

EMPIRICAL ARTICLE

# Advice taking when the stakes are high: Evidence from a game show

Erik Løhre<sup>1</sup> and Torleif Halkjelsvik<sup>2</sup>

<sup>1</sup>Department of Leadership and Organizational Behaviour, BI Norwegian Business School, Oslo, Norway and <sup>2</sup>Department of IT Management, Simula Metropolitan Center for Digital Engineering, Oslo, Norway

**Corresponding author:** Erik Løhre; Email: [erik.lohre@gmail.com](mailto:erik.lohre@gmail.com)

**Received:** 15 September 2023; **Revised:** 2 February 2024; **Accepted:** 03 February 2024

**Keywords:** advice; weight of advice; judge–advisor systems; high stakes; estimation; wisdom of the crowd

## Abstract

Research on advice taking has demonstrated a phenomenon of egocentric discounting: people weight their own estimates more than advice from others. However, this research is mostly conducted in highly controlled lab settings with low or no stakes. We used unique data from a game show on Norwegian television to investigate advice taking in a high stakes and highly public setting. Parallel to the standard procedure in judge–advisor systems studies, contestants give numerical estimates for several tasks and solicit advice (another estimate) from three different sources during the game. The average weight of advice was 0.58, indicating that contestants weighted advice more than their own estimates. Of potential predictors of weight of advice, we did not detect associations with the use of intuition (e.g., gut feeling, guessing) and advice source (family, celebrities, average of viewers from hometown), but own estimation success (the proportion of previous rounds won) was associated with less weight of advice. Solicitation of advice was associated with higher stakes. Together with the relatively high weight on advice, this suggests that participants considered the advice valuable. On average, estimates did not improve much after advice taking, and the potential for improvement by averaging estimates and advice was negligible. We discuss different factors that could contribute to these findings, including stakes, solicited versus unsolicited advice, task difficulty, and high public scrutiny. The results suggest that highly controlled lab studies may not give an accurate representation of advice taking in high stakes and highly public settings.

## 1. Introduction

Decision makers often seek advice from others, but do not always follow their advisors' suggestions. In fact, the general conclusion from studies of advice taking is that people give too little weight to advice and instead stay closer to their own initial opinion (Bonaccio and Dalal, 2006). This phenomenon of *egocentric discounting* has been demonstrated repeatedly but mostly in laboratory experiments with student participants and with low or no monetary or social stakes (Kämmer et al., 2023). In this study, we use unique data from a game show on Norwegian television to investigate whether findings from carefully controlled laboratory settings can be generalized to advice taking in a high-stakes and highly public setting.

### 1.1. Advice taking and judge–advisor system research

Advice has been studied by scholars in many different fields, from communication and psychology to business and medicine; in a wide variety of contexts, from close relationships and health to

© The Author(s), 2024. Published by Cambridge University Press on behalf of Society for Judgment and Decision Making and European Association of Decision Making. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

organizational decisions; and using a large range of different methods, from qualitative analysis to surveys and experiments (MacGeorge and Van Swol, 2018). In this paper, we focus on research in the judge–advisor systems (JAS) tradition (Sniezek and Buckley, 1995). In JAS studies, a judge is tasked with giving an estimate of some unknown quantity. After first giving their initial judgment, the judge receives advice in the form of an estimate from another person, a group, or an algorithm, and can then provide a final, revised judgment. This has become one of the most common paradigms in research on advice, especially within social and organizational psychology (Kämmer et al., 2023).

Since judges in the JAS approach are tasked with estimating a numerical quantity, advice use can be measured on a continuous scale, in contrast to tasks where the judge chooses one among several categorical options. The most common measure of advice taking is weight of advice (WOA), calculated by the following formula:  $\frac{\text{judge final estimate} - \text{judge initial estimate}}{\text{advisor estimate} - \text{judge initial estimate}}$ . With this measure, 0 indicates that the judge has completely ignored the advice, i.e., stayed with their initial estimate; 1 indicates that the judge has relied completely on the advice, i.e., a final estimate identical to the advisor's estimate; and 0.5 indicates that equal weight was given to the advice and the initial estimate. WOA scores may fall outside of the 0–1 range, for example, if a judge adjusts their estimate away from both the initial estimate and the advisor's estimate, or adjusts in the direction of the advisor but beyond the advisor estimate. However, such cases are rare, perhaps less than 5% (Bonaccio and Dalal, 2006, p. 141).

In their early review of the advice taking literature, Bonaccio and Dalal (2006) concluded that studies within the JAS approach show that people do not heed advice sufficiently, with a WOA of 0.2 to 0.3, indicating that judges shift about 20% to 30% toward the advisor's estimate. A recent meta-analysis (Bailey et al., 2023) showed a similar tendency toward egocentric discounting, with an average WOA of 0.39, 95% CI [0.37, 0.42]. Such underweighting of advice is generally suboptimal: when judges are equally (in)accurate, averaging the two estimates will usually give the most accurate result (Soll and Larrick, 2009).

The power of averaging is based on the statistical principle that a combination of imperfect estimates can reduce error (e.g., Stroop, 1932), i.e., the absolute distance between the estimate and the true value. However, this reduction in error is most effective when the errors in the estimates are independent. For instance, if two estimates have the same kind of bias, such as both being overestimates, the benefit of averaging diminishes. The same principle is evident for the wisdom of the crowd (Galton, 1907; Surowiecki, 2004). The average estimate of a group is often more accurate than that of a typical group member, but when the independence of individual estimates is reduced, for instance through social influence (Lorenz et al., 2011), the crowd average becomes less accurate. Bracketing, the occurrence of estimates on opposite sides of the true value, plays a crucial role. A higher bracketing rate, indicating a higher frequency of estimates straddling the true value, increases the benefits of averaging (Larrick et al., 2012).

While averaging is generally a solid strategy, it is not necessarily one that people intuitively appreciate (Larrick and Soll, 2006). Indeed, when receiving advice, people will often use a strategy of choosing, attempting to identify whether their own or the advisor's estimate is the best rather than compromising (Soll and Larrick, 2009). Thus, an average WOA of 0.39 does not mean that people consistently move 39% toward the advice. Instead, the distribution of WOA is often found to be trimodal, with modes at 0 (choosing one's own estimate), 1 (choosing the advisor's estimate) and 0.5 (averaging between the two; Himmelstein, 2022).

In typical JAS studies, judges receive advice from a single advisor. However, when advice stems from a group, averaging would imply underweighting of the advice (Mannes, 2009). Imagine for instance that you receive advice in the form of the mean estimate of a group of four independent individuals. Equal weighting here would mean that each estimate should contribute 20% to the final judgment. Thus, if you weight your own estimate as equal to the group average, i.e., a WOA of 0.5, you are egocentrically discounting the information in the group advice, while a WOA of 0.8 could be called ego-neutral. Mannes (2009) compared weight on advice for groups and individuals and found that people put more weight on advice from groups, but that group size did not have

sufficient influence, such that when group size increased, weight on advice did not increase to the same degree. This could lead to severe underweighting of group advice, especially from large groups.

## 1.2. Limitations of the lab

Most research on judge-advisor systems has used relatively artificial tasks, with low monetary and social stakes, in laboratory settings. Study 1 from Yaniv and Kleinberger (2000) serves as a typical example. In this study, undergraduate respondents gave an initial answer to 15 questions about the dates of historical events, e.g., “In what year were the Dead Sea scrolls first discovered?”. They then received another estimate as advice and gave a final estimate for each question. All participants received 12 shekels (\$3.6) and in addition 1 shekel as a bonus for each estimate that had a better than average accuracy score, such that they could receive up to 15 shekels (\$4.5) in bonus payment. Participants received advice whether they wanted it or not, and had no opportunity to interact with the advisor. Some recent research has expanded the classical paradigm, for instance by allowing participants to sample as many advisory estimates as they want before providing a final estimate (Hütter and Ache, 2016), but still most of the research on advice taking has similar features: students responding to general knowledge questions in the lab, with low or no monetary stakes, with little interaction with the advisors, and with low reputational stakes since answers are not publicly available to anyone except the researcher (Kämmer et al., 2023).

It is a natural question whether similar results are found in situations with different features than the typical lab study. One concern, often raised by economists, is whether findings from psychological experiments apply when stakes are higher (see e.g., Thaler, 2016). For instance, researchers have tested the ultimatum game (Andersen et al., 2011) and classic heuristics and biases-tasks (Enke et al., 2023) with very high stakes by recruiting participants from low-income countries, where a relatively small number of dollars has a larger value. For advice taking, research from organizational contexts with real life high stakes decisions could potentially be relevant, but such studies mostly rely on survey responses to questions about how often someone in an organization, for example a CEO, seeks advice from others (e.g., Alexiev et al., 2020; Vestal and Guidice, 2019) and thus do not provide a measure comparable to WOA.

