

# Supporting Information

Bhatt et al. 10.1073/pnas.1200738109

## SI Materials and Methods

**I. Behavioral Analysis. a. Stability of  $\psi$ .** To assess whether our measure of baseline suspicion,  $\psi$ , was, in fact, a baseline measure and stable over the course of the task, we reestimated  $\psi$  ( $R^2$  from the regression of prices on suggestion modulo and the SD of buyer suggestions multiplied by  $-1$ ) separately for the first one-half and second one-half of the experiment ( $\psi_{\text{late}}$  and  $\psi_{\text{early}}$ ). These independent estimates are highly correlated ( $r = 0.74, P < 10^{-13}$ ), and a paired  $t$  test shows no significant difference ( $|t| < 10^{-13}, P \sim 1$ ), indicating that this measure is fairly stable over the task. To further test the possibility that, despite regressing out effects of the spread of buyer suggestions, this measure might still reflect an increase or decrease in suspicion as a function of buyer behavior, we compared the difference between early and late measures of suspicion ( $\psi_{\text{early}} - \psi_{\text{late}}$ ) with the SD of buyer suggestions and found no significant correlation ( $r = -0.08, P = 0.48$ ). A scatter plot is shown in Fig. S1.

**b. Full cognitive hierarchy model.** In the full cognitive hierarchy (CH) model, we followed the model established in the work by Bhatt et al. (1), which is restated from that paper below with the inclusion of the mathematical descriptions of levels 2 and 3 seller behaviors. In cognitive hierarchy models, players perform different numbers of steps of thinking to form beliefs. Zero step or level 0 thinkers behave naïvely—randomly in the original formulation—and lack a model of how other players with behave. Level 1 thinkers assume that they are playing level 0 players and best respond to naïve behavior. Level  $n$  thinkers think that they are playing a mixture of players from all of the  $n - 1$  levels below them and best respond to that mixture.

In our model, we assume that level 0 buyers have a fixed type,  $\alpha$ , distributed  $u(0, 1)$ . Level 0 buyers will, thus, send suggestions  $s$  according to

$$s = \min(10, \max(1, [\alpha v + \varepsilon])) \quad \text{[S1]}$$

and  $\varepsilon \sim N(0, \sigma^2)$ . Level 0 sellers are assumed to be naïve and respond to a buyer suggestion  $s$  with a price quote  $p$  according to

$$p = \min(10, \max(1, [s + \varepsilon])) \quad \text{[S2]}$$

and  $\varepsilon \sim N(0, \sigma^2)$ . Here,  $[x]$  is the nearest integer function. We add the maximum/minimum operations to account for the fact that both price and suggestion must be integer-valued and between 1 and 10. Let  $F$  denote the cumulative normal distribution function with mean = 0 and variance  $\sigma^2$ . Then, these assumptions translate to the conditional distributions

$$P_{0\text{-buyer}}(s|v, \alpha) = F(s + .5 - \alpha v) - F(s - .5 - \alpha v) \quad \text{[S3]}$$

for  $1 < s < 10$ ,

$$P_{0\text{-buyer}}(1|v, \alpha) = F(1.5 - \alpha v), \quad \text{[S4]}$$

and

$$P_{0\text{-buyer}}(10|v, \alpha) = 1 - F(9.5 - \alpha v) \quad \text{[S5]}$$

and

$$P_{0\text{-seller}}(p|s) = F(p + .5 - s) - F(p - .5 - s) \quad \text{[S6]}$$

for  $1 < p < 10$ ,

$$P_{0\text{-seller}}(1|s) = F(1.5 - s) \quad \text{[S7]}$$

and

$$P_{0\text{-seller}}(10|s) = 1 - F(9.5 - s). \quad \text{[S8]}$$

Intuitively, level 0 buyers suggest a price that is a fraction  $\alpha$  ( $< 1$ ) of their value plus noise. This price represents shaving bids, a behavior commonly observed in auctions of this type (2).

