The Social Psychology of the Wisdom of Crowds

Richard P. Larrick

Duke University

Albert E. Mannes

Carnegie Mellon University

Jack B. Soll

Duke University

Picture yourself taking part in a classroom psychology experiment on perception. You and nine other subjects are being asked to judge the lengths of lines in a vision test. The experimenter holds up a large card that contains three lines of different length, marked A, B, and C, and a target line. The subjects are asked: Which of A, B, or C is the same length as the target? For each card, all subjects take turns reporting which line matches the target. The first few sets of cards are unremarkable. Everyone states what is clear to the eye. On the fifth trial, the obvious answer to you is B, but the first person says C (you find this mildly amusing). However, the second person also says C. The third says C, and then the fourth. Everyone says C. When it is your turn to answer, what do you say? Before hearing the response of others, you thought the answer was obviously B. But the unanimous opinion of others is C.

All students of social psychology recognize the famous Asch (1955) experiment, and most remember that 3 of 4 subjects at some point conform to a group answer that defies their own perceptions. Deutsch and Gerard (1955) introduced the term "normative social influence" to characterize the tendency for people to give public responses that allow them to fit in with others even when they privately disagree. However, Asch's classic study has come to symbolize more. In its own time—an era of totalitarian governments that repressed individualism and McCarthyite pressures in the US that did likewise—Asch's study reinforced a suspicion of groups and a celebration of the lone, independent individual.

Early conformity research was criticized for leaving the widely-held but misleading impression that people yield too much to the judgments of others, thereby harming themselves. Some authors (Allen, 1965; Campbell, 1961) suggested that this pessimistic conclusion is misleading because it was foreordained by the experimental design. In the Asch study, the other nine group members were only pretending to be subjects and had been planted to give consistently false answers before the true subject responded. Allen (1965, p. 136) observed, “[M]ost psychological experiments in this area have been designed in such a manner that conformity was by necessity maladaptive: factually incorrect, detrimental to the group and the individual, or simply dishonest.”

In response, some authors stressed the adaptive value of conformity. Under uncertainty, people look to others for information (Deutsch & Gerard, 1955; Festinger, 1954), and groups are usually a valid source of information (Campbell, 1961). Allen (1965, p. 136-137) observed that “a person may go along with beliefs expressed by most of the other people around him because he realizes that opinions shared by many are often more likely to be correct than the opinions held by a single individual…. In some situations conformity is constructive and appropriate; in other situations it is not" (see Kameda & Tindale, 2006, and Krueger & Massey, 2009 for recent analyses of the adaptiveness of conformity).

This chapter builds on the questions raised by classic social psychology experiments on conformity to examine the wisdom of relying on crowds. We review two general sets of questions. The normative questions ask what is rational or optimal (Bell, Raiffa, & Tversky, 1988): How accurate are the judgments of collectives? When individuals disagree with the judgments of others, should they change their judgment or hold firm to their initial opinion? The descriptive questions ask what people actually do when they have access to the judgments of others (Bell et al., 1988): Do individuals understand why they should listen to others? Do they effectively decide when to listen to others?

To address these questions, we build on a research tradition that differs markedly from the Asch tradition. The Asch situation was by nature misleading. In more mathematical terms, the answers from the false group members were unrepresentative of true answers. Brunswik

(1955) famously argued that decision makers need to be presented with representative stimuli to assess how well they use information (Dami, Hertwig, & Hoffrage, 2004). To understand whether groups yield accurate answers and whether subjects are able to benefit from group accuracy, one needs to study processes in which representative judgments are elicited and shared with decision makers.

The remainder of the chapter is divided into two main sections. First, we describe a recent research on the "Wisdom of Crowds". This innovative literature combines decades of research in different fields showing the benefits of combining information across people. We analyze the conditions that make crowds effective to answer the normative questions of how and when crowds are wise. We then examine the descriptive question of how people use the judgments of others. We review a growing body of research showing that people are egocentric in their use of judgments: They rely too much on their own judgments and miss the opportunity to learn from others.

The Wisdom of Crowds

Some of the earliest studies in social psychology examined whether groups were fundamentally different from individuals. In the 1920s, researchers examined whether groups were smarter than individuals (see Larrick & Soll, 2006, for a historical review). In one early study students estimated the temperature in a classroom. When the estimates were averaged, the result was more accurate than the estimate of a typical group member. It should be noted that individual members of this "group" never interacted with one another. Where did the benefits come from? Early authors were surprised by this result and attributed it to some mysterious group property. As one writer put it, "In every coming together of minds… [t]here is the Creative Plus, which no one mind by itself could achieve" (Overstreet, 1925, cited in Watson, 1928). In time, researchers recognized that the power of groups came from something much simpler, but still elegant: Combining judgments takes individual imperfection and smoothes the rough edges to isolate the collective's view of the truth. Or, to put it more mathematically and mundanely, averaging cancels error.

Subsequent research in the forecasting literature demonstrated that simple combination methods that weight people equally, such as averaging judgments, often perform as well as more sophisticated statistical methods of combination (Armstrong, 2001; Clemen, 1989; Makridakis & Winkler, 1983). The power and simplicity of averaging was summed up in the title of James Surowiecki's 2004 best-selling book, "The Wisdom of Crowds."