One recent study (Pálfi et al., 2022) used the JAS framework but employed realistic descriptions of patients consulting a doctor with symptoms that could be suggestive of cancer. A group of general practitioners provided their initial estimates of the probability of cancer, received realistic advice from an algorithm, and provided an updated estimate. For this consequential judgment, the GPs on average weighed their own estimate and the advice approximately equally,  $M_{WOA} = 0.54$ . This suggests less egocentric discounting when the task concerns more important domains, but even this study employed hypothetical scenarios, with no real stakes involved, and involved advice from a validated cancer risk algorithm. Thus, to enrich our knowledge about advice taking, it would be beneficial with studies outside of the typical lab setting.

The Norwegian game show “Alle mot 1” (“All against 1”) provides a unique opportunity to study advice taking in a real-world setting. The main features of the game show correspond closely to the procedure used in lab studies in the JAS-tradition but involves much higher stakes. The context is further distinguished by the participants being in a highly public setting, where they interact with their advisors. Furthermore, the tasks are more difficult, or involve a higher degree of uncertainty than many laboratory tasks. The approach of studying behavior in game shows is not new. TV shows like Jeopardy, Deal or No Deal, Golden Balls and The Price is Right have been analyzed and used to draw conclusions about decision making under uncertainty with high stakes (Jetter and Walker, 2017; Post et al., 2008; Teeselink et al., 2022; Van Den Assem et al., 2012), but to our knowledge, no such study has looked at advice taking. Similar to these previous studies investigating game shows, the present context implies that the data is observational, lacking the rigorous control typical of lab studies.

### 1.3. Contextual factors associated with advice taking and advice solicitation

Advice taking can be dependent on contextual factors. Below we briefly describe some of the factors that can be explored within the context of the game show under study.

Social factors are of obvious importance in advice taking. Harvey and Fischer (1997) found that even experienced participants followed advice from novices to some extent. This was interpreted as showing a reluctance to reject help, or in other words, a norm to make at least *some* use of the advice you are offered. Such a norm would presumably be of even higher importance when judges and advisors interact face to face, as they do in the game show. In such contexts, building or maintaining relationships, and concerns about politeness or impression management may become front and center (Blunden et al., 2019; MacGeorge and Van Swol, 2018).

People may take advice from novices, but they heed advice from experienced advisors more (Harvey and Fischer, 1997). Expertise is thus one source factor that determines advice taking (Feng and MacGeorge, 2010). In a similar vein, judges seem sensitive to advisors' past performance, with greater weight given to advice from advisors with better performance (e.g., Yaniv and Kleinberger, 2000). In Bailey et al.'s (2023) meta-analysis, the only unique predictor of weight of advice was information suggestive of advice quality, with mean WOAs of 0.32 for low quality advice, 0.37 for medium quality, and 0.48 for high quality advice. In the context of the game show we study, contestants' and advisors' past estimation success can be used as a measure of their performance.

Another source factor that has received attention is advice from humans (experts) vs. from algorithms. Some studies find algorithm aversion, i.e., that people are more willing to take advice when it is said to come from humans than from an algorithm (Burton et al., 2020; Dietvorst et al., 2015), other studies present evidence for algorithm appreciation, i.e., higher weight on advice from algorithms (Logg et al., 2019). In the game show studied in the current article, the contestant can solicit statistical advice, i.e., the average answer of app players in their hometown, or personal advice from family and friends or celebrities present in the studio. Differences between these three sources could shed light on the importance of the social dimension.

Finally, advice may be given in different ways or be based on different types of processing. Previous research suggests that people are more willing to follow advice based on thorough deliberation than advice based on intuition and "gut feelings" (Tzioti et al., 2014), but this may depend on the decision domain and on the need to express authenticity in a choice (Oktar and Lombrozo, 2022). By observing how the game show contestants and advisors deliberate or justify their estimates with reference to intuition or guessing, we can assess the weighting of intuitive estimates.

### 1.4. The present research

We analyze advice taking behavior in five seasons of the television game show "Alle mot 1", launched by the Norwegian public broadcaster NRK in 2018. The format of the show closely mirrors typical studies on advice taking. The show involves high stakes (up to NOK 100 000, or about €9300/\$10,000), has participants giving numerical estimates for different challenging questions, and crucially, also has each participant receive advice from friends and family, celebrities, or an aggregated estimate from their hometown, at three self-chosen points in the game.

First, we will explore the characteristics of the game in terms of wins and losses, stakes, distribution and accuracy of judgments, and solicitation of advice. Second, we will analyze the weighting of advice. We compare contestants' weight on advice in this high stakes, highly public context with the patterns usually found in low stakes experimental laboratory research, and we investigate contextual factors, including the source of advice (friends and family, celebrities, average estimate by hometown), the contestant's own estimation success, and the use of intuition (contestant or advisors expressing the use of intuition or guessing rather than a more analytical or logical approach). This second part of the analyses is preregistered. Third, we explore the usefulness of advice and various strategies by comparing the actual and hypothetical outcomes of (a) the factual (revised) estimates, (b) estimates

ignoring advice (where weight of advice, WOA, equals 0), (c) estimates that directly adopt the advice (WOA = 1), and (d) estimates averaging between the initial estimate and the advice (WOA = 0.5).

## 2. Method

### 2.1. Participants and data

We analyzed data from 49 participants who made 382 judgments, and received advice for 147 of the judgments, in 5 seasons of the show (from 2018 to 2022). Additionally, 263 estimates were given by the celebrities and 213 by family and friends, after contestants' estimates were locked in. These estimates could not be used by contestants, and were labeled as *counterfactual advice*. The contestants were 26 women and 23 men aged 19 to 59 years ( $M = 34$ ). Contestants applied for participation in the show and were chosen based on interviews with the producers of the show and their responses to some example tasks. We contacted the production company who explained that they aim for a diverse participant group but that the most crucial factor is that participants should enjoy taking part in the game and being on TV. This probably means that the contestants score higher on extraversion and openness than a random or representative sample of the Norwegian population.

### 2.2. Open science statement

The project received approval from Sikt, Norwegian Agency for Shared Services in Education and Research (reference number 289231) and from the Ethical Review Board at BI Norwegian Business School. Although the information used to create the data is sourced from a public game show, the combination of this information into a data set constitutes personal data under Norwegian legislation. It is not possible to anonymize the data because the data can be linked to public video recordings, and we are therefore not authorized to share the original data. We do, however, share our R data analysis syntax and a datafile where the order of observations has been shuffled independently for each variable, providing observed distributions for each individual variable while ensuring anonymity. Furthermore, we preregistered some of the analyses reported here and report which analyses were preregistered. The preregistration, the data analysis syntax, the permuted data set, and the Supplementary materials with additional visualizations are available on <https://osf.io/rkpy3/>.

### 2.3. Procedure

#### 2.3.1. Advice taking in “Alle mot 1”

In the game show “Alle mot 1”, participants are given a series of numerical estimation tasks where it is hard to know the answer in advance, e.g., how long will it take a skateboarder to go down a bobsleigh track during summer; how many meters will a cow walk in 24 hours. Supplementary Table S1 describes 10 example tasks (two tasks from each of the five seasons) along with the scale, Norway's estimates, and outcomes. There is one contestant per episode, competing with the viewers who submit their estimates in an app specially designed for this purpose. Contestants and viewers provide their estimate by moving a slider along a scale, with the minimum and maximum number given by the show producers (e.g., “How many balloons will be popped after one minute? The answer is between 0 and 120.”). Participants win if their answer is closer to the actual outcome than “Norway's” answer, calculated simply as the average answer from app players (the arithmetic mean). In other words, the aim of the game is to beat the wisdom of the crowd. The contestant, the viewers, and the advisors all provide their estimates independently: after a question has been introduced, there is a 30 second time limit for everyone to provide their answers. This prohibits viewers and advisors from anchoring on the contestant's answer. The participant and “Norway” accumulate money won throughout the game in separate pots, and the final round decides whether the participant or a randomly chosen app player will receive their respective amounts. The amounts at stake increase for each successive question in

the game show and were (in NOK) 4,000; 6,000; 10,000; 15,000; 25,000; 40,000 in the first season. In the second season, they were 1,000; 3,000; 6,000; 10,000; 15,000; 25,000; and 40,000. In the three last seasons, the amounts were 5,000; 8,000; 10,000; 12,000; 15,000; 20,000; and 30,000. In theory, participants can win up to NOK100,000 (approximately €9,300 or \$10,000 at the time).

