The Accuracy of Small-Group Estimation and the Wisdom of Crowds

Michael D. Lee (mdlee@uci.edu)
Jenny Shi (jshi1@uci.edu)
Department of Cognitive Sciences, 3151 Social Sciences Plaza A
University of California, Irvine, CA 92697-5100 USA

Abstract

We measure the ability of people to estimate the price of familiar household items in a variety of contexts. We manipulate whether estimation is done alone or with others, whether it is done independently or with the knowledge of the estimates of others, and whether it is done in a cooperative or competitive environment. From these basic estimation data, we construct a series of aggregated group estimates, exploring the conditions under which a small group of three people provide the most accurate information. We compare the performance of various small-group estimates to standard Wisdom of Crowds analysis, and find that priming people, or placing them in a cooperative group setting, is less effective than averaging the independent estimates of individuals. We also find, however, that it is possible to extract relatively more information from the decisions people make in a competitive group setting, using cognitive models of their decision-making.

Keywords: Wisdom of crowds, group estimation, Price is Right, game show, cooperative vs competitive decision-making

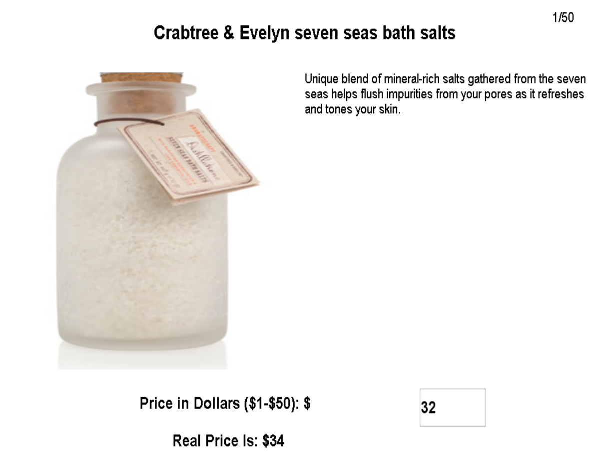

Figure 1: Basic experimental interface. On each trial, a picture and description of an item is shown. Once an estimate has been made, the true price is presented.

Introduction

A basic question for cognitive and social psychology involves how best to extract information from people. There is a large literature on the performance of groups in reaching good decisions in various contexts (see Kerr & Tindale, 2004; Hastie, 1986, for reviews), with accompanying theoretical positions ranging from believing in the robust effectiveness of group decision-making (e.g., Hastie & Kameda, 2005) to the destructive possibilities of "group think" (e.g., Moscovici & Zavolline, 1969).

A recent contribution to the issue of whether and how groups of people make effective decisions involves the "Wisdom of Crowds" phenomenon (Surowiecki, 2004). This refers to the empirical finding that an aggregated decision, made by combining the individual decisions of many people, can often perform as well as or better than the majority of the individual decisions themselves.

In this paper, we examine group decision-making and the Wisdom of Crowds phenomenon in a simple estimation setting. We ask people to estimate the price of everyday household objects, with which they people are familiar, but are unlikely to have exact price knowledge. We ask for these estimates in a wide variety of individual and group settings. These settings manipulate whether estimation is done alone or in the presence of others, whether it is done independently or with the knowledge of the estimates of others, and whether group estimation is done in a cooperative or competitive environment.

To examine how these manipulations affect the accuracy of small-group estimation, we focus on a specific research question. The question is: how well do different ways of using the knowledge of just three people to estimate the price perform, and how does this level of performance relate to standard Wisdom of Crowds aggregation with more people?

Experiment

Materials

Stimuli We used two sets of 50 household items, with pictures and descriptions sourced from on-line shopping websites. Both stimulus sets followed the same price distribution, with totals approximately uniformly distributed between $5 and $45.

