Two to Tango : Effects of Collaboration and Disagreement on Dyadic Judgment

Julia A. Minson, Varda Liberman and Lee Ross

Pers Soc Psychol Bull 2011 37: 1325 originally published online 1 June 2011
DOI: 10.1177/0146167211410436

Personality and Social Psychology Bulletin
37(10) 1325–1338

DOI: 10.1177/0146167211410436
http://pspb.sagepub.com

Two to Tango: Effects of Collaboration and Disagreement on Dyadic Judgment

Julia A. Minson[1], Varda Liberman[2], and Lee Ross[3]

Abstract

Four studies examined dyadic collaboration on quantitative estimation tasks. In accord with the tenets of "naïve realism," dyad members failed to give due weight to a partner's estimates, especially those greatly divergent from their own. The requirement to reach joint estimates through discussion increased accuracy more than reaching agreement through a mere exchange of numerical "bids." However, even the latter procedure increased accuracy, relative to that of individual estimates (Study 1). Accuracy feedback neither increased weight given to partner's subsequent estimates nor produced improved accuracy (Study 2). Long-term dance partners, who shared a positive estimation bias, failed to improve accuracy when estimating their performance scores (Study 3). Having dyad members ask questions about the bases of partner's estimates produced greater yielding and accuracy increases than having them explain their own estimates (Study 4). The latter two studies provided additional direct and indirect evidence for the role of naïve realism.

Keywords

collaborative judgment, dyads, naïve realism, disagreement, judgment aggregation

Received March 11, 2010; revision accepted March 19, 2011

Popular wisdom holds that "two heads are better than one"—presumably because individuals can correct each other's errors and excesses and benefit from each other's insights. However, converging evidence from studies on the performance of dyad members who exchange estimates (Liberman, Minson, Bryan, & Ross, 2011; Soll & Larrick, 2009) and of individuals responding to the input of "advisors" (see the review by Bonaccio & Dalal, 2006; also see Yaniv, 2004; Yaniv & Kleinberger, 2000) suggests that the benefits of exposure to such input may be smaller than one might anticipate on logical and statistical grounds. In particular, individuals consistently fail to achieve the full benefits of access to each other's assessments because they give those estimates too little weight.

The research cited above has also made it clear that, as one would expect based on purely statistical considerations, the potential benefits of averaging, and the costs of doing otherwise, are substantially greater when the dyad members' initial estimates "bracket" (i.e., fall on opposite sides of) the correct answer. One implication of this state of affairs is that the greater the extent to which individuals disagree in their estimates, the more likely they are to "bracket" the answer and thus benefit from moving toward their partner's estimates. However, as the tenets of "naïve realism" (Ross & Ward, 1995, 1996) suggest, the more people disagree, the more inclined they are to attribute that disagreement to the existence of bias or error on the part of those with whom they disagree (Pronin, Gilovich, & Ross, 2004). As a result, we suggest, the very dyad members who stand to gain most from giving each other's inputs due weight will be least inclined to do so—and thus will be the ones who pay the heaviest price in terms of estimation error for that failure. The present research was designed to test the conditions under which dyad members making numerical estimates succeed or fail to maximize the benefits of collaboration, focusing in particular on the role of naïve realism.

Conceptual Framework

The growing literature on advice utilization (Larrick & Soll, 2006; Soll & Larrick, 2009; Vul & Pashler, 2008; Yaniv, 2004; Yaniv & Kleinberger, 2000) as well as a classic literature addressing appropriate use of multiple pieces of evidence (Dawes, 1979; for reviews, see Lorge, Fox, Davitz, &

[1]University of Pennsylvania, Philadelphia, PA, USA
[2]Interdisciplinary Center Herzliya, Herzliya, Israel
[3]Stanford University, Stanford, CA, USA

Corresponding Author:
Julia A. Minson, University of Pennsylvania, 550 Jon M. Huntsman Hall, 3730 Walnut St., Philadelphia, PA 19104-6340
Email: jminson@wharton.upenn.edu

Brenner, 1956; also see Surowiecki, 2004) have offered substantial conceptual and empirical proof for the merits of giving equal weight to divergent opinions in making numerical judgments. Many factors may contribute to dyad members' willingness to adopt this strategy versus other strategies for improving the accuracy of their judgments. In trying to examine such factors, it is useful to sort them by characteristics of the estimation items, characteristics of the estimators, and characteristics of the estimation process.

More specifically, estimation items may pertain to familiar domains (e.g., the weight of pictured individuals; Gino & Moore, 2007), in which errors are likely to be bounded by common knowledge or to lesser known ones with potential for large errors (e.g., the distance between two cities; Soll & Larrick, 2009). They may require estimates of concrete facts, or estimates of peer consensus, the latter being subject to the well-documented false consensus bias (Ross, Greene, & House, 1977). And, as shown by Soll and Larrick (2009), the distribution of estimates for a given item also plays a role. That is, the potential benefits of consistent averaging, and the costs of doing otherwise, are substantially greater when the dyad members' initial estimates "bracket" the correct answer.

Also of importance are the characteristics of the estimators themselves. To date, research on dyadic judgment and utilization of advice has been conducted using participants with no specialized knowledge in the relevant domain. Furthermore, participants generally receive input that has either been computer generated or has been produced by a partner with whom they have no interaction beyond an exchange of estimates. Although some studies have begun to address the way expert advice is treated by novices (e.g., Harvey & Fischer, 1997), no study we know of has invited dyad members who are both highly experienced in the relevant domain to make use of each other's estimates.

Finally, and perhaps of greatest theoretical and applied interest, are factors or procedures that increase the weight participants' give to their partner's judgments and other factors that improve accuracy. For example, Larrick and Soll (2006) showed that presenting estimates in pairs rather than sequentially makes participants more sensitive to high bracketing rates. More relevant to present concerns, Liberman et al. (2011) showed that compelling dyad members to arrive at jointly agreed on estimates through discussion increases the accuracy of both these joint and subsequent individual estimates.

Overview of Studies

The four studies presented here explore dyadic collaboration in the making of numerical estimates, and the consequences of obliging dyad members to agree on joint estimates, under a variety of conditions. The studies differ in the types of estimation items utilized, the relevance of the domain to the dyad members, and the dyads' prior history of collaboration. In the course of these studies we compare accuracy improvement achieved by reaching agreement through a mere exchange of numerical "bids" with that achieved through discussion, the impact of feedback regarding prior accuracy, the effect of long-term collaboration, and the effects of different types of information exchanges about the bases for own and partner's estimates.

All four studies address the role of naïve realism in limiting the benefits of collaboration and/or strategies for overcoming this barrier. Studies 1–3 tested the hypothesis, based on our analysis of the role played by naïve realism and prior research on that phenomenon (Liberman et al., 2011), that underweighting of partner input would be most pronounced when on statistical grounds it is apt to be most costly—that is, where the dyad members initial estimates diverge most sharply. Study 3 allowed us to seek direct evidence for the role of naïve realism among long-term expert collaborators by examining the attributions participants make for the bases of their own estimates versus those of their partners. It also explored the possibility that a *shared* bias among long-term collaborators—in this case, the tendency of competitive ballroom dancers to overestimate the quality of their performances—may keep them from reaping the benefits of collaboration. Study 4 tested the prediction derived from a naïve realism analysis that having dyad members ask questions about the bases for each other's initial estimates would benefit the dyad members more than having them explain the bases of their own estimates.

In our studies we use a four-round estimation procedure introduced in Liberman et al. (2011). This procedure features an initial round of individual estimates, followed by a round of revised estimates made after seeing the estimates of a dyad partner, then a round of agreed-on "joint" estimates, and a final round of individual estimates in which the dyad members are free to give each other's prior estimates as little or as much weight as they wish. This series of estimates allows us to investigate the conditions under which the requirement to reach agreement yields accuracy benefits above and beyond that of simple exposure to a partner's estimates.