Parallel to what is commonly done in JAS studies, participants provide an initial numerical estimate, and can solicit advice in the form of a numerical estimate from three different sources: 2–3 friends and/or family, 2–3 celebrities, or the average answer given by app players in their hometown<sup>1</sup>. The overall number of app users is generally above 200,000. We do not have data on the number of app players involved in the hometown average, but rough calculations show that hometown averages will usually be based on answers from hundreds or thousands of viewers.<sup>2</sup> Each aid can be used only once during an episode of the show. After receiving advice, participants are asked whether they want to change their answer, and then provide a final estimate.

### 2.3.2. Coding, preregistration, and analysis

The External Affairs Department of the Norwegian Broadcasting Corporation provided text files with subtitles for all episodes. A research assistant coded the data based on the subtitles. When the subtitles lacked information about estimates, the video recordings of the show were consulted (available at nrk.no from any Norwegian IP address). When advisors provided an interval instead of a point estimate, we coded the advice as the midpoint of the interval. When the advisor provided more than one estimate as counterfactual advice, i.e., after the contestant's answers were locked in, we used the first estimate mentioned by the advisors, or their actual submitted estimate if it became clear from the context.

After the first season was coded, we inspected the data and derived hypotheses based on exploratory analyses. We preregistered promising analyses and analyses for variables on which we believed the remaining seasons would provide enough data for meaningful results.

In the preregistered analyses, we excluded data from the first season which were used to generate hypotheses, but used the complete data for descriptive analyses, for tests marked as exploratory, and in analyses of combination strategies. In analyses that involve multiple judgments by each participant, confidence intervals and *p*-values are based on cluster-robust standard errors grouped on participant.

## 2.4. Measures

### 2.4.1. Scaled estimates and actual outcomes

As the estimation tasks are given on completely different scales (from kilograms and minutes to number of items), we scaled the estimates and the actual outcomes by the response scale in the game show (0 = minimum of response scale; 1 = maximum of response scale). Specifically, the lower limit of the response scale is subtracted from the estimate and the maximum of the response scale, and the first is divided by the second to obtain the proportion of the maximum response,  $\frac{\text{estimate} - \text{lower limit}}{\text{upper limit} - \text{lower limit}}$ . For example, if the response can be given on a scale from 10 to 90, an estimate of 30 is 25% of the maximum response,  $\frac{30 - 10}{90 - 10} = 0.25$ . In a few cases when tasks were performed live, outcomes were outside of the predefined maximum bound. In such cases they were treated as the maximum, both in the show and by us. For example, if the outcome was 110 when the scale was from 10 to 90, we coded this as 1, not as 1.25.

<sup>1</sup>Friends and family are chosen by the contestant, but the production company encourages contestants to choose people who would have fun guessing and would enjoy being on TV. Celebrities are chosen by the production company based on their ability to express themselves, and have a wide range of backgrounds, with comedians, TV personalities, politicians, athletes and influencers as some examples. Given the idiosyncratic nature of the tasks, there is little reason to believe that celebrities or friends and family possess greater expertise than contestants.

<sup>2</sup>The smallest hometown in the seasons analyzed here was Sommarøy, with 304 inhabitants, and the largest was Oslo, with 709,037 inhabitants according to Wikipedia. Assuming an app usage of 3.6% (about 200,000 app users in total, with the Norwegian population being approximately 5.48 million), this gives a potential range from 11 to 25,525 for the crowd involved in the hometown average.

#### 2.4.2. Weight of advice

WOA was coded as  $\frac{\text{judge final estimate} - \text{judge initial estimate}}{\text{advisor estimate} - \text{judge initial estimate}}$ . As this formula inherently standardizes the WOA, we used the original estimates (not scaled estimates as described above). When the WOA was negative (revising in the opposite direction of the advice) or larger than 1 (adjusting beyond the advice), we trimmed the data to 0 and 1, respectively. There were 10 observations like this, 4 smaller than 0 and 6 larger than 1. Manual inspection of these 10 cases did not reveal any striking patterns, except that neither kind of adjustment outside of the 0–1 range seemed to be a successful strategy, as contestants only won two of these rounds, Norway won seven times, and there was one tie. In three instances, the contestant's initial estimate was identical to the advisors' estimate, making it impossible to calculate WOA since the denominator then is 0. These three cases were excluded, giving 144 valid observations of WOA.

#### 2.4.3. Estimation error

As a measure of estimation error, we computed the standardized percentage point deviation for each task. This entails standardizing the response scale for each question so it varies between 0 and 100, and computing the absolute deviation between the estimate and the correct answer. A lower score on this variable shows that the response was closer to the correct answer. We opted for this measure for three reasons: it makes answers on different scales comparable, it does not differentially punish small and large errors as, for example, mean squared error would do, and it does not introduce asymmetries depending on the correct answer like percentage-wise mean absolute deviation would do.

#### 2.4.4. Past estimation success

The game show proceeds over several rounds (7 in the first season, 8 in the proceeding seasons), and the success rate at different points in the game is highly variable. We calculated a measure representing the proportion of past winning judgments made without any aid. The first judgment was omitted, the second judgment could take the values 0 and 1, the third additionally could take the value 0.5 if no advice has been used so far and the contestant won once and lost once, and so on for the next judgments (e.g., success rate at fourth judgment could take values 0, 0.33, 0.67, and 1).

#### 2.4.5. Intuition

A research assistant coded the use of intuition for contestants' and advisors' answers, with 0 meaning no mention or information about intuition, and 1 indicating that the words intuition, gut feeling, or guessing were mentioned. The inter-rater correlation between the coding of the assistant and one of the authors for the first season data was  $r(68) = 0.76$  for participants' estimates, and  $r(17) = 0.62$  for advisors' estimates;  $r(87) = 0.72$  combined.

### 3. Results

#### 3.1. Dynamics of the game

In this section, we explore the dynamics of the game in terms of actual wins, losses, stakes, and advice solicitation, as well as the relation between stakes and advice solicitation. The analyses in this section were not preregistered. [Table 1](#) shows who won the rounds according to question number for seasons 2–5, with season 1 omitted from the table as there was one fewer round in the game that season. Although there appeared to be an advantage for Norway in terms of total wins, the number of wins for round 8, which determines who wins money, was about the same. In the first season, the final round was the 7<sup>th</sup>, where Norway won in 7 out of 10 episodes (no ties).

[Figure 1](#) shows the distribution of the estimates from contestants, three different types of advisors, and the average of the viewers, and the actual outcomes, as well as their correlation. From [Figure 1](#), it is notable that the hometown estimates, as well as Norway's estimates, include few values close to the minimum or maximum, which is natural as they are based on the average of many judgments. Note that

**Table 1.** Number of wins according to question number for seasons 2–5.

	1	2	3	4	5	6	7	8	Total
Tie	3	3	3	2	0	6	3	4	24
Participant	14	22	13	16	17	13	19	17	131
Norway	22	14	23	21	22	20	17	18	147

there are few observations of hometown estimates because these data were only revealed when chosen as aid, i.e., were not presented as counterfactual advice like celebrities' and family and friends' advice.

