12-1-2009

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

Andrew Healy
Loyola Marymount University, ahealy@lmu.edu

Recommended Citation

Healy, Andrew. 2009. "How Effectively Do People Learn from a Variety of Different Opinions?" Experimental Economics 12(4): 386-416.

How Effectively Do People Learn from a Variety of Different Opinions?

Andrew J. Healy*

Loyola Marymount University

June 20, 2008

Abstract

This paper presents a model of information aggregation in which individuals can draw upon information from heterogeneous sources to improve decisions and then tests that model using experimental data. In the experiment, Thai subjects observed the opinions of Americans and other Thais that they could use to help them answer a series of general knowledge questions. Despite listening too little to either group, subjects demonstrated a significant amount of statistical sophistication in how they weighed observed American information relative to observed Thai information. The data indicate that subjects understood that outside information has extra value because people from the same group tend to make the same kinds of mistakes. The results illustrate the importance of forming diverse groups to solve problems.

---

*E-mail: ahealy@lmu.edu; address: University Hall 4229, Los Angeles, CA 90045; phone: 310-338-5956. I thank Victor Chernozhukov, Esther Duflo, and Sendhil Mullainathan for their advice and support. Special thanks go to Adis Israngkura, Chayun Tantisidivakarn, and Somchai Jitsuchon for helping to arrange the experiment. I thank Ken Chetnakarnkul and Phasook Mengkred for dedicated research assistance. For helpful discussions, I acknowledge and thank Dan Ariely, Gary Charness, Shawn Cole, Ernst Fehr, Jerry Hausman, Dorothea Herreiner, Peter Hinrichs, Dean Karlan, John List, Whitney Newey, Daniel Paravisini, Al Roth, Chinda Wongngamnit, and seminar participants at the University of Essex, MIT, the 2005 ESA meetings, and the 2007 AEA meetings. I thank Russell Sage and the MIT Shultz Fund for financial support, and the NSF for a graduate fellowship. All errors are my own.

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

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

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

The experimental design relates to a variety of real-world examples in which individuals can use information to improve decisions. Information sharing within social networks has been shown to influence agricultural technology adoption, health decisions, and savings behavior, but not always for the better.[1] The research shows that individuals put high weight on information learned from others within their own group and that information sharing between groups often does not occur. Information sharing generally improves decision-making, but the lack of sharing across groups sometimes leads entire groups to choose the wrong decision (Dearden, Pritchett, and Brown, 2004). Would decision-making be improved if individuals had access to a variety of different information sources? Are people generally able to aggregate a variety of different opinions in an intelligent way? The controlled environment of the lab makes it possible to answer yes to both of these questions. In fact, the experimental subjects show an implicit appreciation for somewhat subtle statistical ideas when deciding how to weigh information from a variety of different sources.

In the experiment, Thai subjects considered information that came from sources with different cultural backgrounds. The subjects first answered a series of general-knowledge questions that had correct numerical answers. There were three types of questions: 1) questions about Thailand, 2) questions about the US, and 3) questions about both Thailand and the US. After

[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.

answering the questions on their own, subjects observed randomly selected answers given by Americans and by other Thais, who had answered the same questions at an earlier date, that they could use to help them revise their answers. By looking at how subjects changed their answers, it is possible to estimate the weights that they applied to observed American answers, to observed Thai answers, and to their own initial answers.

The data show that, even for the cases in which 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 were about equally good at answering the questions about both Thailand and the US. The data show that it was 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 Americans tended to make one kind of mistake and Thais tended to make a different kind of mistake. When members of the same group tend to make the same kind of mistake, an agent has more to learn from members of a different group than from other members of her own group.

In general, the subjects appeared to understand this idea, behaving optimally in how they weighed American information relative to Thai information. Subjects achieved this optimal relative weighting despite listening too little to either group. Although they would have benefited by listening more to both groups, subjects significantly improved their performances by correctly weighing observed American answers relative to observed Thai answers. Moreover, the presence of the questions about both Thailand and the US makes it possible to show that subjects appreciated the extra value of an American's independent perspective to a Thai decision-maker.

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

For decision-makers to avoid these sorts of mistakes, the experiment confirms what the anecdotal evidence suggests: access to a diverse set of information is essential. Subjects show the cognitive ability to deal in a sophisticated way with information from a variety of sources. Specifically, subjects understand and can apply the idea of statistical dependence between the mistakes that they make and that their other group members make to problem solving tasks. At the same time, the experimental subjects are constrained by the excessive weight they give to their own initial answers. Improving decision-making thus appears to primarily be about reducing the overconfidence and other causes behind the failure of decision-makers to access independent perspectives. The experiment shows that, conditional on those voices being heard, decision-makers can effectively learn from a variety of different opinions.

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 describes how the experimental design makes it possible to test the hypothesis that subjects appreciate the independence in outside information. Section 8 concludes.