Study 1—The Value of Discussion Versus "Bidding" in Reducing Estimation Error

Previous research (Liberman et al., 2011) linked underweighting of partner judgment to naïve realism by demonstrating that dyad members see their own estimates as more "objective" than those of their partners and that these differences in perceived objectivity predict the degree of underweighting. These studies also demonstrated that the effects of this bias on estimation accuracy can be overcome if dyad members must agree on joint estimates. Study 1 was designed to extend these findings by manipulating the process by which dyad members reached agreement. Although half of the dyads

reached agreement through a discussion in which partners could explain the bases for their estimates (*discussion* condition), the remaining dyads were obliged to arrive at their joint estimates through an exchange of "bids" whereby the dyad members simply exchanged written estimates until they converged on a single number (*Bidding* condition). Dyad members were then freed to make a final set of individual estimates wherein they could give as little or as much weight to their own and their partner's estimates as they wished. This manipulation allowed us to explore the extent to which the previously documented accuracy improvement in the third round resulted from discussion that attenuated the effects of naïve realism by allowing dyad members to form impressions about the basis and relative accuracy of own and partner's estimates, rather than the simple statistical benefit of having to converge on a single joint estimate.

Method

Second year Israeli business school students ($N = 66$), working in dyads made estimates regarding nine statistics relevant to business or political decision making.[1] For example, participants were asked to estimate the amount of money that an average Israeli family of four spends monthly on food (answer: 1,990 shekels). Item order was counterbalanced. Four rounds of estimates were made—an initial round of independent estimates, a second round of individual estimates made with the knowledge of own and partner's initial estimates, a third round of agreed on "joint" estimates, and a final round of individual estimates. Associated confidence ratings were made and shared between Rounds 1 and 2 and between Rounds 3 and 4. To encourage accuracy, participants began each round with a bonus of 200 shekels (about $50) and had that sum reduced by 1 shekel for each percentage point error during that round.

The process by which participants arrived at their Round 3 estimates was varied between conditions. In the Discussion condition dyad members were free to exchange any information they wished in arriving at their joint estimates. In the Bidding condition, dyad members simply exchanged estimates on a worksheet until they reached a single, agreed-on, estimate.

Results and Discussion

The influence of own and partner's initial estimates on second round estimates. In our analyses of weight given to partner's initial estimates we omitted the 30 cases in which the two dyad members offered identical Round 1 estimates. On average, dyad members moved only 30.4% of the distance toward their partner's initial estimates—a distance far short of the halfway point, $t(65) = 9.7$, $p < .001$.[2] Furthermore, they gave greater weight to their own rather than their partner's estimates in 72.7% of cases (Table 1).

To test our prediction regarding the influence of disagreement in initial estimates on participants' willingness to give weight to their dyad partner's judgments we created a dichotomous variable distinguishing cases in which a dyad member moved less than halfway toward the partner's prior estimate from cases in which the dyad member moved at least halfway toward that estimate. We then tested a logistic hierarchical model in Stata with weighting of partner's estimate as a dichotomous dependent variable, and trial-level disagreement between the two estimates (which was group centered) as well as the average disagreement between any two dyad members, as predictor variables.[3] Our analysis revealed no relationship between size of average disagreement between two dyad members and the likelihood of giving greater weight to own rather than partner's initial estimate ($B = –0.692$, $z = –0.35$, *ns*), and the predicted relationship between giving greater weight to own estimate and group-centered disagreement ($B = 1.44$, $z = 2.44$, $p < .02$). In other words, controlling for the average level of disagreement between two dyad members, greater disagreement on a particular item resulted in greater likelihood of participants preferentially weighting their own initial estimate on that item.

Round-to-round changes in estimation accuracy. To allow us to combine accuracy data across items in which both correct answers and errors differed greatly in magnitude, in Study 1 (and also in Study 2), we first transformed both correct answers and dyad members' estimates to base 10 logs. We then treated the absolute difference between the log of the estimate and the log of the correct answer as our measure of estimation error. (Thus, estimation errors for estimates of one tenth or 10 times the correct answer were 1.0, and those for half or twice the correct answers were both 0.301.)

As illustrated in Figure 1, the improvement in accuracy shown by participants from Round 1 to Round 2 (from $M = 0.376$ to $M = 0.325$), $t(32) = 6.43$, $p < .001$, replicated the results reported by earlier investigators.[4] In fact, the latter mean error was very close to the mean error ($M = 0.329$) that would have resulted had members of each dyad simply averaged their two Round 1 estimates. Joint estimates in Round 3 (combining data from both conditions) showed a further reduction ($M = 0.034$) in mean estimation error, $t(32) = 4.6$, $p < .001$, achieving a level of accuracy significantly greater than that available through simple averaging of either Round 1 estimates, $t(32) = 4.16$, $p < .001$, or Round 2 estimates, $t(32) = 2.78$, $p < .01$.

Our main focus in Study 1, however, involved a comparison of the Round 3 results for Bidding versus Discussion condition dyads. Hierarchical linear modeling, using condition as a dummy coded independent variable (*Bidding* = 0, *Discussion* = 1) revealed that reaching agreement through discussion led to greater accuracy improvement than reaching agreement through an exchange of bids ($B = 0.032$, $z = 3.20$, $p = .001$), or averaging. However, it is important to note that even in the absence of discussion, dyad members in the Bidding condition achieved a significant gain in accuracy ($B = 0.016$, $z = 2.07$, $p < .04$), albeit not one significantly different from that achievable through simple averaging.

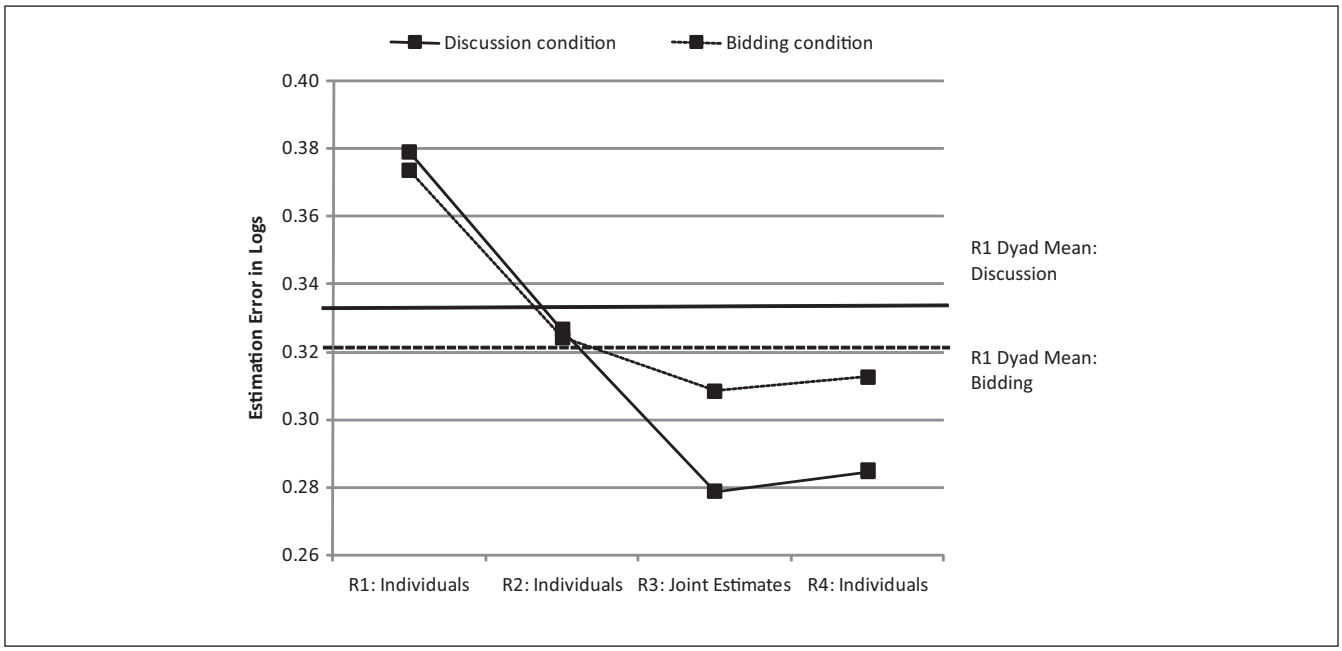

Figure 1. Mean round-by-round estimation errors (in log form) by condition (Study 1)

In further pursuing the implications of these data for understanding when and why reaching agreement improves accuracy, it is instructive to examine separately cases in which the two dyad members' initial estimates "bracketed" the correct answer, and cases in which they did not. From Round 1 to Round 2 the initial bracketers, for whom averaging guaranteed an increase in mean accuracy, reduced their mean error by 0.12 in both conditions (both p values < .005). The initial nonbracketers, for whom improvement depended on discriminating better estimates from poorer ones, showed much more modest (and not statistically significant) reductions in mean error—0.02 in the Discussion condition and 0.03 in the Bidding condition.