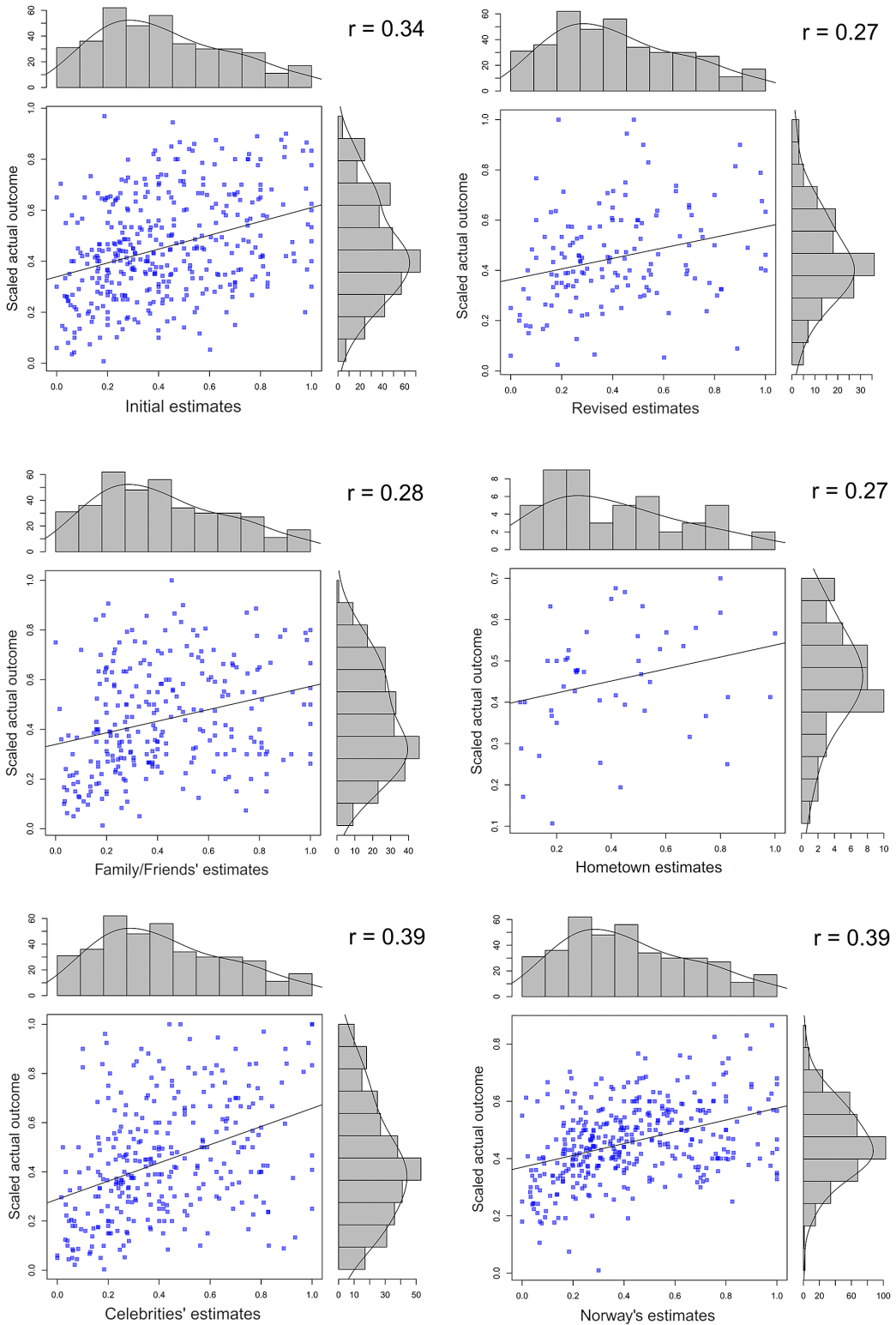
How did the estimates relate to the outcomes? As the tasks involved unusual situations such as spiky bowling balls smashing balloons, one could suspect that the judgments were completely random. However, the correlation between actual outcomes and the estimates of the contestants and the advisors ranged from 0.27 to 0.39, suggesting that the cues presented to the contestants included some valid information, at least for some of the tasks. Note that the outcomes and estimates are calculated as proportions of the response scale, so the correlations should not merely reflect the different units and quantities used in the estimation tasks. Completely random outcomes or estimates would give a correlation of 0. Furthermore, the error scores ranged from 0.20 to 0.22, whereas two random variables on average would give an error score of 0.33, based on a simulation of 100,000 draws of two random uniform variables from 0 to 1. According to exploratory inferential tests, there were no pairwise statistically significant differences between the error scores of contestants ( $M = 0.21$ ,  $SD = 0.16$ ) and the three advisors; family and friends ( $M = 0.22$ ,  $SD = 0.18$ ), celebrities ( $M = 0.20$ ,  $SD = 0.18$ ), and hometown ( $M = 0.21$ ,  $SD = 0.14$ ),  $ps > 0.4$ . However, Norway's estimation error ( $M = 0.19$ ,  $SD = 0.14$ ) was lower than the contestants' error: the viewers on average were two percentage points closer to the actual outcomes than the contestants,  $B = -0.02$ ,  $t(48) = -2.866$ ,  $p = .006$  (exploratory inferential test).

In the game show, stakes increased with each subsequent question and there was a strong tendency to save the advice for the final questions. Table 2 shows there was no advice solicitation in the first round, when stakes were low, and there was most solicitation in the final two rounds, when stakes were the highest. In the first season, not shown in the table as there were only seven rounds, participants in the final round solicited advice from celebrities two times, hometown four times, family one time, and no aid three times. If the solicitation of advice is treated as a purely random choice, the probability of solicitation is exactly 0.375 (3 sources of advice divided by 8 rounds) in round 1 and becomes either 1 or 0 in the final round depending on previous choice; however, the expectation or average probability is 0.375 for all rounds. The actual proportions of advice solicitation from round 1 to round 8 were 0, 0.026, 0.103, 0.385, 0.256, 0.538, 0.846, and 0.846, which clearly refutes the notion of random choice.

One could suspect that people are saving advice for later rounds because the questions become harder, as it is a common feature of game shows like *Jeopardy!* and *Who Wants to Be a Millionaire?* that questions become increasingly difficult with higher stakes. However, we do not find evidence for this being the case for the current game show by using error scores as a measure of task difficulty. The error scores for the participants' initial estimates were similar for first three questions ( $M = 0.21$ ,  $SD = 0.16$ ) and the last three questions ( $M = 0.20$ ,  $SD = 0.15$ ), suggesting that the objective difficulty of the tasks did not increase in later rounds. Relatedly, the probability of solicitation did not increase with higher error scores ( $r = -0.02$ ), neither did WOA correlate with error ( $r = -0.07$ ), see Supplementary Figure S1. Altogether, this suggests that the solicitation of advice in later rounds is not driven by increased difficulty, and points toward stakes as a more likely explanation.

To quantify the increase in solicitation according to stakes, we regressed solicitation on the amount at stake. Each NOK 1000 increased the probability of soliciting advice by 2 percentage points on average,  $p < .001$ . To further investigate whether higher stakes influenced advice taking, we focused on the final round, which determines whether the participant wins. The average stake accumulated to the last round





**Figure 1.** Distributions and associations between actual outcomes and estimates.

**Table 2.** Advice solicitation in seasons 2–5, according to question number.

	1	2	3	4	5	6	7	8
No advice	39	38	35	24	29	18	6	6
Family/friends	0	0	0	7	5	5	14	8
Celebrity	0	1	4	8	3	13	10	0
Hometown	0	0	0	0	2	3	9	25

was 46,490 ( $SD = 22,313$ ; range 3,000 to 90,000), which corresponds to about one month's salary in Norway (median monthly income was NOK 47,680 in 2022).<sup>3</sup> The amount at stake was not associated with WOA,  $r(37) = 0.001$ , 95% CI [-0.31, 0.32],  $p = .996$ , see Supplementary Figure S2.

Table 2 shows that advice from celebrities was usually used first (mean question number = 5.4), advice from friends and family second ( $M = 6.3$ ), while advice from the hometown was typically used in the last rounds ( $M = 7.5$ ). Exploratory tests of pairwise differences (in ranks) between the types of advice obtained  $ps < 0.05$ . This suggests that participants valued the advice from the hometown most highly, perhaps indicating some intuitive appreciation of the wisdom of crowds.

### 3.2. Weight of advice

In this section, we analyze the distribution of WOA and its correlates. Unless specified as exploratory, the analyses below are preregistered. The mean WOA in this setting was higher than that is often found in the advice taking literature,  $M_{WOA} = 0.58$ , 95% CI [0.52, 0.64]. This was higher than equal weighting, WOA = 0.5, according to an exploratory  $t$ -test on subject means,  $t(48) = 2.63$ ,  $p = 0.011$ . Thus, participants on average weighted advice more than their own estimates, showing no egocentric discounting.

One possibility is that the high WOA in the present context is due to the voluntary solicitation: in JAS studies participants often receive advice whether they want it or not, while contestants in the game show choose themselves when to solicit advice and could therefore in principle use it when needed for particularly difficult questions. To explore this possibility, we compared WOA in different rounds. The WOA in rounds one to five ( $M = 0.54$ ,  $SD = 0.30$ ), where choice could be solicited if needed, was not higher than the WOA in the last ( $M = 0.59$ ,  $SD = 0.35$ ) or two last rounds ( $M = 0.59$ ,  $SD = 0.35$ ), where advice was typically “forced” if the participant had saved the advice (40 out of 49 participants).

As in laboratory studies, the distribution of WOA was roughly trimodal (see Figure 2), with 22% fully adopting advice (WOA = 1), 12% fully rejecting advice (WOA = 0), and 22% averaging (WOA between 0.4 and 0.59). However, in laboratory studies the most frequently chosen option seems to be to reject advice completely: for example, in Soll and Larrick (2009) and in three studies reviewed by Himmelstein (2022), participants fully rejected advice between 25% and 60% of the time, and fully adopted advice 2%–13% of the time. The rate of averaging in the game show does not seem too different from lab studies, e.g., in Soll and Larrick (2009), the rate of averaging was between 17% and 25%. Distributions for each advice source are shown in Supplementary Figure S3.

We assumed there would be differences in weighting according to the three sources of advice, but with no clear directional hypothesis. Relying on the wisdom of the crowd would imply a higher weighting of the hometown estimate. On the other hand, participants may tend to give more weight to advice from individually identifiable advisors such as friends and celebrities present in the room rather than a statistical figure like the average answer from their hometown. Table 3 shows the WOA for the different sources.