We assume that level 1 buyers respond (optimally with noise) to level 0 sellers according to the softmax distribution:

$$P_{1\text{-buyer}}(s|v) = \frac{\exp(\lambda \pi_{1\text{-buyer}}(s, v))}{\sum_{s'} \exp(\lambda \pi_{1\text{-buyer}}(s', v))}. \quad \text{[S9]}$$

In this expression,  $\pi_{1\text{-buyer}}(v, s)$  is the expected payoff given value  $v$  and suggestion  $s$  if you are a level 1 buyer:

$$\pi_{1\text{-buyer}}(v, s) = \sum_{p < v} P_{0\text{-seller}}(p|s)(v - p). \quad \text{[S10]}$$

Level 1 sellers are assumed to also respond optimally to level 0 buyers. However, they have the extra computational challenge of updating their priors about the value of  $\alpha$ , and therefore, their choice of price depends on the current suggestion and also the entire history of suggestions (and the sensible inference of how much information about value that the suggestions have generally implied). They respond according to the distributions

$$P_{1\text{-seller}}(p|\{s_t\}) = \frac{\exp(\lambda \pi_{1\text{-seller}}(p|\{s_t\}))}{\sum_{p'} \exp(\lambda \pi_{1\text{-seller}}(p'|\{s_t\}))} \quad \text{[S11]}$$

and

$$\begin{aligned} \pi_{1\text{-seller}}(p|\{s_t\}) &= p \cdot P(v \geq p) \\ &= p \sum_{v \geq p} P_{0\text{-buyer}}(v|\{s_t\}). \end{aligned} \quad \text{[S12]}$$

We find the distribution  $P_{0\text{-buyer}}(v, \{s_t\})$  in two steps. First, we determine the posterior over  $\alpha$  at time  $t$  using Bayesian updating iteratively:

$$P(\alpha|\{s_t\}) = \frac{P(s_t|\alpha)P(\alpha|\{s_{t-1}\})}{\int_0^1 P(s_t|\alpha')P(\alpha'|\{s_{t-1}\})d\alpha'} \quad \text{[S13]}$$

where

$$P(s|\alpha) = \sum_{v=1}^{10} P_{0\text{-buyer}}(s|v, \alpha). \quad \text{[S14]}$$

Second, we use this posterior over  $\alpha$  to find a posterior over values

$$P_{0-buyer}(v|\{s_t\}) = \frac{\int_0^1 P(v|\alpha, s_t)P(\alpha|\{s_t\})d\alpha}{\sum_{v'=1}^{10} \int_0^1 P(v'|\alpha, s_t)P(\alpha|\{s_t\})d\alpha}, \quad [\text{S15}]$$

where

$$P(v|\alpha, s) = \frac{P_{0-buyer}(s|v, \alpha)}{\sum_{v'=1}^{10} P_{0-buyer}(s|v', \alpha)}. \quad [\text{S16}]$$

Level 2 buyers respond optimally to a 50/50 mixture of levels 0 and 1 sellers. Importantly, level 2 buyers anticipate how their suggestions will change a level 1 seller's posterior over  $\alpha$  and how these changes will affect payoffs in the next period.\* To get probabilities of buyer's suggestions, we plug expected payoffs given these beliefs into a softmax function like we did for level 1 buyers and sellers:

$$P_{2-buyer}(s_t|v, \{s_{t-1}\}) = \frac{\exp(\lambda\pi_{2-buyer}(s_t|v, \{s_{t-1}\}))}{\sum_{s'} \exp(\lambda\pi_{2-buyer}(s'|v, \{s_{t-1}\}))}. \quad [\text{S17}]$$

