

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



David E. Levari¹, Daniel T. Gilbert², and
Timothy D. Wilson³

¹Harvard Business School, Harvard University; ²Department of Psychology, Harvard University; and

³Department of Psychology, University of Virginia

Psychological Science

1–14

© The Author(s) 2022

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/09567976211054089

www.psychologicalscience.org/PS



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.

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

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.

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

Answer	Participants who chose this 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: $M = 34.42$

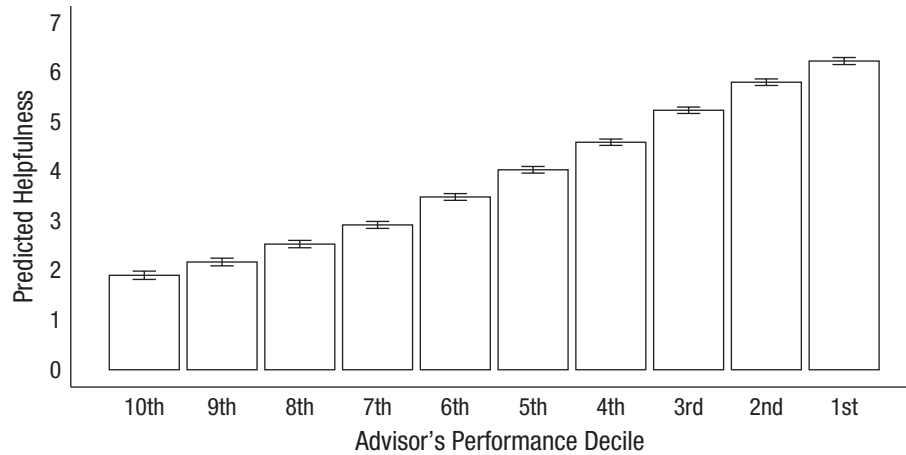


Fig. 1. Mean participant estimate of advice helpfulness in the decile condition of Study 1. The x-axis shows the decile of an advisor's performance (10th = worst, 1st = best), and the y-axis shows participants' estimates of the helpfulness of an advisor's advice. Error bars represent ± 1 SE.

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?

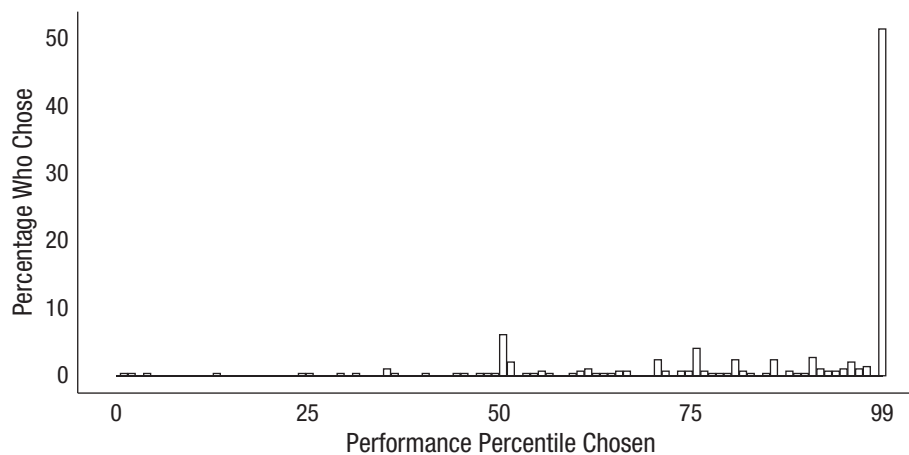


Fig. 2. Participants' preferences for advisors in the percentile condition of Study 1. The x-axis shows advisor performance in percentiles (1st = worst, 99th = best), and the y-axis shows the percentage of participants who reported that they would prefer to receive advice from an advisor in that percentile.

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.

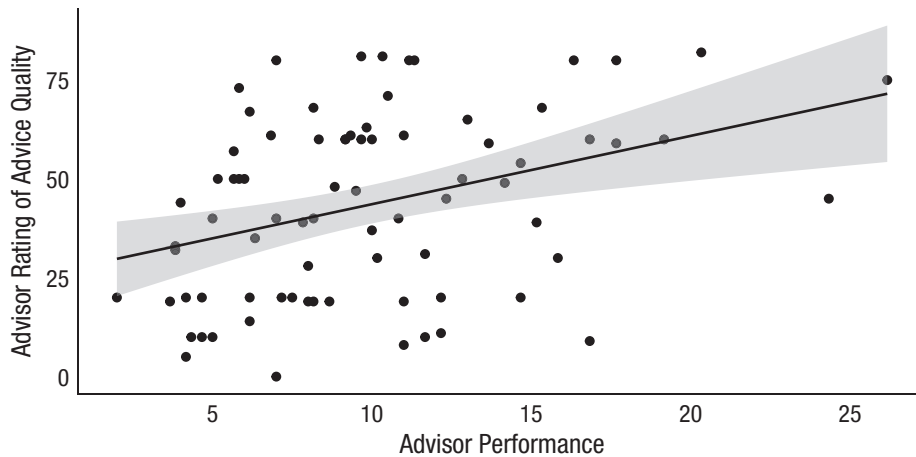


Fig. 3. Scatterplot showing the relationship between advisor performance and advisor rating of advice quality in Study 2. The *x*-axis shows each advisor's performance, the *y*-axis shows the advisor's estimates of the quality of their advice, and the error band shows the 95% confidence interval around 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, $b = 0.77$,

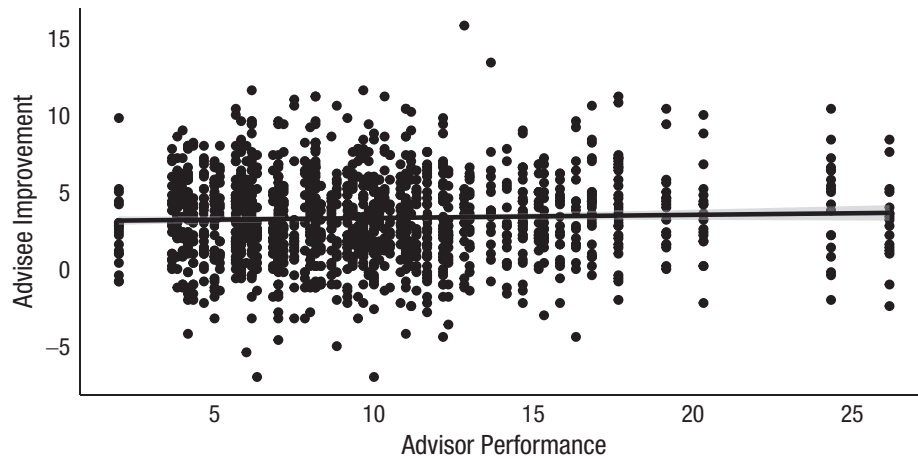


Fig. 4. Scatterplot showing the relationship between advisor performance and advisee improvement in Study 2. The x -axis shows advisor performance (average number of words found across six rounds), the y -axis shows advisee improvement, and the error band shows the 95% confidence interval around the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

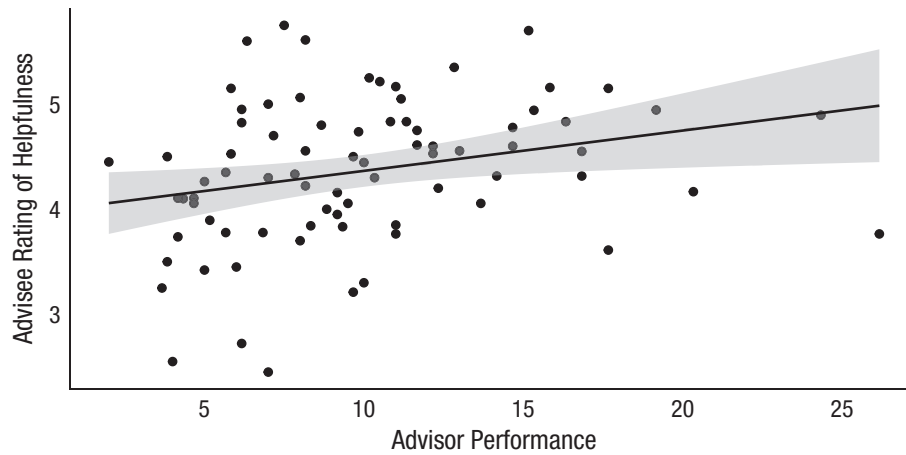


Fig. 5. Scatterplot showing the relationship between advisor performance and advisee rating of advice helpfulness in Study 2. The x -axis shows advisor performance (average number of words found per round), the y -axis shows advisee ratings of helpfulness, and the error band shows the 95% confidence 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, $r = .34$, $p < .01$

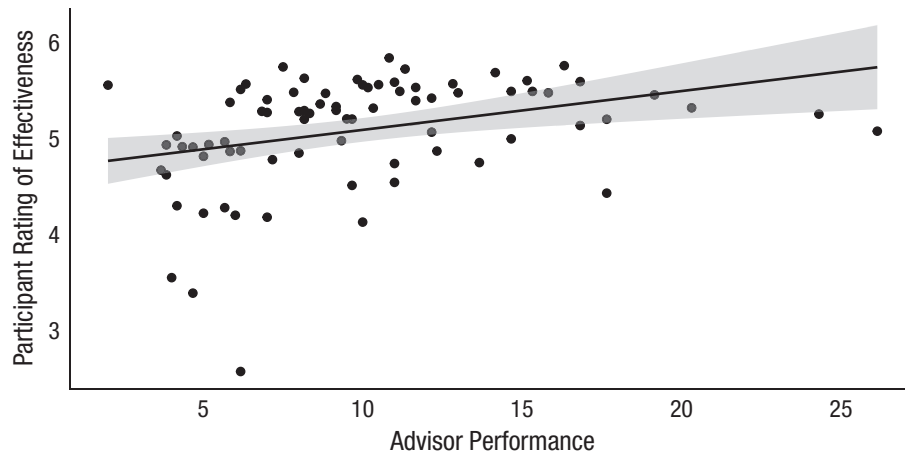


Fig. 6. Scatterplot showing the relationship between advisor performance and participant estimates of advice effectiveness in Study 3. The *x*-axis shows advisor performance (average words found per round), the *y*-axis shows participant ratings of advice effectiveness, and the error band shows the 95% 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 α s = .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	<i>t</i>	<i>p</i>
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

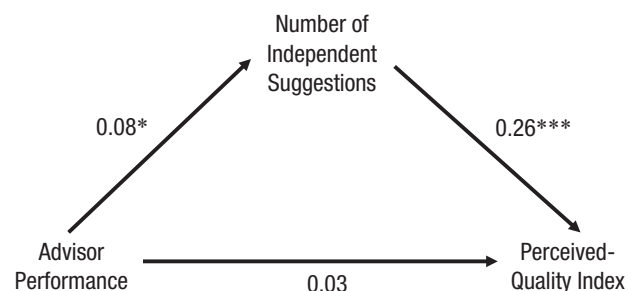


Fig. 7. Mediation analysis in Study 4: effect of advisor performance on perceived-quality index via number of independent suggestions. Unstandardized regression coefficients are shown from the mediation and the outcome regression models. Asterisks indicate significant paths (* $p < .05$, *** $p < .001$).

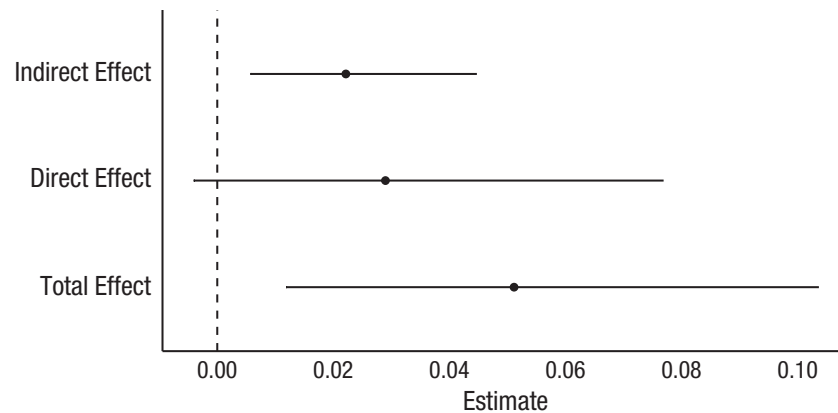


Fig. 8. Effect sizes from the mediation analysis in Study 4: indirect, direct, and total effect for a bias-corrected, bootstrapped mediation analysis examining whether number of independent suggestions mediates the relationship between advisor performance and perceived-quality index. Error bars represent 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 Chmielinski, Ashley Collinworth, 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.

SUPPLEMENT

Table of Contents

The entries below are hyperlinked to subsequent pages in this document. Control-click on an entry to navigate to its page.

(Section 1.1) Exploratory Items in Study 1	2
(Section 1.2) Categorizations in Study 1	4
(Section 2.1) Exploratory Items for Advisors in Study 2	5
(Section 2.2) Exploratory Items for Advisees in Study 2	8
(Section 2.3) Additional Studies S1 and S2.....	11
(Section 2.4) Tests for Chance Performance in Studies 2, S1, and S2	25
(Section 2.5) Advice Offered by Advisors in Studies 2, S1, and S2	28
(Section 2.6) Analyses of Curvilinear Relationships in Studies 2, S1, and S2.....	50
(Section 2.7) Analyses of the Relationship Between Measures in Studies 2, S1, and S2	52
(Section 2.8) Analysis of Advisor Improvement in Studies 2, S1, and S2.....	57
(Section 3.1) Exploratory Items in Study 3	58
(Section 4.1) Additional Studies S3a and S3b.....	61
(Section 4.2) Additional Studies S4a and S4b.....	66
References.....	73

(Section 1.1) Exploratory Items in Study 1

1. The game you played today was called Word Scramble, also known as "Boggle." Have you played this game before today, either on or off of the computer?

- Yes, I've played some version of this game before
- No, I've never played any version of this game before
- I'm not sure

2. [if yes to Q1] You indicated that you've played some version of Word Scramble before. Before today's study, when was the most recent time that you played a version of Word Scramble?

- Within the last week
- Within the last month
- Within the last 6 months
- Within the last year
- Over a year ago

3. [if yes to Q1] As best as you can recall, when was the first time that you played a version of Word Scramble?

- Within the last month
- Within the last 6 months
- Within the last year
- Within the last 3 years
- Within the last 5 years
- More than 5 years ago

4. Approximately how often do you play word-finding games like Word Scramble? (others include Scrabble, Bananagrams, etc.)

- A few times a week
- A few times a month
- A few times a year
- Less than once a year
- I never play word-finding games

5. How skilled do you think you were with Word Scramble before you took this study, compared to most people?

- I was less skilled than most people
- I was as skilled as most people
- I was more skilled than most people

6. How experienced do you think you were with Word Scramble before you took this study, compared to most people?

- I was less experienced than most people
- I was as experienced as most people
- I was more experienced than most people

7. How are you taking this survey right now?

- With a computer mouse
- With a trackpad
- Pointing Stick
- Touch Screen
- Other

8. How old are you?

9. Please indicate your gender.

- Male
- Female
- Prefer not to answer

10. Previous work in behavioral decision-making has shown that people vary in the amount they pay attention to these sorts of surveys. If you're actually reading this question, please answer it by selecting the option "other" and typing in the word "potato." Thanks for your help, and for taking the time to read all of the questions.