<sup>3</sup>Source, Statistics Norway, the national statistical institute of Norway: <https://www.ssb.no/arbeid-og-lonn/lonn-og-arbeidskraftkostnader/artikler/hva-er-vanlig-lonn-i-norge>

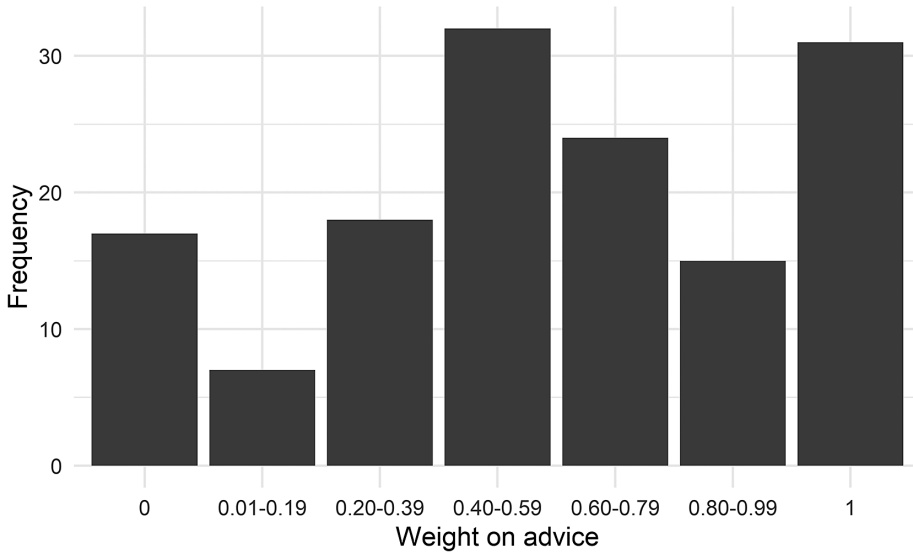


Figure 2. Distribution of weight of advice.

Table 3. Weight of advice according to three sources of advice for seasons 2–5.

	N	M	Median	WOA = 0	WOA = 0.5	WOA = 1
Family/friends	38	0.53	0.58	16% (6)	21% (8)	13% (5)
Celebrities	38	0.63	0.62	5% (2)	18% (7)	26% (10)
Hometown	38	0.61	0.64	11% (4)	26% (10)	24% (9)

Note: WOA = 0.5 includes 0.4 to 0.60.

In a regression with WOA as the outcome and indicators for advice source as predictors, controlling for question number, we found no differences between source of advice,  $\chi^2(df = 2, N = 39) = 1.67, p = 0.43$ . There was no tendency to weight the hometown estimates less (suggesting some kind of algorithm or statistical combination aversion) or more (following the wisdom of the crowd) than the other estimates.

In Table 3, the means and medians do not directly reflect the weighting of those who actually combine their estimate with the advisors' ( $0 < WOA < 1$ ). However, if we calculate means and medians after omitting judgments where WOA was 1 or 0 the results were similar to Table 3: family and friends  $M = 0.56, Mdn = 0.59$ ; celebrities  $M = 0.53, Mdn = 0.58$ ; hometown,  $M = 0.57, Mdn = 0.54$ .

The next hypothesis was that participants would weight advice less after successful own predictions in the preceding tasks. A linear regression where WOA was regressed on the proportion of past success, controlling for question number and type of advice, gave an estimated effect very close to zero,  $B = -0.03, SE = 0.15, z = -0.18, 95\% \text{ CI } [-0.33, 0.27], p = 0.86$ . We also preregistered an analysis of own estimation success using a within-subject model. This model showed an effect of the proportion of past wins,  $B = -0.93, SE = 0.45, t(38) = -2.21, p = 0.044, 95\% \text{ CI } [-1.84, -0.03]$ , suggesting that within subjects, rounds that were preceded by wins from one's own unadvised estimates reduced the weight of advice. Although the above analysis identifies an effect, the effect size is not easily interpretable within subjects using the original scaling of the variables. It suggests that an increase from zero wins in the past rounds to 100% wins in past rounds, within an individual, changes the WOA from 1 to almost 0. As the number of past wins never goes to 100% after initial losses, a more sensible way to frame the result is that for example a 50 percentage point increase in the proportion of past wins decreases

**Table 4.** How different strategies for weighting of advice would have affected estimation error, number of rounds won, average amount at stake, and average amount won [95% CI].

	Absolute proportion error		Outcomes and hypothetical outcomes				
	M (SD)	Range	Won	Tie	Lost	Pot <sup>a</sup>	Amount Won <sup>a</sup>
Observed WOA	0.22 (0.16)	0.00, 0.81	58	13	76	46.5	21.0 [14.3,29.2]
WOA = 0	0.22 (0.16)	0.00, 0.78	59	5	83	45.3	19.6 [12.5, 27.4]
WOA = 1	0.22 (0.18)	0.00, 0.80	54	25	68	45.1	22.2 [16.3, 29.3]
WOA = 0.5	0.21 (0.15)	0.00, 0.73	61	3	83	46.9	20.0 [13.1, 27.8]

<sup>a</sup>Reported in NOK1000.

the WOA by approximately 0.5. Exploratory tests of the *number* of past wins were consistent with the analyses of *proportion* past wins.

The final preregistered research question regarding weight of advice was whether people incorporate advice to a greater degree if their own or the advisors' estimates are made by intuition. If a participant gives an answer that is just a guess or based only on a gut feeling, they may be more likely to follow advice, while if advisors state that their answer is a guess or an intuitive response, participants may give less weight to the advice (Tzioti et al., 2014). In a regression with WOA as the outcome, predicted by participant's intuition and advisor's intuition, controlling for question number and type of advice (celebrity versus family and friends), we did not detect any effect of participant's intuition,  $B = -0.16$ ,  $SE = 0.17$ ,  $z = -0.97$ ,  $p = 0.33$ , nor an effect of advisor's intuition,  $B = -0.13$ ,  $SE = 0.12$ ,  $z = -1.11$ ,  $p = 0.27$ . Similar results were obtained in a within-subject regression,  $B = -0.22$ ,  $SE = 0.20$ ,  $t(38) = -1.10$ ,  $p = 0.28$ ;  $B = -0.20$ ,  $SE = 0.21$ ,  $t(38) = -0.97$ ,  $p = 0.34$ .

### 3.3. Usefulness of advice and best strategies

In the following section, we provide preregistered descriptive analyses of the usefulness of advice as actually implemented by the participants, compared to three hypothetical weighting strategies<sup>4</sup>. The average estimation errors for the judgments made with advice ( $N = 146$ ) are presented in Table 4. The first row of data shows the estimation error for the estimates observed in the game show, with weighting of advice as reported in Table 3, the second row gives the estimation error if no advice were taken, the third if advice were fully followed, and the fourth gives the error of an equal weighting strategy. As observed in the table, there were no or very small differences in estimation error.

To explore why advice did not lead to a higher gain in accuracy, we calculated the number of times the advisors' error had the same sign as the contestants' error. The error was positive for both contestants and advisors 71 times, negative 46 times, and diverged only 30 times. In other words, the estimates of judges and advisors were systematically biased in the same directions, with the bracketing rate as low as 20%, and thus the potential gain in accuracy from incorporating advice was limited (Soll and Larrick, 2009).

As the aim of the contestant is to beat Norway's estimates we also provide information in Table 4 about the number of rounds the contestants would have lost and won given the different weighting schemes. There was no strategy that produced consistently better outcomes. However, a larger number of ties were observed for the actual estimates given in the show and for WOA = 1. The high number of ties is largely due to the influence of hometown estimates. The hometown estimates were equal to Norway's estimates 16 out of 49 times, but also the celebrities' estimates tied with Norway 8 out of 49 times. Family and friends' estimates were identical to Norway's only once.

<sup>4</sup>Note that the preregistration also proposed to include analyses of those who consistently followed advice, did not follow advice, and compromised with advice, but few participants were consistent. We therefore limited analyses to the actual use of advice and the three hypothetical strategies.

How did the different weighting schemes relate to the amount of money in the pot in the final round and the final amount of money won by the contestants? To calculate the final pot as it would have turned out given a specific weighting strategy, we also needed to include the judgments made without advice. Thus, the hypothetical outcomes of the pot and the amount won reported in Table 4 are based on a combination of actual estimates made without advice and hypothetical estimates based on a particular weighting strategy. As observed by the hypothetical amounts won and the 95% confidence intervals (derived from bias-corrected and accelerated bootstrapping), there was little to gain from using advice.