Interface An example of the basic experimental interface is shown in Figure 1. On each trial, a picture and

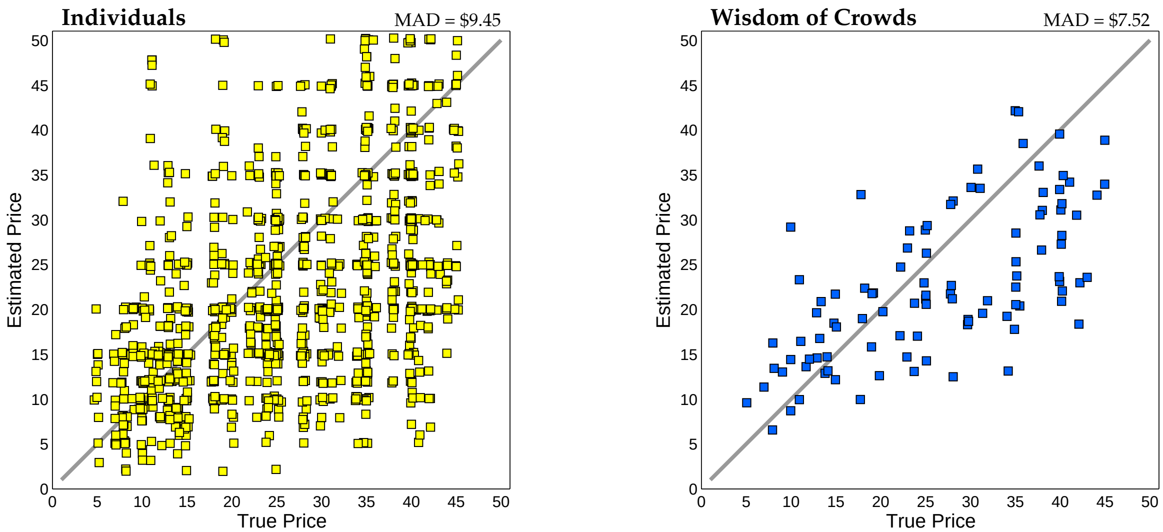

Figure 2: Relationship between true item prices and individual estimates (left panel), and Wisdom of Crowds estimate formed by averaging over all individuals (right panel). (MAD=Mean Absolute Deviation)

description of a prize is shown. Once an estimate has been made, the true price is presented. A counter shows how many of the 50 trials have been completed.

Methods

Using the same sets of items and basic interface, we collected price estimates under a variety of experimental conditions. These conditions manipulated whether estimation was done in an individual or group setting, whether estimates were done independently or with knowledge of other estimates, and whether estimation was done in cooperative or competitive setting.

Individual Estimates The simplest experimental condition just collects individual estimates for each of the 50 items, presented in a random order. A total of 22 participants completed this condition.

Primed Individual Estimates The 'primed' or 'calibrated' condition was the same as the individual condition, except that when each item was presented, the estimates of two other people were also presented. These estimates were drawn at random from the estimates made for the same prize in the individual condition. A total of 25 participants completed this condition.

Cooperative Group Estimates In the cooperative group condition, three people were co-located, and viewed the same experimental interface. They were asked to provide estimates sequentially, hearing the earlier estimates. After all three estimates had been made, the group was asked to form a consensus estimate, through unstructured discussion. The same three people completed all 50 trials, and the order in which they estimated was rotated between each trial. A total of 15 people completed this condition, forming 5 groups.

Competitive Group Estimates In the competitive group condition, three people played a version of the "Price is Right" game show, which has been used previously as a formalism to study competitive decision-making (e.g., Berk, Hughson, & Vandezande, 1996). They were asked to provide bids sequentially, hearing the earlier bids, with the goal of bidding *as close as possible to the true price without exceeding the true price*. People were not allowed to repeat an earlier bid, and the order of making bids was again rotated systematically after each trial. A total of 15 people completed this condition, forming 5 groups.

Basic Results

Bounds on Performance There are two worthwhile preliminary analyses that can serve to give bounds on the accuracy of estimation. The first of these simply considers each individual estimate, and is shown in the left panel of Figure 2. The mean average deviation between the estimated and true price is $9.45. This serves as a sensible baseline for accuracy, since it represents what how well a single person will perform on average.