Consider a brief example that illustrates the power of averaging. Imagine two professors estimating the number of students who are likely to apply to a program in neuroscience. Their goal is to get as close to the truth as possible. Being too high or too low is equally bad (e.g., being 5 high or 5 low will be treated as a miss of 5). Professor L estimates 40. Professor H estimates 60. Their average guess is 50. If the truth turns out to be 47, the judges have missed by 7 and 13, respectively, and their average miss is 10. But the average of the judges' guesses, 50, missed the true value of 47 by only 3—a substantially smaller error than the average miss of 10. Why does this happen? The judgments of 40 and 60 "bracket" the truth: The high error and low error offset each other. It is bracketing that gives averaging its power (Larrick & Soll, 2006; Soll & Larrick, 2009).

Now suppose that the truth is 37. Both professors have overestimated (by 3 and 23 respectively), and the average performance of the professors has missed by (3+23)/2 = 13. The average guess of 50 also misses by 13. In cases like this—where both judges fall on the same side of the truth—averaging "locks in" the average individual error. This is the worst case scenario for averaging. With bracketing, averaging will be more accurate than the average individual error.

Of course, Professor L is more accurate than Professor H in both examples,. Picking Professor L's lone judgment would have done well in the second scenario although fallen short of the average in the first scenario. It is tempting to declare Professor L smarter than the average. Research on groups frequently compares group performance to the performance of the "best member." It is important to point out, however, that in judgments under uncertainty the best member standard can be misleading when defined *post hoc*. Even when there is skill in judgment, there is luck as well, and it takes a large sample to know whether one judge is truly more accurate than another. The critical questions are whether a) one judge is *reliably* better than another judge by *a substantial margin* over time and b) whether the difference in ability is *detectable* from available cues in advance. In this example, could one know anything in advance about Professors L and H (or their answers) that would lead one to heavily favor the judgment of one over the other *before* knowing the outcome? Averaging is powerful because, thanks to bracketing, it must perform better than the average judge and can often perform much better. This performance makes it superior to choosing individual judges when judges are roughly similar in ability or when it is hard to distinguish their ability in advance.

The idea of bracketing generalizes to larger crowds.[1] To illustrate, the *Wall Street Journal* surveys a panel of about 50 economists every 6 months to make macroeconomic forecasts about inflation, unemployment, etc. The resulting diagram of forecasts often looks like a reasonable approximation of a normal distribution (see three hypothetical examples in Figure 1). When the truth is added to the plot six months later, it is often the case that the crowd of economists is roughly centered on the truth (top panel). Even when the distribution is not centered on the truth, it often “brackets” the truth (middle panel). Given some degree of bracketing, the average of the crowd is by mathematical necessity more accurate than the average individual.[2] When there is no bracketing—that is, the whole crowd is biased strictly above or below the truth as in the bottom panel—the average of the crowd is still as accurate as the average individual. Averaging the answers of a crowd therefore ensures a level of accuracy no worse than the average member of the crowd and, in some cases, a level better than nearly all members.

Conditions for Crowds to Be Wise

Crowds, of course, are not always wiser than individuals. The degree to which crowds are more accurate than individuals is a function of two factors: expertise and diversity. First, the crowd needs to consist of individuals with some knowledge or expertise about the issue in question. This could be based on past education or past experience. In quantitative judgments, expertise allows individuals to use imperfect evidence to make predictions that will fall close to the truth over many judgments.

Second, the crowd needs to hold diverse perspectives on the judgment in question. As a result of holding diverse perspectives, different individuals will bring different areas of expertise to bear on a judgment and therefore make different mistakes. For example, imagine two professors predicting the academic ability of graduate school applicants. One professor tends to focus on past research experience; the other focuses on grades. Both cues are valid—both cues are predictive of future performance. Moreover, the cues are not perfectly correlated, so over many cases, they often lead to conflicting conclusions. Thus, when one professor is optimistic about a student who has a good deal of research experience and the second professor is pessimistic because the student has low grades, their average assessment is more complete and likely to be more accurate than their individual assessments. Relying on a subset of cues introduces errors that can be offset by including the additional valid cues used by other judges.

To appreciate the effect of diversity on judgment, consider its absence. Imagine a marketing team evaluating the revenue potential of new possible products. If all of the marketers have worked in the same company on the same past products at the same period time, they have developed a shared set of experiences that guide their judgment about which new products are best. As a result, they are likely to hold shared opinions such as “customizability is more attractive to people under 40 than to people over 40.” The consequence is that they share the same expertise—and the same blindspots. Forecasting researchers term this pattern “positively correlated forecast errors” and show mathematically that it reduces the value of aggregating judgments across people because similar errors cannot cancel each other. Sociologists who study social networks refer to shared patterns of knowledge as “redundant” (Burt, 1992, 2004). The word redundant evokes the right image: In the extreme, the knowledge of one person is a pretty good substitute for another person’s. From a wisdom-of-crowds perspective, it is as if you do not have a crowd; your crowd is effectively 1.

There are two ways to foster diversity. Differences in perspective are created through who is in the group (composition) and how they share information (process).