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

2 Experimental Design

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

2.1 The Questionnaire

The questionnaire consisted of fifteen questions covering a range of topics. For example, one question asked about the January temperature in Bangkok, the January temperature in Boston, and the sum of those temperatures. Another question asked for the number of Thai prime ministers since 1960, the number of American presidents since 1960, and the sum of those two numbers. A third question asked about the 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 one of the questions. The appendix contains all of the questions.

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

Figure 1: Sample question from the experiment

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

For this reason, the presence of the sum questions is a key design feature in the experiment. The intuition is that there are two pieces to the sum question: the Thai part and the American part. As the data will show, Americans and Thais are equally good at answering the questions about the sum. Still, Thais have more to learn from American answers about the sum question because Americans know the piece of the puzzle about which Thai subjects generally have less knowledge. I demonstrate these ideas formally in Section 7.

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

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

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

2.3 Stage 2: Showing American and Thai information to subjects

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

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

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

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

Answers given by Thai students

1. __18__

Answers given by American students

1. __7__
2. __9__
3. __10__

Figure 2: Sample of information that subjects observe

The subjects in this group observed that one randomly selected Thai student thought the number of American presidents was 18. They also observed that three randomly selected American students thought the number of American presidents was 7, 9, and 10, respectively.

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. Second, I randomly chose how many Thai answers that subjects observed (up to three). Third, I randomly chose how many Thai and American 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 20 sets of information that subjects could observe.

In total, each experimental session took approximately 50 minutes. I conducted 25 sessions, 17 at Thammasat and 8 at NIDA. Payments averaged approximately 280 baht ($7) per subject.

2.4 Controlling for anchoring

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

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. The data show that anchoring had a small and statistically insignificant effect on subject behavior in the experiment. Details on the test for anchoring can be found in the appendix.

3 A Model of Information Aggregation

Here, I describe a model of information aggregation that shows how the Thai subjects should weigh the information they see to minimize the mean-squared error (MSE) of their answers. The model applies to each question type (Bangkok/Thailand, Boston/US, or sum) separately. For a question q, take individual i in group j (either A for American or T for Thai), to have a private belief x_{ijq} about the correct answer for the question.[3] The MSE, Δ^2_{jq}, for a group j for question q is then

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

[3]The empirical results indicate that there are no significant differences between individuals with different personal characteristics (i.e. gender) or socioeconomic status in how they treat the information available to them (results available upon request). Therefore, in the model, I will treat all subjects to have the same objective function.

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 estimators of the population variance for the group (s_{jq}^2) and the squared group bias (α_{jq}^2), 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 the appendix.* ■

This decomposition reflects the fact that the total error made by the group consists of individual and group components. The individual component, s_{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}:

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

If individuals in a group tend to make the same kind of mistake, ρ_{jq} will be high since 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, the maximum-likelihood estimators (MLE) for these three 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 the appendix.* ■

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 implicit perceptions of $\frac{\Delta_T^2}{\Delta_A^2}$, ρ_T, and ρ_A determine the average weights that she will apply to the information that she observes. The actual values of the parameters determine what she would optimally do. Without loss of generality, I ignore heterogeneity among Americans and heterogeneity among Thais. The model can be expanded to accommodate this heterogeneity, and the average weights across subjects are identical to the expressions derived below. For the sake of simplicity, the model assumes that subjects treat each American answer that they see in the same way and each Thai answer that they see in the same way, for any given question. As will be discussed, the regression used to estimate subject behavior can allow for subjects putting different weight on outliers and taking into account the spread in the answers that they see. Expanding the regression in this way does not significantly affect the estimates for the average weights that subjects apply to the information that they observe.

To accurately describe subject behavior, it is necessary to model subjects as perceiving their own answers as being of different quality than those of the other Thais they observe. I model a subject to perceive her own MSE to be a fraction c of another Thai's. Where Δ_{Sq}^2 is a subject's perceived MSE for question q,

$$\Delta_{Sq}^2 = c\Delta_{Tq}^2\ . \tag{1}$$

If subjects are overconfident, they perceive themselves to be better than other Thais, so that $c < 1$.[4]

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:

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

where

λ_S = weight for own information

λ_T = weight for any one piece of Thai information

λ_A = weight for any one piece of American information

Assuming independence between the American and Thai group biases and that the weights given to all information sum to one, the expressions in Proposition 3 below capture the weights that subjects would optimally use. Both of these assumptions are confirmed by the experimental data. The expressions below define optimal behavior conditional on any level of overconfidence.