After the manipulation, in Round 3, the initial bracketers showed further reductions in mean estimation error in both conditions—somewhat greater in the Discussion condition, $B = 0.067$, $z = 4.5$, $p < .001$, than in the Bidding condition, $B = 0.073$, $z = 1.66$, $p < .10$. In the case of initial *non*bracketers, by contrast, only the Discussion condition dyads significantly reduced their mean error ($B = 0.042$, $z = 5.21$, $p < .001$). Among the Bidding condition dyads, by contrast, this decrease was close to zero ($B = 0.004$, $z < 1$, *ns*). This between-condition difference was statistically significant, $B = 0.038$, $z = 3.56$, $p < .001$. It thus appears that the benefits of discussion over the exchange of bids were the result of the ability of nonbracketing dyad members to use discussion for identifying the more accurate estimates.

Round 4. To what extent were dyad members able to retain the benefits of agreement when they were once again free to offer their own individual estimates? When we collapse across experimental conditions, hierarchical linear modeling confirmed that although in Round 4 mean estimation error remained lower than it had been in Round 2, $B = 0.029$, $z = 4.50$, $p < .001$, this overall difference was moderated by condition. The final estimates offered in the Bidding condition were no more accurate than those offered in Round 2 ($B = 0.011$, $z = 1.24$, *ns*). By contrast, participants' final estimates in the Discussion condition were significantly more accurate than their prior estimates ($B = 0.042$, $z = 5.15$, $p < .001$).

Study 2: Effects of Feedback

Beyond again exploring dyad member's ability to benefit from collaboration and the requirement to offer joint estimates, Study 2 tested the effects of providing participants with ongoing accuracy feedback throughout the estimation task. Such feedback allowed them to observe both their own and their dyad partner's prior accuracy, and potentially even to recognize the statistical advantages of averaging.[5] The empirical question addressed was the extent to which the "lessons" provided by the relevant feedback regarding *previous* estimates would overcome naïve realism in deciding what to do when faced with a *new* pair of own and partner's estimates.

Beyond replicating the finding of Study 1 regarding the effect of disagreement in initial estimates on underweighting of partner input, in this study we also examine the effect of disagreement on the accuracy improvement shown in final individual estimates, made after dyad members were compelled to reach joint estimates through discussion.

Method

Participants. A total of 76 participants (38 dyads) provided estimates in Study 2. All were Israeli business school

Table 1. Relative Frequency of Round 2 Estimation Strategies (Study 1)

Estimation strategy	Average relative frequency (%)
Stood pat	26.1
Total less than halfway to partner	72.7
Exactly halfway to partner	7.5
All the way to partner	5.4
Total more than halfway to partner	19.8
Percentage of distance yielded	30.4

Table 2. Relative Frequency of Round 2 Estimation Strategies (Study 2)

	Control condition (%)	Feedback condition (%)
Stood pat	37.3	40.8
Total less than halfway to partner	77.4	75.4
Exactly halfway to partner	5.7	4.7
All the way to partner	6.3	8.6
Total more than halfway to partner	16.9	19.9
Percentage of distance yielded	26.6	27.5

students who received credit in a statistics course for their efforts.

Procedure. As in Study 1, dyad members made four rounds of estimates regarding 10 statistics of potential relevance to business or political decision making. For example, participants were asked to estimate the size of the Druze population in Israel (answer: 115,000). Participants offered all four rounds of estimates for each item before proceeding to the next item and in all cases made their Round 3 estimates with the benefit of discussion. Participants in the *Feedback* condition were provided with the correct answer to each question at the conclusion of the Round 4 estimates for that question; those in the *Control* condition received no such feedback. Participants worked with the same partner on all 10 items and had access to their prior estimates throughout the task.

Results and Discussion

The influence of own and partner's initial estimates on second round estimates. After eliminating cases in which dyad members gave identical first round estimates, we found that Feedback condition participants moved a mean of 27.5% of the distance to their partner (vs. 26.6% in the Control condition) and gave greater weight to their own prior estimates in 75.4% of all estimates (vs. 77.5% in the Control condition; both t-values < 1). To address the possibility that the lessons learned from performance feedback required some experience, we repeated these analyses using only the last five items used in the study, with similar results. Clearly, providing dyad members with the correct answer to each item before moving on to the next one, and in so doing obliging them to recognize that their own *past* estimates (on average) had not been more "realistic" than those of their partners, did not prompt them to give substantially more weight to each other's *subsequent* estimates (see Table 2).

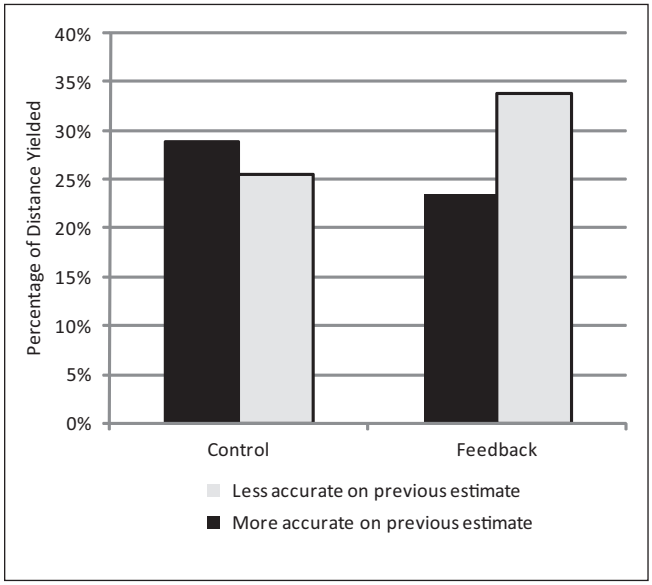

Figure 2. The percentage of distance yielded to partner by condition in cases when participant was more versus less accurate on the previous estimate (Study 2)

Is it possible that, accuracy incentives notwithstanding, participants simply failed to attend to the feedback being offered? A closer look suggests that this was not the case. When we contrasted the weight given to partner estimates by dyad members who were more versus less accurate than their partner on the previous estimation item (dummy coded as *less accurate* = 0, and *more accurate* = 1), hierarchical linear modeling revealed that there was a significant interaction effect between that variable and experimental condition (dummy coded as *Control* = 0, *Feedback* = 1), $B = –0.143$, $z = –2.81$, $p < .01$. Specifically, on items immediately following one on which their partner had proven to be more accurate, participants in the Feedback condition gave greater weight to their partner's estimates (M = 33.9%) than participants in the Control condition, whereas the reverse was true (M = 23.6%) when the participant's own estimate on the previous item was the more accurate one (Figure 2).

To address the influence of disagreement in initial estimates on weight given to partner's estimates, we repeated the analysis used in Study 1, adding experimental condition as a predictor variable. Again, we found that although overall, participants favored their own initial estimate over that of their dyad partner ($B = 1.01$, $z = 4.73$, $p < .001$), this tendency increased as a function of group-centered disagreement ($B = 1.81$, $z = 3.12$, $p = .003$), with no significant effect of average disagreement between two dyad members ($B = 2.25$, $z = 1.32$, *ns*), or

of experimental condition ($B = –0.080$, $z = –0.31$, *ns*). Thus, consistent with a naïve realism analysis, disagreement regarding the initial estimates exacerbated the underweighting bias across both conditions.

Round-to-round changes in estimation accuracy. Feedback led dyad members to give greater weight to their partner's estimate in cases when their partner's accuracy on the previous item had exceeded their own, and less weight to partner estimates when their own previous estimate had been the more accurate one. However, given that relative accuracy on a given item did not predict relative accuracy on the *next* item, no between-condition differences in accuracy improvement were observed in any of the four estimation rounds. Thus, in addressing round-to-round accuracy changes we combine data across conditions.

