Judgment and Decision Making (2025), **20**:e8 1–28
doi:10.1017/jdm.2024.43

EMPIRICAL ARTICLE

The development of a task to study advice taking across nations and its application in a China-Germany comparison

Thomas Schultze [1,2] and Zhijun Chen[3]

[1]Insitute of Psychology, Otto-Friedrich-University Bamberg, Bamberg, Germany; [2]School of Psychology, Queen's University Belfast, Belfast, UK and [3] School of Management, Fudan University, Shanghai, China

Corresponding author: Thomas Schultze and Zhijun Chen; Emails: thomas.schultze-gerlach@uni-bamberg.de, zhijunchen@fudan.edu.cn

Received: 9 October 2023; **Revised:** 6 December 2024; **Accepted:** 6 December 2024

Keywords: advice taking; cultural differences; quantitative judgment

Abstract

Advice taking is a crucial part of decision-making and has attracted the interest of scholars across the world. Laboratory research on advice taking has revealed several robust phenomena, such as sensitivity to advice quality or a tendency to underutilize advice. Despite extensive investigations in different countries, cultural differences in advice taking remain understudied. Knowing whether such cultural differences exist would not only be interesting from an academic standpoint but might also have consequences for multinational organizations and businesses. Here, we argue that prior laboratory research on cultural differences in advice taking is hindered by confounding factors, particularly the confound between participants' cultural background and task difficulty. To draw a valid conclusion about cultural differences in advice taking, it is vital to develop a decision task devoid of this confound. Here, we develop such a judgment task and demonstrate that the core phenomena of advice taking manifest in a sample of German participants. We then use this task in a cross-national comparison of German and Chinese participants. While the core phenomena of advice taking consistently manifested in both samples, some differences emerged. Most notably, Chinese participants were more receptive of advice, even though they still underutilized it. This greater reliance on advice was driven by Chinese participants' greater preference for averaging their own and the advisor's judgments. We discuss how our findings extend current understanding of the nuanced interplay between cultural values and the dynamics of advice taking.

Irrespective of which culture we grew up in, advice is an integral part of human communication. One question that has intrigued scholars around the globe is to what extent people are willing to take advice. The predominant paradigm used for experimental studies of advice taking is the Judge–Advisor System (JAS, Sniezek and Buckley, 1995), in which one person (the judge) first makes an initial judgment, then receives advice in the form of another person's judgment and, subsequently, makes a final, possibly revised judgment. The judgment tasks employed in the JAS are usually quantitative estimation tasks such as estimating the caloric content of food items, the dates of historical events, or the distances between cities. Using quantitative estimation tasks is convenient because it allows measuring advice taking as a continuous variable, namely as the percent weight of advice when making the final judgment (Harvey and Fischer, 1997).

While most of the research using the JAS stems from Western countries, it is not limited to them. The JAS has been used in Australia (Bailey et al., 2021), Chile (Kausel et al., 2015), China (Tinghu et al., 2018), France (Mercier et al., 2012), Germany (Schultze et al., 2015), India (Tzini and Jain, 2018), Israel

(Yaniv and Kleinberger, 2000), Japan (Mercier et al., 2012), the Netherlands (Tzioti et al., 2014), South Korea (Kim et al., 2020), Türkiye (Önkal et al., 2009), the UK (Harvey and Fischer, 1997), and the US (Van Swol, 2009). The question that arises almost naturally, then, is whether—and if so, how—advice taking differs between countries. This question is relevant because national differences can be viewed as a proxy for cultural differences. Put simply, if there are cultural differences in advice taking, then comparing advice taking between culturally heterogeneous countries should yield differences in advice taking (we will refer to this as the countries-as-proxies approach to studying cultural differences).

The first question is whether we should expect cultural differences in advice taking to exist. In the terms of Brunswik, human behavior is shaped by an ecology, which is defined as the 'natural-cultural habitat of an individual or group' (Brunswik, 1955, p. 198). The mere fact that this definition of ecology entails the cultural component does not necessarily mean that there must be cultural differences in advice taking. However, from a theoretical perspective, it is plausible to assume that they exist. A good starting point for such theoretical considerations is Hofstede's framework which differentiates cultures along six dimensions (Hofstede et al., 2010). One of these dimensions is particularly likely to be relevant for advice taking: individualism versus collectivism. Individualism versus collectivism, also referred to as independent versus interdependent self-construal (Markus and Kitayama, 1991), is by far the most studied of Hofstede's cultural dimensions (Gelfand et al., 2007). Broadly speaking, individualism describes the extent to which a society values the needs of the individual over those of the group. People from individualistic societies tend to value autonomy more than those from collectivist societies, whereas people from collectivistic societies place greater emphasis on harmony with others (Markus and Kitayama, 1991). Another one of Hofstede's cultural dimensions, power distance, could also be relevant for advice taking. Power distance describes the extent to which a society accepts inequalities in power (Hofstede et al., 2010). As such, high power distance also entails the less powerful accepting their lower status and complying more strongly with social norms centered on respect for and compliance with high-power individuals (Fikret Pasa, 2000). In the context of advice taking, power distance would likely influence how people seek advice or to what extent they follow it when there are status differences between judge and advisor (while we will focus exclusively on individualism versus collectivism in our own study—which is common in cross-cultural research on advice—we will return to power distance in the general discussion).

Prior research using the countries-as-proxies approach has already provided some evidence of cultural differences in advice seeking and advice evaluation. People from more collectivistic countries report perceiving advice as more supportive and less intrusive than people from individualistic countries (Hosni, 2020; Tavakoli and Tavakoli, 2010), and this finding is complemented by people from more collectivistic countries offering advice more frequently (Chentsova-Dutton and Vaughn, 2012). In addition, in more collectivistic countries, people seem to be more likely to seek advice as a means to establish good interpersonal relations whereas people from individualistic countries seek it more for its informational value (Ji et al., 2017). Finally, a study on advice giving provides some initial evidence that individualism versus collectivism interacts with relative advisor status (Hosni, 2020). Participants from a more collectivistic country reported giving advice more indirectly when there were status differences between advisor and advisee, perhaps to avoid face loss and maintain good social relations with their advisees. In contrast, participants from a more individualistic country were more likely to report giving direct advice when there were status differences. Note that while the authors interpreted these results in terms of individualism versus collectivism, they could also be plausibly explained via differences in power distance, which tends to be highly correlated with individualism versus collectivism (Schermerhorn and Harris Bond, 1997).

Given that there seem to be cultural differences in advice seeking and advice giving between people from countries that are more collectivistic and those that are more individualistic, it stands to reason that such differences also exist in advice taking, which mainly occurs during interpersonal interactions. Specifically, people from more collectivistic cultures might be more responsive to advice. They should be less likely to perceive it as a threat to their autonomy—especially if the advice is unsolicited—but they might also feel greater normative pressure to heed advice in order to maintain a positive

https://doi.org/10.1017/jdm.2024.43 Published online by Cambridge University Press

relationship with the advisor. Such potential differences in advice taking would not only be interesting from a theoretical perspective but would also have practical implications, for example, in the context of globally acting organizations.

Despite the potential relevance of cultural differences in advice taking, our understanding of those differences is still limited. In fact, we are aware of only a single published study investigating cross-cultural differences in advice taking (Mercier et al., 2012). One possible reason for the lack of research is that studying cross-cultural differences in advice taking is not trivial. Returning to the idea that people's advice taking behavior is shaped by their ecology, a complete understanding of how people use advice would require a representative design approach (Brunswik, 1955; Dhami et al., 2004). What this means in the context of advice taking is that researchers need to draw a representative set of tasks for which people usually seek or receive (unsolicited) advice. Since different cultures create different ecologies, a cross-cultural comparison using a representative design approach is difficult, if not impossible. A design that would be representative for one culture might not be for another. Cultures likely differ with regard to when it is appropriate to ask others for advice, what topics can be discussed with an advisor, and how frequently others provide unsolicited advice. Accordingly, the result of a cross-cultural comparison would be difficult to interpret. Emerging differences in advice taking could reflect true cultural differences, but they might also stem from the selected tasks not being equally representative of the different cultures.

A viable—if perhaps not ideal—alternative approach is to use a controlled design, in which people from 2 or more cultures are given the same task. The assumption here is that if cultural differences in advice taking exist, they should manifest in a controlled setting. This is exactly the approach the above-mentioned study on cultural differences in advice taking used (Mercier et al., 2012). In this study, Mercier et al. used the countries-as-proxy approach, comparing advice taking between student participants from France (a more individualistic country) and Japan (a more collectivistic country). They hypothesized that Japanese students would heed advice more due to a stronger preference for compromising in more collectivist cultures. In the context of the JAS, this means that the authors expected Japanese participants to be more likely than their French peers to adopt a strategy closer to one of equal weighting. Mercier et al. indeed found that Japanese students were more responsive to the advice, but not because of a higher propensity to meet halfway between judges' initial estimates and the advice. Instead, Japanese participants were more likely to adopt the advice completely, thus completely disregarding their own initial estimates.

While the general increase in advice taking reported by Mercier et al., (2012) is consistent with the notion that people from more collectivistic cultures should be more open to the opinions of others, the greater observed likelihood of fully adopting peer advice is more difficult to explain in terms of collectivism. We argue that the study by Mercier et al. suffers from a methodological problem that, on the one hand, makes its results difficult to interpret in terms of cultural differences and, on the other hand, provides a simple alternative explanation for the study's results. For a cross-cultural comparison using a controlled approach with a single task to work, this task needs to be free of confounds, and we believe that the estimation task Mercier et al. used was not. Specifically, it contains a potential confound between participants' nationality and task difficulty. Put simply, this confound threatens the internal validity of the study's results. An inspection of the stimuli suggests that at least some of the estimation problems were more difficult for the Japanese participants because they are more deeply embedded in Western cultural knowledge (e.g., 'In what year was the Empire State Building finished' or 'In what year did Marilyn Monroe die?'). Further supporting this notion, Japanese participants' first estimates were substantially less accurate than those of their French counterparts. Since people use advice more when the task is difficult (Ache et al., 2020; Bogert et al., 2021), it is conceivable that Japanese participants' greater weight of advice and, in particular, their greater propensity to completely adopt the advice reflects systematic differences in task difficulty rather than cultural differences. This is not to say that cultural differences did not play a role in the Mercier et al. study, but if so, the confound with task difficulty makes it impossible to isolate their effects.

https://doi.org/10.1017/jdm.2024.43 Published online by Cambridge University Press

Another approach to uncover potential cultural differences in advice taking is via meta-analyses. However, conclusions drawn from meta-analyses do not solve the problem either. For example, a recent meta-analysis on advice weighting did not find the extent of individualism (versus collectivism) of the country where advice taking studies were run to be significantly related to advice taking (Bailey et al., 2023). Since the individual studies were not designed to be informative in cross-national comparisons, they likely differed in terms of how difficult or engaging participants found the respective judgment task and how judgmental accuracy was incentivized. While those differences are unlikely to result in systematic confounds, they are bound to create unsystematic variation (i.e., noise) that may mask existing cultural effects. In other words, we cannot interpret the absence of evidence of cross-national (as a proxy for cross-cultural) effects on advice weighting as evidence of the absence of such effects. In addition, an informative meta-analytic test of cultural differences in advice-taking might require a psychometric meta-analysis rather than the bare-bones meta-analysis reported by Bailey et al. (2023). The crucial difference is that psychometric meta-analysis considers—and corrects for—measurement error in between-study comparisons (Schmidt and Hunter, 2015). In the context of advice taking this would result in more accurate estimates of moderating variables such as participants' culture.

Based on the current state of knowledge, we argue that the first step to studying cultural differences in advice taking using the countries-as-proxies approach must be the development of judgment tasks that do not confound participants' nationality with task difficulty. Once such tasks have been developed, we can design meaningful studies to investigate cross-national differences in advice taking. In the present research, we develop such a task. We then employ it in an exploratory countries-as-proxies comparison of advice taking between more collectivistic (Chinese) and more individualistic (German) participants.

1. Developing an advice taking task to study advice taking across nations

As already mentioned above, the countries-as-proxies approach to studying cultural differences rests on the assumption that differences in behavior between countries are indicative of cross-cultural differences. In order for this assumption to hold, the tasks we give participants in countries-as-proxies studies must be free of confounds or, in terms of behavioral ecologies, equally (un-)representative for all countries. In other words, we would like all behavioral differences to reflect cultural differences—even if these cultural differences may not be completely covered by the cultural dimensions we considered in a study. As illustrated in our description of the, so far, only published study of cultural differences in advice taking (Mercier et al., 2012), an obvious confounding factor likely to impact advice taking is the difficulty of the judgment task because people tend to heed advice more when they perceive the task as difficult (Ache et al., 2020; Bogert et al., 2021).

