Journal of Behavioral Decision Making

WILEY

RESEARCH ARTICLE OPEN ACCESS

Learning Processes in the Judge–Advisor System: A Neglected Advantage of Advice Taking

Thomas Schultze[1,2] | Alexander Stern[3] | Stefan Schulz-Hardt[3]

[1]Institute of Psychology, University of Bamberg, Bamberg, Germany | [2]School of Psychology, Queen's University Belfast, Belfast, UK | [3]Institute of Psychology, University of Göttingen, Göttingen, Germany

Correspondence: Thomas Schultze (thomas.schultze-gerlach@uni-bamberg.de)

Received: 28 June 2024 | **Revised:** 16 May 2025 | **Accepted:** 24 June 2025

Keywords: advice taking | decision making | estimation accuracy | judgment | learning

ABSTRACT

Previous research in the judge–advisor paradigm has focused on how judges utilize the wisdom of others by taking their advice and on the beneficial effect of receiving advice on judges' postadvice final judgments about the exact same problem. However, a completely different possibility of how judges might benefit from advice has been overlooked so far: Learning processes could improve the accuracy of judges' subsequent *initial* judgments from one problem to another problem on the same type of task as well. Hence, we test the assumption that advice can induce individual performance enhancements that differ as a function of the advisor's judgment accuracy. The results of three experiments support our hypothesis and indicate positive learning, particularly when participants receive high-quality advice. Furthermore, we show that external information about the advisor's accuracy is not crucial for the occurrence of these individual performance enhancements. In general, our results suggest that advice can have a positive effect on judges' subsequent initial judgments.

In recent years, social and organizational psychological research has increasingly studied advice taking (e.g., Soll and Larrick 2009; Yaniv and Kleinberger 2000; for reviews, see Bonaccio and Dalal 2006; Rader et al. 2017; Yaniv 2004a). This topic is predominantly investigated in the *judge–advisor system* (JAS; Sniezek and Buckley 1995). The JAS differentiates between an advisor, who provides information or recommendations, and a judge, who is responsible for the judgment. The judge first makes an initial estimate, receives a recommendation by the advisor, and then makes a final, possibly revised, estimate (e.g., Sniezek et al. 2004; Yaniv 2004b). In this final estimate, the judge combines the initial judgment with the advice—which includes the judge retaining the initial judgment or completely adopting the advice. The majority of JAS studies focus on advice weighting and judgment accuracy, with three particularly pronounced and very robust findings: First, judges are sensitive to various cues of advice quality, leading them to heed better advice more (e.g., Harvey and Fischer 1997; Soll and Larrick 2009; Yaniv and Kleinberger 2000); second, judges overweigh their own opinion compared with the advice, a phenomenon called *egocentric advice discounting* (e.g., Yaniv 2004b; Yaniv and Kleinberger 2000); finally, when the advisor provides an independent benevolent opinion, taking advice increases accuracy (e.g., Soll and Larrick 2009; Sniezek et al. 2004). The reason is that aggregating independent opinions reduces unsystematic or even systematic errors (Soll and Larrick 2009; Yaniv 2004a).

Despite the JAS literature focusing on advice taking and post-advice accuracy, there is also research looking at other effects of advice like sharing responsibility, minimizing effort, or validating judges' initial opinions (Bonaccio and Van Swol 2014). Here, we examine another possible function of advice that has received little attention so far, namely, the opportunity to systematically learn from the advisor. In contrast to previous research

Thomas Schultze and Alexander Stern contributed equally; they appear in alphabetical order.

investigating improvements from judges' initial to their *final* estimates, we aim to focus on changes in the accuracy of judges' *initial* estimates (after having received advice on previous occasions). We argue that, beyond improved *final* judgments, judges' ability to come up with accurate *initial* judgments can improve, too, because of vicarious learning.

1 | Learning Processes in the JAS

Previous research on group judgment using the Delphi technique has already demonstrated that awareness of each other's judgments leads group members to adjust and improve their own estimates (see Rowe and Wright 1999 for a review). Just as in the JAS, group members in Delphi groups are not allowed to communicate directly but only learn what each group member estimated individually after which they can revise their own judgments (this process is then repeated until the group reaches a consensus). Not surprisingly, then, there is ample evidence that advice can improve the quality of subsequent judgments or decisions about the same issue (e.g., Biele et al. 2009; Çelen et al. 2010; Chaudhuri et al. 2006; Kocher et al. 2014). For example, Biele et al. (2009) showed that one-time advice improves individual performance on multiple trials of a multiarmed bandit game beyond mere practice effects. Using a similar task, Hertz et al. (2021) further showed that this performance increase is partly specific to advice as being advised to choose the best option produced stronger and more rapid performance increases than simply observing another actor choose that option. In line with these findings, advice taking in JAS experiments can be interpreted as a form of learning as well. Advice serves as a form of feedback and leads to reconsideration of one's initial judgment, which, in turn, can improve subsequent final judgments (e.g., Farrell 2011; Soll and Larrick 2009; Sniezek et al. 2004). Hence, when using JAS terminology, the performance improvements from initial to final judgments represent the beneficial effect of advice weighting.

However—and this is what we are interested in—advice could also initiate a *transfer* of knowledge from one problem to another problem within the same domain. For example, receiving advice when being asked to judge the distance between London and Rome might not only be beneficial when revising an initial estimate of this exact distance but also when subsequently estimating the distance between Paris and Madrid. In other words, advice might not only affect judges' postadvice final judgments but also improve the accuracy of their subsequent preadvice initial estimates on different tasks of the same type. But why should advice allow for such learning processes?

Previous research on quantitative judgments in groups suggests that frames of reference play an important role when it comes to increases in estimation accuracy (e.g., Bonner and Baumann 2008; Bonner et al. 2007; Laughlin, Bonner, et al. 1999; Laughlin, Gonzalez, et al. 2003). For example, knowledge about the length of Germany from north to south (approx. 900 km) and from east to west (approx. 600 km) should improve accuracy when estimating distances within Germany and should prevent completely implausible judgments. Obviously, these benchmarks can be illustrated easily when people are allowed to communicate and can argue for a particular reference value. For example, there is evidence for *group-to-individual transfer* (G-I transfer) on quantitative estimation tasks similar to those frequently used in the JAS (e.g., Lippold et al. 2021; Schultze, Mojzisch, et al. 2012; Stern et al. 2017). G-I transfer denotes an increase in individual accuracy due to collectively working on—and discussing—the task. Advice in typical JAS experiments might also serve as a frame of reference and lead to similar increases in judgment accuracy. Judges might recognize systematic differences between their own initial estimates and the advice. For example, when a judge's estimates are frequently above (or below) the advice, a process of recalibration might be triggered that leads to lower (or higher) subsequent judgments. On average, such recalibration should lead to more accurate subsequent judgments.

2 | The Moderating Role of Advice Quality

Advice quality should be an important moderator of the learning effects outlined above, as behaviors that seem to be effective for others are favored over behaviors that produce negative outcomes (Bandura 1986). In the context of the JAS, observers should be more willing to learn from a high-performing than from a poorly performing advisor. Previous research found that judges can distinguish between very good and very bad advice even in the absence of any information about advice quality, arguably because they can recognize poor advice as implausible (Yaniv and Kleinberger 2000). However, judges' ability to infer the quality of advice increases with the availability of valid cues such as feedback about past performance (Yaniv and Kleinberger 2000; Soll and Larrick 2009). Therefore, we expect that more accurate advice enables stronger improvements in judges' initial judgment accuracy, particularly when the quality of advice is salient.

2.1 | The Present Research

In three experiments, we investigate learning in a prototypical JAS. We expect that advice quality affects not only postadvice final judgments, as shown in many previous studies, but also the accuracy of subsequent initial (i.e., preadvice) judgments. Hence, our first aim is to clarify whether receiving advice of different quality systematically changes the accuracy of judges' initial estimations on subsequent trials. Because better advisors are more accurate and judges are potentially more willing to learn from them, we postulate

Hypothesis 1. *Judges' initial judgment accuracy will improve dependent on advice quality. Judges' estimation accuracy will increase more strongly when receiving high-quality advice compared with low-quality advice.*

As stated before, information about advice quality should influence the differentiation of advice quality and, thereby, the strength of learning. However, even in the absence of this information, judges are likely sensitive to advice quality and should benefit more from better-calibrated advisors, which should produce an increase in initial accuracy as well. Hence, we hypothesize

Hypothesis 2. *The accuracy of judges' initial estimates after receiving advice differs as a function of advice quality with or without information about the quality of advice. However, the effect of advice quality is stronger if information about advice quality is given.*

If Hypothesis 1 receives support, we can conclude that the already well-known improvements in the accuracy of judges' *final* estimates after receiving advice are, at least partially, the result of learning from the advice, which is reflected in improved *initial* estimates. One important question concerns the relative importance of these learning effects, on the one hand, and the statistical benefits of aggregating one's initial estimate and the advice, on the other. Since we did not have any a priori hypotheses about the relative importance of the two beneficial processes, we address this open question in an exploratory fashion.