As in Study 1, dyad members' mean estimation errors decreased significantly from Round 1 ($M = 0.323$) to Round 2 ($M = 0.282$). This decrease ($M = 0.041$, $SD = 0.045$) yielded a highly significant t-value, $t(37) = 5.56$, $p < .001$, but was more modest than the decrease (0.064) that would have been achievable had the dyad partners simply averaged their initial estimates.

When in Round 3 dyad members were required to reach agreement through discussion, their mean estimation error decreased over the 10 items to 0.256. This further 0.026 reduction was statistically significant, $t(37) = 4.5$, $p < .001$, and slightly greater than the 0.023 reduction they could have achieved through averaging. Thus, as in Study 1, the accuracy of Round 3 joint estimates increased beyond that produced as a result of mere exposure to partners' estimates in Round 2.

Round 4. In arriving at their final individual estimates dyad members offered estimates that were significantly more accurate than those offered in Round 2 (M diff = 0.023, SD = 0.050), $t(75) = 4.07$, $p < .001$. As in Study 1, the requirement to reach agreement through discussion continued to reduce mean error even when the two dyad members once again offered individual estimates.

To address the role of initial disagreement on accuracy improvement we tested a hierarchical linear model using accuracy improvement from the second to the fourth round as the dependent variable and group-centered disagreement, average disagreement between two dyad members, experimental condition, and the interactions between experimental condition and the two disagreement variables as trial-level predictor variables. The analysis revealed a significant effect of group-centered disagreement on accuracy improvement ($B = 0.079$, $z = 2.37$, $p < .02$), with no significant main effects or interaction effects involving the other variables. Thus, as in Study 1 and in line with predictions based on naïve realism, it was the dyad members whose initial estimates had been furthest apart who proved least inclined to give each other's initial estimates due weight in Round 2. And it was these dyad members who benefitted most from having been required to reach agreement—even after they subsequently again became free to offer whatever individual estimates they wished.

Study 3: When Long-Term Partners Share a Source of Bias

Study 3 dealt with the effect of shared history of long-term collaboration on the ability and willingness of dyad members to benefit from each other's inputs. The benefits of such a shared history would seem to be obvious. Veterans collaborators can develop "transactive memory" (Wegner, 1986; Wegner, Giuliano, & Hertel, 1985). They can become able to anticipate what their partners are likely or unlikely to know and can develop mechanisms for effectively sharing and utilizing such knowledge. They may learn to recognize verbal and nonverbal cues signaling their partner's certainty or doubt and come to know how much weight those cues should be given. Longstanding partnerships in particular may afford partners a measure of psychological safety (Edmondson, 1999), giving them greater freedom in asking questions and admitting doubt or error. Finally, as contributors to the group dynamics tradition of an earlier age showed (see Cartwright & Zander, 1968), to the extent that collaborative experience builds cohesiveness, group members become less focused on individual concerns and more focused on joint goals.

However, partnerships in which the individuals have a history of working together toward shared goals in the face of shared norms and values might fall prey to a phenomenon akin to "groupthink" (Janis, 1972). That is, shared motives and perspectives may prevent individuals from expressing, and thus from examining the merits of, dissenting views—especially when the message conveyed by such views is an unwelcome one. As Surowiecki (2004) explains in *The Wisdom of Crowds*, for the crowd to be "wise" the individual judgments must be *independent.* If long-term experience leads partners to rely on similar information and to share sources of bias both in their private judgments and in their communications, such experience may diminish rather than increase the benefits partners derive from collaboration.

Participants in this study were competitive ballroom dancers estimating the scores that their video-recorded waltz performance would receive from professional evaluators. This task was akin to one they had engaged in countless times as they awaited their competition results, with one important exception: In the study the couples would receive scores on a series of suitably labeled absolute scales, rather than a simple ranking relative to other competitors.

Divergences in the two dancers' estimates thus would inevitably pit a more optimistic assessment against a more pessimistic one. To the extent that the dancers' initial estimates deviated from their actual scores in a random fashion, they could be expected to show high rates of bracketing and to benefit from movement toward each other's initial estimates. But, as noted earlier, if they *shared* a bias—in particular, a tendency to overestimate the scores that their performance would receive—they would improve their accuracy only to the extent that the less "realistic" partner disproportionately yielded to

the more realistic one. The literature on self-assessment (e.g., Dunning, Heath, & Suls, 2004; Epley & Dunning, 2000; Kruger & Dunning, 1999) and our informal observations led us to expect that such a shared bias would indeed preclude round-to-round improvements in accuracy.

To more specifically investigate the role of naïve realism in producing the underweighting of partner's initial estimates, and explore the extent to which this bias is present among long-term partners with extensive knowledge of both the subject matter and each other, a questionnaire was introduced immediately after the second round of estimates. This measure asked participants to indicate how similar they perceived their partner's assessments of their dancing to be to their own, and also to estimate the extent to which various factors—some reflective of knowledge and skill, others of bias and error—had affected those assessments (for prior investigations using similar measures, see Kennedy & Pronin, 2008; Liberman et al., 2011; Pronin et al., 2004).

Method

Participants. Participants were 18 couples ($N = 36$) of competitive ballroom dancers who had trained for a mean of 10.1 years. The mean duration of the partnerships at the time of the study was 6.6 years. They received no compensation for their efforts, although the opportunity to receive feedback from the professional judges provided a considerable incentive to participate. It should be noted that although the dancers paid no penalty for estimation error, the ability to accurately assess their performance was emphasized in the instructions as a skill critical to competition success.

Procedure. The experimenter met with the couples one at a time at a local dance studio. Each couple was video-recorded as they performed the waltz routine with which they normally opened their competitive program (typically 1.5 minutes in duration). They then watched the video and estimated how a panel of five judges—all retired world-class dancers and experienced adjudicators—would rate their performance on seven relevant dimensions. Each dimension was specifically defined for the participants (e.g., "Musicality refers to whether the dance was performed in time with the basic rhythm of the music and the extent to which movement expressed the mood and character of the music") Participants were told that judges would use a 100-point scale for each dimension, anchored at 1 = *extremely poor* and 100 = *typical of internationally known professional competitors*.

After viewing the video, the two dancers made individual estimates of the judges' ratings on each dimension and indicated their confidence that their estimate falls within 10 points of the judges' rating. They then exchanged those estimates and confidence ratings with their partners and, after producing a second set of individual estimates (but without viewing the Round 2 estimates made by their partners), completed the questionnaire designed to assess naïve realism. Finally, the couples engaged in discussion to produce a set of joint estimates for each of the seven dimensions. The fourth round of estimates was omitted to make time for the additional questionnaire.

Table 3. Relative Frequency of Round 2 Estimation Strategies (Study 3)

Estimation strategy	Average relative frequency (%)
Stood pat	45.2
Total less than halfway to partner	60.9
Exactly halfway to partner	13.4
All the way to partner	18.3
Total more than halfway to partner	25.7
Percentage of distance yielded	34.0

Results and Discussion

The influence of own and partner's initial estimates on second round estimates. We omitted from our analyses the 14 cases in which the two dancers made identical initial estimates on a given evaluation dimension. Overall, dyad members moved less than halfway toward their partner's estimate 60.9% of the time, and the mean percentage of distance they moved toward their partner's estimates was 34.0%—a distance again significantly short of 50%, $t(35) = 3.49$, $p < .002$ (see Table 3).

Interestingly, postexperimental questionnaire responses revealed a *negative* correlation between amount of yielding and both length of partnership, $r(16) = -.47$, $p < .05$, and the individual partner's years of competitive dancing, $r(34) = -.36$, $p < .03$. Regressing yielding on both years of experience and partnership duration, and grouping the data at the dyad level, we find that although experience remains a marginally significant predictor of underweighing ($B = -0.013$, $z = -1.77$, $p < .08$), partnership duration does not ($B = -0.010$, $z = -0.86$, *ns*). In other words, if experience makes one less willing to take advice, being long-term partners makes one no more so.

Round-to-round changes in estimation accuracy. The expert judges showed a high level of agreement in the scores that they assigned to the couples, averaged across the seven dimensions evaluated ($\alpha = .89$). When we examined the extent of agreement for the corresponding estimates initially offered by the two dyad members, we found a similarly high correlation, $r(16) = .77$, $p < .001$.