Proposition 3 *The following expressions define the MSE-minimizing weights that subjects should*

[4] This modeling of overconfidence corresponds to the idea that a subject perceives her confidence interval to be c times the width of another Thai's, for any given significance level. For experimental evidence on overconfidence, see Gigerenzer, Hoffrage, and Kleinbolting (1991), Griffin and Tversky (1992), Camerer and Lovallo (1999), and Hoelzl and Rustichini (2005). For field evidence on overconfidence, see Barber and Odean (2001), Scheinkman and Xiong (2003), and Daniel et al. (1998).

use to evaluate information:

$$\textit{Weighing self relative to other Thais:} \quad \frac{\lambda_S}{\lambda_T} = \frac{1}{c} + \frac{1-c}{c}\left(\frac{\rho_T}{1-\rho_T}\right)n_T \tag{2}$$

$$\textit{Weighing self relative to Americans:} \quad \frac{\lambda_S}{\lambda_A} = \left(\frac{\Delta_A^2}{\Delta_T^2}\right)\left(\frac{1-\rho_A}{c} + \frac{\rho_A}{c}n_A\right) \tag{3}$$

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

Proof. *See the appendix.* ■

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

Notice also that increases in ρ_A only cause subjects to put less weight on individual American answers when n_A is greater than one, but increases in ρ_T cause subjects to put less weight on observed Thais even when only one Thai is observed. When one Thai is observed, there are two Thai answers to consider: a subject's own answer and the one she observes. As a result, the Thai group bias term enters (4) when $n_T = 1$, but the American group bias term only enters when $n_A > 1$.

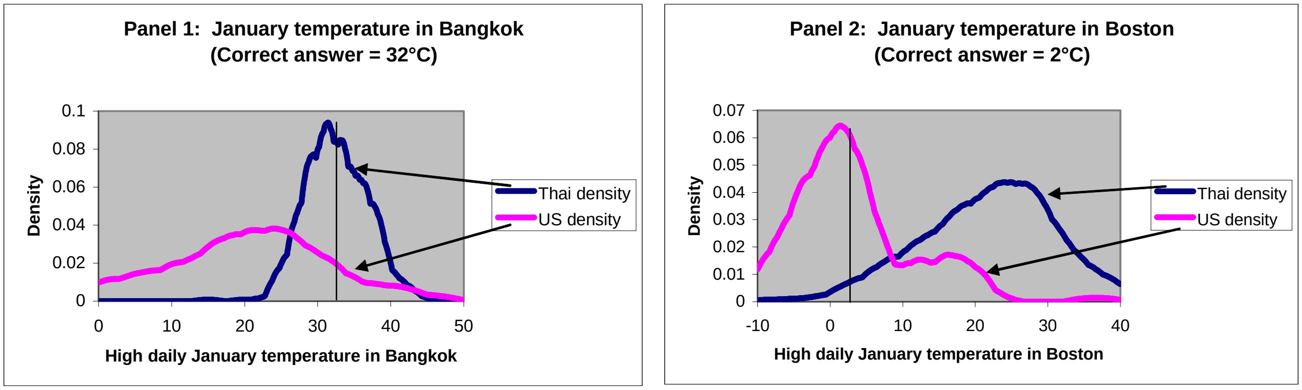
The experimental data on how subjects update their answers provide estimates of the actual weights that subjects use. While the experiment does not directly observe subjects' perceptions of $\frac{\Delta_T^2}{\Delta_A^2}$, ρ_T, and ρ_A, the following sections show how subject behavior makes it possible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

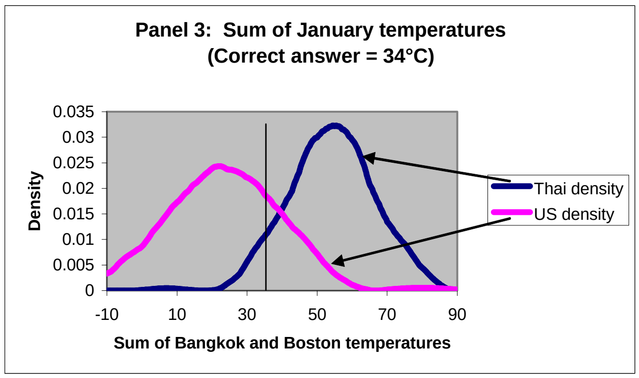
4 Summary Statistics

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

It is important to note that other questions show a different pattern than the question about January temperature. For other questions, the American average is not as 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 a greater opportunity to learn from the other group than from one's own group.

Across questions, the data provide estimates of the average Thai-to-American MSE ratio for each of the three types of questions. For the questions about Thailand, $\widehat{\frac{\Delta_T^2}{\Delta_A^2}}$ is 0.517, meaning that the expected squared distance between a randomly selected American answer and the correct answer is about twice as large as the expected squared distance between a randomly selected Thai answer and the correct answer. I will describe this kind of result as Thais being





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

(2) Vertical lines correspond to the correct answers

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

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' answers.[5]

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

Table 1: Relative accuracy of Americans and Thais

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

Note: Bootstrapped standard errors in parentheses

Consider the second column in Table 1. Thais have about one-half the MSE of Americans for the questions about Thailand. Americans are three times more accurate for the questions about the US and about 1.3 times more accurate for the sum questions. These ratios exactly describe the weights that a subject should use if group bias did not matter. For the sum questions, for example, a subject should put 1.3 times more weight on any observed American answer than she puts on any Thai answer. The data will show that group bias means that a subject should actually put about twice as much weight on American answers than on Thai answers.

