Research Article

Psychological Science
1–14

DOI: 10.1177/09567976211054089
www.psychologicalscience.org/PS
SAGE

Tips From the Top: Do the Best Performers Really Give the Best Advice?

David E. Levari[1], Daniel T. Gilbert[2], and Timothy D. Wilson[3]
[1]Harvard Business School, Harvard University; [2]Department of Psychology, Harvard University; and [3]Department of Psychology, University of Virginia

Abstract
Everyone knows that if you want to learn how to do something, you should get advice from people who do it well. But is everyone right? In a series of studies ($N = 8{,}693$), adult participants played a game after receiving performance advice from previous participants. Although advice from the best-performing advisors was no more beneficial than advice from other advisors, participants believed that it had been—and they believed this despite the fact that they were told nothing about their advisors' performance. Why? The best performers did not give better advice, but they did give more of it, and participants apparently mistook quantity for quality. These studies suggest that performing and advising may often be unrelated skills and that in at least some domains, people may overvalue advice from top performers.

Keywords
advice, performance, learning, open data

Received 3/16/20; Revision accepted 9/15/21

People seek every kind of advice, but not from every kind of advisor. Americans pay companies such as MasterClass and Skillshare hundreds of millions of dollars each year to connect them to "icons, experts, and industry rock stars" who will teach them to write novels, start businesses, play chess, or barbecue brisket, and they pay these premiums because they naturally believe that the best advice comes from the best performers. Personal trainers flex their toned muscles and financial advisors show off their expensive sports cars because they know that potential customers will take these signs of successful performance as robust indicators of the quality of their professional advice. Although advice seekers do not always seek advice from the best performers (Casciaro & Lobo, 2008; Hofmann et al., 2009), they seem to value that advice highly even when they are too shy or too vain to seek it (Hur et al., 2020).

But should they? There are several reasons to suspect that in at least some domains, the best performers may not necessarily give the best advice (Grant, 2018). First, the knowledge that top performers have about their domains of excellence is often implicit (Speelman, 1998). A gifted chanteuse can produce a smoky timbre, but that does not mean she can explain how her brain, ears, and vocal cords work in harmony to achieve the effect (Reed et al., 2010). Second, even when top performers do have explicit knowledge about their domains of excellence, they may not be especially good at communicating that information (Hinds et al., 2001). Singing is a skill, but so are patience, eloquence, and the ability to take a novice's perspective—all of which contribute to the effectiveness of one's advising but do not necessarily come bundled with the ability to hit high C (Nickerson, 1999). Indeed, knowing a lot about a domain can sometimes make it harder to take a novice's perspective (Beilock et al., 2002; Fisher & Keil, 2016). In short, the skills that are likely to make someone an excellent advisor—explicit knowledge of a domain and the ability to communicate that knowledge to someone who does not share it—are not necessarily the same skills that make someone an excellent

Corresponding Author:
David E. Levari, Harvard University, Harvard Business School
Email: dlevari@fas.harvard.edu

performer. Those who can do are not always those who can teach (Hattie & Marsh, 1996; Zhang et al., 2022).

Do people generally believe that the best performers give the best advice? And if so, are they right? And if not, then why do people believe it? We explored these questions in four studies in which participants gave, sought, used, and evaluated advice about how to play a game. We hoped to discover whether people do indeed believe that an advisor's performance is a robust indicator of the quality of their advice (Study 1); whether that belief is warranted (Study 2); and if not, why people mistakenly hold it (Studies 3 and 4).

The protocols for all studies were approved by Harvard University's Committee on the Use of Human Subjects, and the studies were carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki.

Statement of Relevance

Millions of Americans who want to learn to write novels, start businesses, play chess, or barbecue brisket pay hundreds of millions of dollars each year to companies such as MasterClass and Skillshare who connect them to "icons, experts, and industry rock stars," and they pay these premiums because they believe that the best performers obviously give the best advice. But is that true? In a series of studies, we found that in at least some domains, top performers do not give better advice than other performers, they just give more of it—and people seem to mistake quantity for quality. Our studies suggest that in at least some instances, people may overvalue advice from top performers.

Study 1: Do People Expect the Best Performers to Give the Best Advice?

Method

Overview. Participants played a word game and were then asked to predict the relationship between an advisor's performance in the game and the quality of their advice.

Sample size. Prior to collecting data we used the *pwr* package (Version 1.3-0; Champely, 2020) in R (Version 4.0.5; R Core Team, 2021) to determine a minimum sample size for each condition on the basis of our best estimates of likely effect sizes. When that minimum was reached, we automatically closed the survey to further participation. In some cases, there was a brief lag between the time at which the minimum sample size was reached and the time at which the survey was closed, so occasionally we ended up with a few more participants in a condition than we required.

Participants. Participants were 1,103 users of Amazon Mechanical Turk (610 females, 485 males, eight preferred not to answer; age: M = 34.79 years, SD = 16.03) who were paid for their participation.

Procedure. Participants were told that they would be playing a game called Word Scramble and then answering questions about it. Participants were shown a 4 × 4 grid of 16 letters. We will refer to the grid hereafter as a "board." Participants were told that they would have 60 s to find as many words as possible on the board, that they would play three rounds of the game, and that each round would feature a different board. Participants were told the rules of the game, which were that (a) words must be made of letters that appeared in contiguous squares, (b) the same square could not be included more than once in a single word, and (c) any English word was permitted except for proper nouns and words of fewer than three letters. Participants completed a brief training session to make sure they understood the rules of the game and then played three rounds of Word Scramble. Although the boards differed between rounds, all participants saw the same board on each round. We pretested the three boards to ensure that they were of about equal difficulty.

After playing three rounds, participants were told that 100 previous participants had been allowed to play six rounds of Word Scramble and had then been asked to write advice for future players. We will refer to these previous participants hereafter as "advisors." Participants were asked to predict the relationship between an advisor's performance in the Word Scramble game and the quality of their advice. Because the way a question is structured can influence its answer, we assigned each participant to one of five conditions, and in each condition, we asked this question in a different way.

Participants in the *free-choice* condition (n = 151) were asked whether they would prefer to receive advice from the advisor who had scored best in the game, the advisor who had scored worst in the game, an advisor who had scored "the same as you in the game," an advisor who had scored "slightly worse than you in the game," or an advisor who had scored "slightly better than you in the game" or whether they believed that "it won't matter, because there is no relationship between how well someone plays the game and how well they give advice about the game." Participants in the *forced-choice* condition (n = 151) were asked the same question as participants in the free-choice condition but were

Table 1. Preferences for Advisors in the Free-Choice and Forced-Choice Conditions of Study 1

	Participants who chose this answer (%)	
Answer	Free-choice condition	Forced-choice condition
"A participant who scored slightly better than you in the game"	10.00	25.50
"A participant who scored slightly worse than you in the game"	0.67	2.68
"A participant who scored the same as you in the game"	4.00	2.68
"The participant who scored the worst in the game"	2.00	5.37
"The participant who scored the best in the game"	53.33	63.76
"It won't matter, because there is no relationship between how well someone plays the game and how well they give advice about the game"	30.00	—

not given the final option. Participants in the *decile* condition (n = 398) were asked to estimate how helpful advice would be if it came from an advisor whose performance was in the top 10%, the top 20%, the top 30%, and so on through the bottom 10%. Participants made these ratings on a 7-point scale ranging from 1, *extremely unhelpful*, to 7, *extremely helpful*. Participants in the *percentile* condition (n = 300) were asked from whom they would most like to receive advice if they could choose between advisors whose performance ranged from the 1st percentile to the 99th percentile. Finally, participants in the *open-response* condition (n = 103) were asked an open-ended question: "Whose advice do you think would improve your performance the most?" Participants typed their answers into a text box.