What was most notable about the respective assessments of judges and dancers, however, was the large and systematic *gap* between them—a gap reflecting the dancers' "wishful thinking" about the quality of their performance. Although the mean rating by judges was $M = 29.5$, the mean rating initially expected by the two dancers was $M = 50.6$. Thus, as we had anticipated, the dyad members shared an optimistic bias, with the more pessimistic partner's estimate ($M = 43.7$) tending to show smaller error and be more "realistic" than that of the more optimistic partner ($M = 57.6$), $t(17) = 3.83$, $p < .002$.

Thus, subsequent round-to-round improvement in accuracy required that the more optimistic (and almost always less realistic) dancer yield more to his or her less optimistic (and generally more realistic) partner than vice versa. No such net increase in accuracy occurred. The mean estimates over the three rounds were 50.6, 50.6, and 51.4, respectively, and the corresponding mean errors were 23.2, 22.7, and 23.2 points.

Clearly, long-term collaborative experience did not train dyad members to consistently discern more accurate judgments from less accurate ones. In contrast to previous findings, even the requirement to agree on a single estimate produced no round-to-round improvement in accuracy. Rather, what the data reveal is a kind of "dyad-think" whereby the partners shared an unwarranted positive bias about the quality of their performances—one that they did not overcome through access to, or even discussion of, each other's inputs.

Level of disagreement, attributions, and yielding. Although the overoptimism of dyad members limited the gains in accuracy attainable merely by giving weight to partner's inputs, consistent averaging would have resulted in estimates with average error ($M = 22.6$, $SD = 14.3$) that would have been significantly lower than the errors produced in either Round 1, $t(17) = 2.59$, $p < .02$, or Round 2, $t(17) = 2.28$, $p < .04$. Thus, there remains the question of why even these experienced collaborators, like the dyad members in earlier studies, gave so little weight to their partner's estimates. One reason, as suggested earlier, involves the presence, even among long-term, expert collaborators, of "naïve realism"—that is, the tendency to assume that to the extent that others fail to share one's own views, those others are in error (Pronin et al., 2004).

To address this prediction, we again tested a logistic hierarchical model using willingness to give equal or greater weight to partner's estimates as the dependent variable and group-centered disagreement, and the average disagreement for the dyad as predictors. As in the prior studies, the model revealed a significant effect of group-centered disagreement on participants' willingness to go at least halfway to their partner's estimate of their performance ($B = 0.061$, $z = 2.61$, $p < .01$). In this study the average dyad level of disagreement also turned out to be a significant predictor of underweighting ($B = 0.165$, $z = 2.84$, $p < .005$).

To directly address the role of naïve realism, we asked participants to rate the extent to which various factors had influenced their own and their partner's assessments. Some of these factors would be regarded as reasonable or normative (e.g., "in-depth knowledge of dancing technique") whereas others would be regarded as biases (e.g., "wishful thinking"). We then subtracted the average influence attributed to the biases (on scales anchored at 1 = *little if at all* and 4 = *very much*) from the average influence attributed to the valid reasons. Finally, we correlated this difference score (both for self and partner) with the level of disagreement dyad members perceived between their own and their partner's initial assessments about their performance.

The predicted negative relationship between perceived disagreement and the level of objectivity attributed to partner's estimates proved to be statistically significant, $r(34) = -.35$, $p < .04$. By contrast, as is apparent in Figure 3, no such relationship was found with respect to participants' attributions about their *own* performance assessments ($r = .10$). That is, greater perceived disagreement increased perceptions of bias in partner's but not in own estimates. Interestingly, participants judged *both* their own assessments and the partner's assessments to be most objective and free of bias when they found themselves in closest agreement.

Study 4: The Effect of Exchanging Questions Versus Giving Reasons

Our first three studies demonstrated dyad members' unwillingness to give weight to each other's judgments regarding a variety of topics, when given access to accuracy feedback, and in the face of long-term collaborative experience. Moreover, in line with naïve realism, participants were most inclined to underweight their partner's judgments in cases of high disagreement in initial estimates—precisely those cases in which this strategy is likely to be the mostly costly.

Study 4 was designed to examine a simple procedure for reducing the impact of naïve realism on dyad members' ability to benefit from each other's inputs. We predicted that *questioning* the other participant regarding the reasons and bases for their estimates would increase the weight participants gave to partner's estimates and increase the benefits of the subsequent discussion. We compared the results of this procedure with those obtained when, as is more commonly the case, decision makers simply list the bases or reasons for their own judgments.

Recent research by Chen, Minson, and Tormala (2010) demonstrated that asking and receiving "elaboration questions" in the context of contentious debate led to more positive interpersonal inferences and greater receptiveness to the other party's views than did the exchange of counterarguments. Also, earlier work done in the context of work on overconfidence in judgments (Heath & Gonzalez, 1995) and belief perseverance (Ross, Lepper, Strack, & Steinmetz, 1977) showed that discussion of a particular outcome or decision makes people more confident in it.

We propose that *asking questions* regarding the bases of another's estimates versus *explaining* the bases of one's own estimates would produce a similar difference in receptiveness. We hypothesized that asking questions about the reasons behind a peer's estimates reduces one's tendency to simply assume that those estimates are based on biased or erroneous considerations. By contrast, offering reasons for one's own estimate is apt to buttress one's certainty in the face of the differing estimates offered by one's partner.

Before turning to the specifics of the design, it is important to emphasize that participants asking questions did not receive *answers* to those questions. As such, dyad members

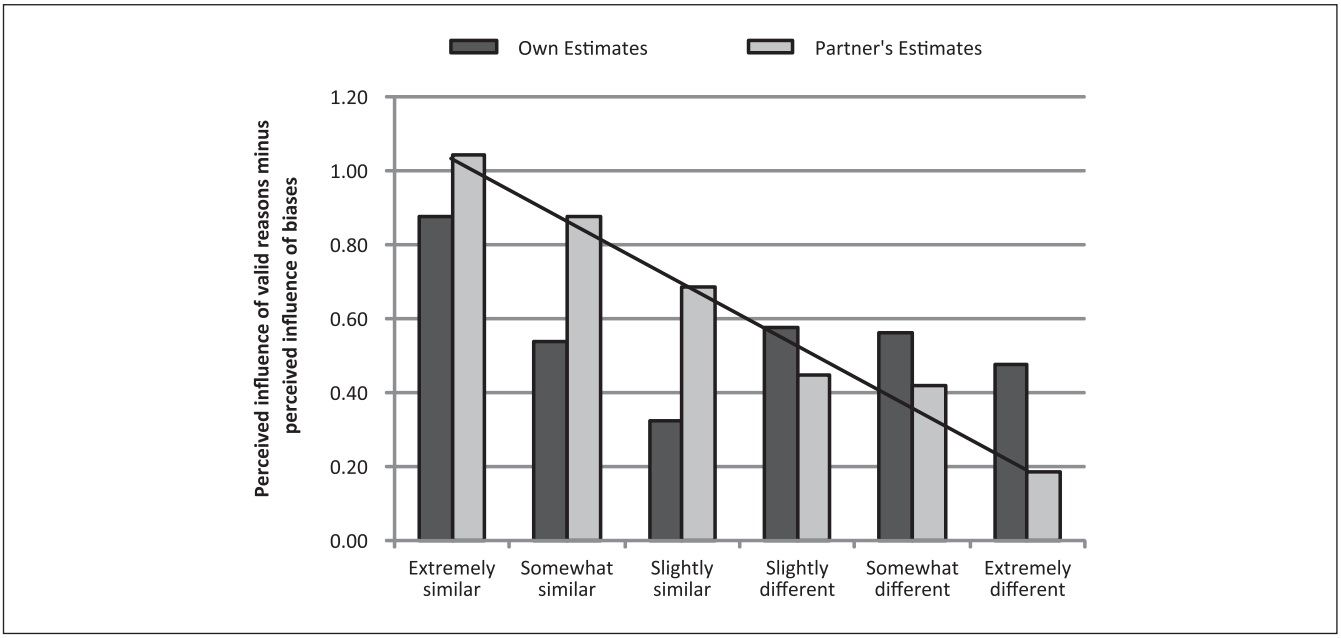

Figure 3. Attributions of own vs. partner objectivity by level of perceived disagreement (Study 3)

in the Questions condition of Study 4 actually ended up with *less* information on which to base assessments about the potential accuracy of their own versus their partners' estimates than dyads in the Reasons condition.