- Gender, marital status, and education
- Sexual orientation, relationship satisfaction
- Hobbies, favorite sports
- Other

11. Is English your native language?

- Yes, I only spoke English growing up
- Yes, I spoke English and these other languages growing up:
- No, English is not my native language. My native language is:

12. If you encountered any errors, glitches or other problems taking this survey, please tell us below. Otherwise, leave this field blank.

13. What do you think this study was about?

14. If you have any other comments about the study you took part in today, please tell us below.

(Section 1.2) Categorizations in Study 1

Participants in the free response condition of Study 1 answered the question: “Whose advice do you think would improve your performance the most?” Two coders who were blind to the hypothesis read all responses and classified each response as belonging to one of 12 categories. Table S1 lists these categories and shows the percentage of participants whose responses were coded as belonging to each category.

Table S1

Categories used in Study 1

Category	All participants	Participants not coded as “No answer/did not understand question”
[No answer/did not understand question]	37.86	N/A
The advisor with the highest performance	19.42	31.25
An advisor with high performance	16.50	26.56
Whichever advisor gave advice “X”	10.68	17.19
The advisor who improved the most over time	2.91	4.69
[Unspecified/unclear]	2.91	4.69
An advisor like myself	1.94	3.13
Myself	1.94	3.13
An advisor who is a woman	1.94	3.13
The advisor who had the worst performance	1.94	3.13
An advisor with experience	0.97	1.56
The advisor who paid the best attention to the directions	0.97	1.56

(Section 2.1) Exploratory Items for Advisors in Study 2

1. The game you played today was called Word Scramble, also known as "Boggle." Have you played this game before today, either on or off of the computer?

- Yes, I've played some version of this game before
- No, I've never played any version of this game before
- I'm not sure

2. [if yes to Q1] You indicated that you've played some version of Word Scramble before. Before today's study, when was the most recent time that you played a version of Word Scramble?

- Within the last week
- Within the last month
- Within the last 6 months
- Within the last year
- Over a year ago

3. [if yes to Q1] As best as you can recall, when was the first time that you played a version of Word Scramble?

- Within the last month
- Within the last 6 months
- Within the last year
- Within the last 3 years
- Within the last 5 years
- More than 5 years ago

4. How experienced do you think you were with Word Scramble before you took this study, compared to most people?

- I was less experienced than most people
- I was as experienced as most people
- I was more experienced than most people

5. How skilled do you think you were with Word Scramble before you took this study, compared to most people?

- I was less skilled than most people
- I was as skilled as most people
- I was more skilled than most people

6. How well do you think the average person does on one round of the game you played today?
Average number of words found in 1 minute: _____ words

7. Approximately how often do you play word-finding games like Word Scramble? (others include Scrabble, Bananagrams, etc.)

- A few times a week
- A few times a month
- A few times a year
- Less than once a year
- I never play word-finding games

8. How are you taking this survey right now?

- With a computer mouse
- With a trackpad
- Pointing Stick
- Touch Screen
- Other

9. What internet browser are you using?

- Safari
- Internet Explorer
- Mozilla Firefox
- Google Chrome
- Opera
- Other:
- I'm not sure

10. Earlier in this study, you wrote advice to a future participant who will play the Word Scramble game. How much do you think your advice will help them improve? [0% improvement (not at all) to 100% improvement or more (a lot)]

11. In the box below, please type the names of as many animals as you can think of in one minute. Please only type those that you can think of yourself, without help. Separate each animal name with a comma.

12. How old are you?

13. Please indicate your gender?

- Male
- Female
- Prefer not to answer

14. Previous work in behavioral decision-making has shown that people vary in the amount they pay attention to these sorts of surveys. If you're actually reading this question, please answer it by selecting the option "other" and typing in the word "orange." Thanks for your help, and for taking the time to read all of the questions.

- Gender, marital status, and education
- Sexual orientation, relationship satisfaction
- Hobbies, favorite sports
- Other

15. Is English your native language?

- Yes, I only spoke English growing up
- Yes, I spoke English and these other languages growing up:
- No, English is not my native language. My native language is:

16. If you encountered any errors, glitches or other problems taking this survey, please tell us below. Otherwise, leave this field blank.

17. What do you think this study was about?

18. If you have any other comments about the study you took part in today, please tell us below.

(Section 2.2) Exploratory Items for Advisees in Study 2

1. The game you played today was called Word Scramble, also known as "Boggle." Have you played this game before today, either on or off of the computer?

- Yes, I've played some version of this game before
- No, I've never played any version of this game before
- I'm not sure

2. [if yes to Q1] You indicated that you've played some version of Word Scramble before. Before today's study, when was the most recent time that you played a version of Word Scramble?

- Within the last week
- Within the last month
- Within the last 6 months
- Within the last year
- Over a year ago

3. [if yes to Q1] As best as you can recall, when was the first time that you played a version of Word Scramble?

- Within the last month
- Within the last 6 months
- Within the last year
- Within the last 3 years
- Within the last 5 years
- More than 5 years ago

4. Approximately how often do you play word-finding games like Word Scramble? (others include Scrabble, Bananagrams, etc)

- A few times a week
- A few times a month
- A few times a year
- Less than once a year
- I never play word-finding games

5. Earlier in this study, you received advice from a past participant who played the Word Scramble game. Compared to most people, how skilled do you think the advice writer was at Word Scramble?

- They were less skilled than most people
- They were about as skilled as most people
- They were more skilled than most people

6. Compared to you, how skilled do you think the advice writer was at Word Scramble?

- They were less skilled than me
- They were about as skilled as me
- They were more skilled than me

7. How helpful was the advice you read about Word Scramble? 1 - Extremely unhelpful, 7 - Extremely helpful

8. Earlier in this study, you read advice from a previous participant who played the Word Scramble game. How much do you think their advice helped you improve? [0% improvement (not at all) to 100% improvement or more (a lot)]

9. How experienced do you think you were with Word Scramble before you took this study, compared to most people?

- I was less experienced than most people
- I was as experienced as most people
- I was more experienced than most people

10. How skilled do you think you were with Word Scramble before you took this study, compared to most people?

- I was less skilled than most people
- I was as skilled as most people
- I was more skilled than most people

11. How well do you think the average person does on one round of the game you played today? Average number of words found in 1 minute: _____ words

12. How well do you think the person who wrote your advice does on one round of the game you played today? Average number of words found in 1 minute: _____ words

13. How are you taking this survey right now?

- With a computer mouse
- With a trackpad
- Pointing Stick
- Touch Screen
- Other

14. What internet browser are you using?

- Safari
- Internet Explorer
- Mozilla Firefox
- Google Chrome
- Opera
- Other
- I'm not sure

15. Earlier in this study, you wrote advice to a future participant who will play the Word Scramble game. How much do you think your advice will help them improve? [0% improvement (not at all) to 100% improvement or more (a lot)]

16. In the box below, please type the names of as many animals as you can think of in one minute. Please only type those that you can think of yourself, without help. Separate each animal name with a comma.

17. How old are you?

18. Please indicate your gender?

- Male
- Female
- Prefer not to answer

19. Previous work in behavioral decision-making has shown that people vary in the amount they pay attention to these sorts of surveys. If you're actually reading this question, please answer it by selecting the option "other" and typing in the word "carrot." Thanks for your help, and for taking the time to read all of the questions.

- Gender, marital status, and education
- Sexual orientation, relationship satisfaction
- Hobbies, favorite sports
- Other

20. Is English your native language?

- Yes, I only spoke English growing up
- Yes, I spoke English and these other languages growing up:
- No, English is not my native language. My native language is:

21. If you encountered any errors, glitches or other problems taking this survey, please tell us below. Otherwise, leave this field blank.

22. What do you think this study was about?

23. If you have any other comments about the study you took part in today, please tell us below.

(Section 2.3) Additional Studies S1 and S2

Below we report the methods and results of two additional studies that were modeled on Study 2 but used different tasks.

STUDY S1:

DO THE BEST-PERFORMERS (IN DARTS) GIVE THE BEST ADVICE?

Method

Overview

We assigned a group of participants (the “advisors”) to play a game of darts, and then to write advice for future participants. Next, we gave that advice to another group of participants (the “advisees”) and measured its impact on their performances.

Procedure for Advisors

Advisors were 83 adults recruited in a public, outdoor area with frequent foot traffic in a university in the Northeastern United States (45 males, 38 females, $M_{\text{age}} = 32.53$ years, $SD = 14.83$ years) who were offered candy for their participation.

Advisors were told that they would be playing a game of darts and then answering questions about it. Advisors were then shown a circular magnetized dartboard that was 40.3 cm in diameter, mounted on a 1.3 m wooden pole, along with a set of magnetic darts. We will hereinafter refer to the dartboard as a “board.” Advisors were told that they would throw twelve darts at the board and each throw would be scored by its proximity to the center of the board, known as the “bullseye” (Critcher & Rosenzweig, 2014). Advisors stood approximately 2 m from the dartboard, completed one practice throw that was not scored, and then completed 12 successive throws that were scored. After each of the scored throws, a research assistant recorded

the position of the dart on the board and then removed it from the board so that it would not obstruct subsequent darts.

Next, advisors were asked to write advice to help a future participant (the “advisee”) who would be playing the same game and would know all the rules. Advisors then estimated how helpful their advice would be using a six-point scale whose endpoints were labeled *extremely unhelpful* (1) and *extremely helpful* (6). Advisors also indicated whether they had played darts before, and if so, how long and how often they had played. They also indicated whether they thought they performed at the game better or worse than the average person taking the study on a five-point scale whose endpoints were labeled *much worse than average* (1) and *much better than average* (5). They also indicated whether they had given someone advice about darts before, and provided their age, whether they were a native English speaker, and their gender. Finally, they answered a free-response question about what they thought the purpose of the study was, and whether they had any additional comments about the study.

Procedure for Advisees

Advisees were 333 adults recruited in a public, outdoor area with frequent foot traffic in a university in the Northeastern United States (173 males, 154 females, 3 other, 3 prefer not to answer, $M_{\text{age}} = 31.01$ years, $SD = 12.74$ years) who were offered candy in exchange for their participation.

Advisees were introduced to the darts game in the same way advisors were. They completed one practice throw that was not scored and six subsequent throws that were scored. Next, advisees were randomly assigned to one of two conditions. Advisees in the *advice* condition ($N = 232$) were shown a piece of advice that had been written by a randomly-selected advisor, and advisees in the *no advice* condition ($N = 101$) were not shown any advice (though

for narrative convenience we will continue to refer to these participants as advisees). Advisees in both conditions then completed six more throws that were scored.

Next, advisees in the advice condition rated the quality of the advice they had received using a six-point scale whose endpoints were labeled *extremely unhelpful* (1) and *extremely helpful* (6), estimated how much the advice had improved their performance by drawing a slash on a horizontal line whose endpoints were labeled *0% improvement (not at all)* and *100% improvement or more (a lot)*, and indicated whether or not they had used the advice they received. Advisees then completed the same basic demographic measures and questions about previous experience playing darts as the advisors did. Advisees also indicated whether they had played darts before, and if so, how long and how often they had played. They also provided their age, whether they were a native English speaker, and their gender. Finally, they answered a free-response question about what they thought the purpose of the study was, and whether they had any additional comments about the study.

Results

Advisors

We excluded data from two advisors who reported not being fluent in English, and three advisors who failed to provide advice to future participants. This left 78 advisors (44 males, 34 females, $M_{\text{age}} = 32.32$ years, $SD = 14.59$ years) in the dataset. Excluding these advisors does not appreciably change the results of any of the analyses we performed on the advisor data.

Advisors’ performances. We measured the distance between the dart and the bullseye in centimeters. We then rescaled these values so that the best throw (0 cm from the bullseye) was worth 10 points and the worst possible throw (missing the dartboard) was worth zero points. The 78 advisors scored an average of 5.29 ($SD = 1.63$) points per throw, meaning their throws landed

an average of 9.52cm from the bullseye. To determine whether advisors improved over time, we fit a linear mixed model to our data in R (R Core Team, 2017) using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015). The dependent measure was the advisor’s performance (i.e., the distance between their dart and the bullseye) and the independent within-participants variable was the number of the throw (1 through 12). As random effects, we included a random intercept term for the advisor, and a random slope term for the number of the throw. The inclusion of random intercepts significantly improved model fit, $\chi^2(2) = 88.08$, $p < 0.0001$, but the inclusion of random slopes did not, $\chi^2(2) = 5.61$, $p = 0.06$. Results showed that the number of the round was positively associated with performance, $b = 0.10$, $SE = 0.03$, $t = 4.16$, $p < 0.001$, 95% CI [0.05, 0.15], which is to say that advisors improved with practice.¹

Advisors’ ratings of their advice. Advisors wrote an average of 36.51 (SD = 23.89) words of advice and expected their advice to be relatively helpful: $M = 3.65$, $SD = 1.10$ on a scale whose endpoints were labeled *extremely unhelpful* (1) and *extremely helpful* (6). The advisors’ ratings of the quality of their advice were correlated with their own performances, $r = 0.37$, $p < 0.001$, which is to say that the best-performing advisors believed they had provided the best advice.

Advisees

We excluded data from five advisees who reported not being fluent in English, eleven advisees who did not complete all the parts of the study due to experimenter error, and nine advisees who refused to follow the task instructions and complete the study. This left 308 advisees (158 males, 144 females, 3 other, 3 prefer not to answer, $M_{\text{age}} = 30.73$ years, $SD = 12.21$

¹ This result is unchanged when random slopes are not included in the model.

years) in the dataset. Excluding these advisees does not appreciably change the results of any of the analyses we performed on the advisee data.