The contestant wins only when providing an estimate that is closer to the actual outcome compared to Norway's estimate. In this zero-one loss situation one strategy could be to provide estimates as close as possible to Norway's estimates to cover a larger interval of potential outcomes. We explored the consequences of a WOA = 0.95 strategy but found that it was not superior to the above (Error,  $M = 0.22$ ,  $SD = 0.18$ ; Amount won in NOK1000,  $M = 20.2$ , 95% CI [13.5, 28.3]). Reasons for this was that (a) only the hometown estimate provided a good proxy for Norway's answer, (b) this proxy was still not a perfect substitute for Norway's estimate, and (c) across the three judgments made with advice, the interval of the response scale (standardized 0–100) that would give a win for the contestant increased by only 1.5 percentage points from WOA = 0 to WOA = 0.95. Thus, the potential for more wins using a strategy that aimed to cover a larger part of the range of potential outcomes was negligible.

## 4. Discussion

We investigated advice taking and solicitation in the high stakes, highly public context of a game show broadcast on Norwegian television. In analyses of potential determinants of advice taking, we found no support for our preregistered hypotheses that advice source and the use of intuition would influence weight of advice, but found that less weight was given to advice after a history of successful unaided judgments. Below, we discuss these results, and then highlight two notable observations: (1) advice was appreciated, as indicated by the high observed weight of advice and its reservation for rounds with higher stakes, and (2) the actual and potential impact of advice on accuracy and outcomes was negligible.

### 4.1. Advice source, own estimation success, and use of intuition

We hypothesized that weight of advice would differ between different advice sources. We had no specified directional hypothesis, as one could make a plausible argument from the existing literature both that statistical advice would receive less weight (Dietvorst et al., 2015) or more weight (Galton, 1907; Logg et al., 2019) than advice from friends and celebrities. Descriptively, advice from friends and family was weighted slightly less on average than advice from celebrities and the hometown, but the difference was small and not statistically significant. This result should be seen in concert with the exploratory analysis of advice solicitation, which found that advice from celebrities was usually used first, friends second, and advice from the hometown was usually saved for the last, highest stake round. This suggests higher appreciation of “statistical” advice or the wisdom of the crowd in this setting, without necessarily translating into higher WOA.

The two regression analyses of the influence of one's own estimation success on subsequent advice taking gave different results. We found no effect in a pooled analysis that exploited both between and within-subject variability, but a within-subjects effect indicated less WOA after previous own success. As the pooled analysis included between-person variability, the discrepancy suggests that contestants who happen to be more successful than others are not less likely to use advice than unsuccessful contestants. Yet, within individuals, the history of success is associated with WOA. These results may warrant a cautious interpretation. The within-subject effects exploit subtle changes in the history of wins and losses. The dependencies in the serial correlation that occur when using the accumulated history of wins can in principle be accommodated by the clustered standard errors. However, the process may depend on mathematical couplings and constraints that we have not fully taken into account, for

example differential weighting of different rounds and increase in granularity due to the accumulation of data over time. In support of the notion that a history of own estimation success can reduce reliance on advice, the results replicated using the number of past unaided wins instead of the proportions.

There was no evidence for any influence of contestants' or advisors' use of intuition on the weight of advice. Answers were coded as based on intuition if the words intuition, gut feeling, or guessing were mentioned. Nevertheless, even in cases when such words were not mentioned, it is likely, given the unusual nature of the tasks, that many of the answers were intuition-based rather than analytical, and that contestants were aware of this. Perhaps the questions about how past performance and the use of intuition influence advice taking can be better answered by research using less obscure tasks (i.e., tasks with less random outcomes).

#### 4.2. *Appreciation of advice*

In a recent meta-analysis of advice taking in the JAS approach, the mean WOA was 0.39 (Bailey et al., 2023), in line with previous conclusions about egocentric discounting (Bonaccio and Dalal, 2006). Participants in the current game show had an average WOA of 0.58, indicating a tendency to put more weight on the advice than one's own estimate. Of course, this is not the only study showing a WOA > 0.5, see for example the forest plot of effect sizes in Bailey et al.'s meta-analysis. Still, the current results provide a datapoint on the higher end of the distribution. Note also that while we found a trimodal distribution of WOA, participants in the game show seemed to choose the advisors' estimate more frequently (22% of the time) than what is often found in laboratory studies (2-13% of the time); and conversely, to stick with their own estimate less often (12% of the time vs. 25-60% of the time in lab studies, see Himmelstein, 2022; Soll and Larrick, 2009).

There are several potential explanations why the game show participants may be more willing to take advice into account. The first explanation is the high stakes. The average at stake in the final round corresponded to about one month's salary, much higher than any other JAS studies we know of. This could have incentivized participants to use advice to a greater degree than in low stakes lab studies. Consistent with the role of stakes, participants generally preferred to save advice for later rounds, i.e., rounds with higher stakes, with each NOK1000 increase of stakes on average increasing the probability of soliciting advice by 2 percentage points. On the other hand, the amount at stake in the final round did not correlate with the WOA. We should also note that even if the stakes involved are the reason for the higher observed of WOA in this context, the trimodal distribution shows that people did not follow the a priori "rational" strategy of averaging. Instead, it was more common than usual for participants to completely adopt their advisors' estimates, perhaps indicating a desire to avoid responsibility for high uncertainty, high stakes judgments (Harvey and Fischer, 1997).

Another potential explanation for the high weight on advice is the highly public setting. In laboratory studies of advice taking, participants' judgments are usually made privately and anonymously. In this case, both successes and failures were immediately obvious for the studio audience as well as for the viewers, with on average 663,000 viewers in the fifth season.<sup>5</sup> This increases the reputational stakes as well as the monetary. Under such close scrutiny, it is reasonable to expect that the pressure to not ignore advice (Harvey and Fischer, 1997) is higher than usual, and participants may fear that fully rejecting advice would make them seem ungrateful or unfriendly. Furthermore, adopting the advisors' estimate allows participants to share responsibility for a potential negative outcome with someone else (El Zein et al., 2019). Note, however, that the same argument about scrutiny has been used toward lab studies (Levitt and List, 2007): participants in an experiment are aware that their behavior is being monitored and might therefore behave differently than they would in private. However, the game show setting arguably provides even higher scrutiny than the lab, and scrutiny is a plausible contributing factor.

Some structural aspects of the game may also be of importance. While advice in JAS studies is often given unsolicited, participants in "All against 1" choose themselves when they want to receive advice.

<sup>5</sup>Source: <https://kampanje.com/medier/2023/02/over-700.000-sa-selma-ibrahims-alle-mot-1-debut/>

This could boost the weight of advice (Bonaccio and Dalal, 2006), as people seem to appreciate and follow solicited advice more than unsolicited, unexpected and imposed advice (Landis et al., 2022; Rebholz and Hütter, 2022; Van Swol et al., 2017). However, exploratory analyses did not show any difference in WOA for early rounds, where advice was freely chosen vs. later rounds, where participants were “forced” to use advice if they had not used it before. An additional potentially important aspect of the game is that unlike many laboratory studies, contestants are allowed to interact with the advisors in the studio. The advisors are usually asked to explain how they landed on a particular estimate, and such justified advice may seem more convincing than bare, unexplained estimates from an anonymous advisor, in line with Yaniv and Kleinberger’s (2000) idea that egocentric discounting occurs due to differential access to the reasoning behind one’s own and the advisor’s estimate. Note however that Soll and Mannes (2011) did not find support for this explanation, so the role of personal interactions in increasing weight of advice should be taken as speculation.

A final explanation for the higher weight of advice is the type of tasks used. Many advice tasks in the lab focus on general knowledge questions, where the answer is in principle knowable in advance. In contrast, the tasks in this game show are deliberately made to be spectacular, entertaining, and unusual, and in many cases the answer is not known in advance, since some tasks are performed live. It is hard for anyone to claim expertise in estimating for instance how many balloons a professional radio-controlled (RC) car driver can pop with his RC car in 90 seconds when driving on a court made to resemble the Pac-Man game. In other words, these are difficult questions involving a large degree of uncertainty, which is known to drive advice seeking (Bonaccio and Dalal, 2006), and perhaps also more external or aleatory uncertainty in addition to internal or epistemic uncertainty (Løhre and Teigen, 2016; Walters et al., 2023). One study varying task difficulty found egocentric discounting for easy tasks, WOA of 0.41 and 0.39 in two studies, but not for difficult tasks, WOA of 0.52 and 0.54 (Gino and Moore, 2007; see also Schrah et al., 2006). Thus, the current results are also consistent with task difficulty playing a role in determining the weight of advice.