Method

A total of 122 participants (47 men and 75 women; all members of a paid participant pool at a large East Coast U.S. university) working in dyads made a series of individual and "joint" numerical estimates following the procedure described in Studies 1 and 2. These estimates pertained to the percentage of individuals in the same participant pool who, in a recent pretest survey, reported holding particular political views—for example, a belief that the words "under God" should be removed from the U.S. Pledge of Allegiance, or a belief that people of the same sex should be allowed to marry. Participants were paid $10 for their efforts and were offered a starting bonus of $30 in each round, which was then reduced by $1 for each percentage point by which their estimates deviated from the correct answer.

After responding to the political belief items, making an initial set of individual estimates regarding peer consensus, and exchanging both sets of responses with their dyad partner, participants in the *Questions* condition (N = 62) were instructed to "write down *three questions* [for the other student] that would help you better understand why they made the estimates that they did." The participants in the *Reasons* condition (N = 60) were instructed to "write down *three reasons* behind your own estimates that would explain to the other student why you made the estimates that you did." All participants were told that their questions or reasons could pertain to all, or any subset of, the estimates. After exchanging the written questions or reasons, all dyad members proceeded to make a second round of individual estimates, a third round of joint estimates reached through discussion, and a final set of individual estimates.

Results and Discussion

The influence of own and partner's initial estimates on second round estimates. After omitting the 74 cases in which dyad members offered identical Round 1 estimates, examination of the remaining estimates revealed a pattern of small but consistent between-condition differences in participants' adjustment strategies. Participants who had posed questions to their dyad partners tended to give their partner's initial estimates more weight in revising their own prior estimates than participants who had given reasons for their estimates.

We tested a logistic hierarchical model using the dyad members' willingness or unwillingness to move at least halfway toward their partner's estimates as a dichotomous trial-level dependent variable, and experimental condition (dummy coded as *Reasons* = 0, *Questions* = 1), question number, and the relevant interaction as trial-level predictors. Our test of this model again revealed evidence of the general tendency for dyad members to move less than halfway toward partner's estimates, $B = 0.915$, $z = 4.77$, $p < .001$. But it also revealed a significant main effect of condition, $B = -0.536$, $z = -1.98$, $p < .05$, indicating that, as predicted, dyad members gave more weight to their partner's initial estimates after posing questions to their partners than after furnishing their partners with reasons for their own estimates. Finally, the test yielded a statistically (and theoretically) significant interaction effect between experimental condition and item order. The effect of the relevant experimental

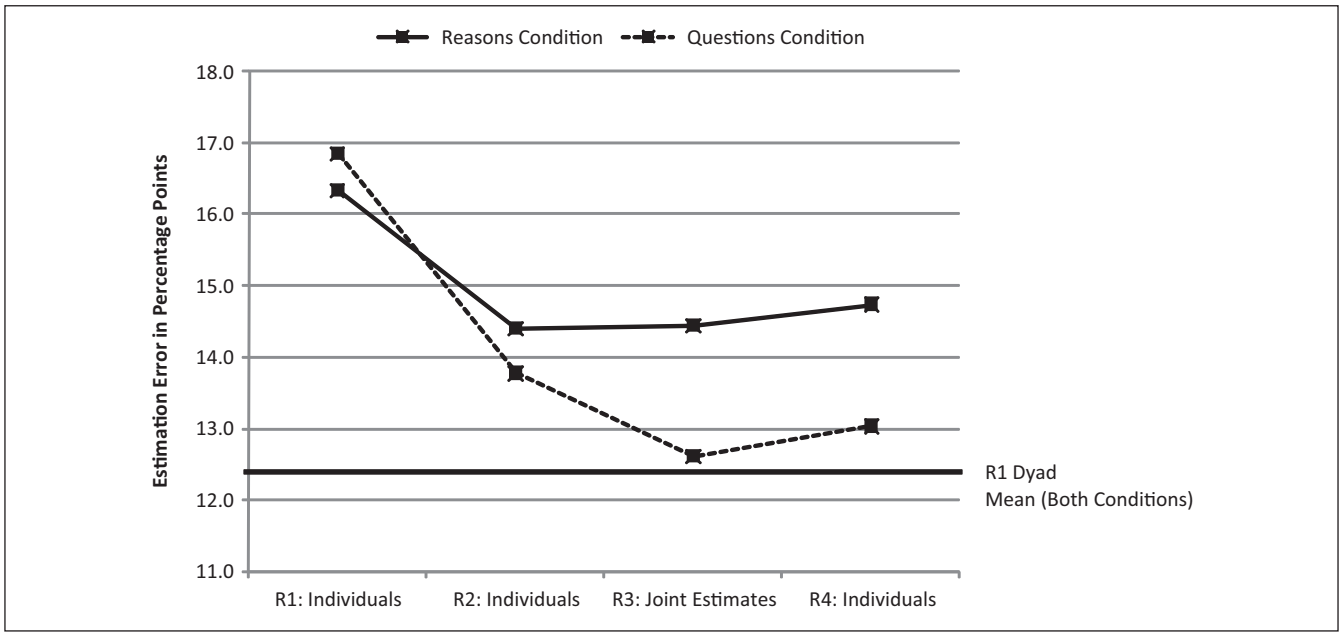

Figure 4. Round-by-round mean estimation error of dyad members' estimates in percentage points (Study 4)

Table 4. Relative Frequency of Round 2 Estimation Strategies (Study 4)

Estimation strategy	Reasons condition (%)	Questions condition (%)
Stood pat	31.1	25.4
Total less than halfway to partner	67.8	63.5
Exactly halfway	7.3	10.0
All the way to partner	9.0	7.4
Total more than halfway to partner	24.8	26.4
Percentage of distance yielded	34.3	36.2

manipulation was stronger for the earlier items (placed sooner after the manipulation) than the later items, $B = 0.079$, $z = 2.00$, $p < .05$ (Table 4).

Round-to-round changes in estimation accuracy. Examination of round-to-round changes in accuracy revealed an interesting, although somewhat complicated, pattern of results. Dyads in both the Reasons condition and the Questions condition significantly improved their mean accuracy from Round 1 to Round 2 (see Figure 4). The *difference* in the magnitude of that improvement between the two conditions did not reach significance ($B = 1.14$, $z = 1.60$, $p = 0.11$). However, when we distinguish cases in which the participants moved less than halfway toward their partner's Round 1 estimates from cases in which they moved halfway or more, and include that variable in a hierarchical linear model, we find a significant effect of condition ($B = 3.11$, $z = 2.77$, $p < .006$) on amount of accuracy improvement from the first to the second round. This analysis also revealed a significant interaction effect ($B = -3.10$, $z = -2.42$, $p < .02$) whereby dyad members who went at least halfway toward their partner's estimates in the Questions condition improved their accuracy, whereas those in the Reasons condition did not.

This interaction suggests that dyad members in the Questions condition were better able to identify and give due weight to the more accurate of the two estimates than those in the Reasons condition. To pursue this reasoning, we distinguished cases in which the participant's own estimate was closer to the correct answer on a particular item from cases in which the partner's estimate was closer. We then tested a logistic model using this dichotomous variable, condition, and the interaction between them as independent variables and the frequency of going at least halfway to partner as a dichotomous dependent variable (see Figure 5).

The test of this model revealed a significant effect of condition ($B = -0.425$, $z = -2.09$, $p < .04$) and a significant interaction between condition and whether the participant's own or the partner's estimate was more accurate ($B = 0.584$, $z = 2.41$, $p < .02$). This result indicates that dyads in the Questions condition did indeed make the more judicious weighting decisions. Interestingly, there was no significant main effect of relative accuracy on frequency of moving halfway to partner's estimate. Thus, overall, willingness versus unwillingness to give equal weight to partner's estimates did not reflect the relative accuracy of the two estimates.