A brief look at the literature on advice taking shows that most studies using the JAS rely on judgment tasks that are based on general knowledge and are, thus, likely to be more difficult for judges from some countries than for others. Examples include estimating the distances between European cities (Schultze et al., 2015, 2018), predicting average salaries of graduates from US business schools (Soll and Larrick, 2009), guessing the date of events from Middle Eastern history (Yaniv and Kleinberger, 2000; Yaniv and Milyavsky, 2007), predicting the outcomes of US college sports games (Soll and Mannes, 2011), or estimating the caloric content of food items typically consumed in Western countries (Schultze et al., 2015; Yaniv and Choshen-Hillel, 2012). While these tasks have several benefits that make them attractive for use in JAS studies (e.g., the true values are known, participants do not require training or specialized knowledge), the general knowledge they draw from is heavily contingent on the country or region of the world, in which participants grew up, were socialized, and went to school. This makes general knowledge tasks a poor choice for countries-as-proxies studies of cultural differences in advice taking.

A smaller subset of JAS studies used tasks that seem better suited for cross-national studies of advice taking because they do not rely on regionally or nationally specific general knowledge but instead have participants make predictions based on some provided information. Examples of those multi-cue judgment tasks include financial forecasts (Önkal et al., 2009) or estimating the severity of disease

https://doi.org/10.1017/jdm.2024.43 Published online by Cambridge University Press

outbreaks (Harvey and Fischer, 1997). One downside to those tasks is that they usually require some training because participants need to learn the cue-target relations. That means, the study needs to include practice trials, in which participants receive immediate feedback on their performance, which can be cumbersome for participants and necessitates the study to run on a computer. Ideally, we would like to capitalize on the multi-cue judgment tasks' relative independence of prior knowledge without the need to train participants on the task. This is exactly what we aimed for here. We decided on a simple multi-cue judgment task, in which participants' aim is to estimate a true value as accurately as possible based on several unbiased but imperfect measurements. These measurements are drawn independently from normal distributions around the true values, satisfying the condition that the cues are unbiased and equally informative.

This type of task has several desirable characteristics. First, the only knowledge requirement is that participants have a secondary school education, as the necessary math to understand the principles of imperfect measurements (i.e., measurement error) and how to deal with them (i.e., by taking the mean of several measurements) is usually covered in secondary school physics classes. This makes the task largely independent of participants' nationality. We use the term largely because there may be populations which do not have access to secondary school education, and for which this type of task may be relatively more challenging. Second, this task allows researchers to specify the exact information basis of judge and advisor. For example, by assigning judge and advisor the same number of independent measurements, they can create a situation, in which the optimal weight of advice is 50%. By assigning the advisor more cues than the judge, one can elevate the advisor to the status of an expert, and so on. Third, it is possible, though not necessary, to inform the judge how many cues the advisor had on any given trial, thus allowing for perfect transparency of the relative quality of the advice. Fourth, the task does not allow for participants to look up the true values, which may be particularly tempting when a study includes monetary incentives for the accuracy of participants' judgments. This feature is particularly relevant for online studies, which provide participants with more opportunities to cheat by researching true values on the internet.

One important aspect to consider when studying cultural differences in advice taking is the scope of the comparison. Previous research focused on only one or two aspects of advice taking such as the mean weight of advice or the frequency of averaging (Bailey et al., 2023; Schermerhorn and Harris Bond, 1997). This approach is efficient but risks overlooking potentially interesting differences. Here, we propose a different way to study cultural differences in advice-taking more holistically. Our approach draws conceptually from Brunswik's idea of a representative design we mentioned above (Brunswik, 1955). The general principle can be described as follows: First, we identify core phenomena of advice taking in a culture that we treat as a frame of reference. We consider core phenomena to be patterns that arise frequently and reliably in different studies on advice taking. This idea of core phenomena that manifest across studies is similar to Cont's concept of 'stylized facts' (Cont, 2001). In other words, we treat the existing research as a facsimile of a representative design. Here, we exploit the fact that published studies on advice taking differ—sometimes arbitrarily—regarding the research question, judgment tasks used, participant demographics, advisor characteristics, manipulated variables, and context variables such as the presence of incentives. With a sufficiently large set of studies that were run in culturally similar countries, we should then be able to look for robust behavioral patterns that emerge across these studies. The resulting core phenomena can then be considered the typical advice taking behavior in one culture that can be compared to the behavior participants from different cultural backgrounds display in the same controlled situation.

In our approach, it is a necessary feature of any task we want to use for countries-as-proxies studies of cultural differences in advice taking that the core phenomena of advice taking manifest when people from the reference culture work on it as judges. Since most JAS studies have been conducted in Western countries (see above), these core phenomena are patterns that we can (at least for now) confidently expect to find in Western samples. Our list of core phenomena included: (a) egocentric discounting, that is, underweighting of useful advice (Harvey and Fischer, 1997; Soll and Larrick, 2009; Yaniv and Kleinberger, 2000); (b) sensitivity to advice quality (Harvey and Fischer, 1997; Lim

https://doi.org/10.1017/jdm.2024.43 Published online by Cambridge University Press

and O'Connor, 1995; Schultze et al., 2017); (c) inability to ignore useless advice (Fiedler et al., 2019; Schultze et al., 2017); (d) a W-shaped trimodal distribution of the weight of advice with modes at 0%, 50%, and 100% weight of advice (Soll and Larrick, 2009; Soll and Mannes, 2011); and (e) a curvilinear relationship between advice weighting and advice distance with low weight of advice both when advice is very near and far away from the initial judgment (Ecken and Pibernik, 2016; Moussaïd et al., 2013; Schultze et al., 2015). This list is not meant to be exhaustive, but we believe that it serves as a good starting point. Note that there are isolated studies showing egocentric advice discounting and a curvilinear relation of advice distance and advice taking in non-Western countries (Du et al., 2019; Wang and Du, 2018). The remaining core phenomena have not yet been investigated in non-Western studies because these studies focused on different research questions.

To summarize, the logic of our research is as follows: First, a newly developed task should ensure that the core phenomena of advice taking typically observed in Western studies also manifest when Western participants work in a JAS using that new task. Once that condition is satisfied, for example, because we can show the phenomena in a German student sample, we can then test whether the effects also occur in non-Western samples, such as a Chinese student sample. If they do, we can further test whether the magnitudes of the effects vary by country. In the following section, we will describe the specific task we developed based on the reasoning above.

2. The rainfall estimation task

In the rainfall estimation task, participants take the role of a meteorologist whose job is to estimate the annual precipitation in different cities or regions as accurately as possible based on several independent measurements. Participants are told that they have a certain number of measurement stations in each region. While the number of available measurements can vary, all measurements are based on the same technology, that is, they measure the true values with equal precision. Participants are informed that the measurement errors are unsystematic and normally distributed, meaning that it is equally likely that the measured values exceed the true value and that they fall below the true value. In addition, small errors of measurement are more frequent than large errors. Participants are asked to form an initial estimate for each city or region based on the respective measurements. For each city or region, participants also receive advice from a colleague who has access to different but equally precise measurements. This advisor tells participants both what their estimate was and how many measurements this estimate was based on. We already know from previous research that participants use information about the number of cues available to an advisor appropriately, that is, they rely more on advisors with a greater information basis (Budescu et al., 2003). After receiving the advice, participants make their final estimates for each city or region.

The choice of the task content—estimating the annual precipitation of different locations—aligns with our key purpose as it would require participants to integrate several imperfect measurements to make a good estimate. The framing of the task as a weather forecasting task was arbitrary, and we chose it because we felt that it would make intuitive sense to participants to base their estimates on the available measurements. Using the same approach, that is, estimating a true value based on some independent measurements, we could have created a similar task by having participants estimate the typical household income in different areas, the typical lifespan of different electric appliances, or the growth rate of different plants. What this means is that researchers who want to adapt our approach have some degree of freedom when choosing the task content. They can copy the rainfall estimation task, but they can also change the content while keeping the underlying logic constant.

In our rainfall estimation task, the advisor is fictional, that is, we generate advice by taking the mean of all measurements drawn for the advisor. The fact that the advisor is simulated is made transparent to participants, that is, we inform them that their advisor is a fictional colleague and that the advice is computer-simulated such that the advisor's behavior reflects that of actual participants. We elaborate further that this means that the simulated advisor uses the available measurements in the exact same

https://doi.org/10.1017/jdm.2024.43 Published online by Cambridge University Press

fashion the average participant would have used them (since, on average, participants, take the mean of their measurements, this information is veridical). Using a simulated advisor comes with the advantage that it allows us to run the study without first conducting a pre-study that aims to gather estimates as advice. By informing participants about the nature of the advice, we further avoid deceiving them. Note that recent research has shown that people may react differently to advice when it is created by an algorithm (Jussupow et al., 2020). We return to this issue in the general discussion.

Regarding the stimuli, the true values can either be derived from actual meteorological data (Study 1), or they can be simulated by drawing uniformly distributed random numbers from a plausible range (Study 2). The individual measurements can be drawn from normal distributions. The only limits for these drawings are that the variance of the normal distributions should allow for meaningful variance in the measurements in order to create at least some distance between initial judgment and advice and that the values drawn from the normal distributions need to be truncated to avoid implausible values such as negative annual precipitation.

As mentioned above, a nice feature of the rainfall estimation task is that we have full control over the relative quality of the advice. For example, we can create situations, in which judge and advisor have the exact same number of measurements. In those cases, the optimal weight of advice is 50%. We can also create situations in which either the judge or the advisor has no valid measurements. If the judge has no valid measurements and the advisor has at least one, the optimal weight of advice is 100%. If, in contrast, the judge has at least one valid measurement while the advisor has none, the optimal weight of advice is 0%, that is, the advice should be ignored.

Finally, since the measurements are unbiased and the advisor integrates them in a rational way, participants' accuracy is contingent on (a) using their own measurements sensibly, for example, by, at least roughly, taking the mean and (b) by weighting the advice according to its relative quality.[1] This allows for studying the benefits of advice in terms of increasing judgmental accuracy (Soll and Larrick, 2009) and also provides an ideal basis for incentivizing effective use of advice by awarding bonus payments based on the accuracy of judges' estimates. We subjected the rainfall estimation task to an initial test in our first study.

3. Study 1

The purpose of Study 1 was to test whether judges working on the rainfall estimation task would display the core phenomena of advice taking outlined above: sensitivity to advice quality, egocentric advice discounting, overutilization of useless advice, a curvilinear effect of advice distance on advice weighting, and a W-shaped distribution of the weight of advice. This was particularly relevant because the rainfall estimation task differs from most previous JAS tasks in 2 ways. First, the rainfall estimation task makes the amount of information available to judge and advisor more transparent and, because the type of information available is the same for judge and advisor, there is less ambiguity about how advisors come up with their estimates. Put simply, although judges don't know the exact measurements observed by the advisor nor which aggregation strategy the advisor uses, they can form reasonable expectations about both. This is relevant because one of the explanations for egocentric advice discounting is differential access to information, which results in judges being better able to justify their own opinions than those of the advisors (Yaniv, 2004). Since this information asymmetry is less pronounced in our task, there is the risk that one of the core phenomena of advice taking, namely egocentrically discounting, might not manifest.

Second, using a simulated advisor might lead to behavior that differs from that observed in JAS studies using (alleged) human advisors. Using a computer-simulated advisor reduces the social element

[1] In fact, since all measurements are unbiased, we can even state how the advice should be weighted given that judges take the mean of their measurements when forming the initial estimates. In that case, the optimal weight of advice is simply the ratio of the available measurements of advisor and judge (Bednarik and Schultze, 2015). For example, if the advisor has twice as many measurements, the weight of advice should be twice that of the weight placed on the initial estimate, that is, 66.7%.

https://doi.org/10.1017/jdm.2024.43 Published online by Cambridge University Press

of the already socially sparse JAS even further. In addition, even despite us explicitly informing participants that the simulated advisor was designed to mimic the behavior of the typical participant, judges might use the computer-generated advice differently when compared to previous studies using human-generated advice (Dietvorst et al., 2015; Logg et al., 2019; Prahl and Van Swol, 2017).

Due to these potential concerns, our rationale was to test, in a first study, whether the core phenomena of advice taking usually observed in studies using Western samples working on general knowledge tasks with (alleged) human advisors still emerge in the rainfall estimation task. Only if these core phenomena manifest, can we consider our study suitable for employment in cross-national studies of advice taking.