The experimental design also enables me to estimate the share of group bias in total MSE for each question type, both for Americans and for Thais. The group bias share expresses what fraction 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 the Thai group bias share and column (2) contains estimates of the American group bias share for each question type.

Table 2: Group effects for Americans and for Thais

Question type	Estimated Thai group bias share (ρ_T)	Estimated American group bias share (ρ_A)
	(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, group bias is responsible for the smallest share of total American MSE for the Boston/US questions and the largest share for the Bangkok/Thailand questions. For example, Thais make small errors about the average high daily January Bangkok temperature and group bias causes a small share of that error. On the other hand, Thais make much larger errors for the January Boston temperature, and a larger share of their mistakes comes from the fact that the group mean for Thais was $20°C$. As I show in Section 6, the fact that Thai group bias is biggest for the Boston/US questions has important implications for how a subject would optimally behave when she sees information for the sum question only.

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

5 Estimating Subject Behavior

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

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

I estimate the following regression equation:

$$y_{iq} = \beta_S x_{iq} + \beta_A \overline{x}_{iAq} + \beta_T \overline{x}_{iTq} + C_q' \beta_q + \varepsilon_{iq} \tag{5}$$

Optimal behavior implies that a subject would choose 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 additional specifications that I consider beyond (5) are contained in the appendix.

The basic results, obtained by estimating (5) for each of the three types of questions, are summarized in Table 3.

Table 3: Summary of estimated weights

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

Notes:
(a) Regression standard errors are in parentheses.
(b) 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 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. To compensate for putting more weight on American answers, subjects put less weight on their initial answers for those questions where American answers are particularly valuable.

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. Apparently understanding this fact, the Thai subjects put higher relative weight on Thai information for the question about Bangkok weather.

It is important to note that subjects improve their earnings considerably by changing their answers after observing information. Compared to what they would have earned with the answers they gave before observing information, subjects earn 14% more on the questions about Thailand when they observe information about those questions, 58% more for the questions about the US, and 24% more for the questions about the sum.

6 Tests for Optimal Behavior

The model described in Section 3 provides estimates of the optimal weights that a subject should use. These estimates come from substituting the estimates of the American-to-Thai MSE ratio, the American group bias share, and the Thai group bias share into equations (2), (3), and (4). Panel A of Table 4 displays these estimates of the optimal weights given that $c = 1$, that subjects on average consider their own answers as being no more or less accurate than the answers given by the Thais they observe.

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

Table 4: Estimating the optimal weights

Optimal weight	Questions about Thailand (1)	Questions about US (2)	Questions about sum (3)
Own initial answer (β_S)	.245	.087	.146
	(.007)	(.008)	(.009)
Thai average (β_T)	.480	.168	.300
	(.014)	(.015)	(.02)
American average (β_A)	.274	.745	.554
	(.021)	(.024)	(.028)

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

6.1 Construction of Confidence Intervals

Simulations using the regression coefficients and their variance-covariance matrix obtained by estimating (5) give a confidence interval for $\frac{\beta_A}{\beta_T}$, the weight ratio that expresses how subjects actually weigh American compared to Thai information. Also, the distributions of Thai and American answers make it possible to generate confidence intervals for weight ratios under different hypotheses that could be driving subject behavior. I focus on confidence intervals for two such weight ratios: 1) the simple weight ratio, which expresses how a subject would behave if she understood each group's accuracy but ignored group bias, and 2) the optimal weight ratio described in equation (4), which expresses how a subject would behave if she correctly perceived each group's accuracy and accounted for group bias. The parameter estimates in Tables 1 and 2 give confidence intervals for the simple weight ratio and for the optimal weight ratio for any value of the overconfidence parameter c.

To summarize, I consider the following three weight ratios:

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

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

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

Table 5 reports confidence intervals for these three ratios for each of the three question types. The p-values in the table correspond to tests that I describe in the next subsection.

Table 5: Comparing how subjects relatively weigh American and Thai information

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

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

The optimal weight ratio increases substantially due to group bias. The optimal weight ratio also increases as overconfidence increases, but not as much. Even for very large overconfidence, 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 by dropping c to 0.25 causes the optimal weight ratio to further rise to 5.04.[6]

[6] The data can also be used to estimate c using a nonlinear model, but the standard errors on the resulting estimates are large. The estimated value of c falls between 0.20 and 0.45, on average, with confidence intervals that are too wide to draw firm conclusions about the true value of c.

6.2 Hypothesis Tests

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

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

Under H_0, as shown in Section 3, subjects will choose

$$\frac{\beta_A}{\beta_T} = \left(\frac{\Delta_T^2}{\Delta_A^2}\right) . \tag{6}$$

Rejection of the prediction (6) implies rejection of H_0.

The second row of Table 5 displays the results of testing for the equality of the ratios in equation (6) for all three types of questions. For the Bangkok/Thailand questions and the sum questions, we can reject equality at the 5% level ($p = 0$ and $p = 0.017$, respectively). For the Boston/US questions, we can reject it at the 10% level ($p = 0.065$). We reject the hypothesis for the Bangkok 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 importance 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 to a Thai subject.