The second preliminary analyses averages *all* of the individuals who gave estimates for each prize. This corresponds to a standard "Wisdom of the Crowds" analysis, and is shown in the right panel of Figure 2. The mean absolute deviation is a much-improved $7.52, and can reasonably serve as an upper bound on performance.

Simple Three Person Estimates

Figure 3 shows the performance of four simple ways to combine the information provided by three people to estimate the prices. These involve, the individual, primed individual, and cooperative group estimation contexts.

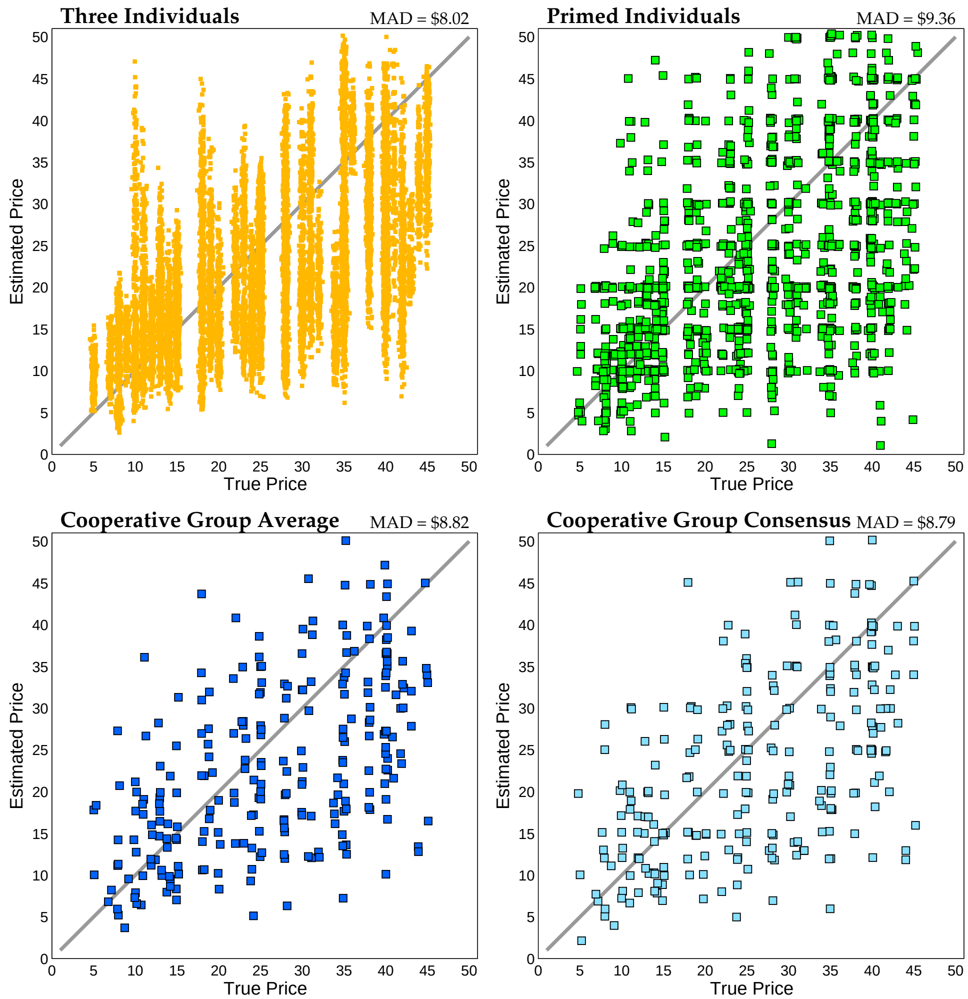

Figure 3: Relationship between true item prices and group estimates, formed from three people, by (top left) averaging the individual estimates of three people, (top right) priming an individual with the earlier estimates of two other people, (bottom right) averaging the estimates of three people made sequentially in a group setting, and (bottom right) the consensus opinion of a group of three people. (MAD=Mean Absolute Deviation)

Three Individuals The most obvious, given the estimates of three people, is simply to average them, as in a standard Wisdom of the Crowds analysis.[1] The performance of this approach is shown in the top left panel of Figure 3, which considers all possible groups of three people using individual estimates. The mean absolute difference is $8.02. As would be expected, this difference lies between that already observed for single individuals, and for all individuals considered together.