After answering one of the five questions described above, participants were shown an item that served as an attention check (i.e., "If you are actually reading this question, please select the option 'Other' and type the word 'potato'") as well as several exploratory items that are described in the Supplemental Material available online (Section 1.1).

Results

We excluded 11 participants who did not pass the attention check, which left 1,092 participants (478 males, 606 females, eight preferred not to answer; age: M = 34.79 years, SD = 16.09) in the data set. Excluding these 11 participants did not appreciably change the results of any of the analyses we performed on the data we collected.

Did participants expect the best performers to give the best advice? No matter how we asked that question, the answer was "yes." As Table 1 shows, 53% of the participants in the free-choice condition (71 males, 79 females; age: M = 33.23 years, SD = 11.13 years) preferred to receive advice from the single best performer. Thirty percent reported that there was "no relationship between how well someone plays the game and how well they give advice about the game," but when this option was removed in the forced-choice condition (57 males, 91 females, one preferred not to answer; age: M = 34.70 years, SD = 12.76), the number of participants who preferred to receive advice from the single best performer increased to 64%. These results suggest that roughly half the participants in these conditions had a strong preference for advice from the top performer and that an additional 10% or so had a weak preference for advice from the top performer.

We analyzed the responses of participants in the decile condition (176 males, 215 females, four preferred not to answer; age: M = 34.53 years, SD = 11.71) by fitting a linear mixed model to our data in R using the *lme4* package (Version 1.1-26; Bates et al., 2015). The dependent variable was the participant's rating of the helpfulness of the advice (ranging from 1–7), and the within-subjects fixed factor was the advisor's performance decile (ranging from 1–10). We included participant as a random-intercept term and advisor's performance decile as a random-slope term in the model. The inclusion of the random-intercept term significantly improved model fit, $\chi^2(2)$ = 2,111.37, p < .001, as did the inclusion of the random-slope term, $\chi^2(2)$ = 2,406.50, p < .001. As Figure 1 shows, the advisor's performance decile was a positive predictor of participants' ratings of the helpfulness of that advisor's advice, b = 0.51, SE = 0.02, t = 33.03, p < .001, 95% confidence interval (CI) = [0.48, 0.54]. In short, participants in the decile condition expected the best performers to give the best advice.

As Figure 2 shows, the same was true for participants in the percentile condition (129 males, 163 females, three preferred not to answer; age: M = 36.09 years, SD = 24.17). Of those participants, 51.53% reported that they would choose to receive advice from an advisor who had scored in the 99th percentile, which is to say from a top performer. In short, participants in the decile condition expected the best performers to give the best advice.

We analyzed the data from participants in the free-response condition (45 males, 58 females; age: M = 34.42

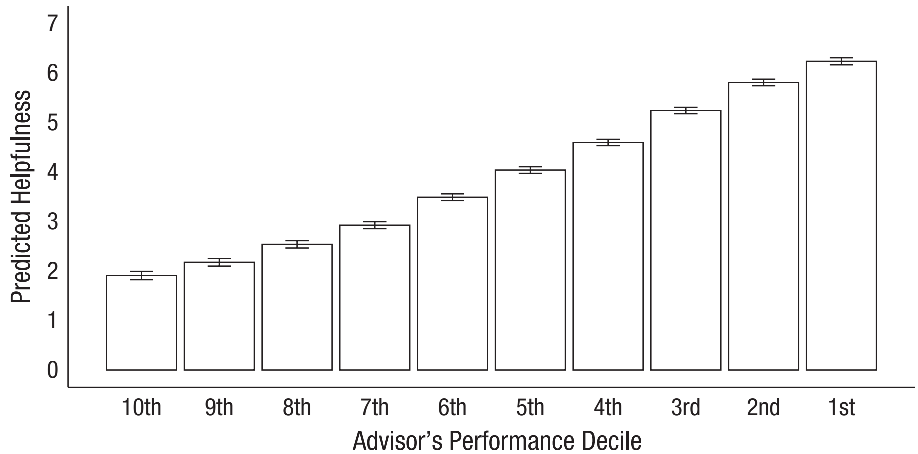

Fig. 1. Mean participant estimate of advice helpfulness in the decile condition of Study 1. The *x*-axis shows the decile of an advisor's performance (10th = worst, 1st = best), and the *y*-axis shows participants' estimates of the helpfulness of an advisor's advice. Error bars represent ±1 *SE*.

years, *SD* = 11.42) by asking two coders who were blind to the hypothesis to read all responses to the question, "Whose advice do you think would improve your performance the most?" and to sort the responses into a set of categories that they developed without consulting with us. (The categories are described in the Supplemental Material, Section 1.2.) After developing the categories together, the coders independently categorized each piece of advice. They agreed on 87% of the categorizations. Coders discussed their disagreements and came to consensus in all cases.

A considerable number of participants (37.86%) either provided no answer or appeared not to understand the question. Of those who both understood and answered the question, 31.25% spontaneously reported that they expected the single best performer to give the best advice, and an additional 26.56% spontaneously reported that they expected a good performer to give the best advice. In other words, the majority of participants who understood and answered the question spontaneously reported that they expected the best advice from people who themselves performed well. (The complete categorization results are described in the Supplemental Material, Section 1.2.) Taken together, the results of Study 1 suggest that people generally expected an advisor's performance in the Word Scramble game to be a robust indicator of the quality of their advice. Were they right?

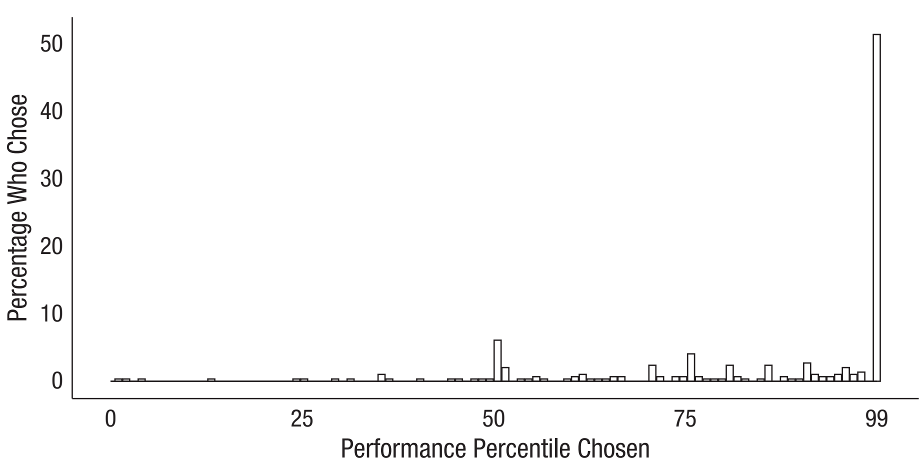

Fig. 2. Participants' preferences for advisors in the percentile condition of Study 1. The *x*-axis shows advisor performance in percentiles (1st = worst, 99th = best), and the *y*-axis shows the percentage of participants who reported that they would prefer to receive advice from an advisor in that percentile.