3 | Experiment 1

In Experiment 1, we tested whether receiving advice improves subsequent initial estimates and whether the advisor's accuracy moderates the extent of these learning effects (Hypothesis 1). Previous studies on advice taking either had judges receive advice from the same (e.g., Soll and Mannes 2011; Minson and Mueller 2012) or from different advisors over the course of the respective study (e.g., Harvey and Fischer 1997; Schultze et al. 2015). Since we do not consider one of the two approaches superior to the other, we applied both of them, allowing us to explore potential differences in learning. Particularly, judges might find it easier to detect systematic deviations between initial estimates and advice when the advisor remains the same across all trials, but being stuck with one advisor means that the potential for learning is limited by that advisor's accuracy. If the advisor's accuracy is low, there might even be the risk of negative learning if the judge adopts the advisor's estimation tendency. In addition, if confronted with the same advisor on each trial, judges might (falsely) attribute systematic differences in opinions to the advisor being biased (Pronin et al. 2004), rendering learning from the advisor unnecessary from the judges' point of view.

In the case of varying advisors, judges cannot infer systematic discrepancies between their own estimates and those of a specific advisor as easily, making learning more difficult. Instead, they need to integrate information over several advisors, for example, by inferring their central tendency. Although this makes the situation with varying advisors more challenging, it also holds the potential for improved learning because the central tendency of a crowd of advisors should be more accurate than a single advisor. Having varying advisors could also foster learning because it helps judges recognize (and correct) their own idiosyncratic biases. In particular, it is more difficult to attribute systematic differences between one's initial estimates and the advice to the advisor (rather than the judge) being biased when advisors differ between trials. We compared the two advice conditions with a control condition, in which participants received no advice at all. This procedure allowed us to control for practice effects and, thus, to attribute stronger increases in initial accuracy in the two advice conditions unequivocally to learning from the advisors.

3.1 | Method

3.1.1 | Participants and Design

The sample size in Experiment 1 and in the following experiments was determined based on a rule of thumb (at least 30 participants per condition to ensure approximate normal distribution of cell means) as well as resource considerations. One hundred and ninety-seven students participated in the experiment. Eight participants were excluded from all analyses: Six reported that they had previously participated in a JAS study and, thus, were familiar with the estimation task, and two were excluded because their initial estimates were unreasonably high (they overestimated the true values by more than 2300%). Of the remaining 189 participants, 133 reported their gender as female, 61 as male, and 3 did not report their gender. Their average age was 23.61 years (SD = 5.15). Experiment 1 is based on a one-factorial design with advisor (same advisor vs. varying advisor vs. no advice) as a between-subjects factor.

3.1.2 | Task and Procedure

In each experimental session, up to 12 people participated. They were seated at separate computers and were informed about the task and the procedure of the experiment. Without any time restrictions, participants had to estimate airline distances between different European capital cities in kilometers, a task where prior studies successfully showed learning in interacting groups (Schultze, Mojzisch, et al. 2012; Stern et al. 2017). The estimates of participants from a pretest ($N = 76$) served as advice in our first experiment. Participants were randomly assigned to one of the three conditions. The experiment consisted of two phases: an individual practice phase with 10 trials and a test phase with 20 trials. The practice phase was identical in all three conditions. Participants first made an initial estimate and indicated their confidence in the accuracy of this estimate on a 7-point scale from 1 (not at all confident) to 7 (very confident). Subsequently, they were instructed to think about their initial estimate and then make a second—and possibly revised—estimate (the final estimate) along with another confidence rating. Participants' accuracy during the practice phase served as a performance baseline.

The test phase differed between conditions. All participants made one initial and one final estimate per trial, along with confidence ratings. However, participants in the two advice conditions were informed that they would receive advice in the form of the estimate of a previous participant (labeled as their advisor) on each trial prior to making the final estimate. In the condition in which the advisor remained the same across all trials, each participant received advice from one specific, randomly drawn pretest participant. In the varying advisor condition, a new advisor was randomly drawn for each trial (with replacement). In both conditions, participants were fully informed about the drawing procedure at the beginning of the experiment, that is, they knew whether they were dealing with a single advisor or with varying advisors. Participants received no advice on quality information in Experiment 1. Participants in the control condition received no advice.

Having participants make one initial and one final estimate in both phases and in all conditions avoids a confound between phases or conditions, on the one hand, and the number of judgments, on the other. All participants were presented the same 30 distances, but their sequence was randomized. Finally, we asked participants to report whether they received advice and, if they did, whether it came from the same advisor or varying advisors as an awareness check for the experimental manipulation. Upon completing the experiment, participants were thanked, debriefed, and paid a compensation of €5.

3.2 | Results and Discussion

3.2.1 | Comprehension Check and Check for Possible Interfering Effects

Prior to the main analyses, we checked whether our advice manipulation was successful. In the no-advice condition, all participants correctly recalled that they did not receive advice. In the same and varying advisor conditions, the majority of participants also correctly recalled their experimental condition (84% and 86%, respectively). Accordingly, there was a significant association between the true and recalled experimental condition, $\chi^2(4) = 277.74$, $p < 0.001$.

As outlined before, advice quality should be a crucial moderator for the occurrence of learning effects. Hence, we ran some preliminary analyses. First, to rule out baseline performance differences, we checked whether judges' initial accuracy was similar in all three conditions. To this end, we calculated the mean absolute percent error (MAPE), a common measure for accuracy in quantitative judgment research (e.g., Sniezek and Henry 1989, 1990), which is particularly useful when the size of judgment errors increases with the magnitude of the targets. An ANOVA with advisor (same advisor vs. varying advisor vs. no advice) as between-subjects factor on participants' MAPE during the practice phase showed no significant differences in baseline performance, $F(2, 186) = 0.20$, $p = 0.815$, $\eta_p^2 < 0.01$. Additionally, we analyzed whether judges' initial estimates and the advice they received were, on average, equally accurate. A *t*-test comparing the judges' MAPE during the practice phase to the MAPE of advice in both advice conditions revealed that advisors were significantly more accurate than the judges ($M = 43.18$, $SD = 16.46$ vs. $M = 63.73$, $SD = 59.30$), $t(229.10) = -4.52$, $p < 0.001$, $d = 0.47$. Furthermore, Levene's test indicated that there was less variance among advisors than among judges, $F(1, 314) = 29.80$, $p < 0.001$. In total, the advisors were superior to the judges in 56% of all cases and outperformed them on average by 20 percentage points. The superiority of advisors mainly derives from the lack of extremely poor judgments. The worst-performing advisor's MAPE score was 146, whereas the worst-performing judge who was not excluded from the analyses had a MAPE score of 458. Finally, there were no significant differences in the MAPE of advice between the constant and the varying advisor condition ($M = 43.04$, $SD = 21.89$ vs. $M = 43.32$, $SD = 8.40$), $t(79.64) = -0.09$, $p = 0.925$, $d = -0.02$.

3.2.2 | Accuracy of Judges' Initial Estimates

Next, we investigated whether individual learning occurs in a prototypical JAS and, if so, how it is related to advisors' judgment accuracy. To this end, we compared the accuracy of the initial estimates between the practice phase and the test phase. We treated the initial estimate of the first trial of the second phase (Trial 11) as part of the individual practice phase because participants provided this estimate prior to receiving any advice and, thus, before any socially induced learning could have occurred. We ran a 3 (*advisor:* same advisor vs. varying advisor vs. no advice) × 2 (*phase*: practice phase vs. test phase) mixed ANOVA with participants' initial accuracy as dependent variable. This analysis revealed no significant main effect of advisor, $F(2, 186) = 1.30$, $p = 0.274$, $\eta_p^2 = 0.01$, but a significant main effect of phase, $F(2, 186) = 7.23$, $p = 0.008$, $\eta_p^2 = 0.04$, that was qualified by a significant interaction of advisor and phase, $F(2, 186) = 5.06$, $p = 0.007$, $\eta_p^2 = 0.05$. We already know from our test of possible baseline differences in accuracy that there were no significant differences between the experimental conditions during the practice phase (see above), but the conditions differed significantly during the test phase, $F(2, 186) = 4.23$, $p = 0.016$, $\eta_p^2 = 0.04$ (see Table 1 and Figure 1, top left panel).

Pairwise comparisons revealed that, during the test phase, participants who received advice from the same and those who received advice from varying advisors made significantly more accurate initial estimates than participants who received no advice, $t(186) = 2.34$, $p = 0.020$, and $t(186) = 2.68$, $p = 0.008$, respectively. The two advice conditions were not

TABLE 1 | Accuracy of initial estimates by advice type in Experiment 1.

	Initial estimates		Final estimates	
	Practice phase	Test phase	Practice phase	Test phase
Advisor	*M* (SD)	*M* (SD)	*M* (SD)	*M* (SD)
No advice	65.38 (65.49)	70.24 (80.86)	69.89 (72.22)	70.64 (80.58)
Same advisor	59.83 (45.66)	48.95 (27.66)	64.29 (66.71)	43.79 (24.79)
Varying advisors	65.97 (65.33)	45.96 (23.88)	63.53 (59.52)	41.34 (21.55)

Note: MAPE scores for initial and final estimates by advice type.