[1] For all of the analyses we present involving the averaging of estimates, we also examined taking the median, or rounding answers to the nearest dollar. Rarely did performance, as measured by the Mean Absolute Deviation, change by more than a few cents, and never did it suggest different conclusions from those we present based on the mean.

Primed Individuals Another estimate based on the information provided by three people comes directly from the primed estimate. This is the estimate of a single individual working along, but in the knowledge of two other people's estimates. The performance of primed estimates is shown in the top right panel of Figure 3. The mean absolute difference is $9.36, which barely improves upon the accuracy of estimates of single non-calibrated individuals.

Cooperative Average The bottom left panel shows the performance of the average of the three people in the cooperative group condition. The mean absolute difference is $8.82. This is better than single individuals, but does not come close to the level of performance achieved by averaging three estimates made independently.

Cooperative Consensus Finally, the bottom right panel of Figure 3 shows the performance of the consensus estimates reached by the groups of three people. The mean absolute difference is $8.79, which is very similar to the average of the group estimates. Taken together, these results suggest that being in a cooperative group setting hinders the generation of accurate estimates.

Competitive Group Analysis

Analyzing estimation performance for the competitive "Price is Right" condition requires more involved inference than averaging. This is because the bids that people make do not necessarily correspond to their actual best estimate of the price of a prize. In the competitive context formalized by the rules of the game, it is often sensible for a player to make a bid that is very different from what they believe the price to be.

This strategic relationship between bids and estimates is most easily seen for the final bid made by Player 3. If the previous bids are $35 and $40, then the best final bid is either $1, $36 or $41. One of these choices is rational, in the sense that it will maximize the probability that Player 3 wins the game. Which choice is rational depends on what Player 3 knows about the price of the prize. If, for example, they believe it is most likely somewhere below $35, then the $1 final bid is optimal.

For this reason, it does not make sense to combine the bids from the competitive group setting as if they were estimates, and just average them. Instead, inferences need to be made about what estimates the players have in their heads, based on their bids. This inference requires a model of decision-making that accounts for how estimates become bids, in the context of the game.

Inferring a Group Estimate from Bids

The decision model we used for inference makes two key assumptions. The first is a representational assumption, which is that all of the players have partial knowledge of the price of a prize, and that their uncertainty can be represented by the same Normal distribution. The second is a decision-making assumption, which is that players make the bid that maximizes their probability of winning the game. Given these assumptions, our inferential goal is to find the mean of the Normal distribution, since it represents the average price, based on the players' knowledge.

Formally, given a Normal distribution with mean μ and standard deviation σ, we can define a 'win' function $w_x(a,b,c,\mu,\sigma)$ for the probability the xth player will win, given bids a, b, c, for Players 1, 2, and 3, respectively. This win probability is just the area under the Normal curve between the bid of the xth player, and the next highest bid (or the maximum of $50, if it is the highest bid). On the basis of this win function, we can formalize what constitutes optimal bidding for each player.

Player 3 Given existing bids a, and b the probability Player 3 will win if they made the bid c is just

$$\pi_c(c \mid a,b,\mu,\sigma) = w_3(a,b,c,\mu,\sigma),$$

and so one way of formalizing what it means to be a rational player, is that they will choose according to these probabilities, so that

$$p_3(c \mid a,b,\mu,\sigma) = \frac{\pi_3(c \mid a,b,\mu,\sigma)}{\sum_{c'} \pi_3(c' \mid a,b,\mu,\sigma)}.$$