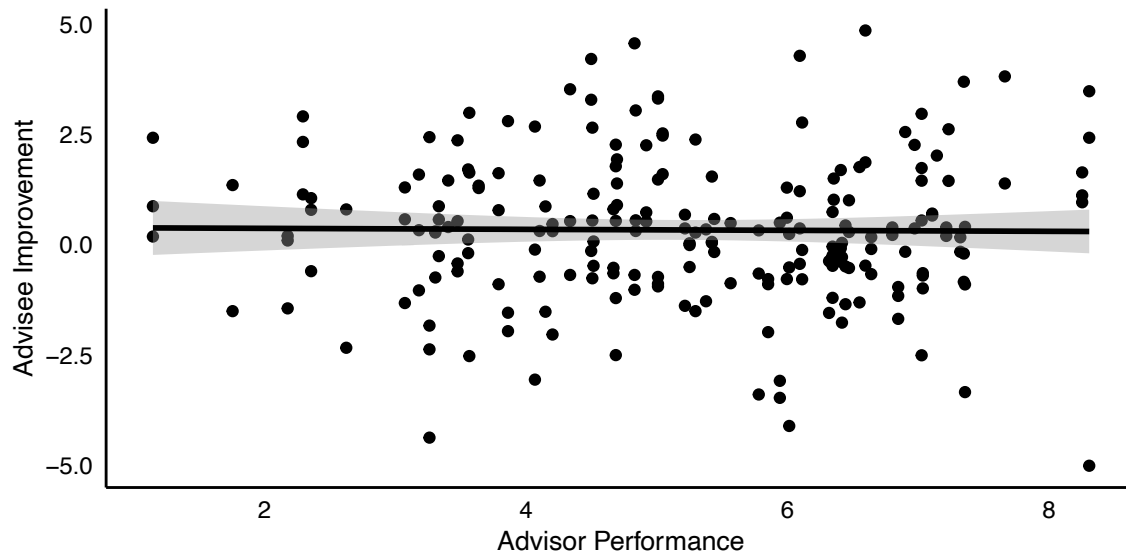
Advisees’ performances. Did receiving advice improve advisees’ performances? To answer this question, we first compared the average performance of advisees in the advice and no-advice conditions on the first six throws, which advisees made before being assigned to condition. As expected, there were no significant differences between conditions on the first six throws, $t(179.67) = -1.21$, $d = 0.15$, $p = 0.23$. Next, we computed each advisee’s improvement by subtracting their average performance on the first six throws from their average performances on the last six throws. Advisees in the advice condition showed no more improvement ($M = 0.34$ points) than did advisees in the no-advice condition ($M = 0.53$ points), $t(201.81) = -1.00$, $d = 0.12$, $p = 0.32$. In other words, on average, receiving advice did not improve advisees’ performances.

Although advice did not help on average, it was still possible to ask whether an advisor’s performance predicted the amount of improvement their advice produced. To answer this question, we analyzed the performances 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 distance from the bullseye); the independent within-participants variable was the number of the scored throw (1- 6); and the independent between-participants variable was the advisor’s average performance across twelve scored throws. An interaction term between the two predictors did not improve model fit, $\chi^2(1) = 0.12$, $p = 0.73$, and was dropped. As random effects, we included a random slope term for the throw and a random intercept term for the advisee. The random slope term significantly improved model fit, $\chi^2(2) = 7.42$, $p = 0.02$, as did the random slope term, $\chi^2(2) = 171.64$, $p < 0.001$.

The analysis revealed that the number of the round predicted advisee performance, $b = 0.07$, $SE = 0.02$, $t = 3.95$, $p < 0.001$, 95% CI [0.03, 0.10], which is to say that on average, advisees in the advice condition improved over time. However, there was no main effect of advisor’s performance on advisee’s performance, $b = -0.08$, $SE = 0.07$, $t = -1.12$, $p = 0.26$, 95% CI [-0.22, 0.06]. Figure S1 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-performing advisors was no more helpful than the advice from other advisors.

Figure S1

Relationship Between Advisor Performance and Advisee Improvement in Study S1



Note: The x-axis shows advisor performance (i.e., the average distance between the dart and the bullseye across twelve throws) and the y-axis shows advisee improvement.

Advisee’s ratings of advice. Advisees believed that the advice they received was generally helpful ($M = 3.77$ on a scale ranging from 1 to 6) and that receiving it had improved their performances ($M = 41\%$ 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. There was a significant correlation between advisor’s performance and the perceived helpfulness of their advice ($r = 0.22$, $p = 0.05$), but not between advisor’s performance and the perceived improvement that advice produced ($r = 0.20$, $p = 0.08$). Advisees mistakenly believed that the advice of the best-performing advisors was more helpful and produced more improvement than did the advice of other performers. It is worth keeping mind that advisees knew nothing about their advisor’s performance.

STUDY S2:

DO THE BEST PERFORMERS (IN COLORLINK) GIVE THE BEST ADVICE?

Method

Overview

We assigned a group of participants (the “advisors”) to play a game called Colorlink, and then to write advice for future participants. Next, we gave that advice to another group of participants (the “advisees”) and measured its impact on their performances.

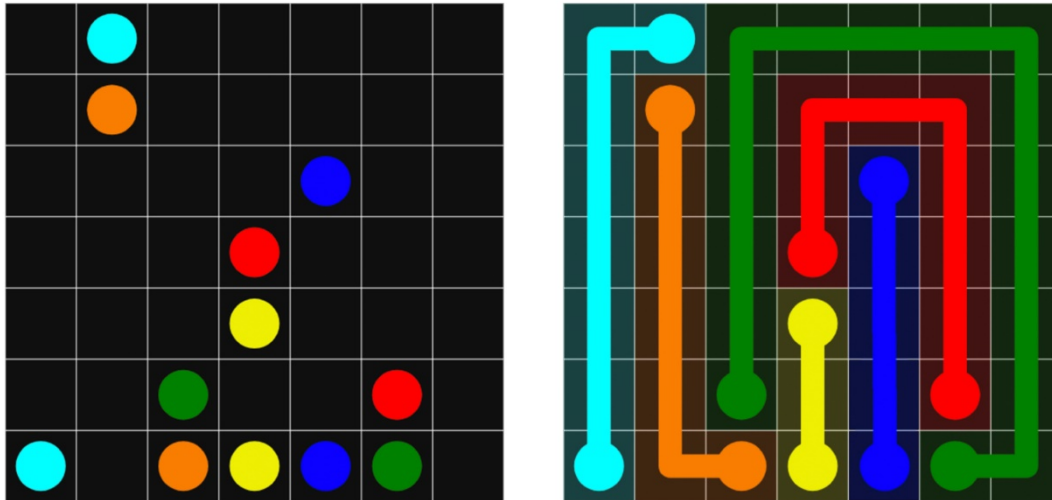
Procedure for Advisors

Advisors were 100 users of Amazon Mechanical Turk (62 males, 38 females, $M_{age} = 31.93$ years, $SD = 8.42$ years) who were paid for their participation.

Advisors were told that they would be playing a game called “Colorlink” and then answering questions about it. Advisors were then shown a grid of 49 squares, as shown in Figure S2. We will hereinafter refer to the grid as a “puzzle.” In each puzzle, 35 squares were empty and the remaining 14 squares were filled colored circles. Advisors were told that they would play 10 rounds of the game, and that each round would feature a different puzzle. Advisors were told that to solve the puzzle they must draw a connecting line between each dot and the other dot of the same color, with the stipulation that (a) the lines they drew could not intersect, and (b) no square could be left empty. Advisors were told that until they completed one puzzle, they could not move on the next puzzle. They were told that they were being timed and that the goal of the game was to complete all ten puzzles as quickly as possible. Advisors completed a practice puzzle and then played ten more rounds. Each advisor saw the same ten puzzles in a different random order. We pretested the ten puzzles to be of about equal difficulty.

Figure S2

Example of a Colorlink puzzle before and after completion



Note: The left panel shows a puzzle before completion and the right panel shows the same puzzle after completion.

Next, advisors were asked to write advice to help a future participant (the “advisee”) who would be playing the same game, knew all the rules, and had 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. Advisors then estimated how much their advice would improve a future participant’s performance on a 101-point scale whose endpoints were labeled *0% improvement (not at all)* and *100% improvement or more (a lot)*. Participants also completed the attention check item used in Study 1, and indicated whether they had played Colorlink or other related games before, and if so, for how long and how often. They also indicated whether they thought they were more or less experienced and more or less skilled at the game than most people. They also estimated how long it would take the average person to complete one Colorlink puzzle, indicated which web browser they were using and whether they were using a computer mouse or

other kind of input device, and reported whether they were colorblind, their age, gender, and native language. Finally, they answered an open-ended question about what they thought the purpose of the study was, and were given the opportunity to make any additional comments they wished.

Procedure for Advisees

Advisees were 2435 users of Amazon Mechanical Turk (1215 males, 1210 females, 10 prefer not to answer, $M_{\text{age}} = 34.53$ years, $SD = 8.09$ years) who were paid for their participation.

Advisees were introduced to the Colorlink puzzle game in the same way advisors were. They completed one practice puzzle and were then assigned to one of two conditions. Advisees in the *advice* condition ($N = 646$) were shown a piece of advice written by a randomly selected advisor, and advisees in the *no-advice* condition ($N = 1789$) were not shown any advice (though for narrative convenience we will continue to refer to these participants as advisees). Advisees in both conditions then completed five additional puzzles.

Next, advisees in the advice condition rated the quality of the advice they had received on a seven-point scale whose endpoints were labeled *extremely unhelpful* (1) and *extremely helpful* (7), and then estimated how much the advice had improved their performance on a 101-point scale whose endpoints were labeled *0% improvement (not at all)* and *100% improvement or more (a lot)*. Advisees in both conditions also completed the same attention check item and exploratory and demographic measures that advisors saw.

Results

Advisors

We excluded data from 27 advisors who reported having played Colorlink before, and one advisor whose time needed to finish the experiment was 4 standard deviations higher than

the mean. This left 72 advisors (44 males, 28 females, $M_{\text{age}} = 32.90$ years, $SD = 8.25$ years) in the dataset. Excluding these advisors does not appreciably change the results of any of the analyses we performed on the data.

Advisors’ performances. We measured how long it took advisors to complete each puzzle. It took the 72 advisors an average of 49.12 seconds ($SD = 22.03$ seconds) to complete each puzzle. To determine whether advisors improved over time, we fit a linear mixed model to our data in R using the lme4 package. The dependent measure was the advisor’s performance (i.e., the time it took to complete all ten puzzles). In this and all other reported trial-level analyses of completion times, we applied a log transformation to the completion-time data to reduce the influence of outliers and skew, as recommended by Whelan (2008). The independent within-participants variable was the number of the puzzle (1 through 10). As a random effect, we included a random intercept term for the advisor, and a random slope term for the number of the puzzle. The addition of random intercepts improved model fit, $\chi^2(2) = 55.35$, $p < 0.001$, but the inclusion of random slopes did not, $\chi^2(2) = 1.99$, $p = 0.37$, so they were dropped from the model. Results showed that the number of a puzzle was associated with performance, $b = -0.04$, $SE = 0.01$, $t = -5.70$, $p < 0.0001$, 95% CI $[-0.05, -0.03]$, which is to say that advisors improved with practice.

Advisors’ ratings of their advice. Advisors wrote an average of 36.08 ($SD = 23.11$) words of advice and expected their advice to improve an advisee’s performance by 48.31% ($SD = 25.21\%$). Advisors’ ratings of the quality of their advice were not significantly correlated with their own performances, $r = -0.05$, $p = 0.66$, which is to say that the best-performing advisors did not believe that they had provided the best advice.

Advisees

Advisees’ performances. No data were excluded. Did receiving advice improve advisees’ performances?. To answer this question, we first compared the average performance of advisees in the advice and no-advice conditions on the first puzzle, which advisees played before being assigned to condition. As expected, there were no differences between conditions on the first puzzle, $t(1083.5) = 0.51$, $d = 0.02$, $p = 0.61$. Next, we computed each advisee’s improvement by subtracting their performance on the first puzzle from the average of their performances on the five subsequent puzzles. Advisees in the advice condition showed no more improvement ($M = -18.65$ seconds) than did advisees in the no-advice condition ($M = -17.77$ seconds), $t(1057.9) = 0.36$, $d = 0.02$, $p = 0.72$. In other words, on average, receiving advice did not improve advisees’ performances.

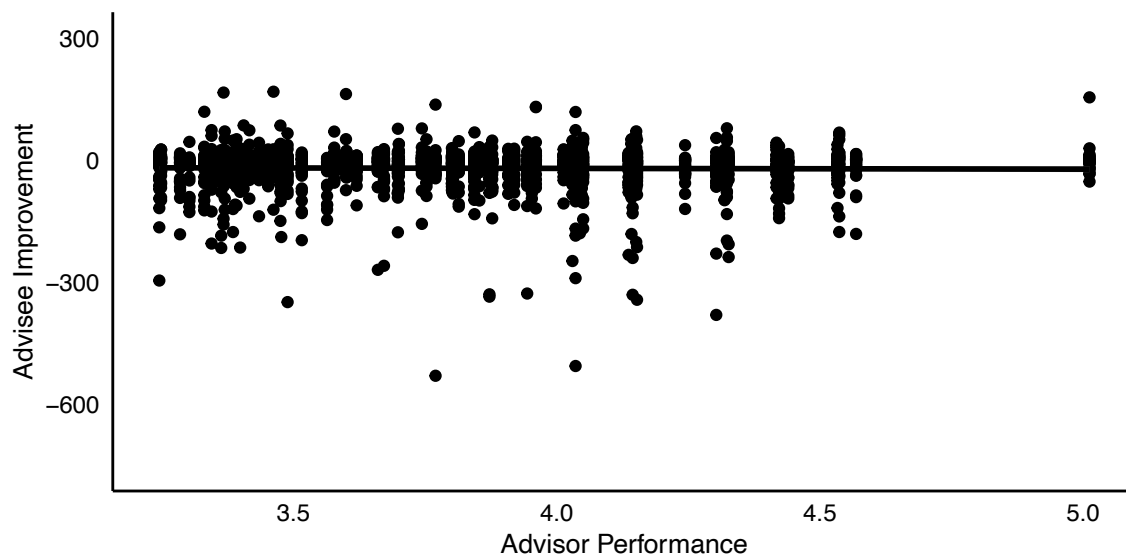
Although advice did not help on average, it was still possible to ask whether an advisor’s performance predicted the amount of improvement their advice produced. To answer this question, we analyzed the performances of advisees in the advice condition by fitting a linear mixed model using the lme4 package. The dependent measure was the advisee’s performance (i.e., the time needed to solve the puzzle); 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 ten rounds. We also included an interaction term between these two independent variables. As random effects, we included a random intercept term for the advisee, and a random slope term for the round number. The random intercept term improved model fit, $\chi^2(2) = 1415.70$, $p < 0.001$, as did the random slope term, $\chi^2(2) = 51.25$, $p < 0.001$.

The analysis revealed that the number of the round predicted advisee performance, $b = -0.12$, $SE = 0.03$, $t = -4.73$, $p < 0.001$, 95% CI [-0.18, -0.07], which is to say that on average,

advisees in the advice condition improved over time. However, there was no main effect of advisor’s performance on advisee’s performance, $b = 0.02$, $SE = 0.04$, $t = 0.53$, $p = 0.59$, 95% CI $[-0.06, 0.10]$, and no interaction between the number of the round and the advisor’s performance, $b = 0.01$, $SE = 0.01$, $t = 1.45$, $p = 0.15$, 95% CI $[-0.004, 0.02]$. Figure S3 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 top performers was no more helpful than the advice from bottom performers.

Figure S3

Relationship Between Advisor Performance and Advisee Improvement in Study S2



Note: The x-axis shows advisor performance (average time needed to complete a puzzle), log transformed, and the y-axis shows advisee improvement. Higher values represent worse performance on the task.

Advisee’s ratings of advice. Advisees believed that the advice they received was generally helpful ($M = 4.30$ on a scale ranging from 1 to 7) and that receiving it had improved

their performances ($M = 35\%$ 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. We found a correlation between advisor's performance (time needed to solve the puzzles) and the perceived helpfulness of their advice ($r = -0.32, p < 0.01$), and with the perceived improvement that advice produced ($r = -0.24, p = 0.04$). Advisees mistakenly believed that the advice of the best-performing advisors was more helpful and produced more improvement than did the advice of other performers. However, as in Study 2, advisees were never told how well their advisors had performed.