Here,

$$\pi_{2-buyer}(s_t|v, \{s_{t-1}\}) = \text{Current Payoff} + \text{Future Payoff}, \quad [\text{S18}]$$

where

$$P_{buyer}(v|\{s_t\}) = \frac{P_{1-buyer}(v|s_t)P(\text{level}-1|\{s_t\}) + \int_0^1 P(v|\alpha, s_t)P(\alpha|\{s_t\})d\alpha}{\sum_{v'=1}^{10} \left( P_{1-buyer}(v'|s_t)P(\text{level}-1|\{s_t\}) + \int_0^1 P(v'|\alpha, s_t)P(\alpha|\{s_t\})d\alpha \right)}, \quad [\text{S24}]$$

$$\text{Current Payoff} = .5 \left( \sum_{p < v} P_{1-seller}(p|\{s_t\})(v-p) \right) + .5 \left( \sum_{p < v} P_{0-seller}(p|s_t)(v-p) \right), \quad [\text{S19}]$$

and

$$\text{Future Payoff} = \frac{\sum_{v_{t+1}} \max_{s_{t+1}} \left( \sum_{p < v_{t+1}} (.5P_{1-seller}(p|\{s_{t+1}\}) + .5P_{0-seller}(p|s_{t+1}))(v-p) \right)}{10}. \quad [\text{S20}]$$

Similarly, level 2 sellers respond optimally to mixture of levels 0 and 1 buyers. However, because the sellers do not need to

\*Because we are assuming limited hierarchical reasoning throughout, we only have level 2 sellers project one period into the future rather than considering the entire experimental run. We consider only one period of forecasting, because adding more periods does not significantly change predicted choices but does become too computationally taxing to estimate. Time to estimate the level 2 model grows exponentially with the number or periods ahead considered.

anticipate the effects of their behavior on the buyers, their computation is extremely similar to the computation for level 1 sellers.

We once again find the distribution  $P_{buyer}(v, \{s_t\})$  in two steps. First, we determine the posterior over buyer type, which now includes potential values of  $\alpha$ , and the possibility that the buyer is level 1 at time  $t$  using Bayesian updating iteratively:

$$P(\text{type}|\{s_t\}) = \frac{P(s_t|\text{type})P(\text{type}|\{s_{t-1}\})}{P(s_t|\text{level}-1)P(\text{level}-1|\{s_{t-1}\}) + \int_0^1 P(s_t|\alpha')P(\alpha'|\{s_{t-1}\})d\alpha'}, \quad [\text{S21}]$$

where

$$P(s|\alpha) = \sum_{v=1}^{10} P_{0-buyer}(s|v, \alpha) \quad [\text{S22}]$$

and

$$P(s|\text{level}-1) = \sum_{v=1}^{10} P_{1-buyer}(s|v). \quad [\text{S23}]$$

Second, we assume an initial distribution of 50% level 1 buyers, with the remaining 50% as level 0 buyers with uniformly distributed  $\alpha$ .

We once again use this posterior over  $\alpha$  to find a posterior over values

which we can then use to compute the expected payoff to each price and the softmax function to derive the predicted distribution of behaviors:

$$P_{2-seller}(p|\{s_t\}) = \frac{\exp(\lambda\pi_{2-seller}(p|\{s_t\}))}{\sum_{p'} \exp(\lambda\pi_{2-seller}(p'|\{s_t\}))} \quad [\text{S25}]$$

and

$$\pi_{2-seller}(p|\{s_t\}) = p \cdot P(v \geq p) = p \sum_{v \geq p} P_{buyer}(v|\{s_t\}). \quad [\text{S26}]$$

Finally, level 3 sellers best respond to a mixture of levels 0, 1, and 2 buyers. This response is, once again, effectively the same computation as the one for level 2 sellers with the addition of the possibility

that buyers could be level 2 and the initial assumption that buyer types were distributed as 40% level 0, 40% level 1, and 20% level 2 (the approximate empirical distribution of incrementalists, conservatives, and strategists on the buyer side of the experiment).<sup>†</sup>

As with the buyers in the work by Bhatt et al. (1), we classified each of the sellers according to this full CH model. To simplify the computations, we approximated the initial uniform distribution over  $\alpha$  with a discrete, even distribution over the points (0, 0.05, ..., 0.95, 1). Because it was difficult to distinguish the two sources of noise,  $\sigma$  and  $\lambda$ , we fixed  $\sigma = 2$  and computed the maximum likelihood estimate for  $\lambda$  for levels 1, 2, and 3 behaviors. The computed log likelihoods and implied types are reported in Table S1.