Player 2 Given an existing bid a, the probability Player 2 will win if they made the bid b, assuming Player 3 subsequently 'behaves optimally' and bids according to $p_3(c \mid a, b\mu,\sigma)$ above, is

$$\pi_2(b \mid a,\mu,\sigma) = \sum_c p_3(c \mid a,b,\mu,\sigma)\, w_2(a,b,c,\mu,\sigma).$$

So, if Player 3 makes their bid decision according to these probabilities, they will choose

$$p_2(b \mid a,\mu,\sigma) = \frac{\pi_2(b \mid a,\mu,\sigma)}{\sum_{b'} \pi_2(b' \mid a,\mu,\sigma)}.$$

Player 1 Player 1 provides the first bid. If they bid a, their probability of winning, assuming subsequent optimal behavior is

$$\begin{aligned}\pi_1(a \mid \mu,\sigma) \quad = \quad & \sum_b p_2(b \mid a,\mu,\sigma) \sum_c p_3(c \mid a,b,\mu,\sigma) \times \\ & w_1(a,b,c,\mu,\sigma).\end{aligned}$$

This gives the bid decision probabilities

$$p_1(a \mid \mu,\sigma) = \frac{\pi_1(a \mid \mu,\sigma)}{\sum_{a'} \pi_1(a' \mid \mu,\sigma)}.$$

Final Inference The joint posterior distribution over the parameters of the Normal representing people's knowledge is given by Bayes Rule

$$\begin{aligned} & p(\mu,\sigma \mid a,b,c) \\ \propto \quad & p(a,b,c \mid \mu,\sigma)\, p(\mu,\sigma) \\ = \quad & p(c \mid a,b,\mu,\sigma)\, p(b \mid a,\mu,\sigma)\, p(a \mid \mu,\sigma)\, p(\mu,\sigma).\end{aligned}$$

We put a simple improper flat prior on $p(\mu,\sigma)$, and all of the other likelihood terms are available from the optimal decision-making analysis.

There are many potential ways the $p(\mu,\sigma \mid a,b,c)$ could be used to estimate the final group price. We use probably the simplest possible approach, and find the mode (i.e., the MAP estimate) $(\mu^*,\sigma^*) \mid a,b,c$, and use μ^* as the price estimate of the competitive group, based on their bids.

Demonstration of Inference

Figure 4 provides a concrete example of the inference process used to estimate the price of a prize from the bidding in the competitive "Price is Right" game. The example relates to one trial for one of our groups, in which the players bid $13, $10 and $1. To find which Normal distribution best explains these bids, under the assumption that people bid to maximize their chance of winning, we

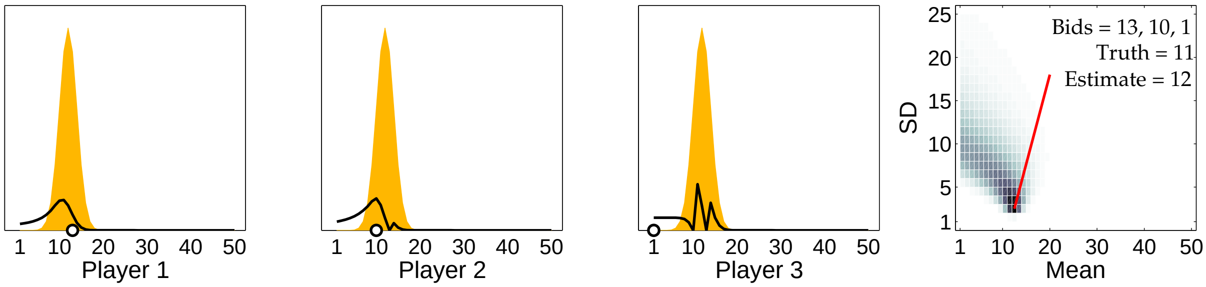

Figure 4: Inference process to find a group estimate from the bids in a competitive “Price is Right” game.

exhaustively test every Normal distribution with a mean of $1, 2, \ldots, 50$ and standard deviation of $1, 2, \ldots, 25$.