Under H_1, subjects will choose the optimal weight ratio

$$\frac{\beta_A}{\beta_T} = \left(\frac{\Delta_T^2}{\Delta_A^2}\right)\left(\frac{1+(n_T-1)\rho_T - c\rho_T^2 n_T}{(1+(n_A-1)\rho_A)(1-\rho_T)}\right) \tag{7}$$

Table 5 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.[7] For $c = 1$, the optimal weight ratio estimates are 4.42 and 1.84, compared to the actual weight ratio estimates of 5.14 and 2.49. Tests of equality give p-values of 0.622 and 0.346, respectively.

Now consider the optimal weight ratios for the Boston/US and sum questions when overconfidence is taken into account. Given $c = 0.25$, the actual and optimal weight ratios match up quite closely. For the Boston/US questions, the optimal weight ratio estimate is 5.13 and the actual weight ratio estimate is 5.14. The test for equality between the two, not surprisingly, gives a p-value of nearly one ($p = 0.986$). For the sum questions, the optimal weight ratio estimate is 2.14, compared to the actual weight ratio of 2.49 ($p = 0.676$). In summary, individuals who weigh their initial answers more heavily than answers given by other Thais, as the experimental subjects do, would optimally weigh American answers relative to Thai answers in a very similar way to how the subjects actually behave.

[7] 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 likely that the Thailand heading for the questions about Thailand causes subjects to underweigh American information. In other words, only when the question is clearly about a Thai's area of expertise do the subjects listen too little to Americans.

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 they incorrectly perceive American answers to be better than they really are.

7 Do Subjects Value the Independence in Outside Information?

The presence of the sum questions makes it possible to distinguish between H_1 and H_2. Specifically, the test that distinguishes between these hypotheses uses the data generated by the case in which subjects observe information about the sum question only. Consider the following example of what one group of subjects observed for the question about political leaders:

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

Answers given by Thai students

	1. 35

Answers given by American students

	1. 10
	2. 12
	3. 17

Figure 4: Sample of information that subjects observe

Subjects who observed the above information had to use it to update their answers for both the question about Thai prime ministers and the question about American presidents. By looking at how subjects *separately* update their answers for the Thailand question and the US question, it is possible to test the hypothesis that subjects ignore group bias when accumulating the knowledge that the observe.

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

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

For example, $\rho_{T,US}$ 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 the Bangkok/Thailand question based on 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 for the Bangkok/Thailand question, she is likely to revise her answer upwards.

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

Proof. See the appendix. ■

Proposition 4 refers to a subject who implicitly perceives there to be no Thai 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.

$$G_0 : \left(\rho_{T,Thai}\right)_{perceived} = \left(\rho_{T,US}\right)_{perceived} = 0$$

This hypothesis states that subjects ignore Thai group bias for both the Bangkok/Thailand and Boston/US questions. Notice that rejection of G_0 would imply rejection of H_2, the hypothesis that subjects perceive Americans to have lower MSE relative to Thais than they actually do and ignore group bias. In other words, since rejection of $\left(\rho_{T,Thai}\right)_{perceived} = \left(\rho_{T,US}\right)_{perceived} = 0$ implies rejection of $\left(\rho_T\right)_{perceived} = 0$ for any question type, rejection of G_0 means we must also reject H_2.

Under G_0, subjects will choose

$$\left(\frac{\phi_A}{\phi_T}\right)_{Thai} = \left(\frac{\phi_A}{\phi_T}\right)_{US} \tag{8}$$

The left-hand side of (8) 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 (8) represents the analogous ratio that subjects use to update for the US questions. Rejection of the equality in (8) would imply rejection of G_0.

The following two regressions give the parameter estimates needed to conduct a test of the equality of the ratios in (8). The first equation expresses the change in a subject's answer for the Thai question as a function of the distance between the average observed American answer for the sum question and her own answer for the sum question and the distance between the average observed Thai answer for the sum question and her own answer for the sum question. The second equation expresses how subjects update their answers for the US question after observing information relating to the sum question.

Where

$$y_{iq,Thai} = \text{final answer given by subject } i \text{ to the Bangkok/Thailand part of question } q,$$
$$x_{iq,Thai} = \text{initial answer given by subject } i \text{ to the Bangkok/Thailand part of question } q,$$
$$y_{iq,US} = \text{final answer given by subject } i \text{ to the Boston/US part of question } q,$$
$$x_{iq,US} = \text{initial answer given by subject } i \text{ to the Boston/US part of question } q,$$
$$x_{iq,Sum} = \text{initial answer given by subject } i \text{ for the sum part of question } q,$$

the regression equations are:

$$y_{iq,Thai} - x_{iq,Thai} = \phi_{A,Thai}(\overline{x}_{iAq,Sum} - x_{iq,Sum}) + \phi_{T,Thai}(\overline{x}_{iTq,Sum} - x_{iq,Sum}) + C_q'\phi_1 + \varepsilon_{iq1} \quad (9)$$
$$y_{iq,US} - x_{iq,US} = \phi_{A,US}(\overline{x}_{iAq,Sum} - x_{iqs,Sum}) + \phi_{T,US}(\overline{x}_{iTq,Sum} - x_{iq,Sum}) + C_q'\phi_2 + \varepsilon_{iq2} \quad (10)$$

Table 6 reports the results from estimating equations (9) and (10). 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 the equality in (8). We can reject equality at a 10% level ($p = 0.058$). At a 10% level, we thus reject G_0, the hypothesis that subjects fail to take group bias into account, regardless of how they perceive American and Thai accuracy.