Study 2: Do the Best Performers Give the Best Advice?

Method

Overview. To find out whether the best performers give the best advice, we assigned a group of participants (the "advisors") to play Word Scramble and then to write advice for future participants. Next, we gave that advice to a new group of participants (the "advisees") and measured its impact on their performance.

Advisors.

Participants. Advisors were 100 users of Amazon Mechanical Turk (49 males, 51 females; age: $M = 36.05$ years, $SD = 12.87$) who were paid for their participation. Sample-size requirements were determined as in Study 1.

Procedure. Advisors were introduced to Word Scramble as in Study 1 and then played six rounds of the game. Pretesting revealed that the six boards used in the six rounds were about equally difficult. Each advisor saw the six boards in a different randomly determined order.

After completing six rounds of the game, advisors were asked to write advice for a future participant (the advisee) who would (a) be playing Word Scramble, (b) know all the rules, and (c) have played one practice round. Advisors were told that they should try to write advice that would help the advisee do as well as possible in the game. When they were finished, advisors rated the quality of their own advice. Specifically, they were asked to estimate how much their advice would improve a future participant's performance on a 101-point scale ranging from 0, *0% improvement (not at all)*, to 100, *100% improvement or more (a lot)*. Participants also completed the attention-check item used in Study 1 as well as several other exploratory items described in the Supplemental Material (Section 2.1).

Advisees.

Participants. Advisees were 2,085 users of Amazon Mechanical Turk (903 males, 1,175 females, seven preferred not to answer; age: $M = 33.33$ years, $SD = 13.12$) who were paid for their participation. Sample-size requirements were determined as in Study 1.

Procedure. Advisees were introduced to Word Scramble as in Study 1 and then played a single round of the game. All advisees were shown the same board in this first round. Next, each advisee was randomly assigned to either the *advice* condition or the *no-advice* condition. Those who were assigned to the advice condition ($n = 1,485$) were shown a piece of advice from a randomly selected advisor and were told that it had been written by "a past participant in the study." Participants who were assigned to the no-advice condition ($n = 600$) were not shown any advice (though for the sake of simplicity, we will continue to refer to these participants as advisees). Advisees in both conditions then played five additional rounds of the game. Advisees in both conditions saw the same boards on each round.

After completing six rounds, advisees in the advice condition rated the quality of the advice they had received on a 7-point scale ranging from 1, *extremely unhelpful*, to 7, *extremely helpful*, and then estimated how much the advice had improved their performance on a 101-point scale ranging from 0, *0% improvement (not at all)*, to 100, *100% improvement or more (a lot)*. Advisees in both conditions completed the attention-check item used in Study 1 as well as several exploratory items described in the Supplemental Material (Section 2.2).

Results

Advisors. We excluded data from one advisor who did not find any words in any of the six rounds, one who did not pass the attention check, three who did not write any advice, three whose task durations were three or more standard deviations from the mean, seven who reported that they were not native English speakers, and seven who reported having computer difficulties. This left 78 advisors (38 males, 40 females; age: $M = 36.38$ years, $SD = 13.13$) in the data set. Excluding these 22 advisors did not appreciably change the results of any of the analyses we performed on the data we collected.

Advisors' performance. The 78 advisors found an average of 9.92 words ($SD = 4.82$) per round of the game. To determine whether advisors improved across rounds, we used a linear mixed model in which the dependent measure was the advisor's performance (i.e., the number of words they found) in a round, and the independent within-participants variable was the number of the round (1–6). We included a random-intercept term for the advisor as a random effect. An additional random-slope term for the number of the round did not improve model fit, $\chi^2(2) = 0.15$, $p = .93$. Results showed that the number of the round was positively associated with performance, $b = 0.32$, $SE = 0.07$, $t = 4.78$, $p < .001$, 95% CI = [0.19, 0.45], which is to say that advisors performed better over time. This result is important because it suggests that an advisor's performance was a result of skill rather than luck: Skill improves with practice, but luck does not. In the Supplemental Material (Section 2.4), we describe an additional analysis that suggests that advisors' performances were not due to luck.

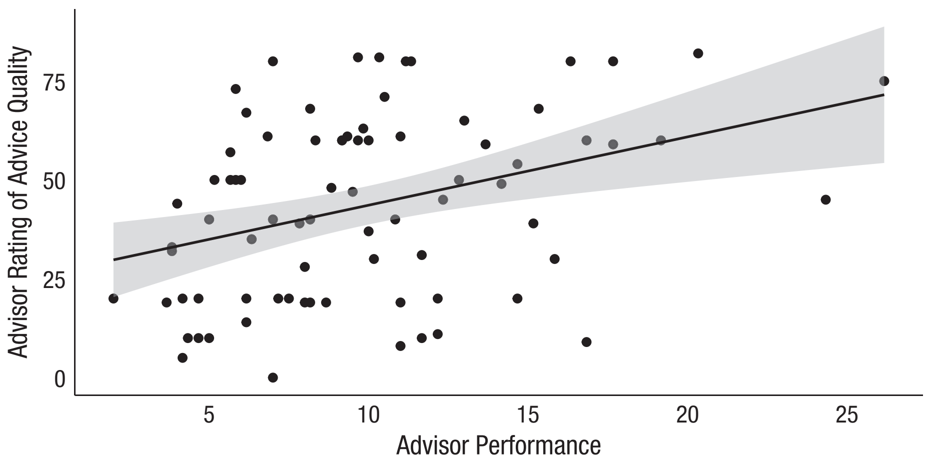

Fig. 3. Scatterplot showing the relationship between advisor performance and advisor rating of advice quality in Study 2. The *x*-axis shows each advisor's performance, the *y*-axis shows the advisor's estimates of the quality of their advice, and the error band shows the 95% confidence interval around the line of best fit.

Advisors' advice. Advisors wrote an average of 40.44 words of advice (SD = 27.31) that described a wide variety of strategies, tactics, tricks, and hints for playing Word Scramble (e.g., "Look for an 's' or an 'ed' to be able to attach onto words to make them plural or the past tense"; "Look for short words, at least three letters, as longer words don't give you more points"; or "Don't just think about a letter a word starts with—think about suffixes and prefixes, for example, words that end in -ix, -ing, or start with re-"). A list of all the advice offered by advisors can be found in the Supplemental Material (Section 2.5).

Advisors expected their advice to improve an advisee's performance by a striking 43.46% (SD = 23.05%), and as Figure 3 shows, advisors' ratings of the quality of their advice were positively correlated with their own performance, r = .36, p = .001. In other words, the best performers believed that they had given the best advice. Were they right?

Advisees. We excluded data from one advisee who reported switching from a computer trackpad to a mouse midway through the game, from four advisees who did not find any words in any of the six rounds of the game, and from 16 advisees who did not complete all six rounds of the game. This left 2,064 advisees in the data set (advice condition: 617 males, 845 females, five preferred not to answer; age: M = 33.19 years, SD = 13.98; no-advice condition: 273 males, 322 females, two preferred not to answer; age: M = 33.73 years, SD = 10.83). Excluding these 21 advisees did not appreciably change the results of any of the analyses we performed on the data we collected.