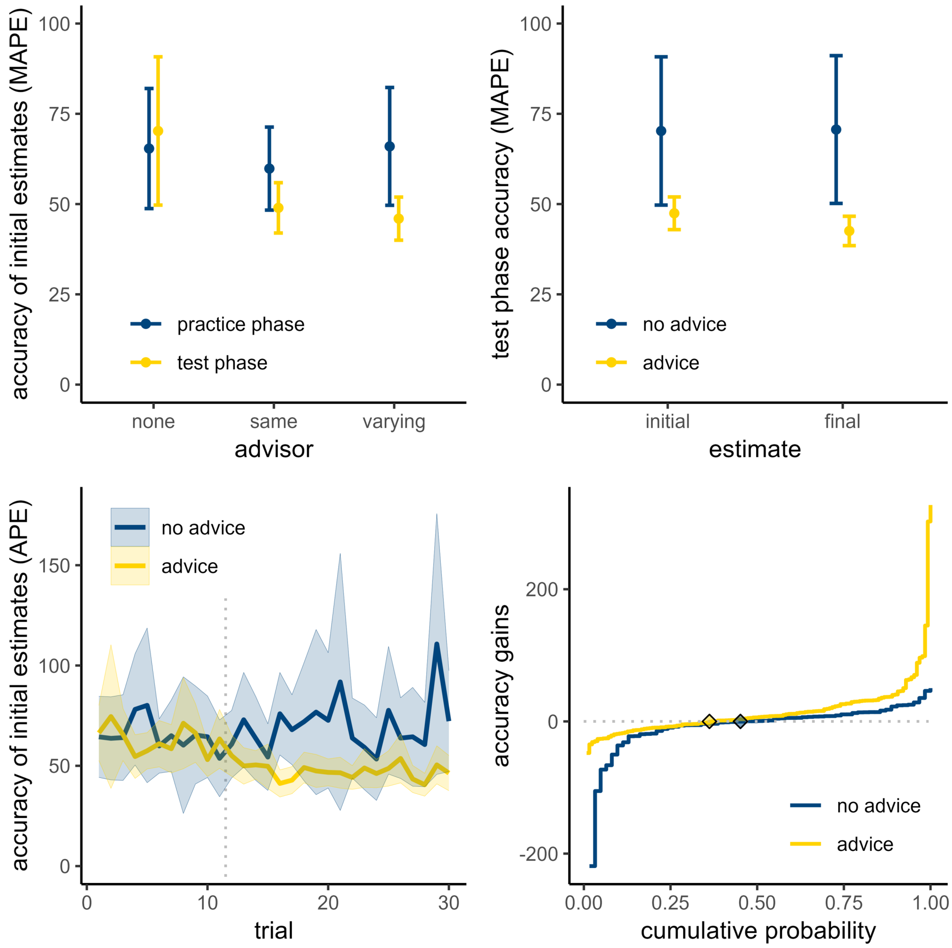

FIGURE 1 | Accuracy of estimates in Experiment 1. *Note:* Since accuracy is operationalized as deviations from the true values, lower values indicate greater accuracy. The top left panel shows participants' accuracy by advisor type and phase, whereas the top right panel shows the accuracy of participants' initial and final estimates during the test phase contingent on whether they received advice or not. Points represent the mean, and error bars denote 95% confidence intervals. The bottom left panel shows judges' mean initial accuracy per trial contingent on whether they received advice or not. The semitransparent ribbons denote the 95% confidence bands. The dotted line separates initial estimates made prior to receiving advice (to the left) from initial estimates potentially influenced by advice (to the right). The bottom right panel shows the empirical cumulative distribution of accuracy gains contingent on whether judges received advice or not. The semitransparent diamonds indicate the percentile within each group at which accuracy gains equal zero. The more shifted to the left these diamonds are, the larger the proportion of participants within the group who had positive accuracy gains between the practice and the test phase.

significantly different from each other, $t(186) = 0.33$, $p = 0.741$. Beyond that, the simple effects of phase showed no accuracy changes for the no-advice condition, $t(186) = -0.87$, $p = 0.388$. In contrast, both in the same and the varying advisor condition, participants' initial judgment accuracy increased after receiving advice, $t(186) = 3.79$, $p = 0.053$, and $t(186) = 3.61$, $p < 0.001$, respectively, even though the former comparison was not statistically significant. Hence, participants' average initial judgment accuracy only improved when they received advice, with slightly (but not significantly) stronger improvement in the varying advisor condition. As can be seen from Figure 1 (bottom right panel), accuracy increased for most but not all participants who received advice. Initial accuracy increased between the two phases for 64% of participants receiving advice compared with 55% in the no-advice condition (see also Figure 1, bottom right panel).

To clarify the role of advice quality and judges' baseline accuracy on the strength of learning, we computed accuracy gains as the difference between judges' MAPEs in the practice and the test phase (positive values indicate increases in accuracy). We predicted these accuracy gains from the advisors' MAPE (for the varying advisor condition, we averaged the advisors' accuracy) and the judges' MAPE during the practice phase in a multiple regression. As the previous analyses had revealed no significant differences between the same and the varying advisor condition at all, we collapsed across the two conditions (the results remain unchanged when running two separate analyses). The two predictors together explained 79.3% of the variance, $R^2 = 0.79$, $F(2, 124) = 237.91$, $p < 0.001$. Judges' errors during the training phase significantly predicted subsequent changes in the accuracy of their initial estimates, $\beta = 0.89$, $t(124) = 21.65$, $p < 0.001$, whereas advice quality did not, $\beta = -0.02$, $t(124) = -0.38$, $p = 0.704$. In other words, judges with low initial accuracy (high MAPE) benefited most from receiving advice, regardless of advice quality. Hence, the results of Experiment 1 do not support Hypothesis 1 to its full extent, but they clearly show learning processes that increase judges' accuracy on subsequent initial estimates.

3.2.3 | Exploratory Analyses

Finally, to put the accuracy gains due to social learning in perspective, we compared them to the accuracy gains resulting from advice weighting when making the final judgment.[1] In other words, we analyzed the accuracy of final judgments in comparison with the accuracy of the already improved initial judgments. To this end, we calculated whether participants' final estimates in the second phase (Trials 12–30) were more accurate after receiving advice compared with the no-advice control condition. Again, we only report one analysis for both advice conditions (results remain unchanged when running separate analyses). Judges receiving advice outperformed participants who did not receive advice by 28 percentage points ($M = 42.55$, SD = 23.16 vs. $M = 70.64$, SD = 80.58), $t(65.97) = -2.69$, $p = 0.009$, $d = -0.47$, replicating the finding that advice weighting increases accuracy (e.g., Gardner and Berry 1995; Soll and Larrick 2009; Sniezek et al. 2004). Furthermore, we assessed to what extent participants in the advice conditions improved in accuracy due to weighting the advice by comparing their final and initial MAPE scores in the test phase. The respective *t*-test showed that judges' final estimates were more accurate than their initial estimates by roughly 5 percentage points ($M = 42.55$, SD = 23.16 vs. $M = 47.45$, SD = 25.77), $t(126) = -6.36$, $p < 0.001$, $d = -0.56$. Thus, out of the total accuracy advantage of 28 percentage points of participants receiving advice, adjustments of the initial estimates toward the advice only accounted for 5 percentage points. In other words, learning processes accounted for roughly 83% of the total beneficial effect of receiving advice, whereas advice weighting (i.e., the integration of advice into one's already improved initial judgments) only accounted for 17% (see also Figure 1, top right panel).

3.2.4 | Conclusions

In Experiment 1, we found evidence of social learning in a prototypical JAS. Receiving advice improved judges' subsequent initial accuracy, no matter whether it came from the same or varying advisors. The most likely explanation for this phenomenon is that judges adjusted their initial judgments to a frame of reference they inferred from the advice. This adjustment accounted for most of the beneficial effect of receiving advice. Contrary to our expectations, the advice quality did not significantly moderate these accuracy gains. One possible explanation is that differences in advice quality were too small for judges to notice absent information about advice quality. This is supported by our preliminary analyses, which showed that the variance in accuracy was substantially smaller among advisors than among judges in Experiment 1. If so, we might be able to find the expected moderating influence when the advisor's accuracy is easier to infer. Furthermore, the advice quality in Experiment 1 was relatively high in general. Therefore, the impact of differences in advice quality on the strength of learning might have been restricted to rather accurate advice. We address these issues in Experiment 2.

4 | Experiment 2

Experiment 2 served as a more stringent test of the idea that advice quality moderates the magnitude of social learning in the JAS. Therefore, we introduced an experimental manipulation of advice quality where participants received advice of either high, moderate, or low quality, or they received no advice at all. Additionally, we aimed to maximize the chances of detecting effects of advice quality if they exist. Therefore, participants received veridical information about the quality of advice in the form of their advisor's accuracy rank during a pretest. Beyond that, we worked with a different estimation task to improve the generalizability of our findings.

4.1 | Method

4.1.1 | Participants, Design, and Task

Participants were 132 students (87 women, 45 men), with an average age of 23.59 years (SD = 5.16). Experiment 2 used a one-factorial design with advice quality (high vs. moderate vs. low vs. no advice) as a between-subjects factor. Participants estimated the weight in grams of small items (e.g., hammer, dustpan, and umbrella) that were present in the room, without being allowed to touch or lift them, a task adopted from Stern et al. (2017).

4.1.2 | Procedure

The procedure of Experiment 2 was similar to that of Experiment 1, with the following exceptions. First, we manipulated whether judges received highly accurate, moderately accurate, or poor advice and compared them with a control condition without advice. Participants were randomly assigned to one of these conditions. Again, we used participants of a pretest ($N = 61$) as advisors. As advisors, we selected the participant with the best average performance, the participant whose performance marked the median of the sample, and the participant with the worst performance (see Yaniv and Kleinberger 2000 for a similar manipulation of advice quality). The advisors' respective MAPE scores were 33 (high advice quality), 150 (moderate advice quality), and 523 (low advice quality). Second, participants received accurate information about their advisor's performance rank during the pretest (1st vs. 31st vs. 61st of 61). In the control condition, participants received no advice at all. Furthermore, we dropped the individual practice phase, because the control condition without advice provides the necessary benchmark to detect learning effects following advice, reducing the number of trials to 20. Because of the reduced duration of the experiment, participants only received a compensation of €4. After estimating all weights, participants were asked to estimate the accuracy of their corresponding advisor (MAPE) as a manipulation check.