(Section 2.4) Tests for Chance Performance in Studies 2, S1, and S2

The best-performing advisors in Studies 2, S1, and S2 provided advice that was no more effective than the advice provided by other advisors. This finding would be unremarkable if performance on these tasks were entirely due to chance. If a group of people were asked to predict a series of coin flips, some would do better than others by chance and the fact that the best performers did not give better “prediction advice” than the worst performers would be unremarkable.

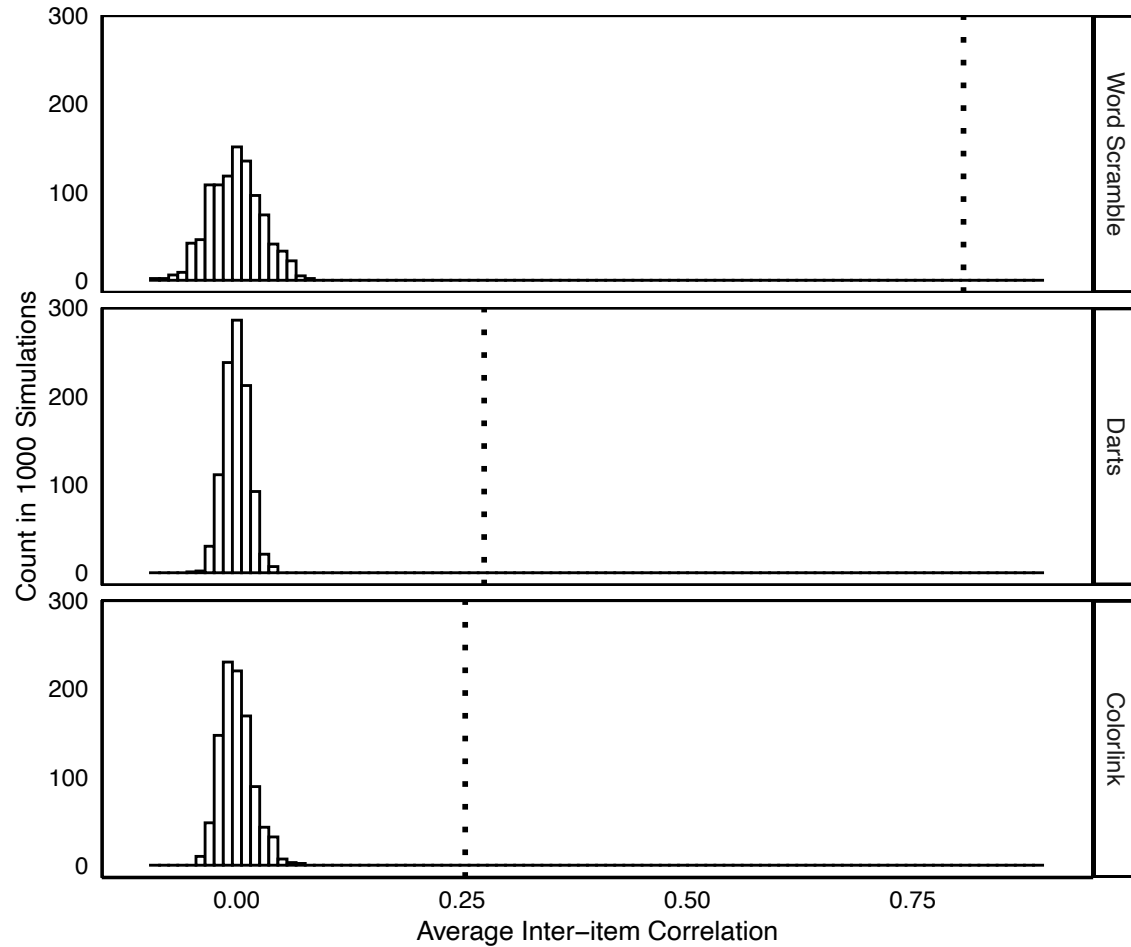
Is it possible that the best-performing advisors in Studies 2, S1, and S2 succeeded by chance alone? On the face it, that seems unlikely. First, one has a 50% chance of succeeding at coin flip prediction by chance alone, but the odds of successfully finding words, throwing bullseyes, or connecting colored dots by chance alone are vanishingly small. Second, advisors in Studies 2, S1, and S2 all improved with practice—and practice improves skilled performance but not chance performance. Third, we performed a permutation test (also known as a randomization test; Dwass, 1957) on the advisor data from Studies 2, S1, and S2. For each study, we created a matrix of pairwise correlations between all advisor performance rounds (Study 2: 6 boards; Study S1: 12 dart throws; Study S2: 10 puzzles). We excluded correlations that were equal to 1, which represented scores correlated with themselves, and we averaged the remaining values together to calculate an average inter-item correlation of advisor performances in each task. We then randomized the advisor data in each task by shuffling the values within each round, and recalculated the average inter-item correlation for the new, shuffled dataset. Repeating this permutation procedure 1000 times allowed us to create a distribution of plausible average inter-item correlations in each task if performance across rounds were random. We then compared this

distribution of simulated correlations to the observed inter-item correlation for actual participants.

Figure S4 shows the results of the permutation tests for advisors in Studies 2, S1, and S2. In Study 2, the permutation test for the Word Scramble Task resulted in a distribution of average inter-item correlations with a mean of 0.001 (SD = 0.03) while the observed average inter-item correlation in the real data was 0.81, $p < 0.0001$ for 1000 simulations. In Study S1, the permutation test for the Darts task resulted in a distribution of average inter-item correlations with a mean of 0.001 (SD = 0.01) while the observed average inter-item correlation in the real data was 0.27, $p < 0.0001$ for 1000 simulations. In Study S2, the permutation test for the Colorlink task resulted in a distribution of average inter-item correlations with a mean of 0.001 (SD = 0.02) while the observed average inter-item correlation in the real data was 0.25, $p < 0.0001$ for 1000 simulations. In short, the performances of advisors in Studies 2, S1, and S2 that was highly reliable across multiple trials and differed significantly from chance.

Figure S4

Permutation tests of advisor task performance in Studies 2, S1, and S2



Note: The x-axis shows the average inter-item correlation of within-subject task performances in one permutation of advisor data from Study 2 (upper panel), Study S1 (middle panel), and Study S2 (lower panel), and the y-axis shows the number of times that correlation was generated by a simulation carried out 1000 times. The vertical dotted line in each panel represents the observed average inter-item correlation for the actual advisor data in that task.

(Section 2.5) Advice Offered by Advisors in Studies 2, S1, and S2

Below is the advice offered by advisors in Studies 2, S1, and S2. Each row in the table represents the complete advice offered by one advisor.

1. Advice Offered in Study 2 for Word Scramble

try to look for three letter words since they are easy to find
Look for an "s" or an "ed" to be able to attach onto words to make them plural or the past tense. Look for "ly" to make a lot of adverbs. If you want to have the most words, focus on shorter words as opposed to longer words (something I didn't really do).
look for easy words, dont try to over think.
Try to focus on the vowels. Notice where the A E I O U's are to help get an idea of what words you can make. Be aware of were S's are, they are a good way of turning 1 word into 2 using a plural.
Just take your time and try to find as many words as possible. Lots of words can be spelled forward and then a new word can be used using the same word backwards.
Look for short words, at least three letters, as longer words don't give you more points. Look around vowels because almost all words have vowels in them, especially short words. The type of words you'll find will generally be consonant-vowel-consonant.
Look for sequences of letters that likely indicate a word can be formed. Try and find what words you can from these sequences. Sequences include such letter combinations as SE, TA, LA, PL, CH, etc. Keeps going from letter sequence to letter sequence. Look to form both singular and plural forms of a word if you can.
My only advice is to make sure that you understand what you need to do with your mouse when clicking and holdling to select the letters.
I started by looking for words that were already spelled out either vertically, horizontally, or diagonally. Once I did that I just started in the top left corner by letter and tried to make words from there. If I could not I moved down to the next letter and repeated that same strategy. Also the game seems to accept the plural and singular form of the words so make sure to use both!
First, look for small (3-letter) words. When you find a word, see if you can extend that word with one more letter, especially with an S; also look for the ED inding to extend verbs. For example, TEA can be extended to TEAR or TEAM; TEAR can be extended to TEARS; EAR

can be extended to SEAR. You can also scramble letters of the same word; for example, EAST can become SEAT and TEAS.

Words are centered around the vowels. Try to use as many plurals as possible.

Just go as fast as possible. Try to remember words you have already entered to save time.

My advice would be to take your time when highlighting letters to make sure they all turn red and you won't have to go back, ultimately wasting time.

Write down the words you have already used. This will make it easier for you not to waste time trying to use the same word when you're stumped and can't find new ones.

Build your vocabulary

look for letters that are next to vowels or common consonant combinations like "st" or "th." Remember that 3-letter words are acceptable and often you can then make 4-letter words with the same 3 letters. Sometimes you can make 5-letter words with 4-letter words you have made by adding an "s." Avoid rarely used letters like "x."

try to find a consonant with a vowel, and work from that.

Go fast.

Look for plurals.

Look for the most obvious straight lines first.

Try to find small words that can be changed slightly to create many words - like pit, pat, pits, pats, etc. Once you click a letter, you can't un-click it, so be careful. If you can't remember if you have already spelled a word, try anyway - you won't be penalized.

Try to look at the whole puzzle and look for bits of words that you may know. Try not to go letter by letter, it will only slow you down.

Look for vowels to begin and then use the letters around the vowels to make words. Use an s around the same words to make the word twice with one being plural.

be precise in how you move your mouse around the boxes. if you hit the wrong box you have to start over in selecting the word. when you are working on a time limit that can hamper you.

Make sure your mouse is on the correct letter before you start making your word. Make sure you look for plurals. For every word you find, see if other words can be found in it, like by reading the word backwards.

I suggest you keep your finger down on the mouse as you are sliding as mine kept slipping and I had to start over. You may also be on the lookout for plurals as I found quite a few. Also move quickly as that 1 minute flies by!!

You can always get two words from something if you have an "S" to plural it first.

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-

Look for plurals and past tenses, words that end in -ed or -s, -es

Don't forget to look diagonally and backwards!

Words within words count! eat, seat, eats, seats

You may get extra points for obscure words. Be well-read to improve your score!

Word length doesn't matter. A three-letter word will give one point, just like a six-letter word. Go for the simpler words. That way you can get more short words in a minute than wasting time looking for longer words.

Look for all versions/tenses of a specific word. For example, if you see "TALK", also look to see if you can find "TALKED". Between scramble boards, look to see if there are repeat words. That is, some of the words you found in board 1 might also be in board 2 or 3. Rather than pick a letter and search for nearby options, loosen your focus and look for general word patterns over the whole board.

Try and make several words out of the word you just got. Also see if it spells something backwards. Making one word seems to help the fluid-ness of finding the next. It will also help to try a couple different strategies while playing and seeing which one works best for you.

Make sure to start with easy words with few letters first and look at the board from every angle.

The tiles are sensitive. If you accidentally click a tile you weren't intending to click, it's better to come up with a new word rather than waiting, and perhaps mis-clicking to get the word you were intending to get.

Know your word before you click, know how to spell it as well. If you don't this slows down your game, reducing the amount of time you have left. Good luck!

1. Look for words all in one row first
2. The same words seem to appear across multiple boards
3. Keep it short, focus on three letter words.
4. Look at the vowels and see if you can build around those

Go for simple words starting with consonants. Move the mouse slowly and consistently. If it highlights a wrong letter, try clicking on the letter to start over.
My advice would be to see if you can make your word plural. so you can make two words
What helped me the most was by looking at the vowels and seeing common constants that I could make alot of words out of and sometimes I would take words and see if I could reverse any of the letters to make a new word.
Don't panic! Find compound words and be sure to look for suffixes and prefixes to help get a lot of words quickly. Also, give yourself a minute to look for opportunities to make a lot of short words together.
My advice is to concentrate on short words, that way you'll find more of them.
When making words try to make longer words, not just 3 letters.
Look the easy words like "sea" "set" "test" sometimes u can click them more then 1.
Be as focused as you can. Quiet room. Start with 3 letter word. Check if a 4th letter like s can be added to make plural, etc.
One things is too look for words and see if you can make them plural. Rat can be come Rats, so you can essentially double up word counts just by adding a letter.
Look for short words. If there is an s, er, or ed available look to make combinations like plurals and tenses. Controls are tricky, be patient.
The only advice I can give is try to think fast and try to think of words that are more than 3 letters. A lot of the words I found were actually very long and take a a lot of thinking to map out on the board. Remember that the letters just need to be touching
When trying to find a word, look for consonants next to vowels. Some 3-letter words can become 4-letter words as well.
My advice and only advice would be to never focus on one letter or set of letters. You have to look at the whole board.
It is difficult to choose letters in a row
Don't release the mouse button, doing so would enter in your incomplete word!
Take your time clicking on the letters to make a word. Be careful not to click on letters that would make your word invalid. Don't pay attention to the clock or the number of words you

have created as it might distract you or make you nervous. Try not to focus on making big words, just try to focus on creating as many words as quickly as possible for points.

- You CAN make the same word twice, just not with the same letters. If you have two of the letter "s" next to each other and can connect them to the same word twice, do it.

- Don't worry about longer words until later in the game, focus on short quick

You should not worry too much about getting longer words, since they don't count for more points. Find three letter words and if possible add an "s" on to the end or beginning. Do a lot of work around the "s". Try to start with one letter until it's exhausted so you don't forget where you've been.

Look for easy pluralizations for that may boost the number of words that you find; for example, 'cat' and 'cats'. Time goes quickly!

Make use of the letter s when it appears. Remember that words can be made in any order.

Think small and don't overthink when it comes to combining words.

My advice to you is just to start picking letters

the words will begin to flow as you start playing the game

don't get distracted by not finding any words to begin with after a few seconds they will begin to pop out at you

I would suggest looking for "s" in the grid so you can make plurals of the same words. Also If you see letters such as "eed" together, you can look for letter to make word with it, like seed, weed, reed, feed. These are a couple quick tips to get more words

Try to focus on a small section of letters to find words for awhile. This can help you spark ideas to find more words without having to search the whole board at once, which can slow you down.

I would say the best bet is to look for the smaller words first, and then build on them. For example, find "eat" and then "eats". Use as many variations of the same small set of letters as possible until you've exhausted all possibilities.

Just clear your mind. Think what letter could come next in a word and look for it. Have fun!

Make small words. Try to use variations of the same word bases for, fore, ford, fords, etc

First, look for basic words. When you find these basic words, look for plurals, past tense, variations, etc. This is a quick way to rack up points.

Look for vowels. You can not make words without them, so find where they are in relation to consonants and map out words accordingly.

Look for popular letters that are combined. Things like TH-, ST-, etc., are the beginnings and ends of a wide variety of words.

look for the vowels

look for words that you can build on

go with your instinct

The only advice I could think of is, don't panic! And if you start making a word that isn't one, just keep on as quickly as you can. If you mess one up, shake it off and you'll do better the next one. It IS only a game.