In contrast, correctly accounting for group bias would lead subjects to behave in a similar way to how they actually do behave. If the perceived group bias share for Thais for the Boston/US questions ($\rho_{T,US}$) is greater than the perceived group bias share for Thais for the Bangkok/Thailand questions ($\rho_{T,Thai}$), then subjects will put a higher relative weight on observed Americans for the US questions than for the Thailand questions. The proof of Proposition 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 question about Thailand after observing information about the sum question. If Thai group bias for the questions about Thailand ($\rho_{T,Thai}$) is high, she should put less weight on Thais relative to Americans because each additional Thai answer contains little new information. On the other hand, if

Table 6: Updating for the Thai and US questions after observing answers for the sum

Dependent variables: (1) Final answer for Thai question
(2) Final answer for US question

Regression weights	Thailand questions (1)	US questions (2)
ϕ_T = Distance between observed Thai average 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:
(a) Regression standard errors are in parentheses.
(b) Regressions include dummies for question categories (meteorology, economic/political, and demography).

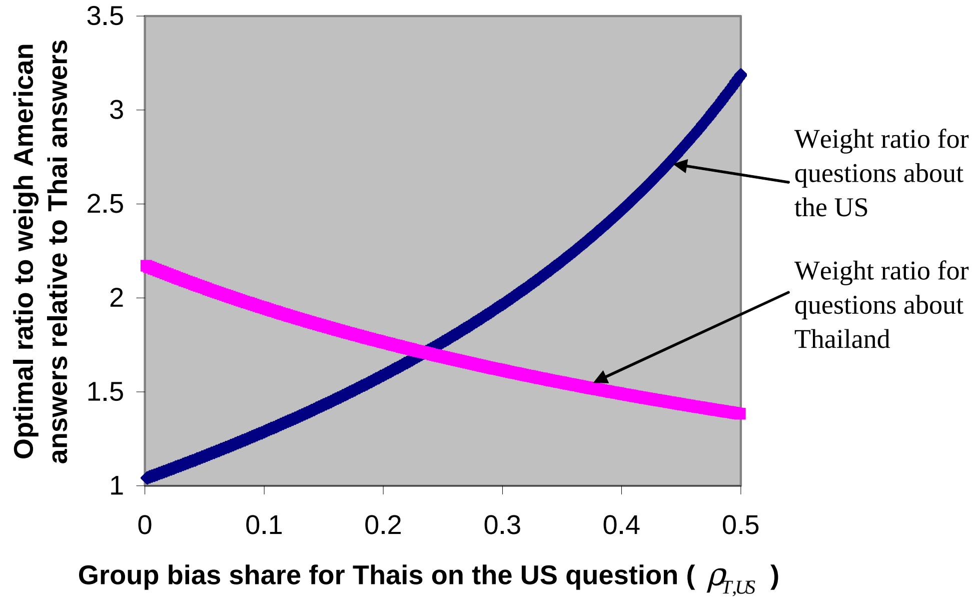


Figure 5: Optimal behavior when information for the sum question is observed

Thai group bias for the Boston/US questions ($\rho_{T,US}$) is high, she should put *higher* weight on observed Thai answers for the sum question. When $\rho_{T,US}$ is high, Thai subjects have a better idea of what other Thai answers about the sum mean for what those observed students believe about the Thai question. For example, if $\rho_{T,US}$ was equal to one, all Thais would give the same answer to the US question. Then, a subject could exactly deduce an observed Thai's private belief about the Bangkok/Thailand question from her answer to the sum question.

The data show that the effects of group bias can explain how subjects actually weigh the information they observe. If subjects applied the 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 $\left(\frac{\hat{\phi}_A}{\hat{\phi}_T}\right)_{US} = 2.53$, similar ratios to the estimates of optimal behavior of 1.34 and 2.79. Figure 4 shows how the optimal weight ratios for the two types of questions vary as a function of $\rho_{T,US}$, holding the other parameters constant at their estimated values. The figure shows that, when updating their answers for the Bangkok/Thailand questions, subjects should put less weight on Americans answers to the sum question relative to Thai answers when $\rho_{T,US}$ is high.

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. The ways in which subjects update for the Bangkok/Thailand and Boston/US questions when they observe answers for the sum question supports the former hypothesis. Subjects appear to appreciate each group's accuracy and the extra value in an American's independent perspective to a Thai subject.