3.1. Method

3.1.1. Participants and design

Participants were 83 German university students (53 identified as female, and 30 identified as male) with a mean age of 23.90 years (SD = 4.10 years), who took part in the study in exchange for either €6 or course credit. In addition, participants could earn a bonus of up to €3 based on the accuracy of their final estimates. We determined the sample size per rule of thumb, aiming for at least 80 participants. We stopped data collection after the experimental session in which we hit that threshold. Since we slightly overbooked that session, the total sample contained 83 participants. The study was based on a 6 (number of judge's cues: 0 to 5) × 6 (number of advisor's cues: 0 to 5) within-subjects design.

3.1.2. Procedure

Study 1 was a computer-based laboratory experiment programmed using the software Alfred (Treffenstaedt and Wiemann, 2018). Participants worked on 72 trials of the rainfall estimation task (2 trials for each combination of the 2 independent variables). Participants received written instructions about the study on their computers. There, we informed them that the study consisted of 2 parts. In part 1, they would estimate the annual precipitation of 72 Asian cities likely to be unknown to them based on available measurements. They were further told that there might be cities for which they could not access any weather data and that they would have to guess the annual precipitation in those cases. We further told participants that, as experienced meteorologists, they knew that annual precipitation in Asia did not exceed 5,000 mm per year and that they should, thus, guess a number between 1 and 5,000 mm per year if they had no valid measurements. As their initial estimates, participants would enter numbers ranging from 1 to 5,000. For each initial estimate, participants would further indicate how confident they were that it was accurate using a 7-point scale ranging from 1 (not at all confident) to 7 (very confident). Figure 1 shows participants' screen for an initial estimate (top panel) and for the corresponding final estimate (top panel).

In part 2 of the study, they would work on all 72 trials again to provide final estimates. On each of these trials, participants would first be presented with their own initial estimate, the number of measurements this estimate was based on, and their initial confidence rating. They would also be shown the advice of their fictional colleague who would have access to the measurements of independent but equally precise weather stations for each city. The advice would consist of that colleague's estimate along with the number of weather stations that the colleague had access to on a given trial. We informed participants that the advisor was computer-simulated such that their behavior mimicked that of the average participant. In case the simulated advisor had no valid measurements for a city, they would randomly pick a number between 1 and 5,000 with all numbers being equally likely (uniform distribution). We chose a uniform distribution here because it aligns best with an advisor who can only guess in the absence of any information. Participants were also to rate their confidence in the accuracy of their final estimates using the same 7-point scale mentioned above. Finally, we explained to participants that they could receive a bonus payment of 1, 2, or 3 Euros depending on the accuracy of their final estimates.

https://doi.org/10.1017/jdm.2024.43 Published online by Cambridge University Press

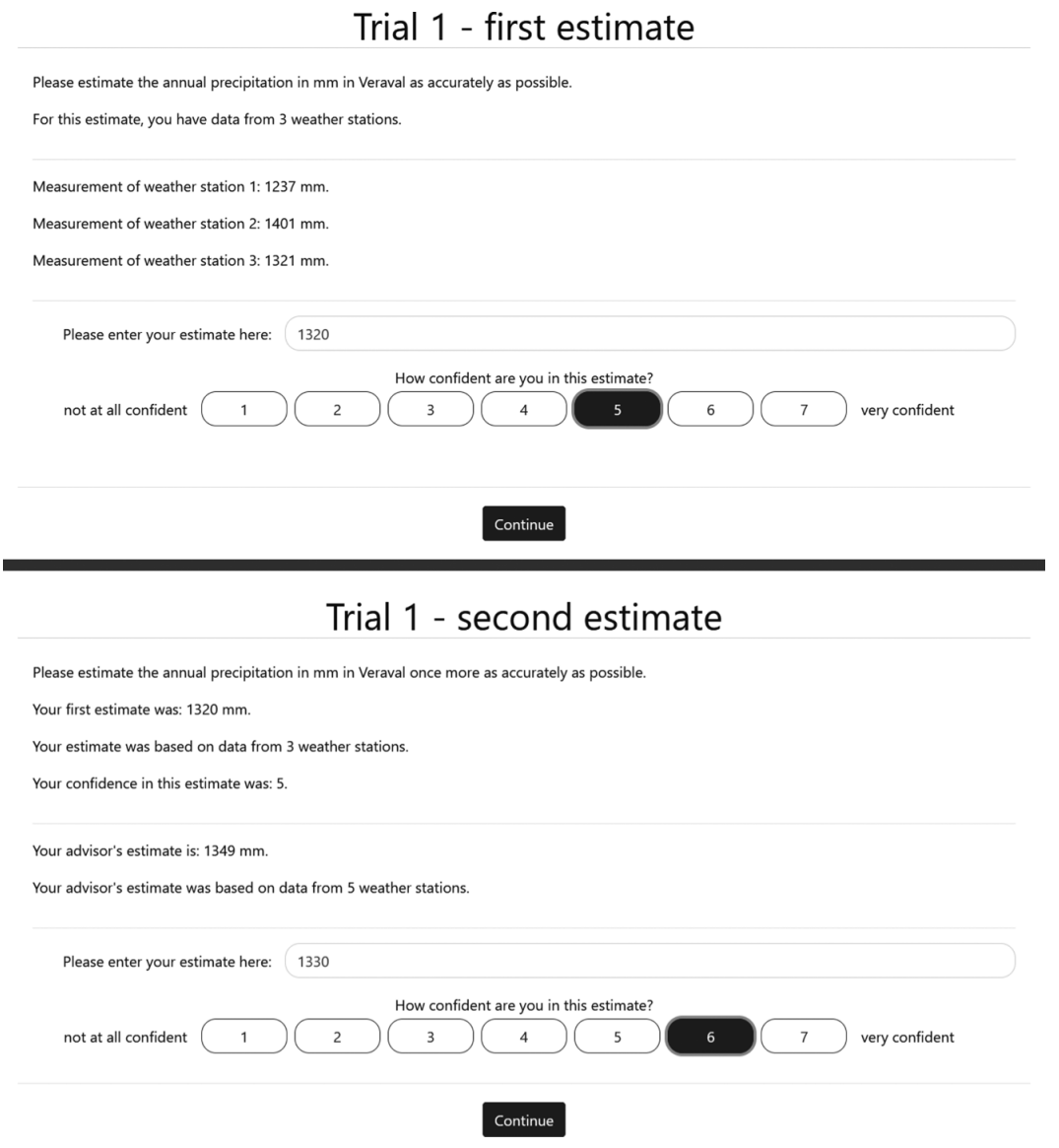

Figure 1. *Example screen of an initial and final estimate in the computerized Experiment 1.*
Note: The text was translated from German to English for this figure for the convenience of the reader.

In Study 1, we used real data on annual precipitation in 72 Asian cities as the targets (at the time the study was run, these data were publicly available from the German Meteorological Service). The measurements available to the judge were generated online by drawing from normal distributions centered on the true value with a standard deviation equaling one-third of the true value. In order to prevent implausible measurements, we limited the range of possible values. The lower bound was 5% of the true value, and values below this threshold we set to 5% of the target value. If the measurement had exceeded the upper limit of 5,000 mm per year, the computer instead drew a random number from a uniform distribution ranging from 4,900 to 5,000. Finally, all measurements were rounded without decimals. The measurements were presented on the screen below a message stating how many valid measurements were available to the judge on that specific trial.

https://doi.org/10.1017/jdm.2024.43 Published online by Cambridge University Press

Advice was generated by first drawing the respective number of measurements from normal distributions centered on the true values with a standard deviation equaling one-third of the target value, then taking their arithmetic mean, and finally rounding it. Advice was limited to values from 10 to 5,000 in order to prevent implausible values. That is, if the advice would have been lower than 10 mm per year it was set to 10, and if it would have exceeded 5,000 mm per year it was set to 5,000 instead. As mentioned above, when the advisor had no valid measurements, advice was generated by drawing a random number between 1 and 5,000 and then rounding it.

Once participants had completed part 2 of the study, they were presented a final questionnaire. In this questionnaire, we first asked them to indicate whether they had worked on the task seriously. We then asked them to rate the average accuracy of their initial estimates, their final estimates, and the advice on 7-point scales from 1 (not at all accurate) to 7 (very accurate). Finally, we asked them to report their age in years and their gender (female, male, or other).

The final page of the experiment consisted of personalized feedback. Here, participants learned how accurate their initial estimates, their final estimates, and the advice had been. To this end, we presented participants with the respective mean absolute percent error (MAPE) scores. We also told them how good their final estimates could have been if they had used the advice optimally. The optimal weight of advice used for this computation assumes that judge and advisor form their estimates by taking the mean of all available measurements. As mentioned earlier, it can be shown that in this case, the expected accuracy of the final estimates is maximal if the ratio of weights placed on the initial estimate and the advice equals the ratio of measurements available to judge and advisor (Bednarik and Schultze, 2015). Finally, participants were told the magnitude of their bonus payment. This bonus payment was determined based on the MAPE of their final estimates. The bonus payment was €3 for a MAPE below 10%, €2 for a MAPE greater than 10% but below 15%, and €1 for a MAPE greater than 15% but below 20%.

Once the experiment was finished, participants were paid and thanked for participation. In addition, we informed them about the aims of our study and answered any questions participants might have had. Finally, we asked participants to keep the aims of our study confidential.

3.1.3. Measures

Advice taking. Our measure of advice taking was the advice taking coefficient (Harvey and Fischer, 1997) which is defined as:

$$AT = \frac{IE - FE}{IE - AD}$$

Here, *IE* is the initial estimate, *FE* is the final estimate, and *AD* is the advice. AT scores are proportional to the percent weight assigned to the advice when making the final estimate. An AT score of 0 means the advice was completely disregarded, an AT of 1 indicates that the advice was fully adopted, and an AT of 0.50 means that the judge averaged the initial estimate and the advice.

It is common practice to winsorize AT scores at 0 and 1, which means that scores lower than 0 are set to 0 and scores exceeding 1 are set to 1 (Bonaccio and Dalal, 2006). Deviating from this common practice, we winsorized AT sores at −1 and 1, instead in Study 1. The reason was that we aimed to test, among other things, whether participants would overutilize useless advice, that is, advice based on 0 measurements. In those cases, winsorizing AT scores at 0 can artificially inflate AT scores, running the risk of overestimating evidence of overutilization of advice (Schultze et al., 2017). In cases where the initial estimate equals the advice, AT scores are not defined. This happened on 4 occasions (0.07% of cases), and the respective trials had to be excluded from data analysis. Due to winsorizing, 168 AT scores were set to 1 (2.81% of cases), and 14 AT scores were set to -1 (0.23% of cases).

Confidence shifts. For exploratory analyses, we computed confidence shifts as the difference between participants' confidence in the accuracy of their final estimates and initial estimates. Positive

https://doi.org/10.1017/jdm.2024.43 Published online by Cambridge University Press

values of this measure indicate that participants were more confident in their final than in their initial estimates.

3.2. Results

3.2.1. Transparency statement

Study 1 was not preregistered, but materials, data, and analysis code are available publicly from the Open Science Framework (https://osf.io/fwbau/).

3.2.2. Egocentric advice discounting

We first tested whether participants working on the rainfall estimation task discounted advice egocentrically. Since judges and advisors had, over the course of the experiment, the same information basis, the mean AT scores of a rational judge would amount to 0.50. In order to test whether mean AT scores differed from this normative benchmark, we analyzed AT scores in an intercept-only multi-level model with random intercepts for participants. The fixed intercept of this model represents the mean level of advice taking. Not surprisingly, the intercept was significantly different from 0 ($B = 0.45$ ($SE = 0.01$), $t(82.02) = 46.36$, $p < 0.001$, 95% CI [0.41; 0.46]). Importantly, the 95% CI excluded 0.50, which means that participants heeded advice significantly less than they should have, given that, on average, advisors were equally knowledgeable as judges.

As a robustness test, we re-ran the model but restricted it to trials in which judge and advisor had the exact same number of available cues (e.g., both judge and advisor had 3 independent measurements). In these trials, it should have been particularly salient to participants that their advisor was equally knowledgeable. The fixed intercept of the respective multi-level model was below 0.50 ($B = 0.44$ ($SE = 0.01$), $t(82.08) = 33.61$, $p < 0.001$, 95% CI [0.41; 0.46]). That is, even when restricting the analysis to cases in which judge and advisor had a similar information basis, participants egocentrically discounted the advice.

