(2) The regression in column (1) includes the data for questions 4 and 8 where subjects saw either 0, 5, 10, or 20 American and Thai answers.
(3) Regressions include dummies for the three question categories (meteorology, economic/political, and social/cultural).
(4) The regressions are estimated under the constraints:
$\beta_{s,1} + \beta_{A,1} + \beta_{T,1} = 1$ and $\beta_{s,j} + \beta_{A,j} + \beta_{T,j} = 0$ for j = 2, 3, or 4.