Ok, first of all pay close attention! It seems easier than it actually is. If you see a word that you can make, but you have to drag back the other way it will NOT let you, so that word will not work. So lets say the word is work and -w-o-r and k is on the other side you cannot drag back. Find one letter and look all ways to see if it can make a word, then move on to another.

Don't be concerned about making long words with more letters, all words are worth 1 point. Still use the plurals though - leap as well as leaps - whenever possible for more points.

Look for words in the puzzles, then drag your mouse to form words. You can not use the same word twice. Each word must be at least 2 letters. You can drag the mouse vertically, horizontally and diagonally to form the words. Your goal is to form as many words in one minute.

1. Look for vowels. All words need one.

2. Work around a single vowel until you have found all words with that vowel.

3. YOu might start with vowel on top of board. Then move to another vowel on bottom of board to make sure you do not duplicate.

Here's some advice for how to do well on the Word Scramble game. Look quickly for the available vowels in the game. Then look for consonants that are adjacent to the vowels. Try to scan the letters as quickly as possible to come up with words. Keep a mental track of the words you have already identified, as you will not get credit for duplicate words. Try to stay calm and not get flustered as the time ticks down. Just keep scanning the letters in all directions looking for new words until time runs out.

The most important thing is to not just sit by and stare at the board. You need to take the first few seconds to look at all the letters but then you need to just begin moving. If you just sit there looking at all the letters you can end up feeling overwhelmed and lose time on making words.

Next, you do not need to make complex, long words. You are looking to get points for creating as many words as possible. Therefore, the best advice is to begin right away by looking at one

letter and branching off to see what words you can create. It is usually better to begin by looking at a consonant and looking at nearby letters to see all the words you can create. And you just begin by creating words.

Overall, the best advice is to not get lost staring at the box. Second, begin right away and start looking for easy words you can create. Get in a rhythm and the words will start flowing out.

Place the cursor on the beginning letter of the word you want to make. Hold down click and drag it to the other letters of the word. Make sure that all of the letters are touching, they have to be together in the game.

You will have a time limit to find words so be as quick as possible.

Even short words qualify.

The same words qualify as long as you use different letters in the block.

Be careful dragging the cursor in this game. It is not very accurate. Also, it does not matter how long the words are. For maximum points, stick to 3-letter words if possible.

Keep your fingers moving; look for common suffixes, e.g., -ear, -art, -are so that you can add a letter before them (e.g., -art, could be dart, part, etc); continuously scan back and forth, up and down. Make sure your touch pad or mouse are smooth and easy to use so you don't accidentally flow onto a letter you don't want because you can't go backwards. If you see a Q but no U, obviously don't click on it. Scan for common letters as well. Look for the most commonly used words, like "and". Most importantly, the more you play, the better you will get. And take your time - if you rush, you'll be putting pressure on yourself and you will not find as many words as you are capable of.

You'll be surprised at how many words you can find. Going diagonally then across and then down, I found up to six letter words. It's pretty fun to try to find the longest word.

Try to focus on small words that are three or four letters long because you should be able to find those quicker. Try to focus on the vowels because if I'm not mistaken most words have vowels. Also, try to scan in all different directions to try to put a short word together.

Look for small words first. Look for words within words. Look for words similar to the ones you just found. Stick to one area as long as you're finding words quickly. Look for words that end in -ed or -s if they're on the board. Stick to areas that have letters that are common.

Try and look for the plural of a word that you find, or if it's possible to add an 'ed' to the end to get multiple points for a the same base word. Also, make sure you don't rush so you're careful in choosing where you click your mouse, because if you rush and start on the wrong letter then it wastes time.

Make sure You click on the corner or edge of the tile, if not the game wont accept it and then You get a sorry and look like a complete idiot for makeing a stupid word.

2. Advice Offered in Study S1 for Darts

Throw the darts as straight as possible, metal part forward.

I found it helpful to lean forward as far as I reasonably could.

eye on the target.

My advice is to lean into the dart throw, releasing the dart when your hand is in alignment with your target. Also, take a moment to make sure you don't have a gust of wind coming!

I used my dominant hand and wasn't very successful so maybe try to use the other hand (which I did and at least all the darts landed somewhere on it).

Concentrate and don't rush, get a good feel for the dart. Try avoiding the 'spin dart' method, the weighted end won't land on the board.

It might help to use your index and thumb lightly rested in the dipped curves of the dart just behind the magnetic weighted end.

Enjoy!

-Because of the magnet in the dart, you need to aim higher than what feels natural, as it will curve down.

-More force in your throw is better than less force.

-If there is wind, wait for it to pass.

-Aim slightly higher than your intended target.

-Make sure to follow through on your throws.

While this game is similar to dart, the darts themselves are different. These darts are less aerodynamic, so count on their weight playing a larger role than normal.

Try to work with the weights at the ends of the darts.

Use your first free throw to see which may not to throw. Try to pay attention to which way you've thrown, and note the difference between the direction of your throw, and where it actually landed.

I did better when I aimed higher, I got closer to the bulls eye.

If you landed near the bulls eye, but lower, aim even higher.

You can learn this, with enough tries. Don't give up. Good luck.
1. Consider wind direction before throwing. 2. Think about where last "dart" thrown was placed on board.
Take your time. Breathe in-out when throwing. Focus mentally.
Be aware of the weight of the dart, keep in mind that the dart makes a slight wave and keep trying.
Adjust your feet positions to correct for how far to the right or left you were on the last throw. Focus on the bullseye. Try to eliminate distractions. Hold the dart close to the magnet. Don't use too much force. You can do it!
-Throw the darts with little force. -Look @ the bullseye when you throw it. -Take your time -Point foot towards target.
-try relasing at different angles. stick to the one where you think you can hit the red dot. -try adjusting the power exactly where the dart flights and hit the board.
Stand perpendiculat to board with leading/dominant side closest. Rely on muscle memory and go as quickly as possible.
There is a significant difference between real darts & the magnetic variety. Rather than imagining throwing a spear, with these magnetic darts it's better to imagine throwing a small ball, trying to figure out at which point the board will intersect the arc when toned.

Darts in an iterative process; the first throw is a datapoint to be built on by adjusting stance + throw for each subsequent turn.
-Take notice of where your practice dart lands and how you threw it. -Keep evaluating your 'performance' as you throw the darts and adjust each throw by using the most effective technique aspects. -Keep your eye on the bullseye and line the dart up
Focus on the spot and once you feel like you have a good throw duplicate it as much as possible.
Hit the bulls eye. Learn from your previous throws. Don't stress too much
If you are left-handed, you will have a tendency to hit the right side of the board, and vice-versa if you are right-handed. So, to play more effectively, try once to see how far to the side you are, then try to adjust by bringing your aim out closer to your side. Same thing goes for right. Throw from your elbow, not your wrist.
Relax- Follow Through
Take your time and not stress over the first few throws. Use those for a way to alter your stance, power, and aim to be accurate. Have fun, it's a game.
Try to aim for the middle. Standing up straight seems to help a little bit. Don't hit the experimenter with darts.
Don't think too much about it. I don't know the mechanics of darts but if you think too much you'll psyche yourself out. That basically applies to anything. In the wise words of Kevin Hart, "Do you, boo boo"
You will stand about 2 meters from the board. I think that you should try to throw a bit higher than you think you should to hit the middle of the board. It's a magnet darts board and all the darts are the same. Put one foot a little further forward than the other foot while throwing.
Try to throw under same conditions (take the wind in account)
Change how hard you throw and the direction of the dart by the previous dart
Practice makes perfect!
-Take your time.

-Breath in slowly, Breath out as you throw. -Relax, have fun. -Aim true! Lean forward. -Trow the dart!
-Be careful of crosswinds -don't let the sun get in your eyes -compensate for which side your focal point is on so you can aim for the center
for best results Hold dart by "its" tail and throw underhand. Toss in a slightly upward underhand motion. Best result is to throw gently.
-TO GET BETTER RESULTS IT'S IMPORTANT THAT YOU MEASURE THE WEIGHT OF THE DARTS WITH YOUR HANDS. ANOTHER KEY THING IS TO GET THE RIGHT AMOUNT OF STRENGTH IN EVERY TROW, SO YOU DO NOT GET TO FAR FROM THE BULLSEYE.
Take your time -hold the dart pins properly Take note of the magnet in the dart pin if you are not accoustomed to pinless darts.
Lean as close to the board as possible Do try to throw against the wind. aim for the center.
Key is to throw darts with some speed directly at bullseye. I tried loffing the darts with less success. Throwing the darts with some speed yielded better results.
Pay attention to your angle to the board. Throw hard. Hire a coach.
-Take time to get used to the touch, weight, etc. of the arrows -Mind the wind that may influence the direction. -Be confident

-Just try the first shots without really thinking just to get used to it and adjust
Once you get used to the weight of the dart, remember how you guided the dart in your best shots. Accompany the dart with the whole body so that deviations made by the inaccuracy of your arm are minimized. Do not shoot too fast because magnet will not stick; a parabola is desired for your shot.
-turn your body sideways -follow through on the throw -aim a little higher than where you expect it to land. -avoid throwing in an arc. - throw straight. -don't pinch the darts too hard.
The darts stick at the end only so try to hit with that. Not much is on the lane so don't worry.
-Pay attention to the weight & balance of the dart during the practice throw. -Make sure you have a good feel for the practice throw to ensure you get the best score during the 'live' game. -Make sure you follow through with your throw. -Aim small,
Aim a little higher than the bulls eye. Keep the "dart" centered within your index, & middle finger and thumb.
Stand and ground through feet. Focus on board. Throw the dart. If it is anywhere near where you want it to be, concentrate, repeat, refine. Be careful not to overcompensate when throwing and you want to get to your mark. Be relaxed and keep it light.

It's a game!
Future player, I guess only with some practice you'll be going to always get the bullseye. So, I'd recommend having an earlier practice before doing it. The thing I noticed when I reached the goal was that I was pointing a little bit above the target. I'm not sure how much, so...
FIRST OF ALL USE THE FIRST PRACTICE SHOOT TO SEE HOW THE DARTS WORKS. THAT MEANS IF YOU THROW IT STRAIGHT ON AND IT GOES ON THE LEFT, YOU HAVE TO AIM MORE RIGHT TO TRY TO GET THE CENTRE OF THE BOARD. TRY TO BE RELAXED WHEN YOU SHOOT THE DART AND DO NO USE ALL OF YOUR POWER BECAUSE MORE POWER MEANS LESS PRECISION. DARTS BY DART TRY TO IMPROVE AND CALIBRATE YOUR NEXT SHOOT.
Throw high. Darts will arc.
-grip the dart on its magnetic base (front end) -don't let it fly the way you would a standard dart- it doesn't fly as well- put more into it -short compact motion is best because the dart does not fly as well -make adjustments & let the wind die d
*try not to be nervous or think too hard- it's just a game! *don't try to compensate for your mistakes. i.e. if you throw one too far left, don't try to throw more to the right next time. Just try to aim for the center each time. *throw hard, if you don't, they won't stick to the board. Even if you're accurate, if it doesn't stick, it doesn't count! *don't get distracted by the fact that you're in a "study"!
-The dart drops a considerable amount so aim higher than you think. -Wait for the board to stop shaking before you throw -Channel your inner spirit for maximum accuracy points.
Think parabollic movement, so aim high (like with an elevation angle of 45*). Speed within each consecutive darts was an issue for me because it seemed the faster I did it the better I got.

The darts don't have the same feeling/weight as normal darts. The magnets tend to slide down if not thrown very straight and don't land flush with the board. Keep a focus on throwing the darts straight as opposed to forcefully.
-Use a little bit of arc. -Don't throw too hard. -Aim slightly above bullseye. -Keep eye on target.
Throw the darts fairly hard. Try to keep your arm in the same place so you can tweak each throw.
Focus. Toss firmly but don't throw too hard.
Darts are lightweight & easily affected by wind. You need to throw w/ a steady power, but too hard & the darts bounce off board. My best option seemed to be to throw them a little higher than you think you need to & they'll drift down.
Work on isolating the movement to as few moving parts as possible. Consistent movement may reduce variability of throws...?
Adjust for wind Aim high Concentrate Don't waste time
Put your dominant foot in front of the other. Put equal weight on each foot & center your centre of gravity. Grab each dart right below the magnet in 2 fingers & your thumb. Throw fairly lightly as bounce is significant. Aim slightly higher than the bullseye as it will slide down after contact. Neglect gravity due to your proximity. Use your arm as a lever to it pointed at the target & parallel to the ground. Do not lean forward on a lot when you throw.
I don't think there is any written advice that can help you; sorry! The best advice is to practice darts in your spare time. Other than that, just relax & try to hit the bullseye.
1) Be aware that there will be a pause after each throw, and a longer pause after 6. Thus you need to stay relaxed and concentrate to get momentum and get in the groove. 2) The darts will fall off the target sometimes (if they hit the metal ribs.) This does not mean

they were bad shots!
3) The target is really close. Just relax and you'll get it! (And don't feel bad if you miss the target completely.)
The darts are weighted from the magnetic tip, aiming higher than the bullseye would improve accuracy.
aim! throw the dart higher than you think you need to, and aim above the bullseye. lean into the throw with your leg opposite your throwing arm.
Aim small, miss small.
SOME DARTS DO NOT STAY ON THE BOARD. THE FEEL OF THE MAGNETIC DARTS IS DIFFERENT THAN NORMAL, AND TAKES SOME GETTING USED TO. MY AIM/FOCUS WAS ON THE CENTER BUT THE BEST THROWS CAME WHEN I LOOKED ABOVE THE TARGET AND ALIGNED MY HAND W/ MY NOSE.
1) step as close to the line as you are allowed to 2) Try to limit motion in parts of your body other than your throwing arm 3) Try not to blink as you throw 4) Stand up straight 5) Relax 6) Establish a rhythm between throws
Every time you throw pay attention to your body not just where the dart lands, try to correct your movements.
- Have good aim or good eyesight. - Extend your arm completely when throwing the dart since you can't step forward.
- Don't throw too forcefully - Focus on the center. I found it helpful. - Closing one eye and throwing was more difficult.
Don't overcompensate for misses. Watch where you release the dart.
Hold the darts where the tip meets the tail.
Hello, (to the one coming after me) I'd recommend you to stay calm, put your strongest foot a little bit forward. Don't throw the "arrows" too strong because it will not stick to the board. However, if you will throw it too

weak from two main reasons: 1. Maybe you will not get to the board. 2. the arrow will have a ballistic path and therefore you will lose accuracy.
- Follow through with arm movement, keeping your hands, elbows, wrists, and shoulders on a straight axis (maintain pointer finger for follow through) - Foot placement is important. Place lead foot towards center of dartboard and adjust back leg accordingly
- Aim a little bit higher than the target (depending on how hard you throw, the dart will drop variable amounts during its flight) - Try to repeat the same motion each time. Any variance in form will affect your throws (and take you longer to get better)
- grab the dart in a specific way that feels most comfortable to you. (Make sure you grab it the same way everytime) - aim for the bullseye on your practice throw. (make sure to fully extend your arm when you release the dart) - after taking note of h
Slightly bend your knees with your feet shoulder width apart and your dominant foot forward of the other. The magnet darts throw more like pebbles than actual darts. Hold the dart with your fingertips of your dominant side and toss (I aim slightly above the target anticipating the drop). When you throw, lock your arm (at elbow) straight so your arm is pointing right where you want the dart to go. No matter how you do, remember that it's only a game of darts and you can go practice at the pub anytime to increase your skill level. Have a nice day. Love [name]
Honesty, don't overthink it. The more I focused, the worse I did. The one tip that I'd give which seemed to help a little bit is pointing your dominant foot straight at the target with the other slightly behind you. You've got this one, champ. Love, A Fellow Participant
Two things- 1. The darts are magnetic- throw holding the blue magnet with your fingers- this will allow for greater accuracy.