4.2 | Results and Discussion

4.2.1 | Comprehension Check and Check for Possible Interfering Effects

We first ran an ANOVA with the three advice conditions (high quality vs. moderate quality vs. low quality) as between-subjects factor and the judges' estimate of the advisors' MAPE as a dependent variable. This analysis revealed significant differences between the advice conditions, $F(2, 96) = 8.79$, $p < 0.001$, $\eta_p^2 = 0.15$. Tukey post hoc tests showed that participants were able to distinguish the advice of moderate and high quality from low-quality advice ($M = 23.09$, SD = 16.34 vs. $M = 134.12$, SD = 191.57), $t(96) = -3.94$, $p < 0.001$, and ($M = 43.78$, SD = 43.38 vs. $M = 134.12$, SD = 191.57), $t(96) = -3.18$, $p = 0.006$, respectively. However, they did not perceive the high-quality advice to be significantly more accurate than advice of moderate quality ($M = 23.09$, SD = 16.34 vs. $M = 43.78$, SD = 43.38), $t(96) = -0.72$, $p = 0.751$, although the means were in the predicted direction. As a consequence, we anticipated finding evidence of the moderating effect of advice quality when contrasting low-quality advice with either moderate- or high-quality advice but not necessarily when contrasting the latter two conditions.

TABLE 2 | Accuracy of estimates by advice condition in Experiment 2.

	Initial estimates		Final estimates	
Advice type	***M***	**SD**	***M***	**SD**
No advice	198.80[a]	172.96	201.11[a]	174.94
Low quality	235.47[a]	140.82	248.30[a]	143.38
Medium quality	119.64[b]	54.10	114.64[b]	47.47
High quality	84.40[b]	56.64	62.14[b]	45.16

Note: MAPE scores for initial and final estimates by advice type. For initial estimates, means with different superscripts are significantly different from each other based on pairwise comparisons using Tukey post hoc tests. The same is true for final estimates.

4.2.2 | Accuracy of Judges' Initial Estimates

We first conducted an ANOVA with advice quality (high vs. moderate vs. low vs. no advice) as a between-subjects factor and the MAPE of initial estimates as the dependent variable. We eliminated the first trial from the calculations, because on that trial, participants had not yet received any advice. We found a significant effect of advice quality, $F(3, 128) = 11.44$, $p < 0.001$, $\eta_p^2 = 0.21$. Tukey post hoc tests showed that the MAPE of the initial estimates was not significantly different in the high and in the moderate advice quality condition, $t(128) = -1.20$, $p = 0.630$, although, descriptively, judges in the high-quality advice condition were more accurate by about 35 percentage points (see Table 2 and the upper left panel of Figure 2).

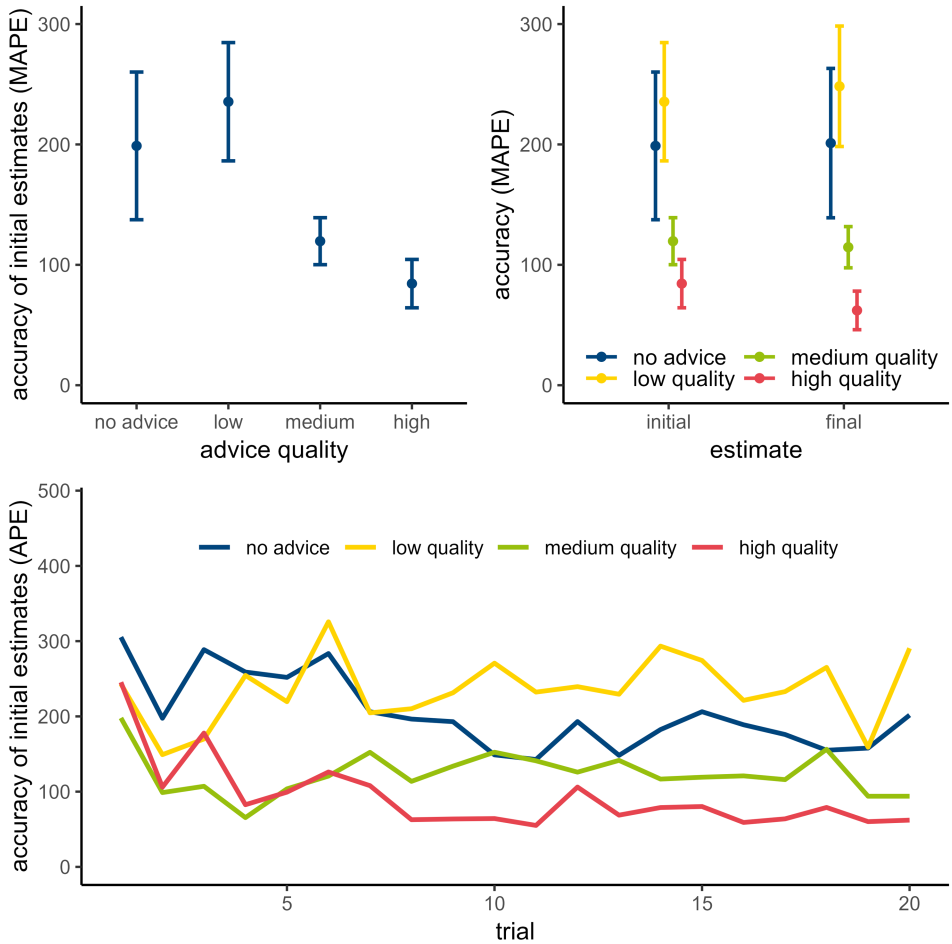

FIGURE 2 | Accuracy of Estimates in Experiment 2. *Note:* Since accuracy is operationalized as deviations from the true values, lower values indicate greater accuracy. The top left panel shows participants' accuracy by advisor type and phase. The top right panel shows the accuracy of participants' initial and final estimates during the test phase contingent on whether they received advice and, if so, the quality of the advice. Points represent the mean, and error bars denote 95% confidence intervals. The bottom panel shows judges' mean initial accuracy per trial contingent on the advice condition.

Initial estimates were significantly more accurate in the high and moderate than in the low advice quality condition, $t(128) = -5.21$, $p < 0.001$, and $t(128) = -3.96$, $p < 0.001$, respectively. Both receiving advice from a highly accurate and from a moderately accurate source led to significantly more accurate initial estimates than receiving no advice at all, $t(128) = -3.91$, $p = 0.001$, and $t(128) = -2.67$, $p = 0.040$, respectively. Finally, participants in the low advice quality condition were (descriptively) somewhat inferior to participants without advice, but this difference was not significant, $t(128) = 1.26$, $p = 0.588$. In total, these results indicate that participants' subsequent initial estimation accuracy increased as a function of advice quality, supporting Hypothesis 1.

4.2.3 | Exploratory Analyses

Similar to Experiment 1, we compared the accuracy gains due to social learning to the accuracy gains from advice weighting. To this end, we calculated an ANOVA with advice quality (high vs. moderate vs. low vs. no advice) as a between-subjects factor and the MAPE of final estimates on Trials 2–20 as a dependent variable. This analysis revealed a significant effect of experimental condition, $F(3, 128) = 16.70$, $p < 0.001$, $\eta_p^2 = 0.28$. Tukey post hoc tests showed that the final estimates were significantly more accurate when receiving high or moderate quality advice as compared with no advice, $t(128) = -4.77$, $p < 0.001$, and $t(128) = -2.95$, $p = 0.020$, respectively. In the low advice quality condition, final accuracy was not significantly different from the control condition, $t(128) = 1.63$, $p = 0.365$ (see Table 2).

We then tested to what extent the benefit of receiving advice was due to social learning. We focused on the medium- and high-quality advice conditions only since in the low advice quality condition, participants' final estimates were descriptively less accurate than their initial estimates. The total accuracy advantage over the control condition was 139 percentage points for high-quality advice and 86 percentage points for medium-quality advice. We then tested to what extent judgment accuracy in the high and medium advice quality conditions improved due to weighting advice. As can be seen from Table 1, in the high-quality condition, participants' accuracy improved significantly by about 22 percentage points, $t(32) = -6.91$, $p < 0.001$, $d = -1.20$. The accuracy gain from taking medium-quality advice was somewhat lower at around 5 percentage points, and this improvement in accuracy fell short of statistical significance, $t(31) = -1.85$, $p = 0.074$, $d = -0.33$. Hence, in the high advice quality condition, learning accounted for 84% of the total beneficial effect of receiving advice, whereas advice weighting only accounted for 16%. Even if we neglect the fact that accuracy gains due to advice weighting were not statistically significant for medium-quality advice, social learning accounted for 94% of the accuracy gain with advice weighting contributing a mere 6% (see also Figure 2, top right panel). Again, the majority of the accuracy advantages of participants who received high- or medium-quality advice over those who did not receive advice already manifested in their initial judgments (i.e., due to learning processes), with relatively minor subsequent changes as a consequence of advice weighting. In sum, this mirrors the corresponding findings of Experiment 1.