Examination of changes in accuracy from Round 2 to Round 3 allows us to tell a simpler story. Overall, participants in the Questions condition showed a further improvement in their mean accuracy following their discussion to reach a joint estimate $t(29) = 3.00$, $p < .006$, whereas those in the Reasons condition did not, $t(30) = .08$, *ns*. This between-condition

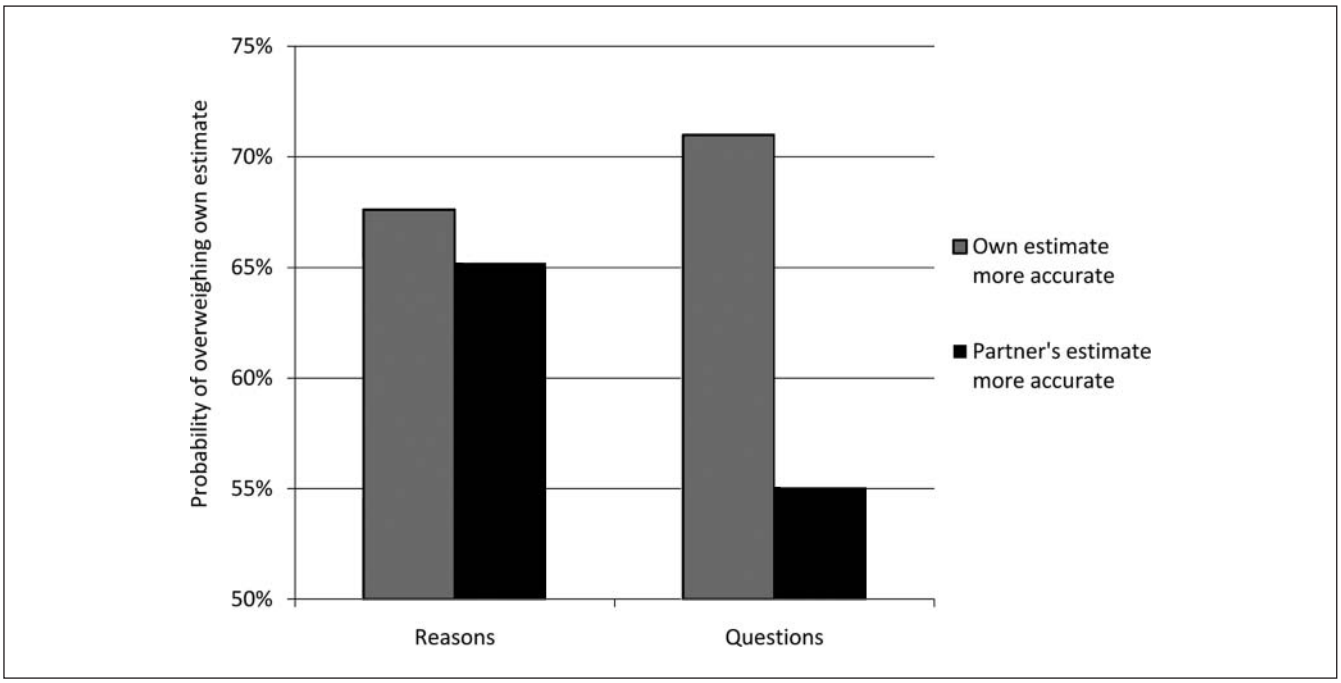

Figure 5. Probability of overweighting own estimate relative to that of partner as a function of condition and relative accuracy of the two estimates (Study 4)

difference in improvement from Round 2 to Round 3 proved to be statistically significant, $B = 1.203$, $z = 1.98$, $p < .05$. Given that participants in our earlier studies showed significant improvement from Round 2 to Round 3, the lack of improvement shown by the dyad members in the Reasons condition prompts a potentially important conjecture. Having given reasons—or one might say *justifications*—for their inputs, individuals may engage in less fruitful subsequent discussion, even when they are obliged to reach agreement.

Examination of fourth round estimates shows that dyad members in both conditions showed a slight, but not significant, decrease in accuracy in that final round of individual estimates. When we control for initial accuracy, dyad members in the Questions condition ended up making more accurate estimates than those in the Reasons condition ($B = –1.70$, $z = –2.56$, $p < .01$). Also, although dyad members in the Questions condition showed a significant decrease in mean error from Round 2 to Round 4, $t(29) = 2.12$, $p < .05$, those in the Reasons condition did not ($t < 1$ in the Reasons condition).

General Discussion

Four studies examined the ability of dyad partners to increase their estimation accuracy through access to each other's input using a variety of items, informational contexts, and levels of collaborative experience. These studies also demonstrated the role of naïve realism in this process and provided evidence for the impact of this bias. In all four studies dyad members tended to give their partner's initial assessments less weight than their own, a tendency that on average has obvious costs in terms of accuracy. In line with naïve realism predictions, this tendency was more pronounced when the initial disagreement was relatively large. Participants underweighted partner input the most in the very cases in which doing so was apt to be the most costly because of increased probability of bracketing.

It is worth pausing to consider that logically, disagreement in initial estimates could have led to at least three possible inferences about the relative merits of own and a dyad partner's estimates. Disagreement might have compelled our participants to question their *own* judgment, or simply to assume that their estimates and those of their partner are based on different, but equally persuasive pieces of evidence. Instead, our data suggest that participants gravitated toward the third inference: namely, that of the two divergent estimates their own was the more accurate and more soundly reasoned.

Studies 1, 2, and 4 showed that the requirement to reach agreement with a dyad partner increases estimation accuracy above and beyond the gains achieved through simple exposure to a partner's estimate, and that these benefits persist even when individuals subsequently make independent estimates. Study 1 clarified this finding by demonstrating that although discussion produces larger gains in accuracy than the mere exchanging of "bids," even the latter produces gains over individual estimates. Study 2 demonstrated the limited impact of accuracy feedback both on the weight that dyad members give to each other's estimates and on the gains in accuracy they achieve over successive rounds of estimates.

Study 3 showed that long-term partners (i.e., competitive ballroom dancers estimating their performance scores) showed

no round-to-round improvement in accuracy, despite the fact that consistent averaging would have led to such improvement. Access to the judgments of a peer, even the requirement to discuss and reach agreement, this study shows, does not attenuate the effects of naïve realism and its role in underweighting peer input, when the individuals in question share the same bias—in this case the tendency to overestimate the quality of their performances. But this study also added to the body of direct evidence of the role of naive realism by showing that it makes its influence felt even in the case of long-term collaborators making judgments in a domain in which they have considerable expertise. That is, the dancers rated their own individual estimates as more "realistic"—that is, based more on valid considerations and less reflective of various biases—than those of their partners.

Study 4 tested a manipulation designed to overcome the tendency for dyad members to attribute other's estimates to bias and error and thus not reap the full benefit of such input. Dyad members who asked each other questions about the bases of their estimates gave greater weight to their partner's estimates and achieved greater round-to-round improvement in accuracy than did those who gave reasons for their own estimates.

Practical and Theoretical Implications

Our studies attest to the benefit that individuals making quantitative assessments derive from mere exposure to the estimates furnished by a peer—provided that he or she has made an independent assessment and brings different sources of accuracy and bias (in particular, self-serving biases) to the task. More importantly, our studies document the *incremental* value of having to reach agreement (by deliberative discussion or, to a lesser extent, even the mere exchanging of bids) in the face of initial disagreement, especially large disagreements of the sort that generally make people dismissive of each other's views. Moreover, our studies also show that these benefits persist even when dyad members are free again to give each other's prior inputs as much or as little weight as they wish in a final set of individual estimates.

Our Study 1 findings regarding the value of discussion over that of mere bidding, and the findings that such discussion was not helpful in Study 3 or the Reasons condition of Study 4, prompt us to comment further regarding the discussion process itself. Dyad members in our studies improved the accuracy of their estimates significantly through such discussion—presumably because at least on some items it helped them to give greater weight to the more soundly based estimate. Nevertheless, the magnitude of their errors remained notably greater than would have been the case had they consistently picked the better of the two estimates (a benchmark that Soll & Larrick, 2009, refer to as "perfect choosing").