The first three panels of Figure 4—which correspond to the decision making of Players 1, 2 and 3—all show the same particular Normal in the background, with a mean of \$12 and a standard deviation of \$3. The black line then shows the probability of each player winning the game, if they made each possible bid between \$1 and \$50. The white circle represents the bid they actually made.

Intuitively, it is easiest to understand this analysis by looking at Player 3. Here, it is already known that the previous bids are \$13 and \$10. The black line shows that the probability of Player 3 winning peaks around \$14 and \$11, one above the earlier bids, and is also high for bids starting at \$1, up until the point where the Normal says it becomes possible the true price might lie.

Looking at all three players, it is clear that this particular Normal distribution gives predictions that are reasonably consistent with the bids actually made. It peaks at the right bid for Player 2, and gives appreciable probably to the bids of Players 1 and 3. In fact, the Normal shown corresponds to the most likely one, out of all the possibilities considered. This result is shown in the rightmost panel of Figure 4. In this plot, each point corresponds to a Normal distribution, and the darker it is shaded, the more probable that Normal made the observed bid data. The mode is at $\mu = 12, \sigma = 3$, and so the final estimate we infer is \$12. As it happens this is very close to the true \$11 price of the prize for this trial.

Notice that simply averaging the bids would not produce the same estimate, because it would treat the \$1 bid as a literal estimate, rather than a strategic attempt to win the game, based on the belief that earlier bids may have been too high.

Results

The performance of the inferred three-person estimates based on the competitive game bids is shown in Figure 5. The mean difference is \$8.05. This is a large improvement on the cooperative group average and consensus estimates, and is comparable to the accuracy obtained by averaging three individual estimates.

The results for all of the three-person estimates, and their relationship to Wisdom of Crowds averaging, are summarized in Figure 6. The curve shows the accuracy of Wisdom of Crowds averages, starting with a single individual and finishing with all individuals. These start- and end-points correspond to the bounds established in Figure 2.[2] A clear and interesting pattern evident in this curve is how quickly including additional independent people in the Wisdom of Crowds average fails to improve accuracy. There is little improvement beyond the fifth or sixth person.

Figure 6 shows the performance of all of the three-person estimates—primed individuals, cooperative average, cooperative consensus, and competitive Price is Right inference—in relation to the Wisdom of Crowds curve. Motivated by a similar analysis presented by Vul and Pashler (2008), we map from the mean absolute difference of each three-person estimate to the Wisdom of Crowd curve, and then down to the number of people. This mapping allows the performance of the various approaches to be conceived in terms of how many independent estimates worth of performance they achieve. The results show that a primed individual is the same as a single non-primed individual, putting three people in a cooperative setting produces the accuracy of about one-and-a-half independent people, but putting three people in a competitive setting constitutes three independent people’s worth of information.

Discussion

There are many analyses besides those reported here that could be pursued with the current data. For example, it would be interesting to compare the accuracy of individuals primed while working alone with those who gave the final estimate in the cooperative group setting. In a sense, these individuals have access (on average) to the same information, and so differences in their accuracy could be attributed to the social setting. We plan to pursue extensions and variants on the cognitive modeling of the competitive setting, including making different assumptions about how homogenous information is across participants, and how bidding decisions might be made.

[2] There were 22 individuals who provided individual estimates, so that 11 completed each of the two stimulus sets. The performance measures shown average over the two stimulus sets.

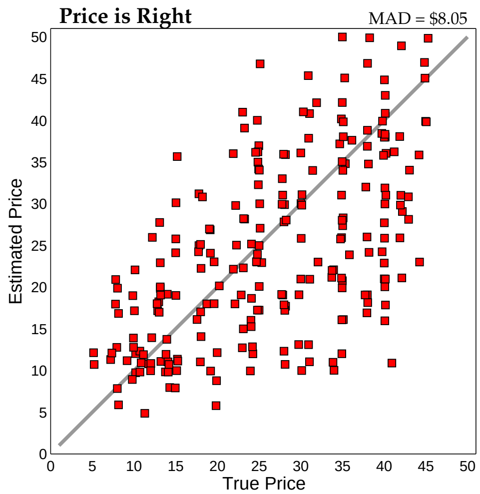

Figure 5: Relationship between true item prices and estimates inferred from an optimal decision-making analysis of three people competing in a “Price is Right” game. (MAD=Mean Absolute Deviation)