Forty of the sellers were classified as level 1; only 11 sellers were classified as level 3. Notice that, of these 11 sellers, it was difficult to distinguish between levels 2 and 3 designations; for 8 of 11 level 3 subjects, the log likelihood that they were a level 2 seller was within one log likelihood that they were a level 3 seller. The three subjects for whom this finding was not the case are highlighted in bold in Table S1. This difficulty arises because of the low probability of level 2 buyers; levels 2 and 3 behaviors generally do not differ significantly. Interestingly, in the most significant case of level 3 seller behavior (134-DH-2), the buyer in question was a strategist who almost exclusively sent suggestions of either 1 or 10, making his deception more obvious to the seller. However, he was one of only two strategists who elicited level 3 behavior from the corresponding seller.

In fact, because of the difficulty in establishing clear differences between levels 2 and 3 seller behaviors along with the similarity between level 1 and low-slope level 0 buyers, most of the important aspects of sophisticated seller behavior are captured by the level 1 model, which was indicated by the fact that the majority of sellers were classified as level 1.<sup>‡</sup> Naïve seller behavior is largely captured by the gross measure of baseline suspicion described in the text.

**c. Simplified model of seller behavior.** To select a single model for seller behavior that addresses the important aspects of inference about buyer types, we used the level 1 model for seller behavior from the full cognitive hierarchy model described. Recall that the level 1 sellers modeled buyer behavior by assuming that they chose suggestions according to Eq. S1 and  $\varepsilon \sim N(0, \sigma^2)$ , where  $\alpha$  is initially assumed to be drawn from the uniform distribution on the points (0, 0.05, 0.1, ..., 0.95, 1). In addition, using the model used in the work by Bhatt et al. (1), we simplify the model by assuming that sellers have limited memory and only used the previous trial along with the current trial's suggestions. For all trials,  $t > 2$ :

<sup>†</sup>This distribution also corresponds to the Poisson distribution with a mean = 1 truncated after  $k = 2$ . The work by Camerer et al. (3) found that truncated Poisson distributions were highly effective for modeling hierarchical beliefs across a variety of economic games, although it generally found higher average levels of thinking (generally around 1.5).

<sup>‡</sup>This finding is in contrast to the buyers where level 2 behavior was very distinct from levels 1 and 0 behaviors, and none of the three behavioral types composed a majority of the subject pool.

1. Bhatt MA, Lohrenz T, Camerer CF, Montague PR (2010) Neural signatures of strategic types in a two-person bargaining game. *Proc Natl Acad Sci USA* 107:19720–19725.  
2. Camerer CF (2003) *Behavioral Game Theory: Experiments in Strategic Interaction* (Princeton University Press, Princeton, NJ).

$$P(\alpha|s_t, s_{t-1}) = \frac{P(s_t|\alpha) \cdot P(\alpha|s_{t-1})}{\sum_{\alpha'} P(s_t|\alpha') \cdot P(\alpha'|s_{t-1})}, \quad [\text{S27}]$$

where

$$s_t|\alpha = \frac{1}{10} \sum_{i=1}^{10} \min(10, \max(1, [\alpha \cdot i + \varepsilon])) \quad [\text{S28}]$$

and  $\varepsilon \sim N(0, \sigma^2)$ . Therefore,

$$P(s_t = 10|\alpha) = \frac{1}{10} \sum_{i=1}^{10} P([\alpha i + \varepsilon] > 10) = \frac{1}{10} \sum_{i=1}^{10} P(\varepsilon > 10.5 - \alpha i), \quad [\text{S29}]$$

and similarly,  $s_t = 1, 2, 3, \dots$

A buyer with  $\alpha = 0$  will choose  $s = \min(10, \max(1, [1 + \varepsilon]))$  in each round (i.e., they will generally send low suggestions), although they may occasionally send a higher suggestion because of noise. More importantly, these suggestions will contain no information about  $v$ . This finding reflects the conservative group observed in buyers. If, however,  $\alpha = 1$ , the buyer would be sending  $s = \min(10, \max(1, [v + \varepsilon]))$  (i.e., they will be sending a highly reliable signal of  $v$ , reflecting the incrementalist group observed in the buyer).