2. Aim a little higher than you might with a normal dart- the darts arc, and I found my throws to be slightly lower than expected.
Make sure you follow through with your throw. Wait for optimal conditions (such as no wind) until you throw. I found it helped when I put my weight on one foot that was slightly behind my other foot and leaned forward to the other foot which was behind the line when throwing.
- If you are tall, squat down to adjust your height versus target - Unlike real darts, more speed (throwing hand) does not help the magnet stick to the board - Watch out for the wind! - Put your leading foot forward (on the same side has throwing ha

3. Advice Offered in Study S2 for Colorlink

Connect the dots closest to each other first.
I would move the balls that are furthest away from each other first. I would also go for the ones near the corner or edges.
The best way to complete the board is by linking the colors that are furthest apart first. This will allow you to see patterns and quickly link the colors that are close together in the best way. If you get stuck and two colored lines seem to be intersecting, try to attach the line from one color to another by taking the line in the opposite direction and coming at it from a different angle. This will usually solve any conflicting lines.
ALways test the longets routes first, the rest will be easy.
Connect the dots with the shortest ling first. Then focus on the dots on the corners. It is easier to work from the inside to the outside.
Complete the colors that are closest to eachother first.
Complete the colors where there is only one way to complete it.
Be sure not to separate two colored dots. If while building a line between two of the same color, make sure to roughly build a line of sight between other colors' dots so that you don't isolate them from each other.
If the dots are close together, they will likely connect via short distance. The ones that are further apart are the ones that will need a more circuitous route. Connect the short ones first and then look at what's available for the ones farther apart.
If you get stuck- see if there are any other ways to make the colors connect than what you are already taking.

My advice for strategy would be to connect the dots farthest apart first, so that you can create paths for the closer dots with the board "map" already mostly in place.
For me the trick to this game was to keep the entire board in mind at all times. When you initially start the most obvious move is to immediately match all the colors that are extremely close to each other. I then look for the next and then the next in closeness. You just have to be mindful of other connections that need to be made
Connect the colors that are closer together first. Then fill up the rest of the space with the farther apart colors. Try to hug the walls whenever possible. Try to leave plenty of space in the middle of the board.
The main thing to focus on first are the dots that are the closest together. Those have the least possible pathways. Also try to do the ones first that you realize can't have very many possible paths.
try starting by connecting the dots that are closest together first. then weave dots together around those. finish around the perimeter of the board last.
Try to stay to the outside as much as possible, and work your way inward
When completing the board its easiest to make all the small simple connections first and then work your way out to the longer ones.
Start with the largest gap between colored dots and move backwards from there.
First start with the dots of the same color that are close to each other and connect those ones first. Then you will be able to see a distinct path for the longer lines of the same color to connect. Once you complete the first few it gets easier each time.
Much easier using a normal mouse on a hard surface. Make sure when you're connecting lines you don't accidentally hit another line or it breaks it up and you have to redo it from where you hit it. Boards aren't too complicated. You just need to take a few seconds at the start and look at the whole board. I found it easier to connect the farthest dots first and work my way down to the ones that were closest to each other. Don't really need any sort of strategy. Just try and draw the lines on the outside of the board and work your way into the middle.
The objective should be to find the pairs of dots that connect in order of decreasing difficulty. That is, connect the two dots which are most difficult to connect first, followed by the next most difficult pair of dots to connect, and so on. If a mistake is made, do not worry about undoing it. Just continue with the next correct action as connecting two dots over an existing path will overwrite any previous paths.
The only piece of advice I can really offer is just go for it. don't try to strategize or think it out. Just start and see if certain moves work.
Puzzles like this are not my strong suit but I still enjoy the challenge, even if they are a bit frustrating for me. What I did was randomly start connecting colors and redoing them over and over as I saw where by connecting one color another would be blocked from being able to connect. By trial and error I was eventually able to connect all the colors with none being blocked and thus cleared the boards. The biggest drawback to my method is that it can be quite

time consuming. I wish I could give better advice but I do not have the most analytical mind. That's the best advice I can give you, other than just relax with it and have fun!
Start with the the ones that are closest to the edges of the board. Get the long ones out of the way and then worry about the ones that are close together. Don't overthink it. It's not too difficult.
Focus on connecting the dots that are the furthest apart first. That way you have a better chance of not having to undo the shorter connections later if your longer paths are blocked. Also, be fluid in drawing your lines. You don't have to go one square by one square. Just draw fluidly to save time.
When it comes to connecting all the colors, set small goals. connect one color and then another, and don't be afraid to mess up. a lot of your solutions will come from failures.
Quickly scan the board for dots that can only be connected in one way and connect them first. The remaining ones will fall into place more easily once those are finished. As you are connecting the remaining dots, pay attention to the path you are taking and make sure it does not block the paths that will need to be used to connect the other dots.
Generally try to start with the color that has the two dots that are closest together, knocking those out will clear up how the board will look at the end. Also try to work from outside in, taking care of all the outside squares first so that you don't finish a level with a random square available.
Always connect the 2 closest dots first and work from there. If you are having trouble thinking of how it should fit together, just start connecting the dots and replace the ones you did incorrectly because it will be easier for you to see what will fit when there aren't as many open spaces left.
Try to connect the dots right off the bat instead of trying to figure it out in your head. Many of the solutions will come when you can visually see them.
It is usually easier if you connect the closest ones first. Start from the inside and work your way out. If you are having a tough time with one color, move on to the next, and come back to it later.
Just go for it, start connecting colors right away and it will quickly become apparent how to complete the board
you will need to use patience. it is important to plan your moves with care. try to connect the close items first, then work your way up to the larger distance items.
Dont try to figure it out all in your head. Go ahead and start connecting immediatly then fix it as you go along
If you can't seem to figure it out then start over and do something different.
When trying to connect colors that look farthest from each other, try and use the route that is least direct. One that uses up most boxes.

Try to connect the harder ones first maybe, like if one of the dot things is somewhere where there is only one way in or out do that one first and the rest later probably?
connect the dots that are closest together first before moving on to dots further apart.
Start by finding colors that have the closest distance and work your way up from there.
The way to do well is by process of elimination. Look for obvious paths first, and then go from there.
I would suggest starting by connecting the dots furthest apart and closest to the edges of the board, then work your way down by then completing the next furthest, until you finish.
Don't always go for the obvious pattern. Try to weed out the ones that only have one way to go first. Or do the difficult ones first.
Find the circles that are farthest apart and connect those first and then the closer ones will be evident of the path you need to connect them.
Start with the ones that will only have one solution possible.
Focus on the shortest links first, then the longest links. The rest will fall into place
I would start by linking the dots that are closest together on the board and then moving outward to the dots that are farther apart, in this way you avoid intersecting lines when you are linking the colors.
Complete the shortest routes first. Trust your intuition, as it is easy to change your mind if you feel you're wrong.
Take your time and think things through. Do not rush through the game it's okay to make mistakes as long as you can quickly adjust them. Be calm don't get frustrated and you will do great.
Just find the ones that are obvious first, do those, then move on to the next one that is obvious. Don't over complicate it, it's really simple.
The best advice I can offer is to make the long connections first; that is. try to link the color that takes up the most squares first. Then you can work around with the colors that take up smaller squares.
Start with the easy connections first, then work your way up to the ones that are more far apart. Usually once you get most of the board it's pretty easy to figure out where everything else should go.
My advice is just to start each board and go with your gut instinct on which color to choose 1st. Try to make the 1st couple colors passes go around the perimeter or the long way, that way the ones towards the middle can be done without any mistakes.
Look for the obvious connections first. Then sort out your pathing from there.

When you have two colors that look like they will have to cross paths, just make them go opposite ways, or around each other kind of like a clock. Also, the shortest distance between pairs is usually the most likely path. When you start, start with the corners or the edges first.
Don't be afraid to make mistakes. It's very easy to undo them than to just try to get it right the first few tries.
Find the most constrained points (those with the fewest directions the line can travel) and start from those.
Find the two same color dots that are furthest apart and work out the path for them. Then the next two furthest apart etc. Remember once you've made a line it is easily rearranged by clicking on it and dragging it.
My advice is to find the ones furthest apart and start there. A lot of them seem to work inside out. Meaning work from the inside of the board and work out.
I found it helps to figure out the outer paths first. As you go along the path, try to check the other color dots to be sure that you're not blocking paths for them.
I found it difficult at first but after a few games I completed them a lot quicker. I learned strategies but it is hard for me to verbalize it. For example, as you are playing, when you see that a line will intersect/cut off a necessary future "link" between other colors, immediately stop the line and look for another possibility. If possible, use up as much space by winding the links (to take up free space).
Look for the farthest away pair then work your way in.
First, make the shortest and most direct. Then look at situations where two different-colored circles are close together. Try to determine which directions you can go so the lines overlap. You could instead try to make the most far apart connections first using the most square as possible, then try to connect the closer together circles of the same color and adjust accordingly.
Start with the dots that are furthest from each other, then work the shortest distance and alternate until the end.
My advice is to not overthink it. Do the shortest legs first and the longer ones will become obvious.
Experiment at first. Try to see where certain color lines would HAVE to fill in the space in order for all the spaces to be filled.
Try to be patient and think about all possible combinations
Best way is to just start connecting the dots right off the bat and correct yourself as you go. You may be wrong the first time but you can easily fix it.
My advice for doing well in the Colorlink game is to first complete the simple color patterns first. If you correctly fill-in the easier patterns, this will make the harder color paths easier to figure out.

Instead of taking time in the beginning to think about how to connect to each dot in the best way, just start connecting dots immediately. Even if it isn't correct, you'll be able to see quicker which way is going to work by already having connections as opposed to trying to form it in your head. I found that this was the best way to get the board completed quicker.

Try to connect dots on the edges first. Try to connect simple paths next. Try to leave openings for complicated paths.

I think the most important advice that I can give is to pinpoint the colors that are far away or what looks like to be the hardest to connect and try to find the right route to connect. You will get the hang of it after awhile even though they get more difficult as they go.

Try to connect to two furthest away first. Then work the shortest distance. You will encounter paths that cross, but that is okay. Once you have a few connected the path reveals itself.

the only advice is keep trying

Map things out in your head or by tracing with your mouse first. If you come to a block, just drag through it and it will overwrite what was there instead of having to manually remove the colors.

(Section 2.6) Analyses of Curvilinear Relationships in Studies 2, S1, and S2

In Studies 2, S1, and S2, we found no evidence of a linear relationship between advisor performance and advisee improvement. But is it possible that these variables have a non-linear, U-shaped relationship? To answer this question, we reanalyzed the data for these three studies using the “Two Lines” method of testing for U-Shaped relationships (Simonsohn, 2018). These analyses revealed no evidence of a non-linear, U-shaped relationship between an advisors performance and advisee improvement in any of the studies. Specifically:

- For Study 2, as described in the main text, we calculated advisee improvement by subtracting advisee performance on the first round of the Word Scramble task from the average of advisee performance on the subsequent five rounds. We then performed the Two Lines test, using advisee improvement as the outcome variable and advisor performance as the predictor variable. The test did not support a U-shaped effect of advisor performance on advisee improvement in the Word Scramble task (break point = 6.17; Slope 1: $b = 0.21$, $z = 1.73$, $p = 0.08$; Slope 2: $b = 0.03$, $z = 1.5$, $p = 0.13$).
- For Study S1, we calculated advisee improvement by subtracting the mean of the six pre-advice rounds on the darts task from the mean of the six post-advice rounds. We then performed the Two Lines test, using advisee improvement as the outcome variable and advisor performance as the predictor variable. The test did not support a U-shaped effect of advisor performance on advisee improvement in the Darts task (break point = 6.11; Slope 1: $b = -0.1$, $z = -0.86$, $p = 0.39$; Slope 2: $b = 0.44$, $z = 0.86$, $p = 0.39$).
- For Study S2, we calculated advisee improvement by subtracting advisee performance on the first round of the Colorlink task from the average of advisee performance on the subsequent five rounds. We then performed the Two Lines test, using advisee improvement as the

outcome variable and advisor performance as the predictor variable. The test did not support a U-shaped effect of advisor performance on advisee improvement in the Colorlink task (break point = 69.67; Slope 1: $b = -0.14$, $z = -1.23$, $p = 0.22$; Slope 2: $b = 0.35$, $z = 2.82$, $p = 0.005$). Although the test suggested that the second line had a significant positive slope, the first line did not have a significant slope in any direction. Further examination of this result suggested that it was driven by a single outlier—an advisor whose performance was more than 4 SD from the mean. When this advisor was excluded, the positive slope of the second line disappeared (break point = 63.6; Slope 1: $b = -0.15$, $z = -1.18$, $p = 0.24$; Slope 2: $b = 0.48$, $z = 1.43$, $p = 0.15$).