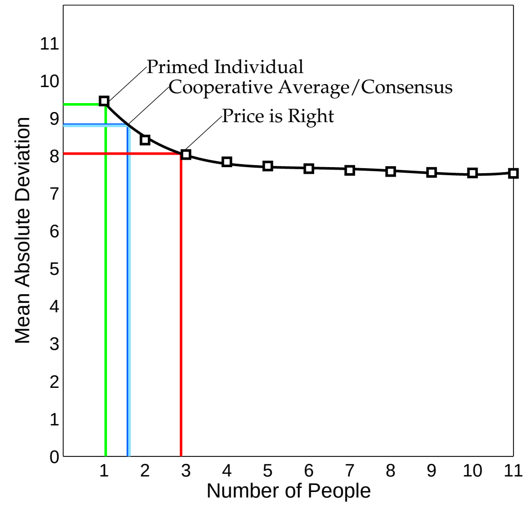

Figure 6: Characterization of the three-person estimates in terms of wisdom of crowds averages including $1, \ldots, 11$ individual estimates.

However, we can draw a number of interesting initial conclusions from the analyses reported here. The first is that the basic Wisdom of Crowds approach performs remarkably well. None of our alternative three-person estimation settings was superior to simply taking the average estimates of three random independent individuals. This averaging corresponds to a very simple generative model, in which each person estimates a signal with the addition of independent noise.

Our second conclusion applies to situations in which aggregate estimation must be done in a group setting, or when individuals share too much knowledge for independent estimates to be possible. These constraints could apply, for example, in situations where the goal is to pool the estimates of domain experts, who have overlapping training and knowledge. Here, our results argue for competitive rather than cooperative or passive approaches to extracting and combining information seem superior. The accuracy of the estimates from the simple “Price is Right” game were far superior to the other estimates we collected in group settings, and justified our effort to develop the much more complicated generative model for that setting.

We think the result highlighting the benefits of competition is suggestive, for both theoretical and applied reasons. Theoretically, it argues for the need to incorporate models of cognition and decision-making within Wisdom of Crowds research, to understand not just final behavior, but the underlying knowledge that generated that behavior. As we pointed out, it does not make sense to average the bids people make in the “Price is Right” game, but it does make sense to aggregate the knowledge they had that led them to decide on those bids. Practically, our results reinforce recent evidence for the effectiveness of competition instruments like prediction markets (e.g., Christiansen, 2007), rather than cooperative or collaborative groups settings, as better ways to combine knowledge across individuals.

References

Berk, J. B., Hughson, E., & Vandezande, K. (1996). The price is right, but are the bids? An investigation of rational decision theory. *The American Economic Review*, *86*(4), 954–970.

Christiansen, J. D. (2007). Prediction markets: Practical experiments in small markets and behaviors observed. *The Journal of Prediction Markets*, *1*, 17–41.

Hastie, R. (1986). Experimental evidence on group accuracy. In B. Grofman & G. Owen (Eds.), *Information pooling and group decision making* (pp. 129–157). Greenwich, CT: JAI Press.

Hastie, R., & Kameda, T. (2005). The robust beauty of majority rules in group decisions. *Psychological Review*, *112*(2), 494–508.

Kerr, N. L., & Tindale, R. S. (2004). Group performance and decision making. *Annual Review of Psychology*, *55*, 623–655.

Moscovici, S., & Zavolline, M. (1969). The group as a polarizer of attitudes. *Journal of Personality and Social Psychology*, *12*, 125–135.

Surowiecki, J. (2004). *The Wisdom of Crowds*. New York: Random House.

Vul, E., & Pashler, H. (2008). Measuring the crowd within: Probabilistic representations within individuals. *Psychological Science*, *19*(7), 645–647.