4.2.4 | Conclusion

Summarizing Experiment 2, we replicated the main finding of the first experiment, in that advice affects the accuracy of judges' subsequent initial estimates. Furthermore, we found that judges benefited only when advice quality was high or moderate. Hence, advice quality seems to be an important moderator for social learning in the JAS. Unexpectedly, we did not find that high-quality advice was more beneficial than advice of medium quality. The differences between these two conditions, albeit in the predicted direction, were weak and statistically insignificant. Hence, as long as it is sufficiently reasonable, advice might have a positive effect on the judges' initial estimate accuracy. Beyond that, one has to take into account that in our advice quality manipulation, the high- and moderate-quality advice was much more similar in terms of accuracy than, for example, the moderate- and the low-quality advice, which might also explain why there was no significant evidence of participants perceiving high-quality advice as more accurate than advice of moderate quality. Low-quality advice, in contrast, did not improve judges' initial accuracy, but neither did it significantly harm it. This supports the idea that judges seem to be rather sensitive to the quality of advice, which prevents them from adjusting their subsequent initial judgments too far toward the poor advice. Furthermore, as in Experiment 1, we observed that learning from the advisor accounts for the largest part of the beneficial effects of receiving advice.

5 | Experiment 3

Experiment 2 provided evidence that advice quality moderates improvements in subsequent initial judgment accuracy. However, the situation created in Experiment 2 might be considered an almost ideal situation for such improvements. The differences in advisors' expertise were very salient since participants received veridical information about the accuracy of the advice. In real-world situations, this degree of salience of advice quality is rather uncommon. Accordingly, in Experiment 3, we manipulated whether participants received information about advice quality. This allowed us to test whether we would still find a moderating effect of advice quality without such information, and, if so, how strong it would be in comparison with a situation with information about the quality of advice.

5.1 | Method

5.1.1 | Participants and Design

Participants were 164 students, four of which were excluded because of unreasonably high estimates either during the practice phase or during the last trials of the test phase (overestimating the true values by more than 1.300%). Of the remaining 160 participants, 104 were female, 59 were male, and one participant did not report their gender. Their average age was 24.00 years (SD = 4.25). Experiment 3 is based on a 2 (advice quality: high vs. low) × 2 (feedback: yes vs. no) factorial design.

5.1.2 | Task and Procedure

In general, the task and basic procedure were the same as in Experiment 1 and largely similar to Experiment 2. Therefore, we report only the changes made compared with the previous experiments. First, we manipulated whether judges received advice from a good or poor advisor in Experiment 3. In the high advice quality condition, subjects received advice from the most capable of 76 participants of a pretest (the same pretest that we referred to in Experiment 1) with an average MAPE score of 22. In the low advice quality condition, the advisor was the least accurate pretest participant, with a MAPE score of 146. Second, half of the participants received accurate feedback about their advisor's performance rank during the pretest (1st vs. 76th), whereas the other half solely received the advice, without any feedback about its quality. Participants were randomly assigned to one of the four conditions. Finally, we re-established the individual practice phase, comparing judges' accuracy in the practice phase to that of the initial estimates in the test phase. Since Experiment 1 showed that there were no accuracy gains in the absence of advice in the distance estimates, we can interpret all changes between the two phases as effects of receiving advice. This allowed us to drop the no-advice control condition, resulting in a straightforward two-factorial design (instead of a design featuring a nonfactorial control group). As in Experiment 1, participants received a compensation of €5.

5.2 | Results and Discussion

5.2.1 | Comprehension Check and Check for Possible Interfering Effects

We first checked whether there were systematic differences in judges' baseline accuracy between the conditions. To this end, we calculated a 2 (*advice quality:* high vs. low) × 2 (*feedback*: yes vs. no) between-subjects ANOVA with participants' MAPE scores during the practice phase as a dependent variable. This analysis revealed no significant main or interaction effect, all Fs < 0.68, all ps > 0.413.

To analyze whether participants were able to assess the quality of advice, we ran another 2 (*advice quality:* high vs. low) × 2 (*feedback*: yes vs. no) between-subjects ANOVA on judges' estimates of the advisors' MAPE. This analysis revealed a significant main effect of advice quality, $F(1, 156) = 68.14$, $p < 0.001$, $\eta_p^2 = 0.30$, that was qualified by a significant interaction of advice quality and feedback, $F(1, 156) = 16.43$, $p < 0.001$, $\eta_p^2 = 0.10$. The main effect of feedback was not significant, $F(1, 156) = 1.01$, $p = 0.318$, $\eta_p^2 = 0.01$. Simple effects analyses using post hoc contrasts revealed that high-quality advice was rated as more accurate when judges received feedback about the quality of advice as compared with when there was no feedback ($M = 22.40$, $SD = 18.12$ vs. $M = 37.02$, $SD = 15.15$), $t(156) = -3.60$, $p < 0.001$, and that low-quality advice was rated as less accurate with feedback than without it ($M = 58.00$, $SD = 21.44$ vs. $M = 49.18$, $SD = 18.07$), $t(156) = 2.14$, $p = 0.034$. Consequently, our feedback manipulation was successful. As expected, participants rated high-quality advice as more accurate than low-quality advice when they received veridical feedback about the quality of advice, $t(156) = -8.65$, $p < 0.001$. However, even without such feedback, high-quality advice was rated as more accurate than the low-quality advice, $t(156) = -2.99$, $p = 0.003$, indicating that judges could infer the quality of advice even without feedback about its quality.

5.2.2 | Accuracy of Judges' Initial Estimates

To test for learning effects, we conducted a 2 (*advice quality:* high vs. low) × 2 (*feedback*: yes vs. no) × 2 (*phase*: practice phase vs. test phase) mixed ANOVA with participants' initial accuracy as the dependent variable. This analysis showed a significant main effect of advice quality, $F(1, 156) = 4.76$, $p = 0.031$, $\eta_p^2 = 0.02$, and a significant main effect of phase, $F(1, 156) = 6.22$, $p = 0.014$, $\eta_p^2 = 0.04$. Both were qualified by a significant interaction of advice quality and phase, $F(1, 156) = 14.08$, $p < 0.001$, $\eta_p^2 = 0.08$. Simple effects analyses using post hoc contrasts revealed no significant differences between the advice quality conditions during the practice phase, $t(156) = 0.151$, $p = 0.880$, whereas participants receiving high-quality advice were significantly more accurate than those receiving low-quality advice during the test phase, $t(156) = -5.52$, $p < 0.001$. Furthermore, participants receiving high-quality advice significantly improved in accuracy from the practice to the test phase, $t(156) = -4.45$, $p < 0.001$, whereas participants who received low-quality advice showed no significant changes in accuracy, $t(156) = 0.88$, $p = 0.378$ (see Table 3 and Figure 3, top left panel).

There was no significant main effect of feedback in the mixed ANOVA, and contrary to our expectations, the interaction of feedback and advice quality was not significant either, all Fs < 1.59; all ps > 0.210. To make sense of this absence of a moderating effect of feedback, we tested whether the interaction of advice quality and phase was significant even without information about the quality of advice (i.e., to check whether advice quality affects accuracy changes even if the judge is not informed about the quality of the advice). Separate 2 (*advice quality:* high vs. low) × 2 (*phase*: practice phase vs. test phase) mixed ANOVAs for the two feedback conditions revealed that judges' initial accuracy improved more strongly after receiving high- compared with low-quality advice, both with feedback about the quality of advice and without it, as indicated by significant interaction effects, $F(1, 77) = 5.82$, $p = 0.018$, $\eta_p^2 = 0.07$, and $F(1, 79) = 12.33$, $p = 0.001$, $\eta_p^2 = 0.13$, respectively. The relatively smaller effect size in the feedback condition mainly derives from one participant overestimating the true values by more than 600% during the practice phase. However, there are no changes in the general pattern of results when excluding this participant.

Again, we also inspected the proportion of judges whose accuracy improved between phases by advice quality. Since we did not find effects of feedback on accuracy gains, we collapsed across the two feedback conditions. Initial accuracy increased for 84% of participants receiving high-quality advice, but only for 55% of participants in the low advice quality (see Figure 3, bottom right panel).

TABLE 3 | Accuracy of initial estimates by advice quality and feedback in Experiment 3.

	Initial estimates		Final estimates	
	Phase 1	**Phase 2**	**Phase 1**	**Phase 2**
Advice	*M* (SD)	*M* (SD)	*M* (SD)	*M* (SD)
High quality				
Feedback	68.03 (98.01)	31.95 (15.05)	60.13 (75.32)	22.69 (10.93)
No feedback	56.16 (45.22)	36.61 (20.24)	56.27 (46.14)	29.24 (16.64)
Low quality				
Feedback	62.96 (51.45)	65.63 (53.90)	62.57 (51.59)	63.62 (50.75)
No feedback	58.18 (47.06)	66.71 (43.14)	58.90 (49.05)	68.51 (38.70)

Note: MAPE scores for initial and final estimates by advice type.