*Composition***.** Research in the forecasting literature (Clemen & Winkler, 1986) has demonstrated the value of combining forecasts from different econometric models based on different economic “schools of thought”. The organizational literature has emphasized “cross-functional” teams as the best source for new ideas (Cronin & Weingart, 2007). A company does not want a new product designed by marketers alone—marketers will give it a fantastic look with many features, but may pay less attention to the cost of production. The company needs expertise in other areas (finance, engineering, etc.) to ensure that the product meets a more optimal set of objectives.

Network sociologists propose enhancing diversity by finding “gaps” in the social network (Burt, 1992). The premise is that in tightly knit groups, people talk a great deal, and come to know and value the same things. They become redundant. From a group’s vantage point, new ideas and new perspectives lie outside their own boundaries—even in other closely knit groups. The beauty of separation is that separated groups are likely to have evolved different views that are redundant *within* a group but non-redundant *between* them (Burt, 2004).

Creating diversity through composition can be challenging. One obstacle is that “in-groups” who share a similar perspective often look down on “out-groups” (DiDonato, Ullrich, & Krueger, 2011). Members of the marketing group suspect that members of the finance group are just “bean counters” who will stifle creativity. A second obstacle is that real differences in language and thinking impede collaboration between groups (Cronin & Weingart, 2007). These obstacles tend to lead people to associate with similar others. Sociologists term this tendency “homophily”; informally, it is the tendency for “birds of a feather flocking together.” Thus, when people assemble teams, they often select members based on common experience, common training, and common attitudes. This facilitates harmony but limits diversity.

Process. The second source of diversity is the process a group or collective uses to elicit judgments. A good group process ensures that individuals think independently before sharing their judgments. Why is independence important? Imagine being in an experiment in which you are asked to estimate the years in which important events happened. Imagine also that before you see a question, you see the real answer from another subject. For example, you see 1300 followed by the question “in what year was the Magna Carta written?” To what extent is your answer influenced by first seeing the answer of another subject? Koehler and Beauregard (2006) found that when answers were given *without* seeing other answers—that is, when answers were independent—they were typically 50 years apart (median value). When answers were given *after* seeing someone else's answer, however, they differed by only 10 years from the answer that was seen. Cognitive psychologists have called this general phenomenon "anchoring" (Tversky & Kahneman, 1974). Anchoring occurs because people start with a number, such as the year 1300, and unconsciously recruit evidence consistent with it (Chapman & Johnson, 1999; Mussweiler & Strack, 1999) and then fail to adjust sufficiently from it (Epley & Gilovich, 2001).

Whether a deliberate strategy or an unconscious anchoring process, people rely on others to form their judgments. Deutsch and Gerard (1955) called this tendency "informational influence" and contrasted it with the "normative influence" illustrated in the Asch study. Recall that participants in the Asch experiment would often publicly report an answer with which they disagreed privately. Informational influence, on the other hand, occurs when people use the judgments of others to reduce their own uncertainty. It results in a new opinion—public and private—that incorporates the beliefs of others.

Because of anchoring and informational influence, single judgments in a group can "propagate" through the judgments of others if the others have not first formed an opinion. For example, having one person start a discussion with an opinion ("we should increase inventory by 20%") tends to lead others to think in small variations of that opinion. The result is that both the initial accuracy and the initial error in the first judgment are spread to those who hear it (economists have studied this phenomenon under the name "information cascades"; see Krueger & Massey, 2009, for a review). Anchoring and informational influence produce "positively correlated error" and "redundancy", which amount to reducing the effective sample size.

Better group processes ensure that members think independently to form their judgments before sharing them. Research on brainstorming has found that working independently leads to a bigger and better pool of ideas than working face-to-face (Paulus & Yang, 2000). This effect occurs, in part, because independent ideas are generated in parallel, whereas face-to-face conversations require taking turns. But some of the effect is due to groups developing a common way of thinking about a question due to anchoring and informational influence. The danger of anchoring during group discussion has led one forecasting expert (Armstrong, 2006) to propose that many decisions would be better made if groups shared information without ever meeting.

Crowds Versus Individual Experts

To say that a crowd is wise invites the question, “compared to what?” The implicit comparison in these examples has been to the average individual in the crowd. Indeed, for quantitative judgments, the average of the crowd can never be worse than the average individual in the crowd. With a high rate of bracketing, averaging will be far superior to the average individual. However, the average individual is not the only standard by which to judge a crowd. One can also ask how well averaging performs compared with the best member of the crowd (Luan, Katsikopoulos, & Reimer, 2011). For this comparison to be informative, it is important to predict beforehand who the best judge will be. As noted earlier, it is always possible to determine the best performer after the fact.

There can be significant barriers to accurate identification of expertise. If there is no reliable track record of performance, one needs to rely on other cues. In a group discussion, members often rely on confidence and verbosity as cues to expertise (Littlepage, Robison, & Reddington, 1997). Unfortunately, cues like confidence often correlate weakly with actual accuracy (Burson, Larrick, & Klayman, 2007). Consequently, group members often inaccurately rank order the expertise of the members of their group (e.g., Miner, 1984).

Even if one has access to past performance, the sample may not be large enough to allow a valid inference of stable ability. If you track the best performers in the *Wall Street Journal* data from one period to the next, they are no more accurate than the average judge in later periods (Mannes, Soll, & Larrick, 2011). Why? Performance is a function of both skill and luck. More formally, apparent experts in one period of time "regress to the mean" of performance in the next period (for a discussion of regression to the mean, see the chapter by Fiedler and Krueger in this volume).[3] As a result, one is better off using the average of the whole sample of economists in the current period than to bet on the winner from the last survey.