As a final test of egocentric advice discounting, we analyzed only trials in which judges had no valid measurements while their advisor had at least 1. In those cases, the optimal weight of advice is 100%. The fixed intercept of the model indicated that judges weighted the advice by about 89% when they had no reliable information ($B = 0.89$ ($SE = 0.02$), $t(82.00) = 54.01$, $p < 0.001$, 95% CI [0.85; 0.92]). However, the 95% CI excluded 1, indicating that participants egocentrically discounted advice even when their own initial opinion was an uninformed guess.

3.2.3. Sensitivity to advice quality

In order to test whether participants were sensitive to advice quality, we predicted AT scores in a multi-level model with the number of judges' and advisor's cues as fixed effects. We included random intercepts as well as random slopes for the number of the advisor's cues (the model did not converge when including random slopes for both predictors). Judges were aware of their own information basis, heeding advice less when they had more information ($B = -0.09$ ($SE = 0.002$), $t(5805.02) = -45.74$, $p < 0.001$, 95% CI [−0.094; −0.087]). Likewise, they heeded advice more, the better the information basis of their advisor was, indicating that judges used information about advice quality in a sensible fashion ($B = 0.10$ ($SE = 0.005$), $t(81.98) = 21.64$, $p < 0.001$, 95% CI [0.09; 0.11]).

3.2.4. Inability to ignore useless advice

We next tested whether judges working on the rainfall estimation task would overutilize useless advice by computing an intercept-only multi-level model of AT scores with random intercepts for participants. We limited the analysis to trials in which judges had at least 1 valid measurement while their advisors had none. From the instructions, participants knew that the advisor would merely guess in those trials, meaning that the advice contained no information at all about the target value and should, accordingly, be ignored. On average, participants heeded advice systematically, indicated by a fixed intercept

https://doi.org/10.1017/jdm.2024.43 Published online by Cambridge University Press

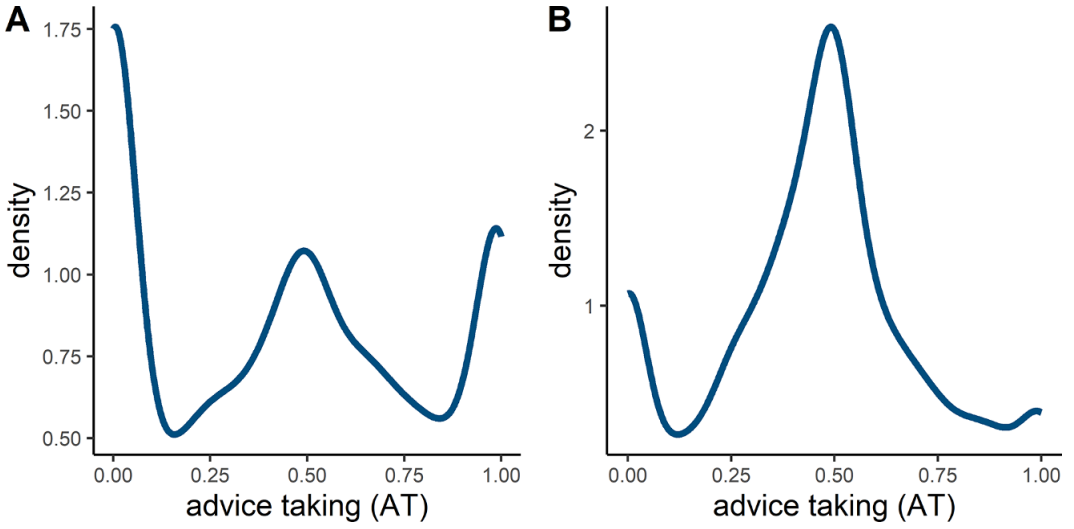

Figure 2. *Density plot of AT scores ranging from 0 to 1.*
Note: The density plot in panel A includes all trials for which AT scores were between 0 and 1. Panel B shows the density of AT scores between 0 and 1 for trials in which the number of the judges' cues was equal to the number of the advisor's cues.

significantly greater than 0 ($B = 0.08$ ($SE = 0.02$), $t(82) = 4.21$, $p < 0.001$, 95% CI [0.04; 0.11]). That is, they were unable to ignore useless advice.

3.2.5. Trimodal distribution of AT scores

Similar to Soll and Larrick (2009), who first reported the W-shaped distribution of AT scores, we relied on a visual inspection of the data in order to determine whether AT scores in the rainfall estimation task showed the trimodal distribution. To this end, we restricted the data to AT scores between 0 and 1 and then plotted their density. As can be seen from Figure 2 (panel A), the W-shape is very pronounced in our data with an unusually high mode at 100% weight of advice. This may be partly because the experimental manipulations created situations in which 0% and 100% weight of advice were the normatively correct responses. As a robustness analysis, we plotted the density of AT scores again for trials, in which judge and advisor had the exact same number of cues. As could be expected, participants were more likely to weight the advice by 50% in those trials. However, despite the fact that 50% weight of advice is the normatively correct choice in those trials, the W-shape is still observable (see Figure 2, panel B).

3.2.6. Curvilinear relation of advice distance and advice taking

We used multi-level modeling to test for a curvilinear effect of advice distance on advice taking. To this end, we first computed the absolute distance between advice and the initial estimate for each trial. We then entered absolute advice distance and its logarithm as fixed effects in the model (see Schultze et al., 2015, for a similar approach). Since models with random slopes did not converge, we only included random intercepts for participants in the model. The model showed a negative effect of absolute advice distance ($B = -0.0001$ ($SE = 0.000001$), $t(5918.49) = -12.05$, $p < 0.001$, 95% CI [−0.0001; −0.00008]), as well as a positive effect of its logarithm ($B = 0.05$ ($SE = 0.005$), $t(5916.57) = 11.14$, $p < 0.001$, 95% CI [0.04; 0.06]). The combination of the 2 effects yields the classic pattern in which advice taking is highest for advice of moderate distance, with lower AT scores for both advice that is either close or far away from the initial estimates (see Figure 3).

As a robustness analysis, we reran the multi-level model while excluding trials in which either the judge or advisor had no valid cues. Those trials might be characterized by both extreme distances

https://doi.org/10.1017/jdm.2024.43 Published online by Cambridge University Press

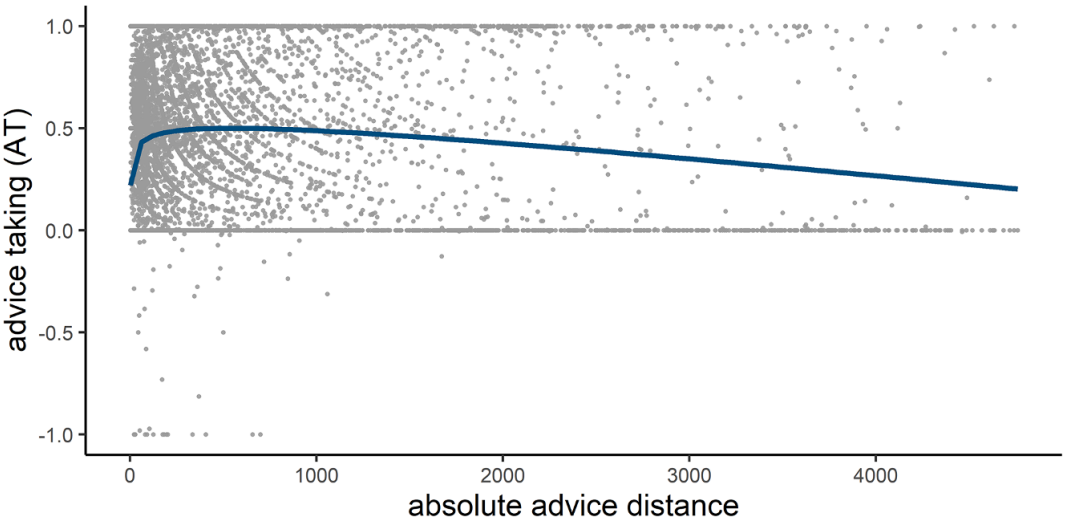

Figure 3. *AT scores as a function of advice distance.*
Note: The grey points represent individual observations. The bold blue line is the regression line derived from the fixed effects of the multi-level model predicting AT scores from absolute advice distance and its logarithm.

between initial estimates and advice and also extreme AT scores. When excluding those trials, the curvilinear relationship still emerged with a negative effect of advice distance and a positive effect of its logarithm ($B = -0.0001$ ($SE = 0.00002$), $t(4082.65) = -2.93$, $p = 0.003$, 95% CI [−0.00008; −0.00002], and $B = 0.03$ ($SE = 0.01$), $t(4082.41) = 6.20$, $p < 0.001$, 95% CI [0.02; 0.04], respectively). To further test the robustness of the pattern and to align the analysis with previous research, we restricted the data to AT scores between 0 and 1. The results were qualitatively similar ($B = 0.0001$ ($SE = 0.00002$), $t(4042.05) = -3.59$, $p < 0.001$, 95% CI [−0.00009; −0.00003] for the linear effect, and $B = 0.03$ ($SE = 0.05$), $t(4042.61) = 6.68$, $p < 0.001$, 95% CI [0.02; 0.04] for the logarithmic effect).

3.2.7. Judges' confidence shifts

While the focus of our study is on advice weighting, we also ran some exploratory analyses of judges' confidence shifts. We know from previous research that judges are, on average, more confident in their final than in their initial estimates, and these confidence gains are moderated by advice quality (Schultze et al., 2015). To test whether confidence shifts behave in a comparable fashion in the rainfall estimation task, we first tested for an overall increase in confidence by predicting confidence shifts from an intercept-only multi-level model with random intercepts for participants. The intercept of this model was positive and significantly different from 0 ($B = 0.74$ ($SE = 0.04$), $t(82.00) = 17.85$, $p < 0.001$, 95% CI [0.66; 0.82]). That means participants were, on average, more confident in their accuracy of their final estimates as is typical in research using the JAS.

We then tested whether the increase in confidence was moderated by advice quality. To this end, we entered confidence shifts as the criterion in a multi-level model with fixed effects of the number of the judge's cues and the number of the advisor's cues. The model contained random intercepts as well as random slopes for the judge's number of cues and the advisor's number of cues. Confidence shifts were lower when the judge had more information initially ($B = -0.22$ ($SE = 0.02$), $t(81.99) = -13.78$, $p < 0.001$, 95% CI [−0.25; −0.19]). They increased by a comparable margin with the number of cues available to the advisor ($B = 0.21$ ($SE = 0.02$), $t(82.00) = 12.92$, $p < 0.001$, 95% CI [0.18; 0.25]). This finding, too, is well in line with previous research using the JAS (Schultze et al., 2015).

https://doi.org/10.1017/jdm.2024.43 Published online by Cambridge University Press

3.3. Discussion

The results of Study 1 show that Western participants working as judges in the rainfall estimation task show the core phenomena of advice taking typically observed in Western samples. Specifically, participants egocentrically discounted helpful advice, but overutilized advice when it was not helpful and should have been ignored instead. Apart from these fallacies, participants were sensitive to advice quality, weighting the advice more the more information it was based on. We also observed the typical curvilinear relation of advice taking and advice distance, with lower weights of advice for both advice that was very close and far away from judges' initial estimates. Finally, the distribution of the weights of advice followed the characteristic W-shape with modes at 0%, 50%, and 100% weight of advice. Our exploratory analyses of judges' confidence also showed results reminiscent of earlier findings, that is, judges' confidence increased after receiving advice, and this increase was more pronounced the more knowledgeable the advisor was in a given trial.

In sum, German participants working on the rainfall estimation task showed all of the core phenomena even though the task differed from most previous JAS tasks in that the advisor was computer-simulated and there was less ambiguity around how the advisors formed their judgments. These findings suggest that the rainfall estimation task is well-suited to studying advice taking cross-nationally because it produces what we termed the core phenomena of advice taking without relying on general knowledge likely to be biased by participants' nationality. Having established the in-principle suitability of the new task, we next used it to explore cultural differences in advice taking using a countries-as-proxies comparison between German and Chinese participants.

4. Study 2

The objective of Study 2 was to compare advice taking between German and Chinese students using the rainfall estimation task. We chose this comparison largely because Germany and China differ in terms of individualism versus collectivism, which is the dimension previous research has focused on when investigating cultural differences in advice taking (Bailey et al., 2023; Mercier et al., 2012). To explore potential national differences, we employed a simplified version of the rainfall estimation task. This simplified task had fewer trials and could be administered using pencil and paper. In Study 2, we were mainly interested in potential differences in the overall weight of advice, egocentric advice discounting, sensitivity to advice quality, the distribution of AT scores, and the curvilinear relationship of advice distance and advice taking.