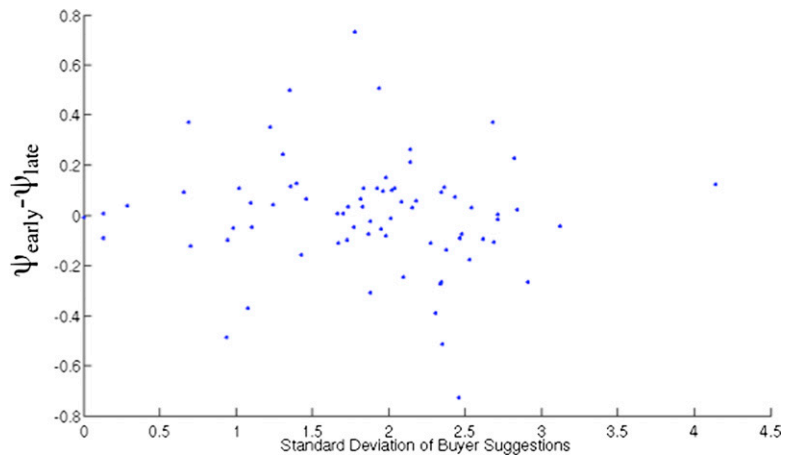
The entropy of the seller's belief distribution gives us a trial-by-trial measure of the seller's strategic uncertainty about the buyer's credibility. The higher the entropy, the less certain that the seller is about the value of  $\alpha$  (i.e., the buyer's type).

**II. Functional MRI Analysis.** We considered two general linear models. Key presses, head motion, and time derivatives were included as nuisance regressors in both models. The first model used a boxcar regressor beginning at trial onset and ending at decision parameterized by the buyer's suggestion and the seller's chosen price. The second model used separate point regressors at trial onset and decision, and both regressors were parameterized by strategic uncertainty as defined by the entropy measure defined above. All regressors were convolved with the standard hemodynamic response function.

**a. Second level analysis for the first model.** At the second level, we regressed the coefficients of the boxcar regressor for each subject on two subject-level parameters: baseline suspicion ( $\psi$ ) and the SD of suggestions ( $\zeta$ ). The results of the whole-brain analysis as well as the  $P$  values corrected in the small volumes around the left and right amygdala [20-mm spheres around the foci identified in the work by Winston et al. (4)] are presented below in Tables S2 and S3.

**b. Second level analysis for the second model.** At the second level, we performed a  $t$  test on the coefficients of the regressor describing strategic uncertainty (entropy of the seller's belief distribution about buyer type) at trial onset. Results from the whole-brain analysis are shown below in Tables S4 and S5. We performed a similar analysis looking at strategic uncertainty at the end of each trial and found no significant correlates.

3. Camerer CF, Ho TH, Chong J (2004) A cognitive hierarchy model of games. *Q J Econ* 119:861–898.  
4. Winston JS, Strange BA, O'Doherty J, Dolan RJ (2002) Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nat Neurosci* 5:277–283.



**Fig. S1.** Scatter plot showing the change in our measure of baseline suspicion from the first one-half to the second one-half of the experiment ( $\psi_{\text{early}} - \psi_{\text{late}}$ ) vs. apparent buyer credibility as measured by the SD of buyer suggestions ( $\zeta$ ). This measure for baseline suspicion does seem to be relatively stable over the course of the experiment.