We cannot conclude from our analyses which (if any) of these explanations matter the most, if they all contribute or perhaps interact in some way. This is a natural consequence of studying the topic of advice taking in this unique setting, as there is no strict experimental control over possibly confounding factors. However, we agree with List (2020) that all settings are unique in some way, and that interesting data should not be dismissed simply due to inherent limitations of the settings. We doubt that future laboratory studies will be able to use incentives at the same level as those in this game show. Nevertheless, it would be a worthy pursuit to attempt to disentangle the different factors involved, for instance by varying the type of tasks, the incentives, the degree of public scrutiny, and/or the interaction with the advisors. However, while task difficulty or public scrutiny might explain the higher observed WOA, it is harder to explain the fact that advice was generally saved for the last, high-stakes rounds, with reference to either of these factors, which are constant throughout the game. Our findings suggest that the phenomenon of egocentric discounting may be more likely in typical laboratory settings, and is less likely to be observed when the stakes and the scrutiny is higher.

### **4.3. On the usefulness of advice and the optimal strategy**

Previous studies have concluded that people do not sufficiently heed advice. Here, we found that people gave more weight to advice than to their own estimate. However, the benefits of this increased weight of advice were minimal, if any. Regardless of whether participants had followed a strategy of always staying with their own estimate, always adopting the advisors’ estimate, or always combining them, the error and the outcomes would largely have been the same. These results illustrate how the advantage of “rational” combination strategies may be negligible when fewer (yet more than a hundred) judgments are made under high uncertainty. Note that there was an advantage for aggregate estimates when comparing the contestants’ and Norway’s estimates across all 382 judgments, with the viewers on average 2 percentage points closer to the actual value and winning a few more rounds than the contestant (24 ties, 131 won by participant, 147 won by Norway, see Table 1). However, the advantage is relatively

small, is based on judgments from a large number of independent judges, and is only apparent on the aggregate. This could indicate that the repeated measurements in laboratory studies make differences in estimation error appear more salient and more important than they are in many applied contexts.

Although the game show setting is artificial, it is not difficult to imagine real life contexts with many similarities. For instance, leaders make a number of judgments and typically solicit advice for high-stakes decisions with considerable uncertainty (Ma et al., 2020). In these contexts, a rational estimation strategy could rightfully be subordinate to social and psychological considerations like impression management, allocation of responsibility for the decision, or feeling of autonomy. Also note that while a by-the-book averaging strategy may have an advantage on the aggregate level, for a single individual the difference between using advice or not is highly unpredictable. In the game show, contestants seem to put a lot of effort into the weighting decision, and could display substantial regret after using or not using advice. Our results indicate that the expected values of their decision options were practically equal.

It is not obvious what represents the normative or rational approach in this setting. Averaging has been discussed as one a priori logical strategy. However, as discussed by Mannes (2009), averaging between your own estimate and an estimate from several other advisors is not an ego-neutral strategy. In fact, an ego-neutral strategy for the hometown advice would lead to a WOA very close to 1. From this viewpoint, the current results imply a severe underweighting of the information contained in the hometown advice, and perhaps also of the advice from friends and family and celebrities.

On the other hand, the game show has some similarities to so-called prediction contests (Pfeifer et al., 2014). In a prediction contest, people attempt to estimate an unknown quantity, and the one who comes closest to the correct answer wins. Theoretical and empirical work (Lichtendahl et al., 2013; Pfeifer, 2016) show that for such contests, participants have an incentive to exaggerate their own private information. In other words, your chances of being the single person who wins the contest increases if you emphasize things that only you know or believe, at the expense of information that is commonly known. In the current context this would imply a WOA close to 0, at least for the hometown advice, but unlike in prediction contests participants in the game show do not have to beat everyone else to win, but rather the average of everyone else. Assuming that the advisor estimates are closer to Norway's estimate than the contestant's estimate because they are based on more than one judge, the most advantageous strategy may be to weight advice strongly, without fully adopting it. Adjusting one's estimate toward the advisors' estimate will increase the winning part of the interval up to the point where it reaches Norway's estimate, where the expectancy becomes 50%, either by tie or by randomly being on one or the other side of the crowd estimate. Indeed, we found that a WOA = 0.95 strategy increased the part of the winning interval for the contestant. However, the practical consequence of covering more of the scale by strategies WOA = 0.95 or WOA = 0.5 were negligible.

Specific features of the game show may have diluted potential advantages of averaging or strongly weighting advice. For example, the either/or (win or lose) components in each round introduce substantial randomness, akin to the dichotomization of continuous variables (MacCallum et al., 2002). Although the "wisdom of the crowd" estimates from the thousands of viewers using the designated app were slightly more accurate, they only won somewhat more rounds than the contestants; and the final round, where contestants could either win the pot or get nothing, introduced even more randomness. Thus, any advantage of being skilled or following a good weighting strategy would be attenuated by the nature of the game show. Similarly, win or lose components in real life, for example, winning or losing the bid on a contract, may reduce the benefits of advice in comparison to continuous outcomes.

## 5. Conclusion

While it is important to consider the unique context of this study, we believe the findings raise important questions about the traditional, low stakes, lab-based JAS approach. It might be that warnings about discounting of advice are overstated, and that people listen more to advice when stakes are higher,



when their judgments are under public scrutiny, and when the tasks are more difficult or involve more uncertainty. Similarly, the purported benefits of consistent averaging strategies were negligible in this setting. As it can be difficult to know when an averaging strategy would be more beneficial than other strategies (e.g., choosing), it might be reasonable for a decision maker to give greater consideration to other factors, even those that do not improve the judgment per se, such as reputational or interpersonal concerns. A provocative conclusion from these findings is that in some high stakes, highly public contexts, advice is more used but less useful than one would come to believe from laboratory studies.

**Supplementary material.** The supplementary material for this article can be found at <https://doi.org/10.1017/jdm.2024.4>.

**Data availability statement.** The original data cannot be shared due to privacy requirements. A datafile where the order of observations has been shuffled independently for each variable (which provides observed distributions for each individual variable while ensuring anonymity) and the data analysis syntax is available on <https://osf.io/rkpy3/>.

**Acknowledgements.** We would like to thank the Norwegian broadcasting company, NRK, for providing the subtitles and Ingebjørg Flaata Bjaaland for coding the estimates and their characteristics. Attendees at the 2023 SPUDM conference in Vienna provided valuable feedback.

**Author contribution.** E.L. came up with the initial idea and developed it in collaboration with T.H. E.L. wrote the first draft of the introduction and discussion sections, contributed to analysis and visualization, and reviewed and edited the manuscript. T.H. developed the idea in collaboration with E.L., had the main responsibility for data analysis, visualization and writing of the methods and results section, and reviewed and edited the manuscript.

**Funding statement.** This research received no specific grant funding from any funding agency, commercial or not-for-profit sectors.

**Competing interest.** The authors declare none.

## References

- Alexiev, A., Volberda, H., Jansen, J., & Van, F. (2020). Contextualizing senior executive advice seeking: The role of decision process comprehensiveness and empowerment climate. *Organization Studies*, 41(4), 471–497. <https://doi.org/10.1177/0170840619830128>
- Andersen, S., Ertaċ, S., Gneezy, U., Hoffman, M., & List, J. A. (2011). Stakes matter in ultimatum games. *American Economic Review*, 101(7), 3427–3439. <https://doi.org/10.1257/aer.101.7.3427>
- Bailey, P. E., Leon, T., Ebner, N. C., Moustafa, A. A., & Weidemann, G. (2023). A meta-analysis of the weight of advice in decision-making. *Current Psychology*. <https://doi.org/10.1007/s12144-022-03573-2>
- Blunden, H., Logg, J. M., Brooks, A. W., John, L. K., & Gino, F. (2019). Seeker beware: The interpersonal costs of ignoring advice. *Organizational Behavior and Human Decision Processes*, 150, 83–100. <https://doi.org/10.1016/j.obhdp.2018.12.002>
- Bonaccio, S., & Dalal, R. S. (2006). Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational Behavior and Human Decision Processes*, 101(2), 127–151. <https://doi.org/10.1016/j.obhdp.2006.07.001>
- Burton, J. W., Stein, M., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making*, 33(2), 220–239. <https://doi.org/10.1002/bdm.2155>
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 114–126. <https://doi.org/10.1037/xge0000033>
- El Zein, M., Bahrami, B., & Hertwig, R. (2019). Shared responsibility in collective decisions. *Nature Human Behaviour*, 3(6), 554–559. <https://doi.org/10.1038/s41562-019-0596-4>
- Enke, B., Gneezy, U., Hall, B., Martin, D., Nelidov, V., Offerman, T., & Van De Ven, J. (2023). Cognitive biases: Mistakes or missing stakes? *Review of Economics and Statistics*, 105(4), 818–832. [https://doi.org/10.1162/rest\\_a\\_01093](https://doi.org/10.1162/rest_a_01093)
- Feng, B., & MacGeorge, E. L. (2010). The influences of message and source factors on advice outcomes. *Communication Research*, 37(4), 553–575.
- Galton, F. (1907). Vox populi. *Nature*, 75, 450–451. <https://doi.org/10.1038/075450a0>
- Gino, F., & Moore, D. A. (2007). Effects of task difficulty on use of advice. *Journal of Behavioral Decision Making*, 20(1), 21–35. <https://doi.org/10.1002/bdm.539>
- Harvey, N., & Fischer, I. (1997). Taking advice: Accepting help, improving judgment, and sharing responsibility. *Organizational Behavior and Human Decision Processes*, 70(2), 117–133. <https://doi.org/10.1006/obhd.1997.2697>
- Himmelstein, M. (2022). Decline, adopt or compromise? A dual hurdle model for advice utilization. *Journal of Mathematical Psychology*, 110, 102695. <https://doi.org/10.1016/j.jmp.2022.102695>