Whether one should pick a single expert or rely on a crowd depends not just on access to valid evidence about the ability of judges, but also the presence of a large difference between the best judge and other judges. The benefits of diversity are so strong that one can combine the judgments from individuals who differ a great deal in their individual accuracy and still gain from averaging. In a two judge case, one judge can be 50% more accurate than another judge and averaging will still outperform the better judge with realistic rates of bracketing (Soll & Larrick, 2009). For example, if over many prediction periods one economist tends to miss true GNP increases by $50 billion on average and a second economist tends to miss by $75 billion, there is still a benefit to combining their judgments. It takes a large difference in ability to justify choosing one judge instead of averaging.

In sum, there are many advantages to averaging a crowd over choosing single experts. First, when judges are similar in ability, it allows their errors to cancel. Second, when judges differ in ability but differences are hard to detect, an averaging strategy is sure to give at least some weight to the best performers; by contrast, trying to pick a single expert based on available cues could put all the weight on a less accurate judge. Finally, an averaging strategy can be implemented even in the absence of evidence about relative expertise.

Types of Crowd Judgments

Most of the examples offered so far have focused on quantitative judgments under uncertainty. But the benefits of crowds can apply to many tasks. When choosing among options, "majority rule" performs well compared to more complex approaches (Hastie & Kameda, 2005; Sorkin, Luan, & Itzkowitz, 2004). In the popular game show "Who Wants to be a Millionaire?" the studio audience answered the trivia questions correctly 91% of the time (Surowiecki, 2004). More qualitative tasks also benefit from crowds. Research on creativity has shown that once a person has pursued one approach to solving a problem, it is hard for the person to generate other approaches. Brainstorming overcomes this problem by tapping the diverse perspectives of a group to generate a larger and more complete pool of ideas. Bond, Carlson, and Keeney (2006) asked business students to generate objectives for their summer job (e.g., pay level, location, growth opportunities, relevance to future career plans, etc.). They found that students listed about seven objectives on average. However, when students were then presented with both their own objectives along with a list of objectives generated by others, they tended to find another seven objectives that were as important as the ones they had generated on their own. Groups are smarter than individuals in creating a wider range of creative ideas, objectives, and alternatives.

Technology has made it easier to draw on the wisdom of crowds. Some tools are as simple as aggregating average ratings on consumer websites. Others are more complex, such as prediction markets, in which people wager real or pretend money to bet on future events, such as sports. (A famous but short-lived Department of Defense market had people betting for and against the timing of future terrorist attacks; betting "for" an attack was perceived as distasteful.) Similarly, corporations have used internal "idea jams" to tap employee perspectives on innovation opportunities and have used external "crowd sourcing" to reach a broad pool of diverse entrepreneurs to address unsolved technological problems they currently face.

The Individual as a Crowd

Research on the wisdom of crowds supports the old saying that two heads are better than one. Interestingly, the insights from this literature have also been used to show that one head can be nearly as good as two (Herzog & Hertwig, 2009; Vul & Pashler, 2008), as long as a judge follows the principles of relevant knowledge and diverse perspectives. The key insight is that people typically rely on only a sample of the evidence available to them at any given time. But what if people had a reset button, so that they could retrieve facts from memory anew or handle the same facts in a new way?

One way to free people from their original answers is to delay a second answer (Vul & Pashler, 2008) so that people forget their initial perspectives and think about the problem differently the second time around. Another way to free people is to have them try to construct a fresh perspective. To demonstrate this possibility, Herzog and Hertwig (2009) had participants make estimates about quantitative values they did not know with certainty (specifically, dates in history). All participants gave two answers to the same question, and the authors constructed an average of the first and second judgments for each individual. In one condition, participants simply gave a second estimate following their first estimate. This condition did little to increase diversity—people simply anchored on their initial opinions—and there was no benefit from averaging. In a second condition, participants were told to assume that their first answer was incorrect, think of some reasons it might be wrong, and then "based on this new perspective, make a second, alternative estimate." (p. 234). This process of "second guessing" increased bracketing; as expected, the average of the first and second judgments was significantly more accurate than the first estimate. A lesson of the Herzog and Hertwig's study is that we each carry around our own crowd, but we gain the wisdom only if we ask different members of the crowd.

How Well do People Use Crowds?

Strategies For Combining Judgments

Given that crowds are often wise, an important question for social psychology is whether people understand the value of combining knowledge across people. How well do people use judgments of others? This has become a growing area of research in recent years. Most research has focused on a very simple version of a crowd: Two-person collectives involving the self (Sniezek & Buckley, 1995; Yaniv & Kleinberger, 2000). In these studies, people make estimates and then learn actual, representative estimates of others. The source of estimates is often called an "advisor" and the literature as a whole has come to be known as "advice taking." The advice taking literature has used a wide range of quantitative stimuli, including estimating ages or weights of people from photographs, years in which historical events occurred, temperatures of cities, and so on. Studies that have focused on accuracy then compare initial and revised estimates to the truth and typically pay subjects for the closeness of their judgment. A smaller amount of research has looked at how people use the estimates of larger collectives of which they might or might not be members. We consider these two areas of research separately.