4.1. Participants and design

Participants were 149 German and 142 Chinese university students of which 101 identified as male (54 in the Chinese and 47 in the German sample), 187 identified as female (86 in the Chinese and 101 in the German sample). One German participant reported their gender as other than male or female, and 2 Chinese participants preferred not to state their gender. Participants were, on average, 22.03 years old (SD = 4.57). The average age for the German sample was 24.56 years (SD = 5.15 years) whereas Chinese participants were, on average, 19.37 years old (SD = 1.10 years). Participants took part in the study in exchange for a fixed participation fee (5 Euros for German and 7 RMB for Chinese participants) or course credit. In addition, we awarded a performance-based bonus (10 Euros and 70 RMB, respectively) to the 10 best performing participants in each sample with performance being determined via the accuracy of their final estimates. We determined the sample size per rule of thumb, aiming for roughly 150 participants per country. Originally, we gathered data from 294 participants (150 from Germany and 144 from China), but we had to exclude data from 2 participants because they did not complete the study, and we excluded 1 further participant who made their first estimates by roughly adding the cues rather than (roughly) averaging them, which led to excessive initial estimates. The study was based on a 2 (number of judge's cues: 3 vs. 6) × 2 (number of advisor's cues: 3 vs. 6) × 2

https://doi.org/10.1017/jdm.2024.43 Published online by Cambridge University Press

(country: China vs. Germany) mixed design with number of judges' cues and number of advisor's cues as experimental within-subjects factors and country as a quasi-experimental between-subjects factor.

4.2. Procedure

The procedure for Study 2 was similar to that of Study 1 with the following exceptions. First, we used a simplified version of the task that could be administered using pencil and paper. This necessitated pre-generating the cue samples for judge and advisor. Instead of using archival data, we generated true values by drawing 12 times from a uniform distribution between 200 and 3,000 and rounding the resulting numbers. We then generated measurements for the judge by drawing randomly from normal distributions centered around the true values with a variance of 30,000 (a standard deviation of roughly 173.2). Instead of a continuous manipulation of the information basis of judge and advisor, we manipulated the number of their cues orthogonally as 2-level factors (3 vs. 6 cues). Specifically, we drew samples of 3 measurements for trials 1 to 6 and 6 measurements for trials 7 to 12. We then generated the advice by drawing from the same normal distribution. On trials 1 to 3 and 7 to 9, we drew samples of size 3, and on trials 4 to 6 and 10 to 12, we drew 6 measurements for the advisor. Advice was generated by taking the mean of the respective measurements and rounding the resulting number. As a consequence of this approach, each participant of Study 2 received the exact same measurements and advice on each trial (this may have had the added benefit of reducing some of the between-participant variance that would have resulted from generating stimuli online as we did in Study 1). The R code used to generate the judges' samples and the advice is available online (https://osf.io/fwbau/).

Second, we no longer used the names of Asian cities in our stimulus materials. While Asian cities were likely to be unknown to judges in the German sample of Study 1, retaining city names in the study material would have created a possible confound with perceived task difficulty (exactly the confound we aimed to avoid with the new task) when studying cross-national differences between German and Chinese students. Thus, we simply referred to Region 1, Region 2, and so on in the study materials.

Third, we changed the way we incentivized good performance by awarding bonus payments to the top 10 performing students in each sample after data collection was complete. The main reason for this change was that unlike in Study 1, there was no computer program to automatically compute the accuracy of the final estimates to determine bonus payments. Determining who would receive the reward for good performance also required us to record student's email addresses so we would be able to contact them if they were among the top performers (participants could indicate that they did not wish to be contacted by us). We made it explicit to participants that their data would be analyzed in a fully anonymized fashion.

Fourth, the written instructions for Study 2 were slightly changed to reflect the changes mentioned above. Apart from those minor changes, instructions were identical to those of Study 1.

Finally, in Study 2, we winsorized AT scores at 0 and 1 since we did not investigate the overutilization of useless advice. In total, 1 AT score (0.03%) was not defined because the initial estimate equaled the advice, 101 AT scores (2.89%) were winsorized at 1, and 61 AT scores (1.75%) were winsorized at 0.

Once participants entered the lab, they were handed out the first part of the material comprising the informed consent, written instructions, and a sheet for the initial estimates. On this sheet, participants saw their available measurements for each of the 12 regions. On that same sheet, they could write down their initial estimate and indicate their confidence in its accuracy on a 7-point scale. Once participants had made all initial estimates, they received a second sheet listing the advice as well as the number of independent measurements available to the advisor for each trial. Participants could then write down their second estimates and rate their confidence on that sheet.

The original material for Study 2 was created in German, translated into English, and finally translated into Chinese. Technically, this may have led to some aspects of the text being lost in translation, but due to the brevity of the instructions and the simple nature of the judgment task, we

https://doi.org/10.1017/jdm.2024.43 Published online by Cambridge University Press

Table 1. *Multi-level model predicting advice taking in Study 2.*

Predictors	B	SE	95% CI	t	df	p
Intercept	0.40	0.02	0.37–0.43	25.03	288.70	**<0.001**
Judge's cues	0.02	0.01	−0.01–0.04	1.03	287.55	0.302
Advisor's cues	0.05	0.01	0.02–0.08	3.37	289.16	**0.001**
Country	−0.07	0.02	−0.11–0.02	−2.93	288.69	**0.004**
Judge cues × Country	−0.05	0.02	−0.09–0.01	−2.60	287.54	**0.010**
Advisor cues × Country	0.04	0.02	0.002–0.09	2.06	289.15	**0.040**
Random effects						
σ^2	0.07					
$\tau_{ID,\ intercept}$	0.02					
$\tau_{ID,\ judge's\ cues}$	0.01					
$\tau_{ID,\ advisor's\ cues}$	0.01					
$\rho_{intercept,\ judge's\ cues}$	−0.08					
$\rho_{intercept,\ advisor's\ cues}$	−0.02					
$\rho_{judge's\ cues,\ advisor's\ cues}$	−0.55					
ICC	0.26					
N_{ID}	291					
Observations	3488					
Marginal R^2/Conditional R^2	0.030/0.281					

Note. Country was dummy-coded (China = 0, Germany = 1). Judge's cues and advisor's cues were dummy-coded as well (3 cues = 0, 6 cues = 1). Bold face indicates $p < 0.05$.

consider it highly unlikely that participants in our 2 samples understood the instructions differently. The English version of the study material is available online (https://osf.io/fwbau/).

4.3. Results

4.3.1. Transparency statement

Study 2 was not preregistered, but materials, data, and analysis code are available publicly from the Open Science Framework (https://osf.io/fwbau/).

4.3.2. Advice taking

We analyzed participants' mean AT scores in a multi-level regression with fixed effects for country, number of the judge's cues, and number of the advisor's cues. The model also contained the interaction of judge's cues and country as well as the interaction of advisor's cues and country as fixed effects. Finally, we included random intercepts as well as random slopes of judge's cues and advisor's cues for participants (the results of the analysis are shown in Table 1).

The first effect of interest in the model is the main effect of country. As can be seen from Figure 4A, Chinese participants heeded the same advice more than German participants ($M = 0.44$, $SD = 0.17$ vs. $M = 0.37$, $SD = 0.15$). In order to gain more insights into this general difference in advice taking, we inspected the distribution of AT scores by country in a graphical analysis. As can be seen in Figure 4B, Chinese participants had higher average AT scores largely because their advice taking strategy differed from that of the German participants. Chinese participants were less likely to ignore the advice and more likely to engage in averaging. Notably, this difference was so pronounced that the 2 samples differed in the modal weight of advice. For German participants, ignoring the advice was the most frequent response, whereas for Chinese participants the predominant strategy was averaging the initial estimate and the advice. Irrespective of these differences, the graphical inspection of the distribution of AT scores shows the classic W-shaped distribution with modes at 0%, 50%, and 100% in both samples.

https://doi.org/10.1017/jdm.2024.43 Published online by Cambridge University Press

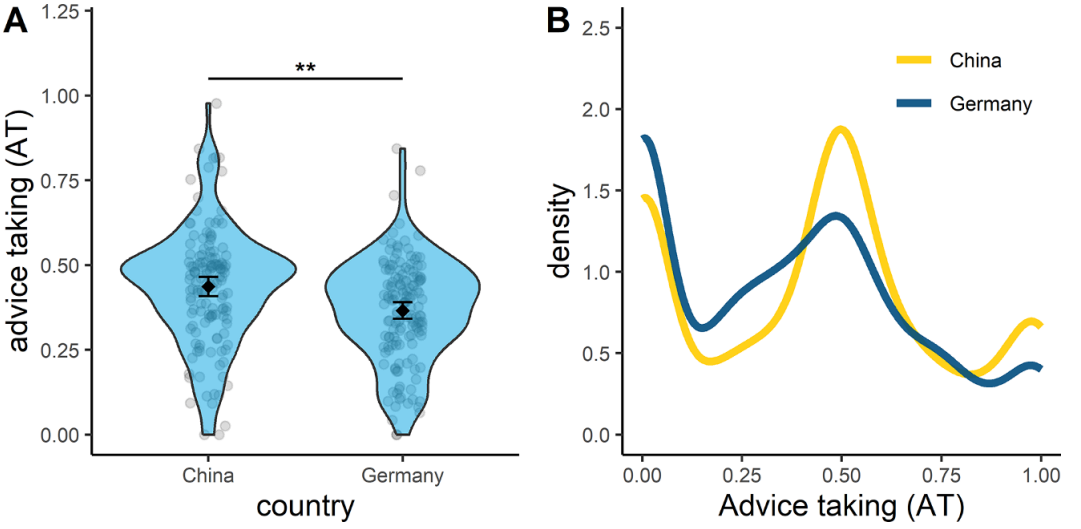

Figure 4. *Advice taking by country.*

Note: Panel A shows participants' mean AT scores across all 12 trials by country (represented by the jittered points in the plot). The violins indicate the density of the data at a given level of advice taking. The diamonds and error bars represent the country means of advice taking and their respective 95% confidence intervals. Panel B shows the density of individual AT scores between 0 and 1 by country.

4.3.3. Egocentric advice discounting

Having found that Chinese and German participants differed in the extent to which they were receptive to advice, we next investigated whether participants in both samples egocentrically discounted advice. Remember that, on average, judge and advisor had the exact same information basis in Study 2, which means that participants should, on average, heed the advice by 50%. Accordingly, average AT scores that fall below 0.50 indicate egocentric advice discounting. We tested for egocentric advice discounting for each sample, using intercept-only multi-level models predicting AT scores. In these models, the fixed intercept represents the average level of advice taking in the respective sample, and we can test for egocentric advice discounting by checking whether its 95% CI excludes 0.50. This was the case for the German but also for Chinese participants ($B = 0.37$ ($SE = 0.01$), $t(148.06) = 29.61$, $p < 0.001$, 95% [0.34; 0.39], and $B = 0.44$ ($SE = 0.01$), $t(141.02) = 30.28$, $p < 0.001$, 95% [0.41; 0.47], respectively). This pattern held true even when restricting the analysis to trials, in which judge and advisor had the exact same number of cues (German sample: $B = 0.38$ ($SE = 0.01$), $t(148.09) = 28.69$, $p < 0.001$, 95% [0.35; 0.40]; Chinese sample: $B = 0.44$ ($SE = 0.01$), $t(141.26) = 28.93$, $p < 0.001$, 95% [0.41; 0.47]). That is, although Chinese participants heeded advice more than German participants, they still discounted it by a small margin.

4.3.4. Sensitivity to advice quality

We next turned our attention to the question of whether participants were sensitive to advice quality. As can be seen in Table 1, our analysis of advice taking showed a main effect of the number of the advisor's cues that was qualified by an interaction of the number of the advisor's cues and country. In addition, there was an interaction of the judge's number of cues and country. To disentangle these interactions, we computed separate multi-level models for each sample to predict advice taking. These models contained the judge's and the advisor's cues as fixed effects as well as random intercepts for participants (when adding random slopes for judge's and advisor's cues, the model only converged for the German sample: thus, we decided to omit them in both models).