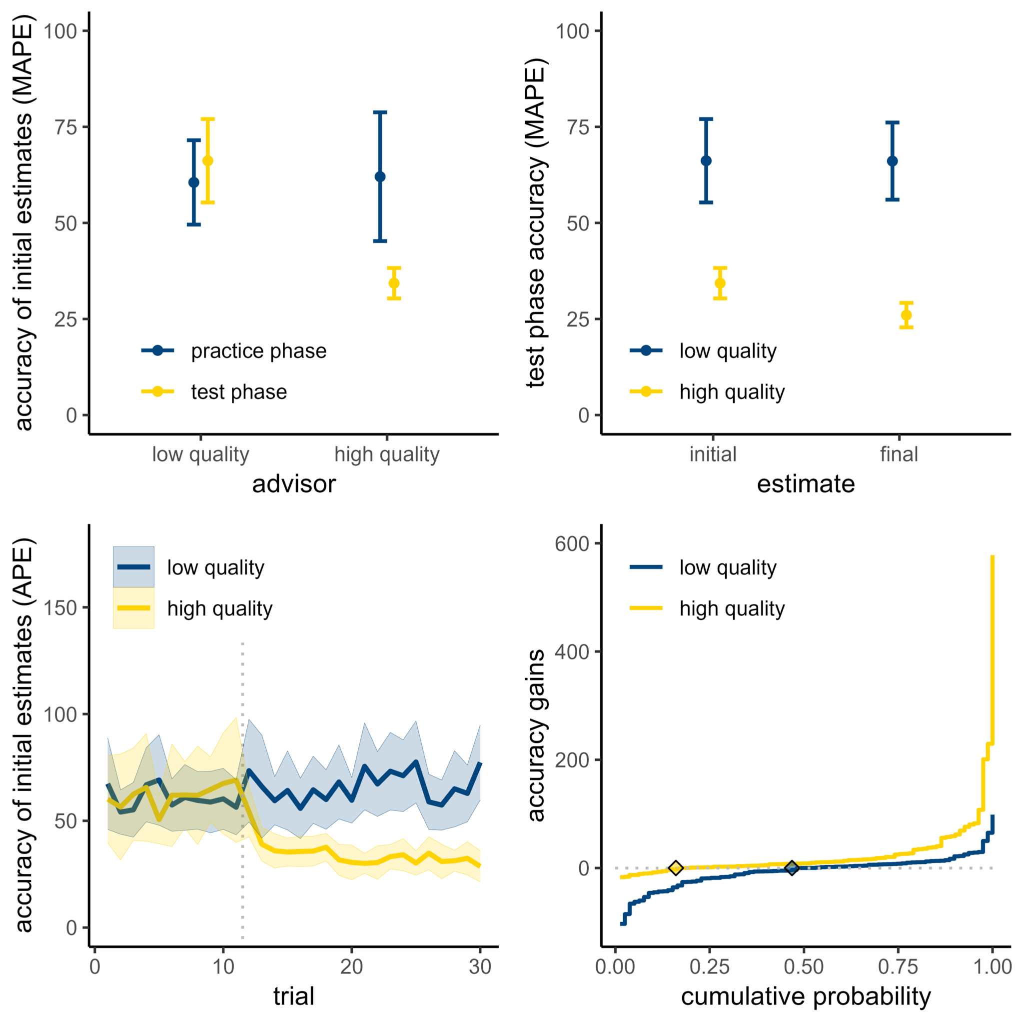

FIGURE 3 | Accuracy of estimates in Experiment 3. *Note:* Since accuracy is operationalized as deviations from the true values, lower values indicate greater accuracy. The top left panel shows participants' accuracy by advisor type and phase, whereas the top right panel shows the accuracy of participants' initial and final estimates during the test phase contingent on whether they received advice or not. Points represent the mean, and error bars denote 95% confidence intervals. The bottom left panel shows judges' mean initial accuracy per trial contingent on whether they received advice or not. The semitransparent ribbons denote the 95% confidence bands. The dotted line separates initial estimates made prior to receiving advice (to the left) from initial estimates potentially influenced by advice (to the right). The bottom right panel shows the empirical cumulative distribution of accuracy gains contingent on whether judges received advice or not. The semitransparent diamonds indicate the percentile within each group at which accuracy gains equal zero. The more shifted to the left these diamonds are, the larger the proportion of participants within the group who had positive accuracy gains between the practice and the test phase.

In sum, judges' initial accuracy improved after receiving good advice, with the result that they outperformed judges receiving poor advice, supporting Hypothesis 1. High-quality advice led to higher initial judgment accuracy even without information about the quality of advice, in line with the first part of Hypothesis 2. However, in contrast to the second part of Hypothesis 2, this performance enhancement was not significantly stronger when participants received information about the quality of advice.

5.2.3 | Exploratory Analyses

Similar to Experiments 1 and 2, we compared the accuracy gains of initial judgments after receiving advice to the accuracy gains resulting from advice taking when coming to the final judgment. In this analysis, we collapsed across the two feedback conditions, since in our main analyses, feedback did not significantly affect accuracy. We first calculated whether participants' final estimates in the test phase (Trials 12–30) were more accurate after receiving advice compared with the uninfluenced final estimates in the training phase (because we had no control condition in Experiment 3, we worked with the training phase as a performance baseline). The corresponding paired-sample t-test revealed that the post advice final estimates were significantly more accurate than the training phase estimates when receiving high-quality advice ($M = 26.01$, $SD = 14.41$ vs. $M = 58.18$, $SD = 61.92$), $t(80) = -4.89$, $p < 0.001$, $d = -0.54$, but not when receiving low-quality advice ($M = 66.09$, $SD = 44.83$ vs. $M = 60.71$, $SD = 50.03$), $t(78) = 1.52$, $p = 0.134$, $d = 0.17$.

Having observed a total accuracy gain of roughly 32 percentage points when participants received high-quality advice, we once again tested to what extent this accuracy gain was due to social learning and advice weighing, respectively. A comparison of judges' initial accuracy and final accuracy showed that their accuracy improved by roughly 8 percentage points ($M = 26.01$, $SD = 14.41$ vs. $M = 34.31$, $SD = 17.91$), $t(80) = -9.30$, $p < 0.001$, $d = -1.03$ (see Figure 3, right panel). Accordingly, learning accounted for roughly 76% of the total beneficial effect of receiving high-quality advice, whereas adjustments toward the advice accounted for roughly 24%.

5.2.4 | Conclusions

The findings of Experiment 3 indicate that good advice leads to stronger improvements in subsequent initial judgments than poor advice, even when the advice quality is not disclosed. High-quality advice without information about the quality of advice led to higher initial judgment accuracy, without significant additional performance enhancement when information about the advice quality was given. Interestingly, low-quality advice did not lead to significant negative transfer, irrespective of whether or not there was information about the quality of advice. This is in line with the idea that judges can identify low-quality advice even without such information (Yaniv and Kleinberger 2000). But why does the learning process after receiving high-quality advice seem to work in a similar way, that is, largely irrespective of information about advice quality? A possible explanation is that judges adjust their own estimation tendency toward that of their advisor as long as they perceive the advice as sufficiently accurate, without substantial differences in the strength of this adjustment. This would be in line with the findings of Experiment 2. Finally, when receiving high-quality advice, judges' postadvice final judgments were superior to the improved initial judgments. However, the major performance enhancement seems to derive from learning processes rather than advice taking, thereby mirroring the general findings of Experiments 1 and 2.

6 | General Discussion

In the present study, we investigated learning processes due to receiving advice, in terms of a generalized transfer from one trial to another of the same type of task. More precisely, we were interested in whether advice would affect not only the accuracy of postadvice final judgments, as shown in many previous studies (e.g., Gardner and Berry 1995; Soll and Larrick 2009; Sniezek et al. 2004), but also the accuracy of the judge's subsequent *initial* judgments. In particular, we expected individual performance enhancements, especially when receiving advice of high quality. Even in the absence of information about the quality of advice, we expected judges receiving high-quality advice to outperform judges receiving low-quality advice, because judges should be somewhat sensitive to advice quality, or because they should, at least, benefit from the superior calibration of their advisors. In addition, we postulated that information about the advisors' accuracy should influence the judges' ability to infer advice quality and, thereby, the strength of learning. In an exploratory manner, we also differentiated between, on the one hand, the performance enhancements manifested in the initial judgments and, on the other hand, the beneficial effect of combining the own initial estimate with advice when coming to a final judgment. To this end, we compared the accuracy of postadvice final judgments to the assumedly improved initial judgments.

In line with our hypotheses, we found evidence that advice of at least moderate quality led to learning processes that manifested as improved accuracy of subsequent initial judgments in two different estimation tasks. In contrast, we found no evidence that poor advice harms judges' subsequent initial judgments (these effects were insignificant throughout our experiments). Hence, at the very least, we can say that the beneficial effect of receiving high-quality advice markedly exceeded the possible detrimental effect of receiving low-quality advice. Surprisingly, information about the quality of advice had no significant additional positive impact on the strength of learning. Even without such information, participants were sensitive to advice quality. Participants receiving high-quality advice but who received no information about its quality benefited roughly similarly to those who did receive such information. In the case of low-quality advice, participants' accuracy did not decrease substantially, and this was similarly true in the absence and in the presence of information about the advisor's accuracy.

Furthermore, even in the absence of information about advice quality, judges' final estimates were still more accurate than their initial judgments when receiving recommendations from a good advisor, which supports the idea of two distinct beneficial effects of receiving high-quality advice. On the one hand, learning processes lead to improved subsequent initial judgments and, on the other hand, advice weighting (i.e., integrating the advice into one's final judgment) improves the accuracy of final postadvice judgments via error cancellation—with the former process being more pronounced than the latter, at least in those cases where both processes can occur simultaneously. In the following section, we discuss these results against the backdrop of previous research, point out limitations of our experiments, and illustrate directions for future research.