*Using Advice from One Other Person***.** One of the most robust findings in the advice-taking literature is that people underweight advice from another and overweight their own opinions (see Bonaccio & Dalal, 2006, for a review). A common result is that, on average, people tend to adjust their estimates 20%–30% of the way toward advice (Harvey & Fischer, 1997; Soll & Larrick, 2009; Yaniv, 2004), a phenomenon that Yaniv and Kleinberger (2000) labeled "egocentric discounting." For example, imagine estimating the age of someone from a photograph. You might make an initial estimate of 42 years old, see an "advisor's" estimate of 50, and adjust your answer to 44. The initial estimate and advice in this example are 8 years apart. The revised answer of 44 reflects a movement of 2/8, which translates to putting 25% weight on advice (WOA) and keeping 75% weight on your initial answer.

Subsequent research has found that 20-30% weight on advice is not descriptive of how people actually revise their judgments. The common pattern of 20%–30% weight on advice is an average result of more extreme underlying behavior. Soll and Larrick (2009) showed that people often either choose one of the two answers (typically their own) or use an equal-weighted average. In their studies, the 30% mean weight on advice reflected a pattern of frequently ignoring advice entirely (0% WOA), sometimes averaging their initial estimates with advice (50% WOA), and occasionally ignoring their own initial estimates and fully accepting advice (100% WOA). The most common response in these studies was 0% weight on advice. Such an extreme strategy takes no advantage of the error cancelling benefits of a two- person crowd and significantly hurts final accuracy. Subjects would have formed more accurate final estimates if they had given *equal* weight to their own estimates and advice (Soll & Larrick, 2009).[4]

Why do people put so much weight on their own estimates? Several explanations have been offered. First, many people have an incorrect theory of combining judgments (Larrick & Soll, 2006), believing that averaging leads to mediocrity. Specifically, they incorrectly believe that the average judgment in a crowd is no more accurate than the average judge. Holding this incorrect belief is significantly related to ignoring advice (Larrick & Soll, 2006).

Other explanations for ignoring advice have focused on more psychological assumptions. Harvey and Harries (2004) proposed that people believe that the advisor is substantially less accurate than oneself. Yaniv and his colleagues (Yaniv, 2004; Yaniv & Kleinberger, 2000) proposed that people weight their own answers more highly because they know the reasons for their own judgments but not those behind the judgments of others. Soll and Mannes (2011) tested both explanations by directly measuring subjects' perceptions of their accuracy relative to their advisor and by systematically providing or withholding cues at the time of revision (where seeing cues can remind subjects of their reasons for their initial answer). They found some evidence for inflated perceptions of the self, but not enough to explain the frequent use of 0% WOA. Moreover, they found no effect of having access to reasons on WOA. They proposed that the tendency to hold on to one's judgment may be less cognitive and more motivational: One's judgments are part of oneself and, like possessions, letting go of them is painful. Moreover, there may be an asymmetry in the regret of changing one's mind. Actively giving up an initially accurate answer for a worse one may lead to more regret than passively holding on to an inaccurate answer and foregoing improvement (see the chapter by Baron in this volume).

The advice-taking literature has found that a number of factors affect the weight placed on advice. Some factors are rational: Subjects weight advice more heavily when advisors are more experienced or knowledgeable (Harvey & Fischer, 1997; Soll & Larrick, 2009), when advisors express greater confidence in the quality of their advice (Sniezek & Buckley, 1995; Sniezek & Van Swol, 2001; Soll & Larrick, 2009), and when the subject finds the task difficult (Gino & Moore, 2008). As long as people use these advisor characteristics according to their validity, shifting weight to more expert advisors is an effective response to these cues. Other factors are more psychological: Decision makers weight advice less heavily when they feel powerful (See, Morrison, Rothman, & Soll, 2011) and when they experience emotions that increase feelings of certainty, such as anger (Gino & Schweitzer, 2008). This research identifies a practical set of factors that reduces the use of advice and can help practitioners in different fields, such as business and medicine, recognize when egocentric discounting will be at its greatest.

Combining Judgments from a Collective. A smaller stream of research has focused on how people combine the opinions of multiple others. Most studies in this area have looked at how people combine judgments across a panel of experts. In these studies, subjects have not made their own initial judgment but are neutral arbiters deciding how to balance the judgments of others (Budescu, 2006; Harvey, Harries, & Fisher, 2000; Yaniv, 1997). Budescu and colleagues (Budescu, 2006; Budescu, Rantilla, Yu, & Karelitz, 2003) have found that people tend to weight expert judgments in proportion to the expertise of different judges.