Chinese participants considered the amount of their advisor's information in their advice taking decisions, weighting the advice more by about 5 percentage points when the advisor's information base was better ($B = 0.05$ ($SE = 0.01$), $t(1558.10) = 3.77$, $p < 0.001$, 95% CI [0.02; 0.08]). However, there

https://doi.org/10.1017/jdm.2024.43 Published online by Cambridge University Press

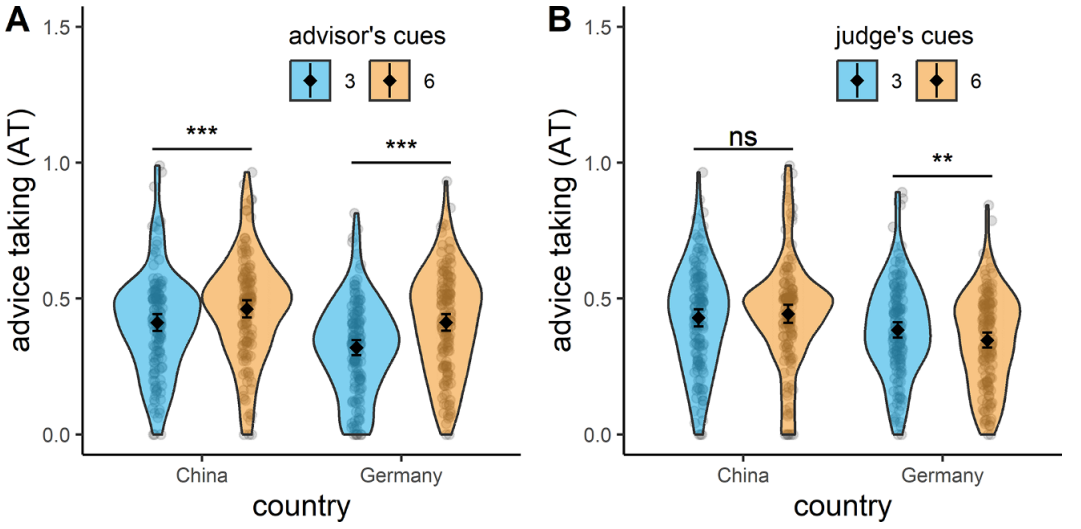

Figure 5. *Advice taking by advisor's cue and country as well as judge's cues and country.*
Note: Panel A shows the mean AT scores for participants by advisor's cues and country while Panel B shows the mean AT scores by judge's cues and country. Mean AT scores for each participant are represented by the jittered points in the plots. The violins indicate the density of the data at a given level of advice taking. The diamonds and error bars represent the country means of advice taking and their respective 95% confidence intervals.

was no evidence of sensitivity to their own information basis ($B = 0.02$ ($SE = 0.01$), $t(1558.10) = 1.13$, $p = 0.2578$, 95% CI [−0.01; 0.04]). German participants were more sensitive to advice quality than Chinese participants, increasing the weight of advice by roughly 9 percentage points when the advisor had 6 cues instead of only 3 ($B = 0.09$ ($SE = 0.01$), $t(1635.15) = 7.37$, $p < 0.001$, 95% CI [0.07; 0.12]). In contrast to their Chinese counterparts, German participants also considered how good their own information basis was, weighting advice less by 4 percentage points when they had 6 cues as opposed to 3 ($B = -0.04$ ($SE = 0.01$), $t(1635.15) = -3.00$, $p = 0.003$, 95% CI [−0.06; −0.01]). The results are also shown in Figure 5.

4.3.5. Curvilinear relation of advice distance and advice taking

In order to test whether the curvilinear effect of advice distance on advice taking was moderated by country, we predicted AT scores from absolute advice distance and its logarithm in a multi-level model. In addition to fixed effects for the 2 advice distance terms, the model also contained a fixed effect for country, as well as fixed effects for the interaction of country and advice distance, and country and logarithmic advice distance. Since the model did not converge when adding random slopes for advice distance and its logarithm, we only included random intercepts for participants.

The analysis revealed the classic inverse U-shaped relationship between advice distance and advice taking, characterized by a negative linear effect of advice distance and a positive effect of logarithmic advice distance ($B = -0.001$ ($SE = 0.0002$), $t(3306.31) = -5.26$, $p < .001$, 95% CI [−0.0012; −0.0006], and $B = 0.10$ ($SE = 0.02$), $t(3357.82) = 6.22$, $p < 0.001$, 95% CI [0.07; 0.13], respectively). However, country did not interact significantly with advice distance nor with its logarithm ($B = 0.0005$ ($SE = 0.0002$), $t(3300.06) = 1.94$, $p = 0.052$, 95% CI [−0.000004; 0.0009], and $B = -0.04$ ($SE = 0.02$), $t(3342.24) = -1.78$, $p = 0.075$, 95% CI [−0.08; 0.003], respectively). A graphical analysis suggested that there were a few outliers in terms of advice distance, and these outliers may have distorted the results slightly (see Figure 6). However, since we did not preregister the study and, thus, had no a-priori rule for handling outliers, we refrained from removing outliers, and instead treated the analysis regarding a moderating effect of country on the curvilinear relationship of advice distance and advice taking as inconclusive.

https://doi.org/10.1017/jdm.2024.43 Published online by Cambridge University Press

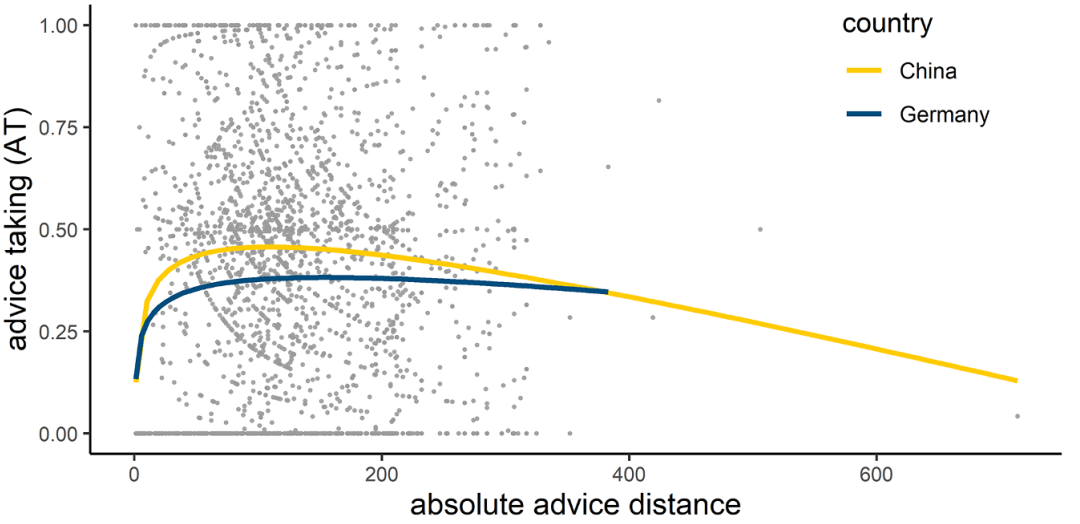

Figure 6. *AT scores as a function of advice distance and country.*
Note: The gray points represent individual observations. The bold lines are the regression lines derived from the fixed effects of the multi-level model predicting AT scores from country, absolute advice distance, its logarithm, as well as interactions of country with the 2 advice distance terms.

Table 2. *Multi-level model predicting confidence shifts in Study 2.*

Predictors	B	SE	95% CI	t	df	p
Intercept	0.25	0.07	0.12 – 0.38	3.87	650.00	**<0.001**
Judge's cues	0.28	0.05	0.18 – 0.39	5.24	3190.62	**<0.001**
Advisor's cues	0.14	0.05	0.03 – 0.24	2.52	3190.63	**0.012**
Country	−0.01	0.09	−0.19 – 0.17	−0.12	648.73	0.905
Judge's cues × Country	0.03	0.08	−0.12 – 0.18	0.36	3190.49	0.717
Advisor's cues × Country	0.13	0.08	−0.01 – 0.28	1.77	3190.49	0.078
Random effects						
σ^2	1.24					
$\tau_{ID;\ intercept}$	0.30					
ICC	0.19					
N_{ID}	291					
Observations	3485					
Marginal R^2/Conditional R^2	0.022/0.212					

Note. Country was dummy-coded (China = 0, Germany = 1). Judge's cues and advisor's cues were dummy-coded as well (3 cues = 0, 6 cues = 1). Bold face indicates $p < 0.05$.

4.3.6. Judges' confidence

Finally, we explored whether there were any national differences in terms of judges' confidence shifts after receiving advice. To this end, we predicted confidence shifts in a multi-level model with judge's cues, advisor's cues, country, and the interactions of country with judge's cues and country with advisor's cues as fixed effects. The model contained random intercepts only, since it did not converge when adding random slopes for the judge's and the advisor's number of cues. The results are displayed in Table 2.

As could be expected, judges' confidence increased significantly more when their advisor had more information. Confidence gains were also greater when judges had more information initially, which seems to contradict earlier research and the results of Study 1. Interestingly, neither the effect of country,

https://doi.org/10.1017/jdm.2024.43 Published online by Cambridge University Press

nor the interaction effects were statistically significant, that is, despite the pronounced differences in advice taking between the 2 countries described above, there was no evidence for differences regarding confidence shifts.

4.3.7. Exploratory analysis of participants' mathematical abilities

One pervasive stereotype is that Asians are better at math than Westerners (Cvencek et al., 2015). Cross-national comparisons between China and Germany using standardized math tests lend at least some credence to this stereotype with high school students from Hong Kong and Macao substantially outperforming their German counterparts (OECD, 2023). This holds some relevance for the results of our second study because the finding that Chinese participants heeded advice more than German participants might reflect their greater mathematical abilities rather than cultural differences. Put simply, Chinese participants might simply be better at computing the mean between their own estimates and those of the advisor, which might explain why averaging is their dominant advice taking strategy. Fortunately, we could test whether participants' ability to compute averages differed between countries. Recall that participants had to form their initial estimates on each trial based on either 3 or 6 independent and unbiased measurements. The strategy that maximizes accuracy is to take the mean of these measurements. We can infer participants' understanding of the average as an effective strategy as well as their ability to compute the average by inspecting how far the initial estimates deviate from the arithmetic mean of the available measurements. If Chinese participants' estimates were closer to the mean than those of the German participants, then our findings might simply reflect their greater mathematical ability.

To test this alternative explanation, we first computed the absolute difference between the initial estimates and the average of the available cues for each trial. We then entered this absolute deviation from the mean as the criterion in a multi-level model with country (dummy-coded) as a fixed effect and random intercepts for participants. The model's intercept was significantly different from 0, ($B = 38.72$ ($SE = 1.65$), $t(289) = 23.45$, $p < 0.001$, 95% CI [35.48; 41.96], indicating that Chinese participants' initial estimates deviated from the average of the available cues by about 39 units. The effect of country was not significant ($B = -0.98$ ($SE = 2.31$), $t(289) = -0.43$, $p = 0.671$, 95% CI [−5.50; 3.54]. That is, there was no evidence that German participants' estimates deviated more from the mean of the available measurements (their average deviation was about 1 unit lower). The median absolute deviation from the cue sample means was 23.50 units for Chinese participants with 0.17 units and 123.21 units marking the 5th and 95th percentile, respectively. Those numbers were comparable in the German sample with a median of 24.33 units, a 5th percentile of 2.00 units, and a 95th percentile of 121.00 units. In sum, while participants did not perfectly average the cue samples, there was no evidence of differences in the ability to compute the mean between countries. It is conceivable that differences in mathematical abilities between high school students from Germany and China would be less pronounced when comparing university student samples, or perhaps these differences are less pronounced because averaging is a relatively simple operation. Nevertheless, differences in mathematical abilities are unlikely to explain Chinese participants' greater propensity to use an averaging strategy.

4.4. Discussion

Study 2 shows that the rainfall estimation task can be successfully applied to study cultural differences in advice taking using a countries-as-proxies approach. The comparison between a German and a Chinese sample working on the exact same task revealed both interesting differences and commonalities in what we consider some of the core phenomena of advice taking. The most notable difference in advice taking between the two samples was that Chinese students weighted the same advice more strongly, largely because they were more likely to engage in an averaging strategy and less likely to ignore the advice. In addition, Chinese students tended to react less strongly to differences in advice quality, which implies that they avoided extreme weights in favor of averaging. These findings are what we would expect when comparing collectivistic with individualistic cultures. Interestingly, the

https://doi.org/10.1017/jdm.2024.43 Published online by Cambridge University Press

differences in advice taking were not accompanied by comparable differences in confidence shifts. Of course, it is not possible to interpret the absence of evidence for such differences as evidence of their absence. One viable interpretation is that, if those differences exist, they are likely much smaller than the ones we observed for advice taking. This interpretation rests on the assumption that participants in both countries used the confidence scale in the same fashion when making their confidence judgments. However, we know from previous research that participants from many Asian countries—notably excluding Japan—tend to make more extreme confidence judgments compared to Westerners, which also results in greater overconfidence (Wright et al., 1978; Yates et al., 1998; Yates and de Oliveira, 2016). Thus, we cannot rule out that the confidence measure we used is simply unable to capture existing differences in confidence gains because an increase in confidence by one scale point might simply mean different things for German and Chinese participants. Regarding commonalities, we found a tendency to egocentrically discount advice, sensitivity to advice quality, the classic W-shaped distribution of AT scores with modes at 0%, 50%, and 100% weight of advice, and a curvilinear relationship of advice distance and advice taking in both samples. While there were differences between German and Chinese students in the magnitude of some of these phenomena, such as German students being more resistant to advice, these commonalities are interesting because they provide some preliminary evidence for the possibility that some of the core phenomena in advice taking may be qualitatively invariant between cultures.