(Section 2.7) Analyses of the Relationship Between Measures in Studies 2, S1, and S2

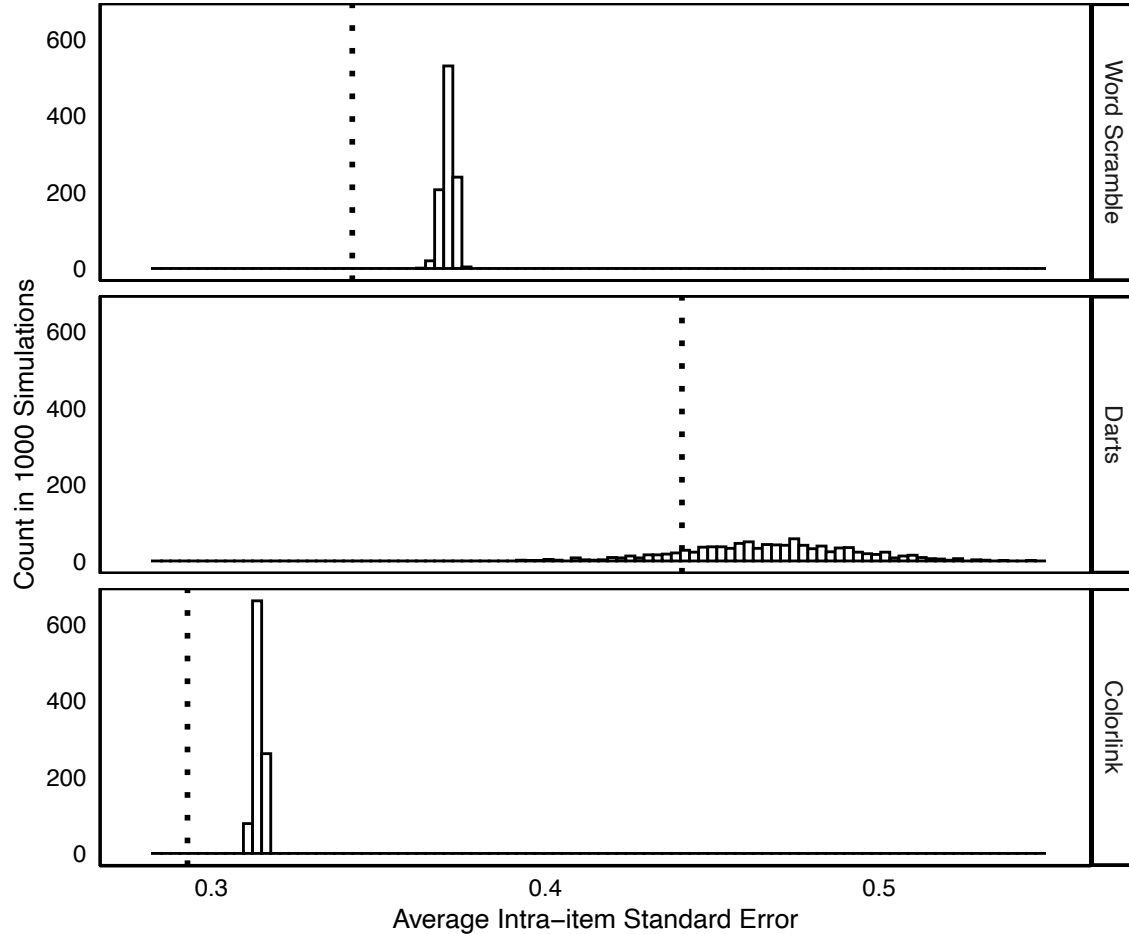
- In Studies 2, S1, and S2, we measure the perceived quality of advice by asking advisees (a) how helpful they thought the advice was and (b) how much they thought it had improved their performances. In all three studies, these two measures were highly correlated with each other (Study 2: $r = 0.72$, $p < 0.001$; Study S1: $r = 0.58$, $p < 0.001$; Study S2: $r = 0.74$, $p < 0.001$).
- In Study 3, participants rated multiple pieces of advice, which allowed us to measure how much they agreed on the helpfulness of different pieces of advice. We used the psych package in R to calculate a Cronbach’s Alpha for all ratings in Study 3, with 500 bootstrapped iterations. The agreement among raters was very high, $\alpha = 0.95$, 95% CI [0.94, 0.96]. Rather than being random, the degree to which a piece of advice in the Word Scramble task seemed helpful or unhelpful had a high degree of consensus in our sample.
- We conducted a permutation test (Dwass, 1975) on advisees’ perceived helpfulness and perceived improvement ratings in Studies 2, S1, and S2. The goal of this test was to see if different advisee’s ratings of the same advice from the same advisor were more similar to each other than to the ratings of other advice by other advisees. To quantify the degree of advisee agreement in each study, we created a vector of standard errors of all the perceived helpfulness and perceived improvement ratings that advisees made for each piece of advice. In three instances we excluded advice from Study S1 that did not receive enough ratings to calculate a standard error value. We averaged the remaining values to calculate the average intra-item standard error of perceived helpfulness and perceived improvement ratings on each task. We then randomly shuffled the advisee data in each task and recalculated the average intra-item standard error for the new shuffled dataset. Repeating this permutation

procedure 1000 times allowed us to create a distribution of plausible average intra-item standard errors in each task if performance across rounds were random. We then compared this distribution of simulated standard errors to the observed intra-item standard error for real participants. Figures S5 and S6 show the results of the permutation tests for advisee ratings of perceived helpfulness and perceived improvement, respectively, in Studies 2, S1, and S2. In Study 2, the permutation test for the Word Scramble Task helpfulness ratings resulted in a distribution of average intra-item standard errors with a mean of 0.37 ($SD = 0.002$), which was higher than the observed average intra-item standard error of 0.34 in the real data, $p < 0.001$ for 1000 simulations. The permutation test for the Word Scramble Task improvement ratings resulted in a distribution of average intra-item standard errors with a mean of 6.23 ($SD = 0.03$), which was higher than the observed average intra-item standard error of 5.98 in the real data, $p < 0.001$ for 1000 simulations. In Study S1, the permutation test for the Darts task helpfulness ratings resulted in a distribution of average intra-item standard errors with a mean of 0.47 ($SD = 0.03$), which was not statistically different from the observed average intra-item standard error of 0.44 in the real data, $p = 0.14$ for 1000 simulations. The permutation test for the Darts task improvement ratings resulted in a distribution of average intra-item standard errors with a mean of 1.38 ($SD = 0.07$), which was not statistically different from the observed average intra-item standard error of 1.46 in the real data, $p = 0.88$ for 1000 simulations. In Study S2, the permutation test for the Colorlink Task helpfulness ratings resulted in a distribution of average intra-item standard errors with a mean of 0.31 ($SD = 0.001$), which was higher than the observed average intra-item standard error of 0.29 in the real data, $p < 0.001$ for 1000 simulations. The permutation test for the Colorlink Task improvement ratings resulted in a distribution of average intra-item standard errors with a

mean of 5.68 (SD = 0.02), which was higher than the observed average intra-item standard error of 5.46 in the real data, $p < 0.001$ for 1000 simulations. In short, advisees in Studies 2 and S2 displayed consensus on advice helpfulness and improvement that was reliable across multiple trials and significantly larger than chance. While this pattern was not evident in Study S1, we suspect that this was due to a lack of statistical power: Study S1 had far fewer ratings (~4 ratings per item) than Study 2 (~27 ratings per item) and Study S2 (~34 ratings per item).

Figure S5

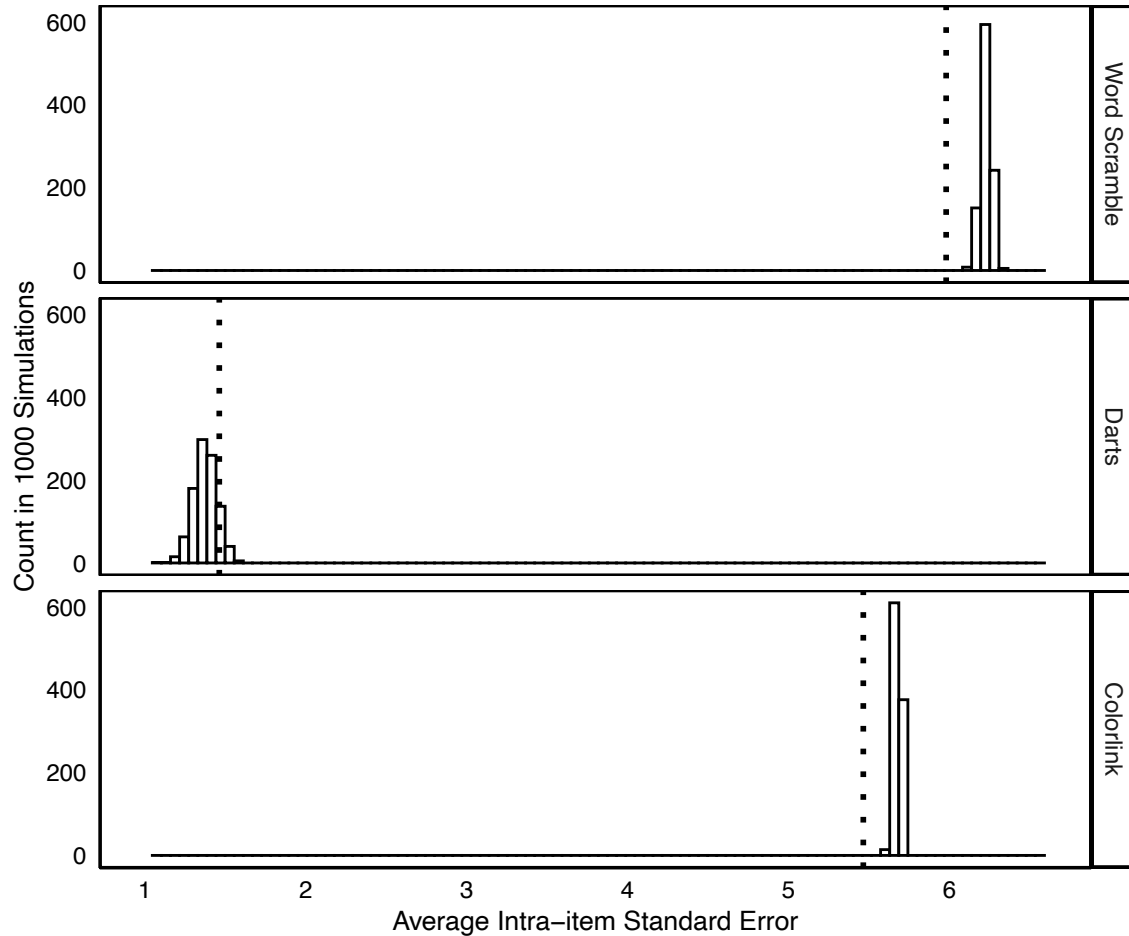
Permutation tests of advisees’ perceived helpfulness ratings in Studies 2, S1, and S2



Note: The x-axis shows the average intra-item standard error of all within-item perceived helpfulness ratings in one permutation of advisee data from Study 2 (upper panel), Study S1 (middle panel), and Study S2 (lower panel), and the y-axis shows the number of times that standard error was generated by a simulation carried out 1000 times. The vertical dotted line in each panel represents the observed average intra-item standard error for the actual advisee data in that task.

Figure S6

Permutation tests of advisees' perceived improvement ratings in Studies 2, S1, and S2



Note: The x-axis shows the average intra-item standard error of all within-item perceived improvement ratings in one permutation of advisee data from Study 2 (upper panel), Study S1 (middle panel), and Study S2 (lower panel), and the y-axis shows the number of times that standard error was generated by a simulation carried out 1000 times. The vertical dotted line in each panel represents the observed average intra-item standard error for the actual advisee data in that task.

(Section 2.8) Analysis of Advisor Improvement in Studies 2, S1, and S2

We found no evidence in Studies 2, S1, and S2 of a relationship between advisor performance and advisee improvement. Is it possible that advisor improvement (rather than average performance) did have a relationship with advisee improvement?

To answer this question, we reanalyzed the data for these three studies using advisor improvement (rather than the advisor performance) as a predictor of advisee improvement. For all three studies, we analyzed the performances of advisees in the advice condition by fitting a series of linear mixed models to our data in R using the lme4 package. The dependent measure was the advisee’s performance (Study 2: the number of words the advisee found in a round; Study S1: the distance between the dart and the bullseye on one throw; Study S2: the time needed to solve a puzzle); the independent within-participants variable was the number of the round (Studies 2 and S2: 1- 6; Study S1: 1-12); and the independent between-participants variable was the advisor’s improvement performance across rounds (average performance across the final two rounds minus average performance on the first two rounds). We also included an interaction term between these two independent variables. As random effects, we included a random slope term for the round and a random intercept term for the advisee.

For all three studies, the analysis revealed no interaction between the number of the round and the advisor’s performance (Study 2: $b = -0.002$, $SE = 0.007$, $t = -0.34$, $p = 0.74$, 95% CI [-0.02, 0.01]; Study S1: $b = 0.003$, $SE = 0.006$, $t = 0.55$, $p = 0.59$, 95% CI [-0.008, 0.02]; Study S2: $b = -0.0006$, $SE = 0.006$, $t = -0.13$, $p = 0.90$, 95% CI [-0.01, 0.01]). In all three studies, this pattern of results held even when we controlled for advisor performance by adding it to the model as an additional between-participants independent variable. In short, our studies provide no evidence of a relationship between advisor improvement and advisee improvement.

(Section 3.1) Exploratory Items in Study 3

1. The game you played today was called Word Scramble, also known as "Boggle." Have you played this game before today, either on or off of the computer?

- Yes, I've played some version of this game before
- No, I've never played any version of this game before
- I'm not sure

2. [if yes to Q1] You indicated that you've played some version of Word Scramble before. Before today's study, when was the most recent time that you played a version of Word Scramble?

- Within the last week
- Within the last month
- Within the last 6 months
- Within the last year
- Over a year ago

3. [if yes to Q1] As best as you can recall, when was the first time that you played a version of Word Scramble?

- Within the last month
- Within the last 6 months
- Within the last year
- Within the last 3 years
- Within the last 5 years
- More than 5 years ago

4. Approximately how often do you play word-finding games like Word Scramble? (others include Scrabble, Bananagrams, etc)

- A few times a week
- A few times a month
- A few times a year
- Less than once a year
- I never play word-finding games

5. How experienced do you think you were with Word Scramble before you took this study, compared to most people?

- I was less experienced than most people
- I was as experienced as most people
- I was more experienced than most people

6. How skilled do you think you were with Word Scramble before you took this study, compared to most people?

- I was less skilled than most people
- I was as skilled as most people
- I was more skilled than most people

7. How well do you think the average person does on one round of the game you played today?

Average number of words found in 1 minute: _____ words

8. How are you taking this survey right now?

- With a computer mouse
- With a trackpad
- Pointing Stick
- Touch Screen
- Other

9. What internet browser are you using?

- Safari
- Internet Explorer
- Mozilla Firefox
- Google Chrome
- Opera
- Other
- I'm not sure

10. Earlier in this study, you wrote advice to a future participant who will play the Word Scramble game. How much do you think your advice will help them improve? [0% improvement (not at all) to 100% improvement or more (a lot)]

11. In the box below, please type the names of as many animals as you can think of in one minute. Please only type those that you can think of yourself, without help. Separate each animal name with a comma.

12. How old are you?

13. Please indicate your gender?

- Male
- Female
- Prefer not to answer

14. Previous work in behavioral decision-making has shown that people vary in the amount they pay attention to these sorts of surveys. If you're actually reading this question, please answer it by selecting the option "other" and typing in the word "trout." Thanks for your help, and for taking the time to read all of the questions.

- Gender, marital status, and education
- Sexual orientation, relationship satisfaction
- Hobbies, favorite sports
- Other

15. Is English your native language?

- Yes, I only spoke English growing up
- Yes, I spoke English and these other languages growing up:
- No, English is not my native language. My native language is:

16. If you encountered any errors, glitches or other problems taking this survey, please tell us below. Otherwise, leave this field blank.

17. What do you think this study was about?

18. If you have any other comments about the study you took part in today, please tell us below:

(Section 4.1) Additional Studies S3a and S3b

STUDIES S3A AND S3B:

DO MORE SUGGESTIONS MAKE ADVICE SOUND BETTER?

Method

Overview

Study 4 suggests that the number of suggestions contained in a piece of advice is related to the perceived quality of that piece of advice. But is it *causally* related? The mediation analysis in Study 4 suggests so, but to be sure, we ran an experiment. We gave participants pieces of advice that varied in the number of suggestions they contained, and then asked them to estimate either the effectiveness of the advice (Study S3a) or the skill level of the advisor (Study S3b).