8 Conclusion

This paper demonstrates that economic agents can learn effectively from a diverse set of information. 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 information sources, but also the value of drawing on a variety of independent perspectives.

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).[8] It may be the case that, in general, Indonesian decision-makers would have a more difficult time optimally using information from Americans due to their 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.

The results illustrate the importance of forming diverse groups to solve problems. Unfortunately, the desire for cohesiveness often prevents diverse groups from being formed. Psy-

[8]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.

chological evidence suggests that homogeneous groups become close-knit more easily (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 information. 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 is to make sure that decision-makers hear those independent voices.

References

[1] Anderson, L., and Holt, C. 1997. Information cascades in the laboratory. *American Economic Review* 87: 847-862.

[2] Austen-Smith, D., and Banks, J. 1996. Information aggregation, rationality, and the Condorcet Jury Theorem. *American Political Science Review* 90, 34-45.

[3] Barber, B. and Odean, T., 2001. Boys will be boys: gender, overconfidence, and common stock investment. *Quarterly Journal of Economics* 116, 261-292.

[4] Camerer, C. and Lovallo, D., 1999. Overconfidence and excess entry: an experimental approach. *American Economic Review* 89*,* 306-318.

[5] Daniel, K., Hirshleifer, D., and Subrahmanyam, A., 1998. Investor psychology and security market over- and under-reactions. *Journal of Finance* 53, 1839-1886.

[6] Dearden, K., Pritchett, L., and Brown, J., 2003. Learning from neighbors: social learning about child feeding during dysentery episodes. working paper.

[7] Duflo, E. and Saez, E., 2003. The role of information and social interactions in retirement plan decisions: evidence from a randomized experiment. *Quarterly Journal of Economics* 118, 815-842.

[8] Feddersen, T., and Pesendorfer, W. 1997. Voting behavior and information aggregation in elections with private information. *Econometrica* 65, 1029-1058.

[9] Forsythe, R. and Lundholm, R., 1990. Information aggregation in an experimental market. *Econometrica* 58: 309-347.

[10] Foster, A. and Rosenzweig, M., 1995. Learning by doing and learning from others: human capital and technical change in agriculture. *Journal of Political Economy* 103, 1176-1209.

[11] Gigerenzer, G., Hoffrage, U., and Kleinbolting, H., 1991. Probabilistic mental models: a Brunswikian theory of Confidence. *Psychological Review* 98, 506-528.

[12] Goeree, J., Palfrey, T., Rogers, B., and McKelvey, R., 2007. Self-correcting information cascades. *Review of Economic Studies* 74, 733-762.

[13] Granovetter, M., 1973. The strength of weak ties. *American Journal of Sociology* 78, 1360-1380.

[14] Griffin, D. and Tversky, A., 1992. The weighting of evidence and the determinants of confidence. *Cognitive Psychology* 24, 411-435.

[15] Hellwig, M. F., 1980. On the aggregation of information in competitive markets. *Journal of Economic Theory* 22, 477-498.

[16] Hoelzl, E. and Rustichini, A., 2005. Overconfident: do you put your money on it. *Economic Journal* 115, 305-318.

[17] Janis, I.L., 1972. *Victims of groupthink: a psychological study of foreign policy decisions and fiascoes.* Boston, MA: Houghton-Mifflin.

[18] Kraemer, C., Noth, M., and Weber, M., 2006. Information aggregation with costly information and random ordering: experimental evidence. *Journal of Economic Behavior and Organization* 59, 423-432.

[19] Lohmann, S., 1994. Information aggregation through costly political action. *American Economic Review* 84, 518-530.

[20] Miguel, E. and Kremer, M., 2004. Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica* 72, 159-217.

[21] Munshi, K., 2004. Social learning in a heterogeneous population: technology diffusion in the Indian Green Revolution. *Journal of Development Economics* 73*,* 185-213.

[22] Pesendorfer, W. and Swinkels, J., 1997. The loser's curse and information aggregation in common value auctions. *Econometrica* 65: 1247-1281.

[23] Pesendorfer, W. and Swinkels, J., 2000. Efficiency and information aggregation in auctions. *American Economic Review* 90, 499-525.

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

[25] Piketty, T., 1999. The information-aggregation approach to political institutions. *European Economic Review* 43, 791-800.

[26] Plott, C. and Sunder, S. 1988. Rational expectations and the aggregation of diverse information in laboratory security markets. *Journal of Economic Behavior and Organization* 56, 1085-1118.

[27] Putnam, R., 2000. *Bowling alone*. New York: Simon & Schuster.

[28] Scheinkman, J. and Xiong, W., 2003. Overconfidence and speculative bubbles. *Journal of Political Economy* 111, 1183-1220.

[29] Surowiecki, J., 2004. *The wisdom of crowds.* New York, NY: Doubleday.

[30] Tversky, A. and Kahneman, D., 1974. Judgment under uncertainty: heuristics and biases. *Science* 185*,* 1124-1131.