5. General discussion

In the present research, we aimed to develop a task suited to study advice taking between different nations which can serve as a proxy for inferring cultural differences in advice taking. Developing such a task seemed necessary because previous cross-cultural comparisons of advice taking using the countries-as-proxies approach likely suffered from confounds between nationality and task difficulty because the judgment tasks used relied heavily on region-specific general knowledge. The newly developed task requires participants to estimate the annual rainfall in different cities or regions based on varying numbers of independent measurements, and they receive advice in the form of the estimate a simulated advisor made based on another set of independent measurements. The rainfall estimation task only requires secondary school-level knowledge about mathematics and measurement theory. Our first study suggests that the new task is suitable to study advice taking as participants working on the rainfall estimation task showed what we consider core phenomena of advice taking: egocentric discounting of helpful advice, sensitivity to advice quality, overutilization of useless advice, a trimodal distribution of the weights of advice, and a curvilinear relationship of advice taking and advice distance.

We then used the new task to compare some of the core phenomena of advice taking mentioned above between two countries, Germany and China, which differ in terms of the most well-studied cultural dimension: individualism versus collectivism. This comparison showed that the core phenomena emerged in both samples: participants in both countries egocentrically discounted advice, were sensitive to advice quality, showed a trimodal distribution of the weights of advice, and reacted similarly to variations in advice distance. The differences we found between the two countries were mostly quantitative in nature, that is, German participants egocentrically discounted advice more than Chinese participants, and they reacted more strongly to differences in advice quality. Most notably, German and Chinese participants differed in their preferences for the most common advice taking strategies, with Chinese participants showing a stronger preference for averaging their initial estimates and the advice, whereas German participants were more likely to retain their initial estimate and completely ignore the advice.

The differences between German and Chinese participants in the overall level of advice taking and in their choice of advice taking strategies are particularly interesting with regard to the, so far, only published study directly comparing advice taking between cultures in a countries-as-proxies approach (Mercier et al., 2012). Mercier et al. hypothesized that individuals from a collectivistic culture (Japan) should heed advice more than those from a more individualistic culture (Canada) due to a greater

https://doi.org/10.1017/jdm.2024.43 Published online by Cambridge University Press

emphasis on compromising in collectivistic cultures, but they did not find evidence supporting this hypothesis—Japanese participants heeded advice more than French participants, but not due to a greater appreciation for an averaging strategy. As we have laid out in our introduction, the best explanation is that the task Mercier et al. used, was simply more difficult for Japanese students, thus prompting greater weights of advice. In contrast, the results of our second study strongly support the hypothesis put forth by Mercier et al., and we are confident that Mercier et al. would have also found evidence for their hypothesis themselves if only their judgment task had been confound-free. This highlights how important careful task development is for cross-national studies of advice taking. In the absence of such carefully designed and (largely) confound-free tasks, we risk overlooking or misinterpreting potentially interesting cultural differences.

In a similar vein, we argued initially that meta-analytic inquiries into the association of individualism versus collectivism and advice taking may be subject to considerable noise stemming from differences between studies, for example, in terms of task difficulty, advisor expertise, or incentivization. This noise may make it difficult to detect existing cultural differences in advice taking when the database is still somewhat limited. To illustrate, consider that mean levels of advice taking can vary considerably between studies within the same culture. For example, among published advice taking studies using German samples, the average weight of advice ranges from less than 20% (Hütter and Ache, 2016, Study 1) to almost 80% (Rakoczy et al., 2015, Study 2). The within-culture variability of weights of advice is somewhat smaller when considering published studies with Chinese samples, but is still considerable, ranging from 24% (Wang and Du, 2018, Study 1) to 65% (Tinghu et al., 2018). In the light of such variability within cultures on the one hand, and the still relatively low number of published studies on advice taking in more collectivistic cultures on the other, the mean difference in advice taking we found between cultures in our second study (ca. 7 percentage points) may seem small, and meta-analyses may simply lack the power to detect them reliably within the current body of evidence.

The critical question now is how to interpret the differences observed between German and Chinese participants. The logic inherent to the countries-as-proxies approach is to attribute the differences between countries to differences in the focal cultural dimension. For example, if two countries were chosen for comparison because they differ in terms of a certain cultural dimension, usually individualism versus collectivism, then any difference arising between countries is interpreted as a correlate of that cultural dimension (Chentsova-Dutton and Vaughn, 2012; Hosni, 2020; Ji et al., 2017; Tavakoli and Tavakoli, 2010). If we apply the same logic to the results of our Study 2, we can conclude that participants from more collectivistic cultures are less prone to discount advice egocentrically than participants from more individualistic countries because they are less likely to fully ignore the advice and more likely to weight it equally to their own initial estimates. This pattern fits well with the individualism versus collectivism narrative. The higher frequency of cases in which judges completely ignored the advice in the German sample can be interpreted as an expression of greater concerns for autonomy or, put simply, an individualistic cultural norm of not conforming (Deutsch and Gerard, 1955). Similarly, the greater propensity of Chinese participants to use an averaging strategy could reflect a collectivistic cultural norm to embrace compromise, as suggested by Mercier et al. (2012).

However, we need to consider that the countries-as-proxies approach allows for alternative interpretations, especially if a study only focuses on a single cultural dimension. One simple alternative explanation is that something other than participants' background differed between the samples. The confound between task difficulty and nationality in previous cross-national studies of advice taking (Mercier et al., 2012) is a good example. Although we designed the rainfall estimation task to avoid this confound, we can never rule out other confounds in a correlational setting. We will return to this issue in the section on limitations and future directions. Importantly, even if we accept observed differences in advice to be correlates of participants' cultural background, we need to consider that they may not reflect differences in individualism versus collectivism. For example, Hofstede's framework currently comprises five additional cultural dimensions: power distance, masculinity versus femininity, uncertainty avoidance, long-term orientation, and indulgence (Hofstede et al., 2010). According to Hofstede's framework, Germany and China are very similar in terms of masculinity, long-term

https://doi.org/10.1017/jdm.2024.43 Published online by Cambridge University Press

orientation, and indulgence, but they differ with regard to both uncertainty avoidance (higher in Germany) and power distance (greater in China).

The differences in uncertainty avoidance between Germany and China allow for a plausible explanation of our findings. Uncertainty avoidance entails an aversion to ambiguous situations. Being confronted with a diverging opinion may create ambiguity, and people from countries high in uncertainty avoidance may be prompted to resolve this ambiguity by simply ignoring the other opinion. In contrast, people from countries characterized by low uncertainty avoidance might feel more comfortable with the idea that both opinions have equal merit, leading to a higher proportion of averaging.

At first glance, differences in power distance are unlikely to have had an impact on our study because there were no power or status differences between judges and advisors in our study (such differences could be easily implemented, though, if one were interested in studying power distance as a cultural correlate of advice taking). However, it is possible that differences in power distance resulted in a confound. Chinese participants might have felt motivated to work harder for the experimenter. Assuming that it is more cognitively demanding to take the mean between the initial estimate and the advice than to simply retain the initial estimate, Chinese participants might have felt more compelled to exert this extra effort than German participants. While we consider this confound relatively unlikely because of the brevity of the study and the presence of incentives for accurate final estimates, we cannot rule it out.

That said, power distance can plausibly be expected to play a role in real-life advisor–advisee relations as they are often characterized by power or status differences. Examples include people receiving advice from their parents (Goldsmith and Fitch, 1997) or managers consulting their co-workers before making a decision (See et al., 2011). It is conceivable that there are cultural norms along the dimension of power distance regarding the proper way to use advice from those of higher (or lower) status or power than oneself. For example, in societies characterized by lower power distance, decision-makers might feel more comfortable ignoring advice from a more senior colleague, and people from cultures high in power distance might feel more compelled to heed their parents' advice—even if they feel that it does not align with their own preferred choice of action.

We now turn to the commonalities we observed in both samples. First of all, even though Chinese students were generally more receptive to the advice, there was evidence of egocentric advice discounting among both German and Chinese participants. This finding was robust even when putting it to the strictest possible test by restricting the analysis to trials, in which judge and advisor had a similar information basis and it should have, thus, been transparent to judges that the optimal weight of advice was 50%. Second, judges in both samples were sensitive to advice quality, which they could infer from the size of the advisor's cue sample in each trial. The only difference here was that German students reacted somewhat more strongly to differences in advice quality. Third, even though the modal strategy of advice taking differed between German and Chinese students, the distribution of AT scores within each sample followed the typical W-shape (Soll and Larrick, 2009). Finally, we observed the classic curvilinear relationship of advice distance and advice taking in both samples. That is, both German and Chinese participants weighted the advice less when it was either very close to the initial estimate or far away from it. Taken together, these findings suggest that the commonalities in advice taking behavior between people from different countries and, arguably, different cultural backgrounds may exceed their differences, at least within the confines of the JAS as a laboratory paradigm to study advice taking.

If we contrast the differences and similarities summarized above, an interesting notion is that the behavioral patterns point toward differences of degree and not of kind. In other words, the core phenomena of advice taking we investigated in our cross-national comparison occurred in both samples. This points to an interesting possibility, namely that of cultural invariants in advice taking. Admittedly, some of these invariants may be more interesting than others. For example, it is not very surprising that people in both countries relied more on the advice when it was based on more information. It is interesting to see, however, that participants in both samples relied heavily on the choosing strategy (weighting advice either by 0% or by 100%), using it in almost 30% of the trials (27% of all trials for

https://doi.org/10.1017/jdm.2024.43 Published online by Cambridge University Press

Chinese and 29% of all trials for German participants). Choosing between the initial estimate and advice is rarely ideal, especially if it is transparent to judges that both their initial estimates and the advice are grounded in some independent information (Bednarik and Schultze, 2015). Accordingly, understanding why choosing is such a pervasive strategy—even across nations—and how to help people engage in more effective strategies such as averaging or differential weighting may be important avenues for future research.

Of course, a single study comparing only two countries based on the notion that they should differ on the individualism–collectivism dimensions hardly suffices to establish cultural invariants. All we can show here is a single failure to reject the idea of the universality of some core phenomena of advice taking in a countries-as-proxies study of advice taking. In order to entertain the idea of cultural invariants of advice taking behavior seriously, we would require a large database involving samples from many different countries that vary across multiple cultural dimensions. While this database does not exist yet, and will likely take considerable time to accumulate, we are confident that the task we developed and tested here provides researchers with a useful (and highly customizable) tool to study differences as well as commonalities in advice taking between cultures.

5.1. Limitations and directions for future research

Naturally, our research has limitations that need to be addressed. We already touched on the most obvious limitation of our study above, namely comparing samples from only two countries which differ in terms of one cultural dimension. While the cross-national comparison of Germany and China as a proxy for a cross-cultural comparison of individualism versus collectivism closely mirrors the approach of the only other published study on cultural differences in the JAS (Mercier et al., 2012), it can only be considered a first step into a more comprehensive investigation of how cultural differences are reflected in differences in advice taking.

A second limitation is that our samples in both countries were convenience samples comprised of university students, that is, participants were selected from young, urban, and highly educated subpopulations. While this is not unusual for advice taking research (or for most of social psychology, for that matter), it might have artificially limited our ability to find existing cultural differences. In other words, the fact that we found strong similarities in advice taking between the two samples might have been a direct result of similarities in participants' life circumstances. As such, it would be desirable to compare advice taking between cultures with more representative samples. As we have mentioned above, our task has the advantage that participants cannot cheat by looking up the true values. Thus, our task would be well-suited for online studies which would allow recruiting more diverse samples. One caveat is that countries may differ regarding which of their inhabitants have access to the internet. In some countries, internet access may be linked more strongly to socio-economic status, and other countries may have better internet coverage in rural areas. Researchers who want to use online studies for cross-cultural comparisons using countries-as-proxies approach should therefore be mindful of potential confounds with variables that determine internet access within the different countries.