Participants

Participants were 714 users of the Amazon Mechanical Turk (Study S3a: 197 males, 161 females, 2 prefer not to answer, $M_{\text{age}} = 34.56$ years, $SD = 10.13$ years; Study S3b: 175 males, 178 females, 1 prefer not to answer, $M_{\text{age}} = 36.01$ years, $SD = 10.71$ years) who were paid for their participation.

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 1-3 and then played a single round. All participants were shown the same board. After playing a round, participants were shown advice that had either one, three, or seven independent suggestions.

Where did that advice come from? Two coders read all the advice written by advisors in Study 2 and listed each suggestion contained therein. We then selected 11 suggestions that were not among the most or least frequently offered. For each participant, we created a small bundle

of advice by randomly selecting one suggestion from this pool of 11, a medium bundle of advice by randomly selecting three suggestions from the remaining pool of 10, and a large bundle of advice by selecting the remaining seven suggestions. The order in which participants read the small, medium, and large bundles of advice was randomly determined for each participant, and the order in which the suggestions that constituted each bundle of advice were presented was also randomly determined for each participant. It is worth noting that because each participant read three bundles of advice, all participants read all 11 of the suggestions, which just happened to appear in three different-sized bundles whose contents were randomly determined.

After reading a bundle of advice, participants in Studies S3a and S3b answered different questions. Participants in Study S3a estimated the effectiveness of the bundle by answering the question “How would receiving this advice affect the person's performance on the task?” They made this estimate on a seven-point scale whose endpoints were labeled *it would make their performance much worse* (1) and *it would make their performance much better* (7). Participants in Study S3b estimated the skill of the advisor who had purportedly offered that advice by answering the question “How skilled do you think the person who wrote this advice was at the game?” They made this estimate on a 101-point scale whose endpoints were labeled *not at all skilled* (0) and *extremely skilled* (100). After answering these questions, all participants were given the attention check used in Study 1. They also indicated whether they had prior experience with the Word Scramble task, how well they thought the average person did at one round of the task, and the same demographic measures used in Study 1.

Results

We excluded data from 7 participants in Study S3a and 7 participants in Study S3b, all of whom failed the attention check, leaving 700 participants (Study S3a: 193 males, 159 females, 1

prefer not to answer, $M_{\text{age}} = 34.60$ years, $SD = 10.18$ years; Study S3b: 170 males, 176 females, 1 prefer not to answer, $M_{\text{age}} = 36.06$ years, $SD = 10.73$ years) in the dataset. Excluding these participants does not appreciably change the results of any of the analyses we performed on the data we collected.

Study S3a

Did the number of suggestions that a bundle of advice contained influence participants' estimates of its effectiveness? To find out, we fit a linear mixed model to our data in R using the lme4 package. The dependent measure was the estimated effectiveness of each bundle of advice, and the independent within-participants variable was the number of suggestions that each bundle of advice contained. As a random effect, we included a random intercept term for the participant.

The analysis revealed that the more suggestions a bundle of advice contained, the more effective it was expected to be. The medium bundle of advice ($M = 5.01$, $SD = 1.20$) was estimated to be more effective than the small bundle of advice ($M = 4.80$, $SD = 1.25$), $b = 0.21$, $SE = 0.08$, $t = 2.53$, $p = 0.01$, 95% CI [0.05, 0.38], and the large bundle of advice ($M = 5.33$, $SD = 1.21$) was estimated to be more effective than both the medium bundle of advice, $b = 0.32$, $SE = 0.08$, $t = 3.74$, $p < 0.001$, 95% CI [0.14, 0.48] and the small bundle of advice, $b = 0.53$, $SE = 0.08$, $t = 6.27$, $p < 0.001$, 95% CI [0.37, 0.69].

Study S3b

Did the number of suggestions that a bundle of advice contained influence participants' estimates of the skill of the advisor who had purportedly offered that advice? To find out, we once again fit a linear mixed model to our data in R using the lme4 package. The dependent measure was the estimated skill of the advisor, and the independent within-participants variable

was the number of suggestions that each bundle of advice contained. As a random effect, we included a random intercept term for the participant.

The analysis revealed that the more suggestions a bundle of advice contained, the more skilled the advisor was estimated to be. Participants estimated that the advisor who had purportedly offered the medium bundle of advice ($M = 56.70$, $SD = 23.93$) was more skilled than the advisor who had purportedly offered the small bundle of advice ($M = 46.33$, $SD = 27.80$), $b = 10.39$, $SE = 1.66$, $t = 6.25$, $p < 0.001$, 95% CI [7.12, 13.64], and that the advisor who had purportedly offered the large bundle of advice ($M = 73.13$, $SD = 21.03$) was more skilled than both the advisor who had purportedly offered the medium bundle of advice, $b = 16.43$, $SE = 1.66$, $t = 9.90$, $p < 0.001$, 95% CI [13.18, 19.69] and the advisor who had purportedly offered the small bundle of advice, $b = 26.82$, $SE = 1.66$, $t = 16.14$, $p < 0.001$, 95% CI [23.56, 30.07].

Between-Participants Analyses

Participants in Studies S3a and S3b were presented with several bundles of advice. Is it possible that this within-participants design created a demand for participants to make different estimates about the different bundles of advice? To rule out this possibility, we performed a set of between-participants analyses that used participants' estimates about the first bundle—and only the first bundle—of advice they read. Specifically, for Study S3a we ran a one-way ANOVA using the number of suggestions that the first bundle of advice contained as the independent variable and participants' ratings of advice effectiveness as the dependent variable, and for Study S3b we ran a similar ANOVA using estimates of the advisor's skill as the dependent variable. These between-participants analyses confirmed the results of our within-participants analyses: The number of suggestions that the first bundle of advice contained

influenced participants' ratings of advice quality, $F(2,1056) = 5.13, p < 0.01, \eta^2 = 0.01$, as well as participants' estimates of the advisor's skill, $F(2,1035) = 45.58, p < 0.01, \eta^2 = 0.08$.

(Section 4.2) Additional Studies S4a and S4b

STUDIES S4A AND S4B:

CAN ADVICE IMPROVE PERFORMANCE IN WORD SCRAMBLE?

Study S4 had two parts. In Study S4a, we examined the data from Study 2 to determine whether advisees’ performances were associated with the particular “kinds” of advice they received. In Study S4b, we gave a new group of advisees a kind of advice that was associated with superior performance by advisees in Study 2, or a kind of advice that was not associated with superior performance by advisees in Study 2, or no advice, and then measured their performances in Word Scramble.

STUDY S4A:

**WERE THE PERFORMANCES OF ADVISEES IN STUDY 2 ASSOCIATED WITH THE
KINDS OF ADVICE THEY RECEIVED?**

Method

Procedure

As a reminder, coders in Study 4 were asked to read the advice offered by advisors in Study 2 and to report how many independent suggestions each bundle of advice contained. These coders identified 101 unique and independent suggestions. In Study S4a, we showed these 101 independent suggestions to two new coders and asked them to indicate which of six “kinds of advice” each of these independent suggestions represented. The six kinds of advice were: (1) *letters advice* (coders were told that this was advice about “which letters or kind of letters to look for in the puzzle.”); (2) *mechanics advice* (coders were told that this was “advice about how to

physically play the game,” including how to “to click and drag letters to spell words... how to select words or letters, how to use the mouse, or other issues about the game’s interface itself.”; (3) *spatial advice* (coders were told that this was advice about “where or how on the board to look for words.”); (4) *speed/pacing/mindset advice* (coders were told that this was any advice about “how to manage time in the game... the pace of the game” or “any advice about mindsets” or “handling pressure”); (5) *word length advice* (coders were told that this was advice about “the length of words one should look for” including “whether it’s better to look for short or long words.”); and (6) *word type advice* (coders were told that this was advice about “what kinds of words to look for... about common or rare words, or whether words can repeat”). Once coders had coded each of the 101 independent suggestions, they shared their judgments with each other and discussed any differences until they reached consensus.

Results

We used the judgments of the coders to determine the kinds of advice each of the advisees in Study 2 had received, and then fit a linear mixed model to our data in R using the lme4 package to examine the relationship between the kind of advice they received and their performance after receiving it. The dependent measure was the advisee’s performance (i.e., the number of words the advisee found in a particular round); the independent within-participants variable was the number of the round (1-6); and the independent between-participants variables were the kinds of advice received (*letters advice*, *mechanics advice*, *spatial advice*, *speed/pacing/mindset advice*, *word length advice*, and *word type advice*). As random effects, we included a random slope term for the round and a random intercept term for the advisee. The random slope term significantly improved model fit, $\chi(2) = 153.48, p < 0.001$, as did the random

intercept term, $\chi(2) = 3651.97, p < 0.001$. The independent between-participants variables (kinds of advice) also improved model fit, $\chi(6) = 25.46, p < 0.001$.

Was receiving a particular kind of advice associated with better performance in Study 2? As Table S2 shows, advisees performed better on later rounds ($b = 0.83, SE = 0.02, t = 47.07, p < 0.001, 95\% CI [0.80, 0.86]$), indicating that their performances improved with practice. More importantly, advisees performed better after receiving letters advice, spatial advice, speed/pacing/mindset advice, and word length advice, but they did not perform better after receiving mechanics advice, and performed marginally worse after receiving word type advice. In short, advisees in Study 2 performed better after receiving some kinds of advice but not others.

Table S2*Linear Mixed Model of the Effects of Six Kinds of Advice on Advisee Performance in Study S4a*

Effect	Estimate	95% CI	SE	t	p
Board number	0.83***	[0.80,0.86]	0.02	47.07	<0.001
Letters advice	0.60**	[0.06,1.15]	0.28	2.17	0.03
Mechanics advice	-0.02	[-0.60,0.55]	0.29	-0.07	0.95
Spatial advice	0.77**	[0.12,1.43]	0.33	2.32	0.02
Speed/Pacing/Mindset advice	0.66**	[0.16,1.16]	0.26	2.56	0.01
Word length advice	0.66**	[0.16,1.16]	0.25	2.59	0.01
Word type advice	-0.67*	[-1.39,0.05]	0.37	-1.82	0.07
Constant	6.43	[5.82,7.03]	0.31	20.81	<0.001
N _{observations}	8802				
N _{participants}	1467				
Marginal R ² / Conditional R ²	0.08 / 0.82				

Note: * $p < .1$, ** $p < .05$, *** $p < .01$

Of course, correlation does not imply causation, so in Study S4b, we gave a new group of participants a kind of advice that was associated with superior performance by advisees in Study 2, or a kind of advice that was not associated with superior performance by advisees in Study 2, or no advice, and then measured their performance.

STUDY S4B:

DO SOME KINDS OF ADVICE IMPROVE AN ADVISEE’S PERFORMANCE IN WORD SCRAMBLE MORE THAN OTHER KINDS OF ADVICE?

Method

Participants

Advisees were 1420 users of Amazon Mechanical Turk (767 males, 639 females, 14 prefer not to answer, $M_{age} = 38.99$ years, $SD = 11.82$ years) who were paid for their participation.

Procedure

As in Study 2, advisees were introduced to Word Scramble and then played a single round of the game. All advisees were shown the same board in this first round. Next, advisees were randomly assigned to one of three conditions. Those who were assigned to the *helpful advice* condition ($N = 472$) were shown a piece of “letters advice” that in Study S4a was associated with superior performance: “After you find a word, add an ending like -ED, -S, or -LY to get another word.” Participants who were assigned to the *unhelpful advice* condition ($N = 493$) were shown a piece of “mechanics advice” that in Study S5 Part1 was not associated with superior performance: “Be careful with the mouse when spelling words, because you can't undo letters you have selected.” Participants were told that the advice they received had been written by a previous player. Finally, participants who were assigned to the *no advice* condition ($N = 455$) were not shown any advice (though for narrative convenience we will continue to refer to these participants as “advisees”). Advisees in all three conditions then played five additional rounds of the game. Advisees in all conditions saw the same boards on each round.

After completing six rounds of the game, advisees in the two advice conditions rated the quality of the advice they had received on a seven-point scale whose endpoints were labeled *extremely unhelpful* (1) and *extremely helpful* (7), and then estimated how much the advice had improved their performance on a 101-point scale whose endpoints were labeled *0% improvement (not at all)* and *100% improvement or more (a lot)*. Advisees in both conditions completed the attention check item used in Study 1 as well as the same exploratory items used in Study 2.

Results

We excluded data from 71 advisees who did not find any words in any of the six rounds of the game. This left 1349 advisees in the data set (in the helpful advice condition, 236 males, 207 females, 5 prefer not to answer, $M_{\text{age}} = 39.49$ years, $SD = 12.16$ years; in the unhelpful advice condition, 243 males, 221 females, five prefer not to answer, $M_{\text{age}} = 38.72$ years, $SD = 11.94$ years; in the no advice condition, 236 males, 193 females, three prefer not to answer, $M_{\text{age}} = 39.46$ years, $SD = 11.82$ years). Excluding these advisees does not appreciably change the results of any of the analyses we performed on the data we collected.

Did receiving advice improve advisees’ performances more in the helpful advice condition than in the unhelpful advice condition or the no-advice condition? To answer this question, we analyzed the performances of advisees 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 condition (helpful advice, unhelpful advice, or no advice). We also included an interaction term between these two independent variables. As a random effect, we included a random

intercept term for the advisee. The interaction term significantly improved model fit, $\chi^2(2) = 6.93$, $p = 0.03$, as did the random intercept term, $\chi^2(1) = 9442.70$, $p < 0.001$.

The analysis revealed that the number of the round predicted advisee performance, $b = 0.82$, $SE = 0.03$, $t = 31.61$, $p < 0.001$, 95% CI [0.77, 0.87], which is to say that advisees' performances improved with practice. Additionally, there was a significant interaction between the number of the round and condition such that advisees in the helpful advice condition performed better than did advisees in the unhelpful advice condition, $b = 0.07$, $SE = 0.04$, $t = 2.03$, $p = 0.04$, 95% CI [0.003, 0.14], and also performed better than did advisees in the no advice condition, $b = 0.09$, $SE = 0.04$, $t = 2.47$, $p = 0.01$, 95% CI [0.02, 0.16]. The performances of advisees in the no advice condition and the unhelpful advice condition did not differ, $b = 0.02$, $SE = 0.04$, $t = 0.48$, $p = 0.63$, 95% CI [-0.05, 0.09].

References

- Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48.
- Critcher, C. R., & Rosenzweig, E. L. (2014). The performance heuristic: A misguided reliance on past success when predicting prospects for improvement. *Journal of Experimental Psychology: General*, 143, 480–485.
- Dwass, M. (1957). Modified randomization tests for nonparametric hypotheses. *The Annals of Mathematical Statistics*, 181–187.
- R Core Team. (2017). R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Simonsohn, U. (2018). Two lines: A valid alternative to the invalid testing of U-shaped relationships with quadratic regressions. *Advances in Methods and Practices in Psychological Science*, 1, 538-555.
- Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record*, 58, 475–482.