Advisees' performance. Did receiving advice improve advisees' performance? To answer this question, we first compared the average performance of advisees in the advice and no-advice conditions on the first round, which advisees played before being assigned to condition. As expected, there were no differences between conditions on the first round, t(1156.8) = −0.79, d = 0.04, p = .42. Next, we computed each advisee's improvement by subtracting their performance on the first round from the average of their performance on the subsequent rounds. Advisees in the advice condition showed more improvement (M = 3.31 words) than did advisees in the no-advice condition (M = 2.91 words), t(1114.7) = −3.03, d = 0.15, p = .003. In other words, receiving advice improved advisees' performance.

Did an advisor's performance predict the amount of improvement that their advice produced? To answer this question, we analyzed the performance of advisees in the advice condition by fitting a linear mixed model to our data in R using the *lme4* package. The dependent measure was the advisee's performance (i.e., the number of words the advisee found in a round of the task), the independent within-participants variable was the number of the round (1–6), and the independent between-participants variable was the advisor's average performance across six rounds. We also included an interaction term between these two independent variables. We included a random-slope term for the round and a random-intercept term for the advisee as random effects. The random-slope term significantly improved model fit, $\chi^2(2)$ = 150.79, p < .001, as did the random-intercept term, $\chi^2(2)$ = 3,718.4, p < .001. The interaction did not improve model fit, $\chi^2(1)$ = 2.72, p = .10, and the lower-order main effects described below were the same whether or not the interaction term was included.

The analysis revealed that the number of the round positively predicted advisee performance, b = 0.77,

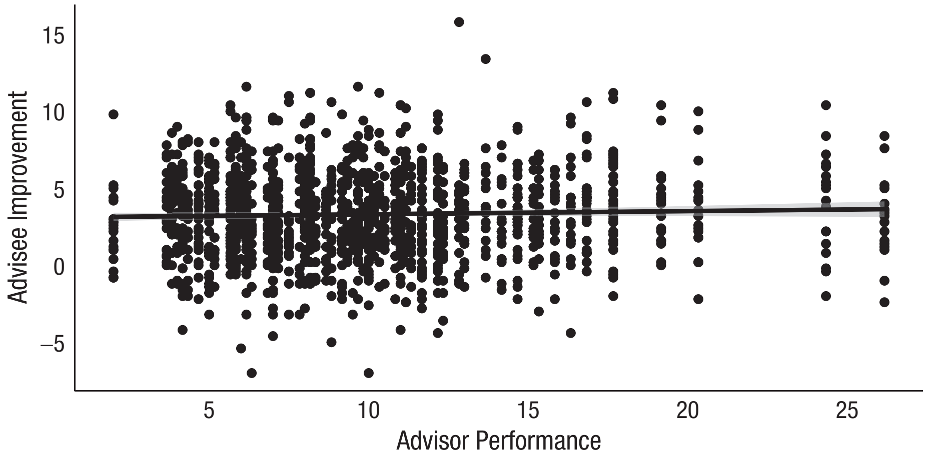

Fig. 4. Scatterplot showing the relationship between advisor performance and advisee improvement in Study 2. The *x*-axis shows advisor performance (average number of words found across six rounds), the *y*-axis shows advisee improvement, and the error band shows the 95% confidence interval around the line of best fit.

$SE = 0.04$, $t = 19.06$, $p < .001$, 95% CI = [0.69, 0.85], which is to say that on average, advisees in the advice condition improved over time, just as their advisors had. However, there was no main effect of advisor's performance on advisee's performance, $b = 0.003$, $SE = 0.03$, $t = 0.11$, $p = .92$, 95% CI = [–0.05, 0.06], and no interaction between the number of the round and the advisor's performance, $b = 0.006$, $SE = 0.003$, $t = 1.65$, $p = .10$, 95% CI = [–0.001, 0.01]. Figure 4 shows the relationship—or more to the point, the absence of a relationship—between the performance of advisees in the advice condition and the performance of their advisors. As the figure shows, the advice from the best performers was helpful—but no more helpful on average than the advice from other performers.

Advisee's ratings of advice. Advisees generally believed that the advice they received was helpful ($M = 4.36$ on a scale ranging from 1–7) and that receiving it had improved their performance ($M = 33.07\%$ on a scale ranging from 0% to 100% or more). We created a *perceived-helpfulness score* for each piece of advice by averaging the helpfulness ratings of the advisees who received it and a *perceived-improvement score* for each piece of advice by averaging the improvement ratings of the advisees who received it. To our surprise, we found that an advisor's performance predicted the perceived helpfulness of their advice ($r = .27$, $p = .02$) and the perceived improvement it produced ($r = .22$, $p = .05$). As Figure 5 shows, advisees mistakenly believed that advice from the best performers was more helpful and that it produced more improvement than did advice from bottom performers. What makes this result so surprising is that advisees were never told how well their advisors had performed. In other words, the advice from the best performers had some objective quality that made it seem better even to advisees who did not know anything about their advisor's performance.

Improvability of performance. The key finding from Study 2 was that advisees who received advice from a top-performing advisor did not perform any better than advisees who received advice from any other advisor. But might that be because Word Scramble is the kind of game in which advice cannot actually help? The fact that advisees who received advice outperformed advisees who did not receive advice suggests it is not, but to be sure, we conducted an additional study in which a new group of participants received a piece of the advice that many of the best-performing advisees in Study 2 had received ("helpful advice"), a piece of advice that many of the worst-performing advisees in Study 2 had received ("unhelpful advice"), or no advice at all. Participants who received helpful advice performed significantly better than participants who received unhelpful advice or no advice, demonstrating that Word Scramble is indeed a game in which some advice—but not all advice—can improve performance. These results are fully described as Studies S4a and S4b in the Supplemental Material (Section 4.2).

Study 3: Why Does Advice From the Best Performers Seem Better Even When It Is Not?

Advisees in Study 2 rated advice from the best performers as better even when it was not—and even when they did not know anything about their advisor's performance. Why? What—if not its effectiveness—distinguished the advice from the best performers? We can imagine two possibilities. One possibility is that the best performers

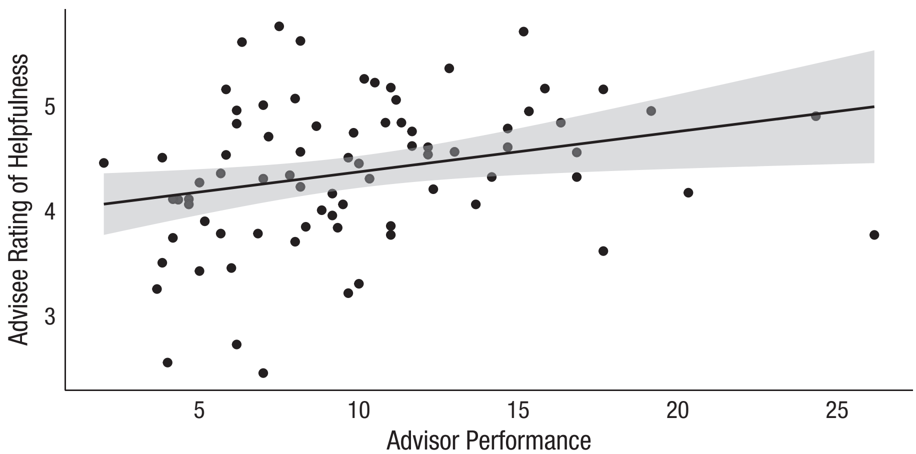

Fig. 5. Scatterplot showing the relationship between advisor performance and advisee rating of advice helpfulness in Study 2. The *x*-axis shows advisor performance (average number of words found per round), the *y*-axis shows advisee ratings of helpfulness, and the error band shows the 95% confidence interval around the line of best fit.