6.1 | Learning From Advice

Previous research already dealt with the question of whether advice enables some kind of learning processes, with the robust result that advice helps to find the right solution more quickly when repeatedly working on one specific problem in prototypical decision-making experiments (e.g., Biele et al. 2009; Çelen et al. 2010; Chaudhuri et al. 2006; Kocher et al. 2014). For example, Biele et al. (2009) found that a single piece of advice can improve the performance on a repeated choice task, such that the decision maker identifies the recommended correct option more quickly and, consequently, chooses this option more often over the course of the experiment. In other words, people seem to learn from specific advice, with the result of improved postadvice decisions on the exact same problem—which is also mirrored by classic judge–advisor experiments showing that advice leads to improved final judgments (e.g., Soll and Larrick 2009; Sniezek et al. 2004). Hence, one can conclude that there is ample evidence for learning that improves the quality of postadvice *final* judgments or decisions. Beyond this positive effect of receiving advice, we found strong evidence that advice can also have a more general beneficial effect on subsequent different problems, in that the judges' *initial* judgments in the problems are already improved. In other words, advice does not only contain information that can increase one's performance on the same problem but can also initiate a transfer from one problem to another problem on the same type of task. We already know somewhat similar learning processes as G-I transfer in interacting groups (Schultze, Mojzisch, et al. 2012; Stern et al. 2017). Interestingly, our findings suggest that interaction with the opportunity to communicate is not a necessary condition for the occurrence of subsequent performance enhancements. Altogether, our findings show that learning, in terms of such a transfer, should be added to the list of positive effects of receiving advice and should be addressed in more detail in future research.

Our study also addressed the moderating effect of advice quality on the strength of learning, with the predicted result of stronger learning after receiving high-quality advice. Judges seem to be sensitive to the quality of advice, in particular when receiving low-quality advice. Accordingly, in Experiment 2, we found differences in the strength of learning between the high and low advice quality conditions as well as the moderate and low advice quality conditions. Beyond that, in Experiment 3, information about the quality of advice had no significant additional effect on the strength of performance changes. Hence, to a certain extent, judges are sensitive to the quality of advice even without any external cues, which is in line with previous research (Biele et al. 2009; Yaniv and Kleinberger 2000). As Yaniv and Kleinberger (2000) discuss, judges might recognize particularly poor estimates as out of bounds, even though the judge cannot generate a correct estimate on his or her own. For example, it should be rather difficult to determine whether the distance between London and Rome is rather 1500 or 2000 km. In contrast, when the advice suggests that this distance is 30,000 km, the judge has a very good chance to notice its low quality and, hence, refrain from using it as a point of reference, even without explicit information about advice quality. Hence, the findings of Yaniv and Kleinberger regarding the influence of advice quality on advice taking can be transferred to learning from advice.

But what exactly is the content of the learning process that allows judges to benefit on future related judgments? In our opinion, the individual performance enhancements mainly derive from judges partially adopting the advisor's calibration. That is, judges recognize systematic differences between their own and the advisor's estimation tendency, and they seek to reconcile this discrepancy by adjusting their judgments toward those of the advisor. In other words, the advice serves as a frame of reference for the judge. Such frames of reference play an important role in the accuracy of quantitative judgments because they allow a person to infer whether their estimates are too high or too low, in general (e.g., Bonner and Baumann 2008; Bonner et al. 2007; Laughlin, Bonner, et al. 1999; Laughlin, Gonzalez, et al. 2003). In this vein, it makes perfect sense that judges treated advice as a frame of reference only when they ascribed high functional value to it (i.e., advice of moderate or high, but not of low quality).[2]

In our data, the learning processes that improved the initial judgment accuracy accounted for at least three-fourths of the total benefit of receiving advice, whereas advice weighting in terms of an adjustment toward the specific advice when coming to a final estimate only accounted for about one-fourth. To put this into perspective: The one process that previous judge–advisor research has heavily focused on as a facilitator of judgment quality, namely, the weighting of advice in the final judgments, did actually play a *minor* role in explaining the facilitation of judgment quality in our experiments, whereas the process that we have introduced and tested here, namely, improved initial judgments due to learning from previous advice, accounted for the *major* parts of the effects.

This is not to say that the accuracy increases stemming from the learning process are necessarily always stronger than those obtained from weighting the advice. It is plausible to assume that there are some tasks in which it is difficult for the judge to learn from the advisor, for example, because both judge and advisor already have a similarly good understanding of the task. In those situations, accuracy gains should mostly (or exclusively) stem from weighting the advice and the resulting cancellation of errors. Nonetheless, our finding suggests that a substantial part of the effect of receiving advice on judgmental accuracy may have gone undiscovered so far. JAS research assumes that judges usually overweigh their own opinion compared with the recommendation of the advisor (e.g., Yaniv 2004b; Yaniv and Kleinberger 2000). In contrast, our results show a substantial adjustment toward the advice in terms of a transfer from one problem to another. Accordingly, judges do not seem to be as resistant to advice as usually presumed. After adjusting their initial judgments toward the advisors, the judges might assume that they have learned sufficiently and, consequently, stick with their judgments. This, in turn, could be a plausible explanation for at least part of the phenomenon of egocentric advice discounting, at least when the task type remains stable and advice is presented sequentially on every trial (e.g., Harvey and Fischer 1997). Further testing this idea might be a fruitful avenue for future research.

One final question worth discussing is whether the learning effects we described here have any relevance in real life. We believe that they do, because they provide decision makers with an effective means to deal with a problem inherent to advice-taking situations: Advice can threaten decision makers' sense of autonomy (Goldsmith and Fitch 1997; Rader et al. 2017). Although no study has directly tested the hypothesis that judges egocentrically discount advice to retain a feeling of autonomy, there is at least indirect evidence for it. For example, people higher in agency use advice less (Schultze et al. 2018), and decision makers receiving advice from algorithms rely on this advice more frequently when they can make even miniscule adjustments to it (Dietvorst et al. 2018). Decision makers who want to retain a feeling of autonomy—or who want to appear autonomous to others—can largely discount the advice but still benefit from it on subsequent occasions via the learning process observed in our Experiments. Real-life examples could be teenagers who reject their parents' advice, asserting that they must make their own mistakes, but take that advice to heart in other situations where it applies. Another example could be a manager who received (excellent) unsolicited advice from a junior colleague and feels that heeding this advice could undermine their status within the organization.

6.2 | Limitations and Directions for Future Research

There are some limitations of our current study that should be considered. First, in Experiments 2 and 3, we employed a rather strong manipulation of advice quality. Hence, it is debatable how accurately people can judge their advisor's competency when it is less evident, and how this affects the learning processes that we investigated. On the one hand, low-quality advice might also negatively affect the accuracy of judges as long as this advice contains a certain degree of plausibility. On the other hand, poor but plausible advice could even have a beneficial effect as long as the advisor's and judge's errors are equally strong and on opposite sides of the target value, with the result that adjusting one's judgment toward the advice leads to more accurate estimates.

Second, we only used two different types of estimation tasks. Although we found structurally similar patterns with both types, our results should be replicated with additional types of tasks. For example, a more complex task (e.g., financial forecasts) might, on the one hand, affect the amount of time needed until the learning process is completed or might even eliminate the performance enhancements in initial estimates, because the exchange of well-calibrated numeric information might not be sufficient to induce a general learning process. On the other hand, on very difficult tasks, negative learning might occur. When the task-specific knowledge of the judge is low, it should be more difficult to assess the quality of advice. Consequently, judges might rely heavily on weak advisors, which, in turn, should lead to a loss in estimation accuracy. Therefore, it is crucial to replicate our findings with different types of quantitative estimation tasks, preferably tasks with a high ecological validity like forecasting tasks, and of tasks of different complexity, to see how the learning processes react to these variations.

Finally, we currently cannot say to what extent the learning process that we demonstrated in our experiments is stable over time. We do, however, consider the temporal stability to be rather high, predominantly for two reasons. On the one hand, in the standard JAS, there is no normative social influence on the judges to adjust their own calibration to that of their advisors. Thus, the advisors' influence on the judges' subsequent initial estimates is informational in nature. Absent new information suggesting that the advice was inaccurate, there is no need for the judge to recalibrate. Second, in the somewhat related field of G-I transfer in group judgment research, there is evidence for stable socially induced improvements of individual accuracy after the group is dissolved (Lippold et al. 2021; Stern et al. 2017). Because of the rather high structural similarity in the utilized tasks and experimental procedures between these group experiments and our current judge–advisor study, there is good reason to expect that the learning process that we found in the present study will also turn out to be rather stable over time.

6.3 | Conclusion

In three experiments, we showed that advice can affect the accuracy of subsequent initial judgments, most likely because of judges adjusting their own general calibration based on the reference values provided by well-calibrated advice. This beneficial learning effect was stronger when advice quality was high, both in the presence and absence of explicit information on advice quality. Interestingly, the increases in initial judgment accuracy after receiving advice accounted for the lion's share of the beneficial effects of advice on the accuracy of the final judgments reported in numerous previous studies. Hence, we can add learning from an advisor on how to perform the task to the list of beneficial effects of advice.

Acknowledgments

Open Access funding enabled and organized by Projekt DEAL.

Data Availability Statement

The data that support the findings of this study are openly available in Open Science Framework at https://osf.io/8s3f5.

Endnotes

[1] We report analyses of advice taking for all three experiments in an online supplement for the sake of brevity (https://osf.io/8s3f5/?view_only=c8f85b32c22e463c80fdb192ed2b50aa). In brief, these analyses replicate two robust findings, namely, that judges egocentrically discount advice and that they place greater weight on advice of higher quality, especially when provided with feedback about advice quality.