- Hütter, M., & Ache, F. (2016). Seeking advice: A sampling approach to advice taking. *Judgment and Decision Making*, 11(4), 401–415. <https://doi.org/10.1017/S193029750000382X>
- Jetter, M., & Walker, J. K. (2017). Anchoring in financial decision-making: Evidence from Jeopardy! *Journal of Economic Behavior & Organization*, 141, 164–176. <https://doi.org/10.1016/j.jebo.2017.07.006>
- Kämmer, J. E., Choshen-Hillel, S., Müller-Trede, J., Black, S. L., & Weibler, J. (2023). A systematic review of empirical studies on advice-based decisions in behavioral and organizational research. *Decision*, 10(2), 107–137. <https://doi.org/10.1037/dec0000199>
- Landis, B., Fisher, C. M., & Menges, J. I. (2022). How employees react to unsolicited and solicited advice in the workplace: Implications for using advice, learning, and performance. *Journal of Applied Psychology*, 107(3), 408–424. <https://doi.org/10.1037/apl0000876>
- Larrick, R. P., Mannes, A. E., & Soll, J. B. (2012). The social psychology of the wisdom of crowds. In J. I. Krueger (Ed.), *Social judgment and decision making* (pp. 227–242). Psychology Press.
- Larrick, R. P., & Soll, J. B. (2006). Intuitions about combining opinions: Misappreciation of the averaging principle. *Management Science*, 52(1), 111–127.
- Levitt, S. D., & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives*, 21(2), 153–174. <https://doi.org/10.1257/jep.21.2.153>
- Lichtendahl, K. C., Grushka-Cockayne, Y., & Pfeifer, P. E. (2013). The wisdom of competitive crowds. *Operations Research*, 61(6), 1383–1398. <https://doi.org/10.1287/opre.2013.1213>
- List, J. A. (2020). *Non est disputandum de generalizability? A glimpse into the external validity trial* (w27535; NBER Working Paper Series, p. w27535). National Bureau of Economic Research. <https://doi.org/10.3386/w27535>
- Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 151, 90–103. <https://doi.org/10.1016/j.obhdp.2018.12.005>
- Løhre, E., & Teigen, K. H. (2016). There is a 60% probability, but I am 70% certain: Communicative consequences of external and internal expressions of uncertainty. *Thinking & Reasoning*, 22(4), 369–396. <https://doi.org/10.1080/13546783.2015.1069758>
- Lorenz, J., Rauhut, H., Schweitzer, F., & Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences*, 108(22), 9020–9025. <https://doi.org/10.1073/pnas.1008636108>
- Ma, S., Kor, Y. Y., & Seidl, D. (2020). CEO advice seeking: An integrative framework and future research agenda. *Journal of Management*, 46(6), 771–805. <https://doi.org/10.1177/0149206319885430>
- MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, 7(1), 19–40. <https://doi.org/10.1037/1082-989X.7.1.19>
- MacGeorge, E. L., & Van Swol, L. M. (2018). Advice across disciplines and contexts. In E. L. MacGeorge, & L. M. Van Swol (Eds.), *The Oxford handbook of advice* (1st ed., pp. 3–18). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190630188.013.1>
- Mannes, A. E. (2009). Are we wise about the wisdom of crowds? The use of group judgments in belief revision. *Management Science*, 55(8), 1267–1279. <https://doi.org/10.1287/mnsc.1090.1031>
- Oktar, K., & Lombrozo, T. (2022). Deciding to be authentic: Intuition is favored over deliberation when authenticity matters. *Cognition*, 223, 105021. <https://doi.org/10.1016/j.cognition.2022.105021>
- Pálfı, B., Arora, K., & Kostopoulou, O. (2022). Algorithm-based advice taking and clinical judgement: Impact of advice distance and algorithm information. *Cognitive Research: Principles and Implications*, 7(1), 70. <https://doi.org/10.1186/s41235-022-00421-6>
- Pfeifer, P. E. (2016). The promise of pick-the-winners contests for producing crowd probability forecasts. *Theory and Decision*, 81(2), 255–278. <https://doi.org/10.1007/s11238-015-9533-9>
- Pfeifer, P. E., Grushka-Cockayne, Y., & Lichtendahl, K. C. (2014). The promise of prediction contests. *The American Statistician*, 68(4), 264–270. <https://doi.org/10.1080/00031305.2014.937545>
- Post, T., Van Den Assem, M. J., Baltussen, G., & Thaler, R. H. (2008). Deal or no deal? Decision making under risk in a large-payoff game show. *American Economic Review*, 98(1), 38–71. <https://doi.org/10.1257/aer.98.1.38>
- Rebholz, T. R., & Hütter, M. (2022). The advice less taken: The consequences of receiving unexpected advice. *Judgment and Decision Making*, 17(4), 816–848.
- Schrah, G. E., Dalal, R. S., & Sniezek, J. A. (2006). No decision-maker is an Island: Integrating expert advice with information acquisition. *Journal of Behavioral Decision Making*, 19(1), 43–60. <https://doi.org/10.1002/bdm.514>
- Sniezek, J. A., & Buckley, T. (1995). Cueing and cognitive conflict in judge-advisor decision making. *Organizational Behavior and Human Decision Processes*, 62(2), 159–174.
- Soll, J. B., & Larrick, R. P. (2009). Strategies for revising judgment: How (and how well) people use others' opinions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(3), 780–805. <https://doi.org/10.1037/a0015145>
- Soll, J. B., & Mannes, A. E. (2011). Judgmental aggregation strategies depend on whether the self is involved. *International Journal of Forecasting*, 27(1), 81–102. <https://doi.org/10.1016/j.ijforecast.2010.05.003>
- Stroop, J. R. (1932). Is the judgment of the group better than that of the average member of the group? *Journal of Experimental Psychology*, 15(5), 550–562. <https://doi.org/10.1037/h0070482>
- Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations*. Doubleday & Co.

- Teeselink, B. K., Van Dolder, D., Van Den Assem, M. J., & Dana, J. (2022). High-stakes failures of backward induction: Evidence from “The Price Is Right.” *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4130176>
- Thaler, R. H. (2016). Behavioral economics: Past, present, and future. *American Economic Review*, 106(7), 1577–1600. <https://doi.org/10.1257/aer.106.7.1577>
- Tzioti, S. C., Wierenga, B., & Van Osselaer, S. M. J. (2014). The effect of intuitive advice justification on advice taking. *Journal of Behavioral Decision Making*, 27(1), 66–77. <https://doi.org/10.1002/bdm.1790>
- Van Den Assem, M. J., Van Dolder, D., & Thaler, R. H. (2012). Split or steal? Cooperative behavior when the stakes are large. *Management Science*, 58(1), 2–20. <https://doi.org/10.1287/mnsc.1110.1413>
- Van Swol, L. M., MacGeorge, E. L., & Prahla, A. (2017). Advise with permission?: The effects of advice solicitation on advice outcomes. *Communication Studies*, 68(4), 476–492. <https://doi.org/10.1080/10510974.2017.1363795>
- Vestal, A., & Guidice, R. (2019). The determinants and performance consequences of CEO strategic advice seeking. *Journal of General Management*, 44(4), 232–242. <https://doi.org/10.1177/0306307019833491>
- Walters, D. J., Ülkümen, G., Tannenbaum, D., Erner, C., & Fox, C. R. (2023). Investor behavior under epistemic vs. aleatory uncertainty. *Management Science*, 69(5), 2761–2777. <https://doi.org/10.1287/mnsc.2022.4489>
- Yaniv, I., & Kleinberger, E. (2000). Advice taking in decision making: Ego-centric discounting and reputation formation. *Organizational Behavior and Human Decision Processes*, 83(2), 260–281. <https://doi.org/10.1006/obhd.2000.2909>