gave advice that sounded better (Schwartz et al., 2017): It may have been more articulate, more authoritative, or given with greater confidence (Carpenter et al., 2013; Gaertig & Simmons, 2018; Price & Stone, 2004). Another possibility is that the best performers gave advice that felt better to implement. It may have been easier to use or more fun to follow, or it may have yielded more rapid improvement. Because advisees in Study 2 both read the advice and had a chance to implement it, we cannot know which of these might have led advisees to rate the advice from the best performers more highly than the advice from other performers.

In Study 3, we asked participants to read the advice given by advisors in Study 2 and to estimate its effectiveness, but we gave them no opportunity to implement it. If advice from the best performers sounded better but did not feel better to implement, then these participants should have rated it more highly—even though they did not know that it came from the best performers and had no opportunity to implement it. On the other hand, if advice from the best performers did not sound better but instead felt better to implement, then these participants should not have rated it more highly.

Method

Participants. Participants were 320 users of Amazon Mechanical Turk (169 males, 148 females, three preferred not to answer; age: M = 33.66 years, SD = 10.54) who were paid for their participation. Sample-size requirements were determined as in Study 1.

Procedure. As in Study 2, participants were told that they would be playing Word Scramble and then answering some questions about it. They were introduced to the game as in Studies 1 and 2 and then played a single round using the same board. Next, all participants were asked to read and rate advice that had been written by advisors in Study 2. We worried that participants who were asked to read all 78 pieces of advice would not read them carefully, so we asked each participant to read just 26 pieces of advice that were randomly drawn from the full set of 78 pieces. These 26 pieces of advice were presented in a random order. After reading a piece of advice, participants were asked to estimate its effectiveness by estimating how it would impact an advisee's performance on a scale ranging from 1 (*it would make their performance much worse*) to 7 (*it would make their performance much better*) with a midpoint at 4 (*it would not affect their performance*). After the 10th and 20th pieces of advice, participants were shown an attention-check item that read, "If you're actually reading this question, please select the middle option." Participants were then shown the attention-check item used in Study 1 as well as several exploratory items that are described in the Supplemental Material (Section 3.1).

Results

We excluded data from 22 participants who did not pass one or more of the attention checks, leaving 298 participants (152 males, 144 females, two preferred not to answer; age: M = 33.81 years, SD = 10.74 years) in the data set. Excluding these participants did not appreciably change the results of any of the analyses we performed on the data we collected.

Was there a relationship between participants' estimates of the effectiveness of a piece of advice and the performance of the advisor who produced it? There was. Participants' estimates of effectiveness were positively correlated with the advisor's performance, $r = .34$, $p < .01$

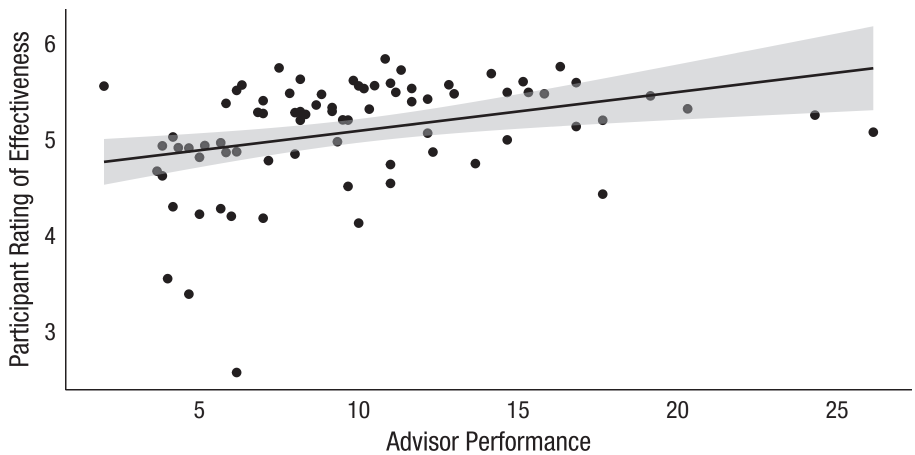

Fig. 6. Scatterplot showing the relationship between advisor performance and participant estimates of advice effectiveness in Study 3. The *x*-axis shows advisor performance (average words found per round), the *y*-axis shows participant ratings of advice effectiveness, and the error band shows the 95% confidence interval around the line of best fit.

(see Fig. 6). In other words, the best performers gave advice that sounded better, even to participants who had no opportunity to implement it. Was there also a relationship between participants' estimates of the effectiveness of a piece of advice and the actual effectiveness of that piece of advice (as measured by the improvement scores of advisees in Study 2)? There was not. Participants' estimates of the effectiveness of a piece of advice were uncorrelated with that piece's actual effectiveness, $r = .003$, $p = .98$. In short, participants who had no opportunity to implement a piece of advice expected advice that happened to come from the best performers to be more effective. Clearly, the best performers gave advice that had some objective quality that made it sound better, even to those who did not have an opportunity to implement it. In Study 4, we sought to determine what that objective quality might be.

Study 4: Why Does Advice From the Best Performers Sound Better?

Method

Overview. We measured seven properties of the advice given by advisors in Study 2 and then sought to determine whether these properties could account for the ratings of advice quality provided by advisees in Study 2.

Procedure. We recruited two undergraduate research assistants (who were blind to the study purposes and hypotheses) and asked them to code detailed descriptions of seven properties that we thought might affect the apparent quality of advice. The first property was authoritativeness. Coders were told that

> authoritative advice is confident, direct, and unqualified. The advisor sounds certain of the advice and does not "hedge." Authoritative advice says, "This is what you should do (or not do)." It implies that the advisor is an "authority" who knows what he or she is talking about. It offers "declarations" rather than "suggestions."

For the second property, actionability, coders were told that

> actionable advice is clear and easy to follow. It often provides a series of steps to take, like a recipe. After reading actionable advice, an advisee should know exactly what to do and how to do it. Actionable advice is specific rather than vague, concrete rather than abstract, and does not require interpretation.

The third property was articulateness. Coders were told that "articulate advice uses appropriate vocabulary to make complete, well-formed sentences, and is generally free of spelling and grammatical errors." For the fourth property, obviousness, coders were told that "obvious advice is advice that just about anyone would probably know without having been told by the advisor." The fifth property was number of suggestions. Coders were told that the number of suggestions

> does not simply refer to the number of words or sentences, but rather to the number of distinct suggestions made. To determine this number, rewrite the advice as a bulleted list in which each bullet is a distinct and separate suggestion.