Two studies in this area have looked at "advice taking" from a group. As in the two-person advice research, these studies involve having people make an initial judgment and then revise it after seeing the estimates of a group (Mannes, 2009; Yaniv & Milyavsky, 2007). Yaniv and Milyavsky (2007) found that people "cherry pick" the advice from a larger crowd, focusing on the judgments most consistent with their own. Subjects effectively use their first guess as a standard of accuracy and dismiss discrepant advice. From a wisdom-of-crowds perspective, this is dangerous: It ignores the benefits of incorporating diverse perspectives to cancel error. Mannes (2009) explored the extent to which people listen to the average judgment from crowds of different sizes. Normatively, subjects would be wise to put less weight on themselves and more weight on the crowd as the crowd grows in size (reflecting a basic principle in statistics known as the “law of large numbers”). In addition, if all judges are expected to be equally accurate in advance, subjects would be wise to put the same weight on themselves as they do on each member of the crowd (i.e., a 1/n weight on their own judgment and a (n-1)/n weight on the crowd’s advice). Mannes (2009) found that people put more weight on larger crowds, as they should from a normative perspective, but not enough weight. They put only 60% weight on advice from a 9-person crowd where equal weighting would require more than 90% weight. The consequence of egocentric weighting is that subjects paid a significant price in the accuracy of their final judgments.

Future Directions for the Psychology of the Wisdom of Crowds

The main conclusion from existing research is that people use crowds too little: They put too much weight on their own judgment and thereby miss out on the benefit of diverse perspectives for reducing error. This is a young and growing area of research. What are some of the unanswered questions? Perhaps a key issue that arises in the existing work is that advice is quantitative, impersonal, and unsolicited—it consists of seeing numerical estimates made by others (including group averages) and making conscious decisions about how best to combine them with one’s own independent estimates. The use of crowd judgments should be explored in other ways.

First, advice *seeking* could be an important variable to study. How often do people seek advice? Perhaps the easiest way to reduce error in judgment is to seek other’s opinions, and the failure to do so has the same negative consequences as ignoring advice. We suspect people seek advice less often than they should. How does seeking advice affect its use? The act of seeking advice is likely to increase the extent to which advice is used (Gino, 2008) and thereby improve judgment. It is also possible that seeking advice will lead people to pick a single expert and put too much faith in that single piece of advice.

Second, other advice-taking contexts may change openness to advice, such as face-to-face interactions or interactions with richer information sharing. Richer information sharing may lead to deeper information processing. Face-to-face interactions may evoke processes, such as empathy and mimicry,that produce greater yielding to others.

Third, the benefits of independence should be carefully examined. Earlier we argued that independence increases diversity of perspective and thereby makes a crowd more accurate. Asch (1955, p 34) captured this idea when writing that “consensus, to be productive, requires that each individual contribute independently out of his experience and insight.” In this view, individual judgments are “inputs” to a combination process and they are more valuable if they are independent. However, individual judgments can also be “outputs” from exposure to the judgments of collectives. Egocentric discounting of advice is an example of an inferior output. In “output” situations there may be a benefit if the individual does *not* have an independent opinion: Anchoring on a crowd may actually yield more accurate judgments because it ensures that individual judges are using a larger sample—even if unwittingly.

Consider a study that predates Asch. Jenness (1932) asked students to estimate the number of beans in a jar. They made their estimates in a prescribed sequence: first individually, then by consensus in a three-person group, and then individually again. The initial individual guesses missed the true value (811) by 305 on average. The group consensus answers missed the value by just 91, producing a substantial improvement over their initial individual guesses. Strikingly, however, their final individual guesses missed by 122 on average—worse than the group answers. What went wrong? By reasserting their independence and deviating from the group answer, the subjects in Jenness's study were less accurate than if they had simply placed their faith in the crowd.

Thus, we can think of crowd processes as having two stages. Independence is important at Stage 1, the input stage, because the effective sample size is increased when different judges make different errors. However, independence may be unimportant or even detrimental at Stage 2, the output stage, when a decision maker uses others' judgments to form a final opinion that he or she is going to act on. Campbell (1961, p. 123) adopted the Stage 2 perspective when he argued that, "[I]n Asch's famous situation, the single true subject might rationally decide that, since everybody's eyes are imperfect, and since it would be so extremely infrequent that so many Swarthmore students would deliberately lie in a situation like this, it is more probable that his own eyes are wrong than that all of the others are wrong. He might, therefore, rationally decide that, if asked to bet, he would bet with the majority." In contrast to the view inherited from the Asch tradition, social influence may actually be beneficial. Normative and informational influence may serve as a "cognitive repair" (Heath, Larrick, & Klayman, 1998) that mitigates depending too much on one's own judgments. Influence ensures that people incorporate the wisdom of the crowd in their own judgments.

References

Allen, V. L. (1965). Situational factors in conformity. *Advances in Experimental Social Psychology, 2,* 133-175.

Armstrong, J. S. (2001). *Principles of forecasting: A handbook for researchers and practitioners*. Boston, MA: Kluwer Academic.

Armstrong, J. S. (2006). Should the forecasting process eliminate face-to-face meetings? *Foresight: The International Journal of Applied Forecasting, 5,* 3-8.

Asch, S. E. (1955). Opinions and social pressure. *Scientific American, 193,* 31-35.

Bell, D., Raiffa, H., & Tversky, A. (1988). *Decision making: descriptive, normative, and prescriptive interactions.* Cambridge: Cambridge University Press

Bonaccio, S., & Dalal, R. S. (2006). Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational Behavior and Human Decision Processes, 101*, 127-151.

Bond, S. D., Carlson, K. A., Keeney, R. L. (2008). Generating objectives: Can decision makers articulate what they want? *Management Science*, *54*, 56-70.

Brunswik, E. (1955): Representative design and probabilistic theory in a functional psychology. *Psychological Review* 62:193-217.