[2] To further investigate this idea, we conducted more detailed analyses where we differentiated between two different sources of estimation

error. Brown and Siegler (1993, see also Brown 2002) argue that people depend on two types of knowledge when generating estimates: *metric knowledge* and *mapping knowledge*. Metric knowledge is a general understanding of the appropriate scaling; it represents one's calibration of judgments. Mapping knowledge involves ordinal relations among individual estimations of the domain, that is, it allows us to put different target values of the same kind in the correct order. The additional analyses support our assumption, as our effects can mostly be traced back to reductions in metric error, whereas the mapping error remains relatively stable in all three experiments. Because of space considerations, we refrain from reporting the detailed results here in the paper, but an online supplement with the analyses of metric and mapping knowledge can be found here: https://osf.io/8s3f5/?view_only=c8f85b32c22e463c80fdb192ed2b50aa.

References

Bandura, A. 1986. *Social Foundations of Thought and Action: A Social Cognitive Theory*. Englewood Cliffs, NJ: Prentice-Hall.

Biele, G., J. Rieskamp, and R. Gonzalez. 2009. "Computational Models for the Combination of Advice and Individual Learning." *Cognitive Science* 33: 206–242. https://doi.org/10.1111/j.1551-6709.2009.01010.x.

Bonaccio, S., and R. S. Dalal. 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. https://doi.org/10.1016/j.obhdp.2006.07.001.

Bonaccio, S., and L. Van Swol. 2014. "Combining Information and Judgments." In *Judgment and Decision Making at Work*, edited by S. Highhouse, R. S. Dalal, and E. Salas, 178–198. New York, NY: Routledge.

Bonner, B. L., and M. R. Baumann. 2008. "Informational Intra-Group Influence: The Effects of Time Pressure and Group Size." *European Journal of Social Psychology* 38: 46–66. https://doi.org/10.1002/ejsp.400.

Bonner, B. L., S. D. Sillito, and M. R. Baumann. 2007. "Collective Estimation: Accuracy, Expertise, and Extroversion as Sources of Intra-Group Influence." *Organizational Behavior and Human Decision Processes* 103: 121–133. https://doi.org/10.1016/j.obhdp.2006.05.001.

Brown, N. R., and R. S. Siegler. 1993. "Metrics and Mappings: A Framework for Understanding Real-World Quantitative Estimation." *Psychological Review* 100: 511–534. https://doi.org/10.1037/0033-295X.100.3.511.

Çelen, B., S. Kariv, and A. Schotter. 2010. "An Experimental Test of Advice and Social Learning." *Management Science* 56: 1687–1701. https://doi.org/10.1287/mnsc.1100.1228.

Chaudhuri, A., S. Graziano, and P. Maitra. 2006. "Social Learning and Norms in a Public Goods Experiment With Inter-Generational Advice1." *Review of Economic Studies* 73: 357–380. https://doi.org/10.1111/j.1467-937X.2006.0379.x.

Dietvorst, B. J., J. P. Simmons, and C. Massey. 2018. "Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them." *Management Science* 64, no. 3: 1155–1170. https://doi.org/10.1287/mnsc.2016.2643.

Farrell, S. 2011. "Social Influence Benefits the Wisdom of Individuals in the Crowd." *Proceedings of the National Academy of Sciences of the United States of America* 108: E625. https://doi.org/10.1073/pnas.1109947108.

Gardner, P. H., and D. C. Berry. 1995. "The Effect of Different Forms of Advice on the Control of a Simulated Complex System." *Applied Cognitive Psychology* 9: 55–79. https://doi.org/10.1002/acp.2350090706.

Goldsmith, D. J., and K. Fitch. 1997. "The Normative Context of Advice as Social Support." *Human Communication Research* 23, no. 4: 454–476. https://doi.org/10.1111/j.1468-2958.1997.tb00406.x.

Harvey, N., and I. Fischer. 1997. "Taking Advice: Accepting Help, Improving Judgment, and Sharing Responsibility." *Organizational Behavior and Human Decision Processes* 70: 117–133. https://doi.org/10.1006/obhd.1997.2697.

Hertz, U., V. Bell, and N. Raihani. 2021. "Trusting and Learning From Others: Immediate and Long-Term Effects of Learning From Observation and Advice." *Proceedings. Biological Sciences* 288, no. 1961: 20211414. https://doi.org/10.1098/rspb.2021.1414.

Kocher, M., M. Sutter, and F. Wakolbinger. 2014. "Social Learning in Beauty-Contest Games." *Southern Economic Journal* 80: 586–613. https://doi.org/10.4284/0038-4038-2010.150.

Laughlin, P. R., B. L. Bonner, A. G. Miner, and P. J. Carnevale. 1999. "Frames of Reference in Quantity Estimations by Groups and Individuals." *Organizational Behavior and Human Decision Processes* 80: 103–117. https://doi.org/10.1006/obhd.1999.2848.

Laughlin, P. R., C. M. Gonzalez, and D. Sommer. 2003. "Quantity Estimations by Groups and Individuals: Effects of Known Domain Boundaries." *Group Dynamics: Theory, Research, and Practice* 7: 55–63. https://doi.org/10.1037/1089-2699.7.1.55.

Lippold, M., S. Schulz-Hardt, and T. Schultze. 2021. "G-I Transfer in Multicue Judgment Tasks: Discussion Improves Group Members' Knowledge About Target Relations." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 47: 532–545. https://doi.org/10.1037/xlm0000947.

Minson, J. A., and J. S. Mueller. 2012. "The Cost of Collaboration: Why Joint Decision Making Exacerbates Rejection of Outside Information." *Psychological Science* 23: 219–224. https://doi.org/10.1177/0956797611429132.

Pronin, E., T. Gilovich, and L. Ross. 2004. "Objectivity in the Eye of the Beholder: Divergent Perceptions of Bias in Self Versus Others." *Psychological Review* 111, no. 3: 781–799. https://doi.org/10.1037/0033-295X.111.3.781.

Rader, C. A., R. P. Larrick, and J. B. Soll. 2017. "Advice as a Form of Social Influence: Informational Motives and the Consequences for Accuracy." *Social and Personality Psychology Compass* 11: 1–17. https://doi.org/10.1111/spc3.12329.

Rowe, G., and G. Wright. 1999. "The Delphi Technique as a Forecasting Tool: Issues and Analysis." *International Journal of Forecasting* 15, no. 4: 353–375. https://doi.org/10.1016/S0169-2070(99)00018-7.

Schultze, T., A. Mojzisch, and S. Schulz-Hardt. 2012. "Why Groups Perform Better Than Individuals at Quantitative Judgment Tasks: Group-To-Individual Transfer as an Alternative to Differential Weighting." *Organizational Behavior and Human Decision Processes* 118: 24–36. https://doi.org/10.1016/j.obhdp.2011.12.006.

Schultze, T., A. F. Rakotoarisoa, and S. Schulz-Hardt. 2015. "Effects of Distance Between Initial Estimates and Advice on Advice Utilization." *Judgment and Decision Making* 10: 144–171.

Schultze, T., T. M. Gerlach, and J. C. Rittich. 2018. "Some People Heed Advice Less Than Others: Agency (But Not Communion) Predicts Advice Taking." *Journal of Behavioral Decision Making* 31, no. 3: 430–445. https://doi.org/10.1002/bdm.2065.

Sniezek, J. A., and T. Buckley. 1995. "Cueing and Cognitive Conflict in Judge-Advisor Decision Making." *Organizational Behavior and Human Decision Processes* 62: 159–174. https://doi.org/10.1006/obhd.1995.1040.

Sniezek, J. A., and R. A. Henry. 1989. "Accuracy and Confidence in Group Judgment." *Organizational Behavior and Human Decision Processes* 43: 1–28. https://doi.org/10.1016/0749-5978(89)90055-1.

Sniezek, J. A., and R. A. Henry. 1990. "Revision, Weighting, and Commitment in Consensus Group Judgment." *Organizational Behavior and Human Decision Processes* 45: 66–84. https://doi.org/10.1016/0749-5978(90)90005-T.

Sniezek, J. A., G. E. Schrah, and R. S. Dalal. 2004. "Improving Judgement With Prepaid Expert Advice." *Journal of Behavioral Decision Making* 17: 173–190. https://doi.org/10.1002/bdm.468.

Soll, J. B., and R. P. Larrick. 2009. "Strategies for Revising Judgment: How (And How Well) People Use Others' Opinions." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 35: 780–805. https://doi.org/10.1037/a0015145.

Soll, J. B., and A. E. Mannes. 2011. "Judgmental Aggregation Strategies Depend on Whether the Self Is Involved." *International Journal of Forecasting* 27: 81–102. https://doi.org/10.1016/j.ijforecast.2010.05.003.

Stern, A., T. Schultze, and S. Schulz-Hardt. 2017. "How Much Group Is Necessary? Group-To-Individual Transfer in Estimation Tasks." *Collabra: Psychology* 3: 1–17. https://doi.org/10.1525/collabra.95.

Yaniv, I. 2004a. "The Benefit of Additional Opinions." *Current Directions in Psychological Science* 13: 75–78. https://doi.org/10.1111/j.0963-7214.2004.00278.x.

Yaniv, I. 2004b. "Receiving Other People's Advice: Influence and Benefit." *Organizational Behavior and Human Decision Processes* 93: 1–13. https://doi.org/10.1016/j.obhdp.2003.08.002.

Yaniv, I., and E. Kleinberger. 2000. "Advice Taking in Decision Making: Egocentric Discounting and Reputation Formation." *Organizational Behavior and Human Decision Processes* 83: 260–281. https://doi.org/10.1006/obhd.2000.2909.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.