Coders were told that the number of suggestions "is the number of bullets in such a list." The sixth property was the number of "should" suggestions, and the seventh property was the number of "should not" suggestions. Coders were told to consult the list they had just created when coding the number of suggestions and ask "How many bullets in the list are suggestions about what the advisee should do . . . and how many are about what the advisee should not do?"

After reading detailed descriptions of each of these seven properties, coders practiced using them by coding 19 pieces of advice about how to write an email. Each coder rated the authoritativeness, actionability, articulateness, and obviousness of a practice item on a scale ranging from 1 (*not at all*) to 7 (*extremely*) and then estimated the number of suggestions, the number of "should" suggestions, and the number of "should not" suggestions that the practice item contained. Coders did this for all 19 practice items. Next, the coders shared their ratings and estimates with each other and with us. When the coders' ratings or estimates diverged, they discussed them with each other and with us until consensus was reached. Finally, each coder was shown each of the 78 pieces of advice generated by advisors in Study 2 and coded each of these pieces for the seven properties.

Results

We calculated the reliability of the coders' ratings and estimates of the seven properties, and all were acceptable (Cronbach's αs = .6–.9) except for articulateness (Cronbach's α = .53), which we therefore excluded from further analyses. For each piece of advice, we averaged the two coders' ratings or estimates of each of the six properties and then entered these averaged ratings or estimates into a multiple regression. We used the perceived-helpfulness score and the perceived-improvement score of each piece of advice (derived from the ratings and estimates of advisees in Study 2) as dependent variables, and we used authoritativeness, actionability, obviousness, number of suggestions, and the ratio of "should" suggestions to "should not" suggestions as independent variables. The ratio of "should" suggestions to "should not" suggestions did not significantly improve model fit when predicting either the perceived-helpfulness score, $F(1, 72) = 0.32$, $p = .58$, or the perceived-improvement score, $F(1, 72) = 0.07$, $p = .79$, and so we removed it from this model and from subsequent analyses.

Did any of the remaining properties predict the perceived-helpfulness score or perceived-improvement score that a piece of advice received? As Table 2 shows, the authoritativeness of a piece of advice did not predict its perceived-improvement score, $b = 0.18$, $SE = 0.10$, $t = 1.86$, $p = .07$, 95% CI = [−0.01, 0.38], or its perceived-helpfulness score, and the actionability of a piece of advice predicted its perceived-helpfulness score, $b = 0.19$, $SE = 0.09$, $t = 2.13$, $p = .04$, 95% CI = [0.01, 0.36], but not its perceived-improvement score. In other words, the authoritativeness and actionability of a piece of advice had, at best, weak and inconsistent effects on its perceived quality. However, the number of suggestions that a piece of advice contained had a strong and consistent effect on both its perceived-helpfulness score, $b = 0.27$, $SE = 0.07$, $t = 4.02$, $p < .001$, 95% CI = [0.13, 0.40], and on its perceived-improvement score, $b = 0.21$, $SE = 0.07$, $t = 2.92$, $p < .01$, 95% CI = [0.06, 0.35]. In short, the more independent suggestions an advisor made, the more helpful and the more likely to produce improvement their advice was seen to be. It is worth noting that the number of independent suggestions an advisor made was not correlated with the actual efficacy of their advice, $r = .13$, $p = .25$.

Of course, making more suggestions typically means using more words. Did the perceived quality of the advice depend on the number of independent suggestions it contained, or did it merely depend on the number of words it contained? To find out, we added the word count for each piece of advice as an additional independent variable to both the multiple regressions described above. Although word count predicted neither perceived helpfulness, $b = -0.001$, $SE = 0.001$, $t = -1.10$, $p = .28$, 95% CI = [−0.02, 0.01], nor perceived improvement, $b = -0.01$, $SE = 0.01$, $t = -1.61$, $p = .11$, 95% CI = [−0.02, 0.002], the number of independent suggestions continued to predict both perceived helpfulness, $b = 0.35$, $SE = 0.10$, $t = 3.46$, $p < .001$, 95% CI = [0.15, 0.55], and perceived improvement, $b = 0.34$, $SE = 0.11$, $t = 3.14$, $p = .002$, 95% CI = [0.12, 0.55]. In other words, the number of words a piece of advice contained did not predict its perceived quality, but the number of independent suggestions it contained did.

Did the relationship between number of independent suggestions and perceived advice quality explain why participants in Studies 2 and 3 believed the advice from the best performers was better, even though they did not know it came from the best performers? To find out, we conducted a mediation analysis. We standardized the perceived-helpfulness scores and perceived-improvement scores for each piece of advice (as determined by advisees in Study 2) and then averaged those two scores to create an index for each piece of advice that we will hereafter refer to as the *perceived-quality index*. We then regressed the perceived-quality index for each piece of advice on the number of independent suggestions that piece of advice offered (as determined by coders in Study 4) as well as on the

Table 2. Results From the Multiple Regression of Four Coded Properties on Perceived-Helpfulness Scores and Perceived-Improvement Scores in Study 4

Outcome and predictor	Estimate	95% CI	*SE*	*t*	*p*
Perceived helpfulness					
Authoritative	0.07	[−0.11, 0.26]	0.09	0.81	.42
Actionable	0.19*	[0.01, 0.36]	0.09	2.13	.04
Obvious	0.08	[−0.17, 0.33]	0.13	0.62	.54
Amount of advice	0.27***	[0.13, 0.40]	0.07	4.02	< .001
Constant	−2.57*	[-4.79,-0.35]	1.11	−2.31	.02
Perceived improvement					
Authoritative	0.18†	[−0.01, 0.38]	0.10	1.86	.07
Actionable	0.04	[−0.14, 0.23]	0.10	0.44	.66
Obvious	−0.10	[−0.37, 0.17]	0.13	−0.72	.47
Amount of advice	0.21**	[0.06, 0.35]	0.07	2.92	< .01
Constant	−1.27	[−3.61, 1.08]	1.18	−1.08	.29

Note: $N = 78$ observations. Estimates shown are unstandardized. For perceived helpfulness, $R^2 = .30$; adjusted $R^2 = .27$; residual standard error = 0.86 ($df = 73$); $F(4, 73) = 7.95$, $p < .001$. For perceived improvement, $R^2 = .22$; adjusted $R^2 = .18$; residual standard error = 0.91 ($df = 73$); $F(4, 73) = 5.15$, $p < .01$. CI = confidence interval.
†$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

performance of the advisor who offered it (as measured in Study 2). The analysis revealed that the number of suggestions a piece of advice offered did indeed predict its perceived quality, $b = 0.26$, $t = 4.24$, $p < .001$, 95% CI = [0.14, 0.38], but that the advisor's performance did not, $b = 0.03$, $t = 1.39$, $p = .17$, 95% CI = [−0.01, 0.07].