A Proofs

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

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

A.2 Proof of Proposition 2

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

$$\widehat{\sigma_{jq}^2} = \frac{1}{N_j}\sum_{i=1}^{N_j}(x_{ijq} - Truth_q)^2 \tag{11}$$

Where $\overline{x}_{jq}$ is the average answer for group j members for question q, the MLE for the sample variance, s_{jq}^2, is

$$s_{jq}^2 = \frac{1}{N_j}\sum_{i=1}^{N_j}(x_{ijq} - \overline{x}_{jq})^2 \tag{12}$$

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:

$$s_{jq}^2 = (1 - \widehat{\rho}_{jq})\widehat{\sigma_{jq}^2}$$

Substitution of (11) into (12) then gives

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

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

A.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

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

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

A.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. When she observes n_A American answers and n_T

Thai answers, her expected mean-squared 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 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

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

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

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

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

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

This equation gives the desired result:

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

B Questionnaire

Below is a list of the fifteen questions used in the experiment. The questions appeared according to the format in Figure 1.

Questions: Group 1

1. At the equator, the sun sets 12 hours after it rises on any day. In other places, the day is longer than 12 hours in the summer and shorter than 12 hours in the winter. Length of longest day in Bangkok? Boston? Sum?

2. In 2002, highest recorded temperature in Bangkok? Boston? Sum?

3. From 1961-1990, average number of days with recordable precipitation (of any type) in Bangkok? Boston? Sum?

4. From 1961-1990, average daily <u>high</u> temperature in Bangkok? Boston? Sum?

5. Distance between Bangkok and Boston?

Questions: Group 2

1. The per capita gross national product (GNP) is a measure of the annual mean income that is earned per person in a country. So per capita GNP is the total income of a country divided by the number of people in that country (including adults and children). In 2002, per capita GNP of Thailand? US? Sum?

2. 2002 population of Thailand? US? Sum?

3. Since January 1, 1960, number of Thai prime ministers? US presidents? Sum?

4. On October 1, 2003, average price of a liter of premium gasoline in Bangkok? Boston? Sum?

5. On October 1, 2003, average Thai baht-to-US dollar exchange rate?

Questions: Group 3

1. In 2002, percentage of Thai women aged 15-24 infected with HIV? US women? Sum?

2. In 2002, percentage of Thai workers who were women (average for the year)? US workers? Sum?

3. In 2002, percentage of Thais aged 25-29 with at least some university education? Americans? Sum?

4. In 2002, percentage of Thais whose primary occupation was in agriculture? Americans? Sum?

5. In 2000, percentage of Thais who reported they were of Chinese ethnicity? Americans of African ethnicity? Sum?

C Additional Regression Results

The regression can be expanded to include terms that account for the number of answers that subjects observe, the accuracy of Americans relative to Thais, and the spread in the observed answers. To account for accuracy, for example, I use $Acc_q = \frac{\widehat{\Delta}^2_{Tq}}{\widehat{\Delta}^2_{Aq}+\widehat{\Delta}^2_{Tq}}$. When Acc_q is high, Americans are better relative to Thais at answering the question q. Table A1 reports the results of expanding the regression in a variety of ways.

Table A1: Expanded regression results

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	.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		.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			-.038	-.04			.041	.045			-.002	.002
			(.019)	(.019)			(.024)	(.024)			(.026)	(.026)
Initial answer * accuracy index				-.142				-.273				-.266
				(.084)				(.127)				(.13)
Thai average	.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		.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			.001	-.002			-.053	-.038			-.024	-.017
			(.025)	(.026)			(.038)	(.041)			(.037)	(.037)
Thai average * accuracy index				.04				-.123				-.151
				(.098)				(.221)				(.204)
American average	.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		-.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			-.036	-.027			-.046	-.053			-.023	-.012
			(.018)	(.018)			(.024)	(.024)			(.025)	(.025)
American average * accuracy index				.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) Regressions include dummies for the three question categories (geography, economics/politics, and demography).

C.1 Testing for Anchoring

To test for anchoring, I consider the following regressions:

$$(y_{iq} - \overline{x}_{iTq})^2 = \theta_{1T} Anchor_i + Version'_{iv}\theta_{2T} + \varepsilon_{iq} \quad (13)$$

$$(y_{iq} - \overline{x}_{iAq})^2 = \theta_{1A} Anchor_i + Version'_{iv}\theta_{2A} + \varepsilon_{iq}, \quad (14)$$

The first (second) regression looks at the distance between subjects' final answers and the Thai (American) answers they observe. $Anchor_i$ is a dummy that is one of the 42 subjects in the anchoring group, and $Version'_{iv}$ is a vector of dummy variables for the 20 sets of information that subjects could have observed. Including the version dummies creates a comparison between the anchoring group subjects and main group subjects who observed the same information.

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

Table A2: Testing for anchoring

Dependent variables: (1) Squared distance from observed Thai answers
(2) Squared distance from observed American answers

Independent variable: Dummy for anchoring treatment group

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

Note: Regression standard errors are reported in parentheses.