**Table S1. Cont.**

Subject	Level 0	Level 1	Level 2	Level 3	Level
134-BM-2	-131.3771	-106.3772	-102.915	-102.7003	3
134-DG-2	<Random	-88.9402	-74.2562	-73.4059	3
<b>134-DH-2</b>	<b>&lt;Random</b>	<b>-136.6534</b>	<b>-114.5722</b>	<b>-111.4</b>	<b>3</b>
134-EG-2	<Random	-127.5317	-118.2782	-117.457	3
134-ES-2	-117.0427	-110.383	-108.3387	-108.0509	3
134-V-2	<Random	-129.5456	-126.5015	-125.9305	3
134-BY-2	<Random	<Random	<Random	<Random	None
134-CF-2	<Random	<Random	<Random	<Random	None
134-CX-3	<Random	<Random	<Random	<Random	None
134-DW-2	<Random	<Random	<Random	<Random	None
134-ER-2	<Random	<Random	<Random	<Random	None

The majority of sellers (54%) were classified as level 1 using the full cognitive hierarchy (CH) model. The three subjects for whom the log likelihood that they were a level 2 seller was not within one log likelihood that they were a level 3 seller are highlighted in bold.

**Table S2. Between-subject correlates to  $\psi$ : Negative correlates to  $\psi$  (whole-brain analysis)**

Region	<i>x</i>	<i>y</i>	<i>z</i>	<i>k</i>	<i>t</i>	<i>P</i>	<i>P</i> (FWE)	SVC <i>P</i> (FWE)
Right amygdala	20	4	-12	18	4.29	2.77E-05	0.24	0.005
Left amygdala	-16	0	-16	11	3.89	1.12E-04	0.57	0.030

Cluster sizes (*k*) are shown at  $P < 0.001$  uncorrected. Corrections for family-wise error (FWE) are shown for the peak voxel in each cluster. Small-volume correction (SVC) was done for the 20-mm spheres around (18, 0, -24) and (-16, -4, -20) for the right and left amygdala, respectively.

**Table S3. Between-subject correlates to  $\zeta$ : Negative correlates to  $\zeta$  (whole-brain analysis)**

Region	<i>x</i>	<i>y</i>	<i>z</i>	<i>k</i>	<i>t</i>	<i>P</i>	<i>P</i> (FWE)
Inferior frontal gyrus	60	-4	28	13	4.06	6.21E-05	0.52

Cluster sizes (*k*) are shown at  $P < 0.001$  uncorrected. Corrections for family-wise error (FWE) are shown for the peak voxel in each cluster. Small-volume correction was done for the 20-mm spheres around (18, 0, -24) and (-16, -4, -20) for the right and left amygdala, respectively.

**Table S4. Within-subject correlates to the entropy of trial-by-trial beliefs about buyer type (strategic uncertainty): Positive correlates to entropy of beliefs about buyer type (strategic uncertainty)**

Region	<i>x</i>	<i>y</i>	<i>z</i>	<i>k</i>	<i>t</i>	<i>P</i>	<i>P</i> (FWE)	<i>P</i> (cluster)
Left parahippocampus	-20	-36	-12	16	4.44	1.58E-05	0.335	0.046
Right parahippocampus	32	-40	-12	28	4.03	6.77E-05	0.843	0.004
Middle temporal gyrus	44	-68	12	11	4.01	7.33E-05	0.861	0.152

Cluster sizes (*k*) are shown at  $P < 0.001$  uncorrected. Corrections for family-wise error (FWE) are shown for the peak voxel in each cluster. Cluster corrections are shown for  $P < 0.001$  and  $k > 5$ .

**Table S5. Within-subject correlates to the entropy of trial-by-trial beliefs about buyer type (strategic uncertainty): Negative correlates to entropy**

Region	<i>x</i>	<i>y</i>	<i>z</i>	<i>k</i>	<i>t</i>	<i>P</i>	<i>P</i> (FWE)	<i>P</i> (cluster)
Brodman area 6	24	8	64	13	4.51	1.20E-05	0.256	0.093

Cluster sizes (*k*) are shown at  $P < 0.001$  uncorrected. Corrections for family-wise error (FWE) are shown for the peak voxel in each cluster. Cluster corrections are shown for  $P < 0.001$  and  $k > 5$ .