Next, we constructed a mediation model with bias-corrected bootstrapping (DiCiccio & Efron, 1996) as implemented in the R package *mediation* (Version 4.5.0; Tingley et al., 2014) with 10,000 bootstrapped samples. The analysis revealed that 43.33% of the total effect of advisor performance on perceived quality was mediated by the number of independent suggestions, $b = 0.022$, $p = .01$, 95% CI = [0.006, 0.04]. (Separate mediation analyses using perceived helpfulness and perceived improvement, rather than combining them into an index of perceived quality, produced the same basic pattern of results.) Figure 7 shows the unstandardized regression coefficients of the mediation and outcome models, and Figure 8 shows the effect sizes of the causal mediation analysis. As these figures suggest, the number of independent suggestions a piece of advice included mediated the relationship between an advisor's performance and the perceived quality of that piece of advice.

In short, the best performers made more independent suggestions, and advisees believed that the advice that included more independent suggestions was better advice. About this they were wrong. In the Supplemental Material (Section 4.1), we describe two additional experiments (Studies S3a and S3b) that demonstrate the causal relationship between the number of independent suggestions a piece of advice contains and the perceived quality of that piece of advice.

General Discussion

Participants in our studies preferred to receive advice from advisors who performed well. They expected that advice to be more helpful before they implemented it, and they believed it had been more helpful after they implemented it, despite the fact that they were told nothing about their advisor's performance. These expectations and beliefs turned out to be wrong: Advice from the best-performing advisors was no more helpful than advice from any other advisors. So why did participants think it was? Although the best-performing advisors did not give better advice, they did give more

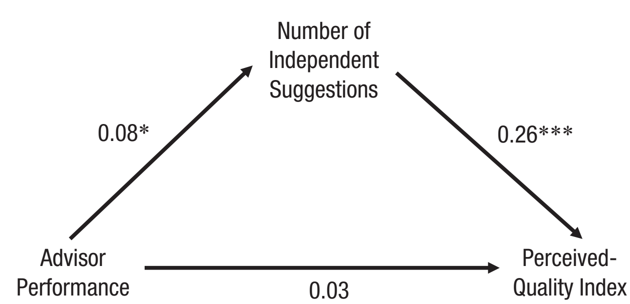

Fig. 7. Mediation analysis in Study 4: effect of advisor performance on perceived-quality index via number of independent suggestions. Unstandardized regression coefficients are shown from the mediation and the outcome regression models. Asterisks indicate significant paths (*$p < .05$, ***$p < .001$).

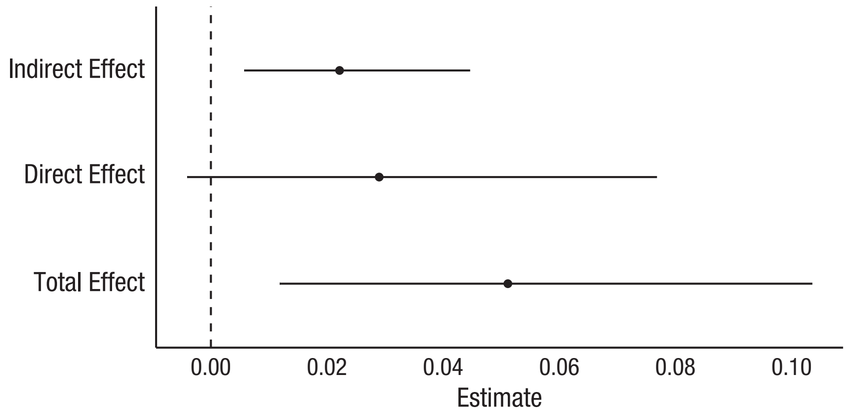

Fig. 8. Effect sizes from the mediation analysis in Study 4: indirect, direct, and total effect for a bias-corrected, bootstrapped mediation analysis examining whether number of independent suggestions mediates the relationship between advisor performance and perceived-quality index. Error bars represent quasi-Bayesian 95% confidence intervals. Estimates shown are unstandardized.

of it, and the number of independent suggestions an advisor made mediated the relationship between that advisor's performance and the perceived helpfulness of their advice. In short, advice from the best performers was not better. It just sounded better because there was more of it.

The fact that advice from the best performers was not especially helpful is counterintuitive. Why was this advice not more helpful? We can think of at least three reasons. First, highly skilled performers in many domains—from orators and artists to entrepreneurs and athletes—often execute their performances without thinking much about them because natural talent and extensive practice have made conscious thought unnecessary (Blessing & Anderson, 1996; Petersen et al., 1998; Tsay & Banaji, 2011) and even self-defeating (Flegal & Anderson, 2008). Consequently, these highly capable performers may not have especially useful things to say about their domains of excellence. A natural-born slugger who has played baseball every day since childhood may not think to tell a rookie about something they find utterly intuitive, such as balance and grip.

Second, even when an excellent performer does have explicit information to share, they may not be especially adept at sharing it. People vary widely in their communication skills, and most find it difficult to take the perspectives of others when trying to convey information (Epley et al., 2004). This may be especially true of skilled performers, who may no longer remember the challenges faced by novices (Chi et al., 1981; Hinds et al., 2001; Nickerson, 1999; Yaniv & Choshen-Hillel, 2012) and therefore may have difficulty taking the novices' perspectives (Fisher & Keil, 2016).

Third, the fact that the advisees in our studies did not benefit more from the advice of the best performers in our studies may say more about the former than the latter. For example, the best-performing advisors provided more independent suggestions than other performers did, and it is entirely possible that if advisees had actually followed these suggestions, they would have experienced dramatic improvement in their performance. But it is also possible that advisees were unable to follow more than one good suggestion at a time. The extensive advice of the best performers may well have been as excellent as it sounded, but its overabundance may have been wasted on those who tried to implement it. The point is that the most capable performers may well give lots of good advice, but if it is more good advice than most people can use, it will not be more effective than the less extensive advice given by less capable performers.

Do our results, then, suggest that people are wasting their time and money when they seek advice from the best performers? Yes, sometimes. But not always. First, participants in our studies were a convenience sample rather than a random sample, and thus our results may or may not generalize to other populations. Second, advice seeking can have benefits that are unrelated to improvements in performance, such as making the advice seeker appear competent (Brooks et al., 2015). Taking cooking lessons from Wolfgang Puck may enhance a person's reputation even if it does not enhance their braising technique and may therefore be worth the price. Third, our studies focused on a particular kind of advice—performance advice or "how to do" advice—which is quite different from decision advice or "what to do" advice (Dalal & Bonaccio, 2010). The best-performing advisors who do not provide superior performance advice may still provide superior decision advice. Even if Warren Buffett cannot effectively teach people how to invest, his stock tips may be worth heeding (or ignoring; see Leong & Zaki, 2018). Fourth, although the best-performing advisors in our studies did not provide better performance advice about Word Scramble, it is not difficult to think of domains in which an advisor's performance would indeed be a good predictor of the quality of their advice. A student who took cello lessons from Yo-Yo Ma would surely learn more than a student who took cello lessons from Warren Buffett or Wolfgang Puck, neither of whom plays the cello. For all these reasons, our results should not be taken to suggest that an advisor's performance is never a useful indicator of the quality of their advice. Rather, they simply suggest that in at least some ordinary situations in which ordinary people expect the best-performing advisors to provide the best performance advice, those ordinary people are likely to be mistaken. Tips from the top are not always worth top dollar.