Identifying conditions that further increase the effectiveness of discussion as a vehicle for arriving at more optimal weighting of inputs would of course be of considerable theoretical and applied interest. The role of bracketing in moderating the benefits of discussion (vs. reaching agreement with no discussion) speaks to related investigations of group decision making (e.g., Gigone & Hastie, 1997). This finding suggests that both advice-taking and group decision-making research may benefit from more closely examining the statistical features of the decision-making environment for clues regarding the likely success of various decision-making strategies. In particular, the benefits of simple averaging are bound to be relatively greater when there is reason to believe that sources of estimation error are many, relatively small, and varied in direction, such that the average estimates of all making them approach the correct answer.

The studies presented here thus serve to focus attention on the relationship between a statistical truism and the benefits versus costs of diversity in forming small work groups, *especially* diversity of the sort that produces sharp disagreement. All things being equal, the frequency of "bracketing" estimates will be highest, and the benefits to be achieved from simple averaging will be greatest, when partners come to their collaboration from different backgrounds or schools of thought, and thus bring to bear different assumptions and models. But to the extent that diversities of bases for assessment are also sources of high disagreement and negative attributions about the other party, mere exposure to the views of the other—especially in the absence of a requirement to converge on a shared judgment through discussion—may prompt the parties to give each other's inputs less weight than would be optimal.

Furthermore, a long history of prior collaboration can be a mixed blessing. In our study of ballroom dancers such history neither prompted individuals to give each other's estimates due weight nor helped them recognize which of their estimates was likely to be more "realistic." But even if these dyad members *had* given each other's inputs greater weight it would have resulted in minimal (albeit statistically significant) improvement insofar as they shared a predictable tendency to overestimate the quality of their performances, and no tendency for the more optimistic partner to yield more than the less optimistic one.

The lesson presented by our present findings speaks to the value of environments and specific processes that encourage or even compel compromise between decision makers who make different estimates about factual matters or about the nature of public beliefs. The fact that dyad members forced to reach agreement on a single estimate showed continuing benefits in terms of accuracy even when subsequently offering their own individual assessments, especially in the face of large initial disagreement, is particularly noteworthy in this regard. Our Study 4 results offer the provocative suggestion that the benefits of dyadic interaction are enhanced when individuals refrain from the standard and seeming logical process of justifying their own views, in favor of asking simple questions about the basis of their partner's views.

The present studies, and in particular our findings regarding the attributions made for disagreement, resonate with prior research on the benefits of task conflict and the costs of relationship conflict in organizational units (DeDreu & Weingart, 2003; Jehn, Northcraft, & Neale, 1999). Although further research is necessary to clearly understand the relationships between these constructs, negative attributions for disagreement of the sort we have documented may be the "missing link" that changes beneficial task conflict into unproductive and damaging relationship conflict. Such attributions may also take us some distance toward explaining the failure of teams to maximize the benefits of diversity documented by many researchers (see Mannix & Neale, 2005, for a recent review). But in any case, the recognition that disagreement, especially when it is great, is both a barrier and a resource to effective partnership promises to be a continuing impetus for research.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received financial support for the research, authorship, and/or publication of this article: This research was supported in part by a grant of the Binational Science Foundation to Varda Liberman and Lee Ross.

Notes

1. For a list of items used in Studies 1–4, please email jminson@wharton.upenn.edu.
2. In all of our studies, movement in the *opposite* direction to a partner's estimates was treated as 0% yielding and movement *beyond* a partner's estimates was treated as 100% yielding.
3. Extent of disagreement was calculated as the percentage of the range of all estimates given for an item by which estimates of any two dyad partners differed.
4. Round-to-round changes over the first three rounds were calculated at the dyad level so we could directly compare Round 3 accuracy of joint agreement with the average Round 1 and Round 2 accuracy of the two dyad members.
5. When Larrick and Soll (2006) asked questions about the best strategy for hypothetical dyad members to adopt under various circumstances, a substantial proportion of respondents did recommend simple averaging—especially when bracketing rates were high and the estimates were presented in pairs along with correct answers in a manner that made the high bracketing rates apparent.

References

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.

Cartwright, D., & Zander, A. (1968). *Group dynamics: Research and theory*. London, UK: Tavistock.

Chen, F. S., Minson, J. A., & Tormala, Z. L. (2010). Tell me more: The effects of expressed interest on receptiveness during dialogue. *Journal of Experimental Social Psychology*, 46, 850-853.

Dawes, R. M. (1979). The robust beauty of improper linear models. *American Psychologist*, *34*, 571-582.

DeDreu, C. K. W., & Weingart, L. R. (2003). Task versus relationship conflict, team performance, and team member satisfaction: A meta-analysis. *Journal of Applied Psychology*, *88*, 741-749.

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workforce. *Psychological Science in the Public Interest*, *5*, 69-106.

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, *44*, 350-383.

Epley, N., & Dunning, D. (2000). The mixed blessings of self-knowledge in behavioral prediction: Enhanced discrimination but exacerbated bias. *Personality and Social Psychology Bulletin*, *32*, 641-655.

Gigone, D., & Hastie, R. (1997). Proper analysis of the accuracy of group judgments. *Psychological Bulletin*, *121*, 149-167.

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

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

Heath, C., & Gonzalez, R. (1995). Interaction with others increases decision confidence but not decision quality: Evidence against information correction views of interactive decision making. *Organizational Behavior and Human Decision Processes*, *61*, 305-326.

Janis, I. (1972). *Victims of groupthink*. Boston, MA: Houghton Mifflin.

Jehn, K. A., Northcraft, G. B., & Neale, M. A. (1999). Why differences make a difference: A field study of diversity, conflict and performance in workgroups. *Administrative Science Quarterly*, *44*, 741-763.

Kennedy, K. A., & Pronin, E. (2008). When disagreement gets ugly: Perceptions of bias and the escalation of conflict. *Personality and Social Psychology Bulletin*, *34*, 833-848.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*, 1121-1134.

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

Liberman, V., Minson, J. A., Bryan, C. J., & Ross, L. (2011). *"Be reasonable, see it my way": Naïve realism as a barrier to capturing "the wisdom of dyads*." Manuscript submitted for publication.

Lorge, I., Fox, D., Davitz, J., & Brenner, M. (1958). A survey of studies contrasting the quality of group performance and individual performance 1920–1957. *Psychological Bulletin*, *55*, 337-372.

Mannix, E., & Neale, M. A. (2005). What differences make a difference? The promise and reality of diverse teams in organizations. *Psychological Science in the Public Interest*, *6*, 31-55.

Pronin, E., Gilovich, T., & Ross, L. (2004). Objectivity in the eye of the beholder: Divergent perceptions of bias in self versus others. *Psychological Review*, *111*, 781-799.

Ross, L. D., Greene, D., & House, P. (1977). The "false consensus effect": An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology*, *13*, 279-301.

Ross, L., Lepper, M., Strack, F., & Steinmetz, J. (1977). Social explanation and social expectation: Effects of real and hypothetical explanations on subjective likelihood. *Journal of Personality and Social Psychology*, *35*, 817-829.

Ross, L., & Ward, A. (1995). Psychological barriers to dispute resolution. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 27, pp. 255-304). San Diego, CA: Academic Press.

Ross, L., & Ward, A. (1996). Naive realism in everyday life: Implications for social conflict and misunderstanding. In E. S. Reed, E. Turiel, & T. Brown (Eds.), *Values and knowledge: The Jean Piaget symposium series* (pp. 103-135). Hillsdale, NJ: Lawrence Erlbaum.

Soll, J. B., & Larrick, R. P. (2009). Strategies for revising judgment. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *35*, 780-805.

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*. New York, NY: Doubleday.

Vul, E., & Pashler, H. (2008). Measuring the crowd within: Probabilistic representations within individuals. *Psychological Science*, *17*, 645-647.

Wegner, D. (1986). Transactive memory: A contemporary analysis of the group mind. In B. Mullen & G. R. Goethals (Eds.) *Theories of Group Behavior* (pp. 185-208). New York: Springer-Verlag.

Wegner, D. M., Giuliano, T., & Hertel, P. (1985). Cognitive interdependence in close relationships. In W. J. Ickes (Ed.), *Compatible and incompatible relationships* (pp. 253-276). New York: Springer-Verlag.

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.