Budescu, D. V. (2006). Confidence in aggregation of opinions from multiple sources. In K. Fiedler & P. Juslin (Eds.), *Information sampling and adaptive cognition*. Cambridge, UK: Cambridge University Press.

Budescu, D. V., Rantilla, A. K., Yu, H.-T., & Karelitz, T. M. (2003). The effects of asymmetry among advisors on the aggregation of their opinions. *Organizational Behavior and Human Decision Processes, 90*, 178-194.

Burson, K.A., Larrick, R.P., & Klayman, J. (2006). Skilled or unskilled, but still unaware of it: how perceptions of difficulty drive miscalibration in relative comparisons, *Journal of Personality and Social Psychology, 90,* 60–77.

Burt, R. S. (1992). *Structural holes*. Cambridge, MA: Harvard University Press.

Burt, R. S. (2004). Structural holes and good ideas. *American Journal of Sociology, 110,* 349-399.

Campbell, D. T. (1961). Conformity in psychology's theories of acquired behavioral dispositions. In I. A. Berg & B. M. Bass (Eds.), *Conformity and deviation* (pp. 101-142). New York: Harper & Brothers.

Chapman, G. B., & Johnson, E. J. (1999). Anchoring, activation and the construction of values. *Organizational Behavior and Human Decision Processes, 79,* 115-153.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting, 5,* 559-609.

Clemen, R. T., & Winkler, R. L. (1986). Combining economic forecasts. *Journal of Business and Economic Statistics, 4,* 39-46.

Cronin, M. A., and Weingart, L. R. (2007). Representational gaps, information processing, and conflict in functionally diverse teams. *Academy of Management Review, 32,* 761-773.

Dami, M. K., Hertwig, R., & Hoffrage, U. (2004). The role of representative design in an ecological approach to cognition, 130, 959-988.

Denrell, J., & Fang, C. (2010). Predicting the next big thing: Success as a signal of poor judgment. *Management Science, 56,* 1653-1667.

DiDonato, T. E., Ullrich, J., & Krueger, J. I. (2011). Social perception as induction and inference: An integrative model of intergroup differentiation, ingroup favoritism, and differential accuracy. *Journal of Personality and Social Psychology, 100,* 66-83.

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *Journal of Abnormal and Social Psychology*, *51*, 629-636.

Epley, N., & Gilovich, T. (2001). Putting adjustment back in the anchoring and adjustment heuristic: Differential processing of self-generated and experimenter-provided anchors. *Psychological Science, 12*, 391-396.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*, 117-140.

Gino, F. (2008). Do we listen to advice just because we paid for it? The impact of cost of advice on its use. *Organizational Behavior and Human Decision Processes*, *107,* 234–245.

Gino, F., & Moore, D. A. (2007). Effects of task difficulty on use of advice. *Journal of Behavioral Decision Making*, *20,* 21–35.

Gino, F., & Schweitzer, M. (2008). Blinded by anger or feeling the love: How emotions influence advice taking. *Journal of Applied Psychology*, *93,* 1165-1173.

Harvey, N., & Fischer, I. (1997). Taking advice: Accepting help, improving judgment, and sharing responsibility. *Organizational Behavior and Human Decision Processes, 70*, 117-133.

Harvey, N., & Harries, C. (2004). Effects of judges' forecasting on their later combination of forecasts for the same outcomes. *International Journal of Forecasting, 20,* 391-409.

Harvey, N., Harries, C., & Fischer, I. (2000). Using advice and assessing its quality. *Organizational Behavior and Human Decision Processes, 81,* 252–273.

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

Herzog, S. M., & Hertwig, R. (2009). The wisdom of many in one mind: Improving individual judgments with dialectical bootstrapping. *Psychological Science, 20*, 231-237.

Jenness, A. (1932). The role of discussion in changing opinion regarding a matter of fact. *Journal of Abnormal and Social Psychology, 27*, 279-296.

Kameda, T., & Tindale, R. S. (2006). Groups as adaptive devices: Human docility and group aggregation mechanisms in evolutionary context. In M. Schaller, J. Simpson & D. Kenrick (Eds.), *Evolution and social psychology*. New York: Psychology Press.

Koehler, D. J., & Beauregard, T. A. (2006). Illusion of confirmation from exposure to another's hypothesis. *Journal of Behavioral Decision Making, 19*, 61-78.

Larrick, R. P., & Soll, J. B. (2006). Intuitions about combining opinions: Misappreciation of the averaging principle. *Management Science*, *52*, 111-127.

Littlepage, G. E., Robison, W., & Reddington, K. (1997). Effects of task experience and group experience on group performance, member ability, and recognition of expertise. *Organizational Behavior and Human Decision Processes, 69*, 133-147.

Luan, S., Katsikopolous, K. V., & Reimer, T. (2011). The “less-is-more” effect in group decision making. In R. Hertwig, U. Hoffrage, & the ABC Research Group (Eds.), *Simple heuristics in a social world*. New York: Oxford University Press.

Mannes, A. E. (2009). Are we wise about the wisdom of crowds? The use of group judgments in belief revision. *Management Science, 55,* 1267–1279.

Mannes, A. E., Soll, J. B., & Larrick, R. P. (2011). The wisdom of small crowds (unpublished manuscript).