Transparency

Action Editor: Kate Ratliff
Editor: Patricia J. Bauer
Author Contributions
All the authors developed the study concepts and contributed to the study designs. D. E. Levari collected and analyzed the data. All the authors drafted and revised the manuscript and approved the final version for submission.
Declaration of Conflicting Interests
The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.
Funding
This research was supported by the Harvard Initiative for Learning and Teaching (HILT).
Open Practices
Data, analysis scripts, and study materials have been made publicly available via Zenodo and can be accessed at https://doi.org/10.5281/zenodo.5121876. The design and analysis plans for the studies were not preregistered. This article has received the badge for Open Data. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

ORCID iDs

David E. Levari https://orcid.org/0000-0002-7953-7999
Daniel T. Gilbert https://orcid.org/0000-0002-9865-4445
Timothy D. Wilson https://orcid.org/0000-0003-3106-1212

Acknowledgments

We thank Mohammed Almadhoun, Christine Looser, Patrick Mair, and Enrique Meneses for technical support and Steven Worthington and Ista Zahn of the Institute for Quantitative Social Science at Harvard University for their statistical support. We also thank Mohin Banker, Sophie Carroll, Petrina Chan, Rachel Chmielinksi, Ashley Collinsworth, Irene Droney, Cailey Fitzgerald, Shannon Ganley, Molly Graether, James Green, Lauren Harris, Uriel Heller, Sarah Hoffman, Tanner Hicks, Emily Kemp, Benny Kollek, Rachel Lisner, Zoe Lu, Brenna Martinez, Tina Murphy, Debbie Park, Yeji Park, Michael Powell, Margo Sanders, Gemma Stern, Robin Stramp, Alexandro Strauss, Angela Wang, Christie Wu, and Yaojia Zheng for their assistance with this research.

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/09567976211054089

References

Bates, D. M., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Beilock, S. L., Carr, T. H., MacMahon, C., & Starkes, J. L. (2002). When paying attention becomes counterproductive: Impact of divided versus skill-focused attention on novice and experienced performance of sensorimotor skills. *Journal of Experimental Psychology: Applied*, *8*, 6–16.

Blessing, S. B., & Anderson, J. R. (1996). How people learn to skip steps. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 576–598.

Brooks, A. W., Gino, F., & Schweitzer, M. E. (2015). Smart people ask for (my) advice: Seeking advice boosts perceptions of competence. *Management Science*, *61*, 1421–1435.

Carpenter, S. K., Wilford, M. M., Kornell, N., & Mullaney, K. M. (2013). Appearances can be deceiving: Instructor fluency increases perceptions of learning without increasing actual learning. *Psychonomic Bulletin & Review*, *20*, 1350–1356.

Casciaro, T., & Lobo, M. S. (2008). When competence is irrelevant: The role of interpersonal affect in task-related ties. *Administrative Science Quarterly*, *53*, 655–684.

Champely, S. (2020). *pwr: Basic functions for power analysis* (R package Version 1.3-0) [Computer software]. https://CRAN.R-project.org/package=pwr

Chi, M., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*, 121–152.

Dalal, R. S., & Bonaccio, S. (2010). What types of advice do decision-makers prefer? *Organizational Behavior and Human Decision Processes*, *112*, 11–23.

DiCiccio, T. J., & Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science*, *11*, 189–228.

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology*, *87*, 327–339.

Fisher, M., & Keil, F. C. (2016). The curse of expertise: When more knowledge leads to miscalibrated explanatory insight. *Cognitive Science*, *40*, 1251–1269.

Flegal, K. E., & Anderson, M. C. (2008). Overthinking skilled motor performance: Or why those who teach can't do. *Psychonomic Bulletin & Review*, *15*, 927–932.

Gaertig, C., & Simmons, J. P. (2018). Do people inherently dislike uncertain advice? *Psychological Science*, *29*(4), 504–520. https://doi.org/10.1177/0956797617739369

Grant, A. (2018, August 28). Those who can do, can't teach. *The New York Times*. https://www.nytimes.com/2018/08/25/opinion/sunday/college-professors-experts-advice.html

Hattie, J., & Marsh, H. W. (1996). The relationship between research and teaching: A meta-analysis. *Review of Educational Research*, *66*, 507–542.

Hinds, P. J., Patterson, M., & Pfeffer, J. (2001). Bothered by abstraction: The effect of expertise on knowledge transfer and subsequent novice performance. *Journal of Applied Psychology*, *86*, 1232–1243.

Hofmann, D. A., Lei, Z., & Grant, A. M. (2009). Seeking help in the shadow of doubt: The sensemaking processes underlying how nurses decide whom to ask for advice. *Journal of Applied Psychology*, *94*(5), 1261–1274.

Hur, J. D., Ruttan, R. L., & Shea, C. T. (2020). The unexpected power of positivity: Predictions versus decisions about advisor selection. *Journal of Experimental Psychology: General*, *149*, 1969–1986.

Leong, Y. C., & Zaki, J. (2018). Unrealistic optimism in advice taking: A computational account. *Journal of Experimental Psychology: General*, *147*, 170–189.

Nickerson, R. S. (1999). How we know—and sometimes misjudge—what others know: Imputing one's own knowledge to others. *Psychological Bulletin*, *125*, 737–759.

Petersen, S. E., van Mier, H., Fiez, J. A., & Raichle, M. E. (1998). The effects of practice on the functional anatomy of task performance. *Proceedings of the National Academy of Sciences, USA*, *95*, 853–869.

Price, P. C., & Stone, E. R. (2004). Intuitive evaluation of likelihood judgment producers: Evidence for a confidence heuristic. *Journal of Behavioral Decision Making*, *17*, 39–57.

R Core Team. (2021). *R: A language and environment for statistical computing* (Version 4.0.5) [Computer software]. R Foundation for Statistical Computing.

Reed, N., McLeod, P., & Dienes, Z. (2010). Implicit knowledge and motor skill: What people who know how to catch don't know. *Consciousness and Cognition*, *19*, 63–76.

Schwartz, H. A., Rouhizadeh, M., Bishop, M., Tetlock, P., Mellers, B., & Ungar, L. (2017). Assessing objective recommendation quality through political forecasting. In M. Palmer, R. Hwa, & S. Riedel (Eds.), *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (pp. 2348–2357). Association for Computational Linguistics.

Speelman, C. (1998). Implicit expertise: Do we expect too much from our experts? In K. Kirsner, C. Speelman, M. Maybery, A. O'Brien-Malone, & M. Anderson (Eds.), *Implicit and explicit mental processes* (pp. 135–148). Psychology Press.

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). Mediation: R package for causal mediation analysis. *Journal of Statistical Software*, *59*(5). https://doi.org/10.18637/jss.v059.i05

Tsay, C. J., & Banaji, M. R. (2011). Naturals and strivers: Preferences and beliefs about sources of achievement. *Journal of Experimental Social Psychology*, *47*, 460–465.

Yaniv, I., & Choshen-Hillel, S. (2012). When guessing what another person would say is better than giving your own opinion: Using perspective-taking to improve advice-taking. *Journal of Experimental Social Psychology*, *48*, 1022–1028.

Zhang, T., Harrington, K. B., & Sherf, E. N. (2022). The errors of experts: When expertise hinders effective provision and seeking of advice and feedback. *Current Opinion in Psychology*, *43*, 91–95.