Miner, F. C. (1984). Group versus individual decision making: An investigation of performance measures, decision strategies, and process losses/gains. *Organizational Behavior and Human Performance, 33*.

Mussweiler, T., & Strack, F. (1999). Hypothesis-consistent testing and semantic priming in the anchoring paradigm: A selective accessibility model. *Journal of Experimental Social Psychology, 35*, 136-164.

Paulus, P. B., & Yang, H. C. (2000). Idea generation in groups: A basis for creativity in organizations. *Organizational Behavior and Human Decision Processes, 82*, 76-87.

See, K. E., Morrison, E. W., Rothman, N. B., & Soll, J. B. (2011). Powerful and unpersuaded: The implications of power for confidence, advice-taking, and accuracy. (Working paper).

Sniezek, J. A., & Buckley, T. (1995). Cueing and cognitive conflict in Judge-Adviser decision making. *Organizational Behavior and Human Decision Processes, 62*, 159-174.

Sniezek, J. A., & Van Swol, L. M. (2001). Trust, confidence, and expertise in a judge-advisor system. *Organizational Behavior and Human Decision Processes*, *84*, 288–307.

Soll, J. B., & Mannes, A. E. (2011). Judgmental aggregation strategies depend on whether the self is involved. *International Journal of Forecasting*, *27*, 81-102.

Soll, J.B., & Larrick, R. P. (2009). Strategies for revising judgment: how (and how well) people use others' opinions, *Journal of Experimental Psychology: Learning, Memory and Cognition, 35*, 780–805.

Sorkin, R. D., Luan, S., & Itzkowitz, J. (2004). Group decision and deliberation: A distributed detection process. In D. Koehler & N. Harvey (Eds.), *Handbook of judgment and decision making* (pp. 464–484). New York: Oxford University Press.

Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations*. London: Little, Brown.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124-1131.

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

Watson, G. B. (1928). Do groups think more efficiently than individuals? *Journal of Abnormal and Social Psychology, 23,* 328-336.

Yaniv, I. (1997). Weighting and trimming: heuristics for aggregating judgments under uncertainty. *Organizational Behavior and Human Decision Processes, 69,* 237–249

Yaniv, I. (2004). Receiving other people's advice: Influence and benefit. *Organizational Behavior and Human Decision Processes, 93,* 1-13.

Yaniv, I., & Kleinberger, E. (2000). Advice taking in decision making: Egocentric discounting and reputation formation. *Organizational Behavior and Human Decision Processes, 83,* 260-281.

Yaniv, I., & Milyavsky, M. (2007). Using advice from multiple sources to revise and improve judgments. *Organizational Behavior and Human Decision Processes, 103,* 104-120.

Footnotes

[1] One can conceive of bracketing in a group as the rate at which randomly-selected pairs bracket the truth or as a ratio of the proportion of the crowd falling on each side of the truth. For example, if 60% of the crowd is high and 40% low, the pairwise bracketing rate is 1-((.6*.6)-(.4*.4)) = .48, and there is a 1.5 (.6/.4) ratio.

[2] This mathematical necessity can be proven by applying Jensen's inequality. Let Judge *i*'s estimate on quantity *j* be represented as $X_{ij} = T_j + D_{ij}$, where T_j is the correct answer and D_{ij} is the deviation, $i = 1, 2, ..., n$, and $j = 1, 2, ..., m$. Let $w_1, w_2, ..., w_n$ be the weights assigned to n judgments, where $\sum w_i = 1$. Accuracy on a given estimate is a function of the deviation, $f(D)$. Typically, $f(D)$ is increasing in $|D|$, so higher scores reflect lower accuracy. For a quantity *j*, the deviation for a weighted average is the weighted average of the deviations of the individual estimates:

$$f(w_1 x_{1j} + w_2 x_{2j} + \ldots + w_n x_{nj} - T_j) = f\left(\sum_{i=1}^{n} w_i D_{ij}\right),$$

The right-hand side of the preceding equation gives the accuracy of a weighted average. If *f* is a convex function, then Jensen's inequality states that

$$\sum_{i=1}^{n} w_i f\left(D_{ij}\right) \geq f\left(\sum_{i=1}^{n} w_i D_{ij}\right).$$

In this chapter, we focus on absolute error as our accuracy standard because it is neutral in punishing larger versus small errors. Past forecast research has focused on squared error as the loss function, which implies that larger errors are worse than smaller ones. When squared error is used as the loss function, averaging is more accurate than the average judge even when there is no bracketing.

Aggregation also improves judgment using other criteria for accuracy, such as correlations with the truth.

[3] Denrell and Fang (2010) showed that when a *Wall Street Journal* economist "wins" one period with more extreme judgments—that is, judgments that are outliers compared to the rest of the group—they are actually less accurate than the average judge in subsequent periods.

[4] In these studies, subjects made judgments independently (yielding bracketing rates of 30% to 40%), were randomly assigned advisors from their own subject population (yielding small average differences in expertise between judges), and given cues to expertise such as self-expressed confidence or a small sample of past performance (which are weak cues). All of these features, which favor averaging, were known by subjects.

Figure 1. Three distributions of individual judgments in which judgments are evenly distributed around the truth (top panel), bracket the truth (middle panel), or are biased below the truth (bottom panel).