Third, the task we used was somewhat artificial. Few people who are not trained meteorologists need to estimate precipitation in their daily lives, and if they had to do it, the available data would be more complex and more varied. When designing the task, we traded off task realism for task simplicity, opting for a task where participants did not need to know or learn target-cue relations and that could be administered to convenience samples rather than requiring experts as participants. One central advantage of our simple task is that it avoids two potential confounds when studying cultural differences in advice taking, namely cultural differences in the ability to infer cue-target relations and in the training of experts. Such differences might affect judges' actual and self-perceived performance and, ultimately, their reliance on peer advice. The downside of our task choice is that it limits the generalizability of our findings, and further studies are required to test whether greater reliance on advice and more frequent averaging can be replicated using more realistic multi-cue judgment tasks. Karelaia and Hogarth (2008) provide an excellent overview of tasks that could be used in such future research.

https://doi.org/10.1017/jdm.2024.43 Published online by Cambridge University Press

A fourth limitation is that our task avoids a confound between task difficulty and participants' national or cultural background for many but not all populations as it requires secondary school knowledge. The fact that, across both studies, only one participant failed to understand that the correct way to integrate the available measurements into an initial estimate is to compute a (weighted) average shows that the task is well-suited to study advice taking among educated participants. However, countries differ with regard to the proportion of the population that has access to the necessary education. When aiming at more representative samples for cross-cultural studies on advice taking, the inability to include less well-educated subpopulations bears the risk of introducing bias or attenuating cultural differences to the extent that education plays a role in advice taking within countries.

Fifth, we used the classic anonymous JAS in our study, which was designed to isolate informational social influence (Sniezek and Buckley, 1995). In addition, we made it transparent to participants that the advisor was computer-simulated, arguably reducing what little normative influence there is in JAS studies with anonymous human advisors even further. Therefore, we can only make statements about differences and commonalities in advice taking between German and Chinese participants in situations where there is no normative pressure to heed the advice. We already know that such normative pressures exist in Western populations. Advisors have certain ideas about how much judges should heed the advice, and violations of these expectations lead advisors to evaluate judges less positively and reduce advisors' willingness to offer advice in the future (Ache et al., 2020). It is highly likely that there are cultural differences in the social norms around advice taking, for example, along the cultural dimension of power distance, regarding to what extent one should heed advice from an equally competent but more senior advisor. Such differences might only emerge in an interactive version of the paradigm where the judge knows that the advisor can see to what extent the advice was followed. Fortunately, it is quite easy to create an interactive version of the rainfall estimation task. All one needs to do is replace the computer-simulated advisor with another participant.

Sixth, as we have briefly touched on in the introduction of Study 1, using a simulated advisor to avoid deceiving participants might have inadvertently led to participants using the advice differently from human advice. Currently, there is a debate about whether people show algorithm aversion or algorithm appreciation, that is, whether they under-rely or over-rely on it (Dietvorst et al., 2015; Logg et al., 2019; for a review, see Jussupow et al., 2020). In the JAS, evidence for algorithm aversion is scarce with most studies either showing algorithm appreciation or failing to find any differences in advice taking between human and algorithmic advice (Himmelstein and Budescu, 2023; Logg and Schlund, 2024; You et al., 2022). Importantly, algorithm appreciation does not occur when the nature of algorithm is simple and transparent. For example, Logg et al. (2019) compared advice taking between a condition in which advice stemmed from an unspecified algorithm, and a condition in which advice was generated algorithmically by averaging across a range of human judgments. Participants heeded advice less when the algorithm was averaging. Since we made the nature of our computer-simulated advice transparent and stressed that it aimed to mimic the behavior of our typical participant, we consider it unlikely that our results are influenced by algorithm aversion or algorithm appreciation. However, future studies can put this to the test by using existing human judgments, for example, those from participants in our experiments, as advice, instead of relying on algorithmic advice.

Finally, we have focused on Hofstede's framework of cultural differences (Hofstede et al., 2010), thus neglecting cultural differences not captured by that framework. One aspect that might be relevant in the context of advice taking is that Westerners and East Asians differ in the way they make social inferences about others, with East Asians relying more on situational information than on dispositional information (Norenzayan et al., 2002). This cultural difference in social inference is unlikely to have played a major role in our study because the advisor was computer-simulated, and there was only very limited information on the advisor. However, in more realistic contexts with richer information, cultural differences in social inferences might affect how judges perceive the competence and benevolence of an advisor, or how they expect an advisor to react to the advice being disregarded. This, in turn, might affect how people from different cultures heed advice, ceteris paribus. Whether differences in

https://doi.org/10.1017/jdm.2024.43 Published online by Cambridge University Press

social inferences play a role in advice taking is, again, an empirical question that can be answered by modifying the rainfall estimation task.

6. Conclusion

In the first published study on cultural differences in advice taking, Mercier et al. (2012) concluded by expressing their hope 'that cross-cultural research will become a more important part of the research on advice taking'. We wholeheartedly agree with this sentiment. As our studies have shown, there are both interesting differences and commonalities in advice taking to be explored between countries. While the countries-as-proxies approach to studying cultural differences has its limits when restricted to only a few countries and a single cultural dimension, it can be highly informative with a large data set. We hope that by introducing the rainfall estimation task as a paradigm that is easy to apply and (mostly) free of confounds between nationality and task difficulty, we can help stimulate cross-national research on advice taking to provide the basis necessary to draw firm conclusions about cultural differences in advice taking.

Data availabiliy statement. Data for both studies as well as the R scripts used to analyse the data are available at the Open Science Framework (https://osf.io/fwbau/).

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

Competing interest. The authors have no conflicts of interest to declare.

References

Ache, F., Rader, C., & Hütter, M. (2020). Advisors want their advice to be used—But not too much: An interpersonal perspective on advice taking. *Journal of Experimental Social Psychology*, *89*. https://doi.org/10.1016/j.jesp.2020.103979.

Bailey, P. E., Ebner, N. C., Moustafa, A. A., Phillips, J. R., Leon, T., & Weidemann, G. (2021). The weight of advice in older age. *Decision*, *8*, 123–132. https://doi.org/10.1037/dec0000138.

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*, *42*(28), 24516–24541. https://doi.org/10.1007/s12144-022-03573-2.

Bednarik, P., & Schultze, T. (2015). The effectiveness of imperfect weighting in advice taking. *Judgment and Decision Making*, *10*, 265–276. https://doi.org/10.1017/S1930297500004666.

Bogert, E., Schecter, A., & Watson, R. T. (2021). Humans rely more on algorithms than social influence as a task becomes more difficult. *Scientific Reports*, *11*(1), Article 1. https://doi.org/10.1038/s41598-021-87480-9.

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.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*(3), 193–217. https://doi.org/10.1037/h0047470.

Budescu, D. V., Rantilla, A. K., Yu, H.-T., & Karelitz, T. M. (2003). The effects of asymmetry among advisors on the aggregation of their opinions. *Organizational Behavior and Human Decision Processes*, *90*(1), 178–194. https://doi.org/10.1016/S0749-5978(02)00516-2.

Chentsova-Dutton, Y. E. & Vaughn, A. (2012). Let me tell you what to do: Cultural differences in advice-giving. *Journal of Cross-Cultural Psychology*, *43*(5), 687–703. https://doi.org/10.1177/0022022111402343.

Cont, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, *1*(2), 223. https://doi.org/10.1088/1469-7688/1/2/304.

Cvencek, D., Nasir, N. S., O'Connor, K., Wischnia, S., & Meltzoff, A. N. (2015). The development of math–race stereotypes: "They say Chinese people are the best at math". *Journal of Research on Adolescence*, *25*(4), 630–637. https://doi.org/10.1111/jora.12151.

Deutsch, M. & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *The Journal of Abnormal and Social Psychology*, *51*(3), 629–636. https://doi.org/10.1037/h0046408.

Dhami, M. K., Hertwig, R., & Hoffrage, U. (2004). The role of representative design in an ecological approach to cognition. *Psychological Bulletin*, *130*(6), 959–988. https://doi.org/10.1037/0033-2909.130.6.959.

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*, 114–126. https://doi.org/10.1037/xge0000033.

Du, X., Ren, Y., Wu, S., & Wu, Y. (2019). The impact of advice distance on advice taking: Evidence from an ERP study. *Neuropsychologia*, *129*, 56–64. https://doi.org/10.1016/j.neuropsychologia.2019.02.019.

Ecken, P. & Pibernik, R. (2016). Hit or miss: What leads experts to take advice for long-term judgments? *Management Science*, *62*(7), 2002–2021. https://doi.org/10.1287/mnsc.2015.2219.

Fiedler, K., Hütter, M., Schott, M., & Kutzner, F. (2019). Metacognitive myopia and the overutilization of misleading advice. *Journal of Behavioral Decision Making*, *32*(3), 317–333. https://doi.org/10.1002/bdm.2109.

Fikret Pasa, S. (2000). Leadership influence in a high power distance and collectivist culture. *Leadership & Organization Development Journal*, *21*(8), 414–426. https://doi.org/10.1108/01437730010379258.

Gelfand, M. J., Erez, M., & Aycan, Z. (2007). Cross-cultural organizational behavior. *Annual Review of Psychology*, *58*(1), 479–514. https://doi.org/10.1146/annurev.psych.58.110405.085559.

Goldsmith, D. J. & Fitch, K. (1997). The normative context of advice as social support. *Human Communication Research*, *23*, 454–476. https://doi.org/10.1111/j.1468-2958.1997.tb00406.x.

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. & Budescu, D. V. (2023). Preference for human or algorithmic forecasting advice does not predict if and how it is used. *Journal of Behavioral Decision Making*, *36*(1), e2285. https://doi.org/10.1002/bdm.2285.

Hofstede, G., Hofstede, G. J., & Minkov, M. (2010). *Cultures and organizations: Software of the mind* (3rd ed.). McGraw Hill Professional.

Hosni, H. R. (2020). Advice giving in Egyptian Arabic and American English: A cross-linguistic, cross-cultural study. *Journal of Pragmatics*, *155*, 193–212. https://doi.org/10.1016/j.pragma.2019.11.001.

Hütter, M. & Ache, F. (2016). Seeking advice: A sampling approach to advice taking. *Judgment and Decision Making*, *11*, 401–415. https://doi.org/10.1017/S193029750000382X.

Ji, L.-J., Zhang, N., Li, Y., Zhang, Z., Harper, G., Khei, M., & Li, J. (2017). Cultural variations in reasons for advice seeking. *Journal of Behavioral Decision Making*, *30*(3), 708–718. https://doi.org/10.1002/bdm.1995.

Jussupow, E., Benbasat, I., & Heinzl, A. (2020, June 15). *Why are we averse towards algorithms? A comprehensive literature review on algorithm aversion*. 28th European Conference on Information Systems - Liberty, Equality, and Fraternity in a Digitizing World (ECIS 2020), Atlanta. https://tubiblio.ulb.tu-darmstadt.de/138565/

Karelaia, N. & Hogarth, R. M. (2008). Determinants of linear judgment: A meta-analysis of lens model studies. *Psychological Bulletin*, *134*(3), 404–426. https://doi.org/10.1037/0033-2909.134.3.404.

Kausel, E. E., Culbertson, S. S., Leiva, P. I., Slaughter, J. E., & Jackson, A. T. (2015). Too arrogant for their own good? Why and when narcissists dismiss advice. *Organizational Behavior and Human Decision Processes*, *131*, 33–50. https://doi.org/10.1016/j.obhdp.2015.07.006.

Kim, H. Y., Lee, Y. S., & Jun, D. B. (2020). Individual and group advice taking in judgmental forecasting: Is group forecasting superior to individual forecasting? *Journal of Behavioral Decision Making*, *33*(3), 287–303. https://doi.org/10.1002/bdm.2158.

Lim, J. S. & O'Connor, M. (1995). Judgemental adjustment of initial forecasts: Its effectiveness and biases. *Journal of Behavioral Decision Making*, *8*(3), 149–168. https://doi.org/10.1002/bdm.3960080302.

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.

Logg, J. M. & Schlund, R. (2024). A simple explanation reconciles "algorithm aversion" and "algorithm appreciation": Hypotheticals vs. real judgments. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4687557.

Markus, H. R. & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, *98*(2), 224–253. https://doi.org/10.1037/0033-295X.98.2.224.

Mercier, H., Yama, H., Kawasaki, Y., Adachi, K., & Henst, J.-B. V. der. (2012). Is the use of averaging in advice taking modulated by culture? *Journal of Cognition and Culture*, *12*(1–2), 1–16. https://doi.org/10.1163/156853712X633893.

Moussaïd, M., Kämmer, J. E., Analytis, P. P., & Neth, H. (2013). Social influence and the collective dynamics of opinion formation. *PLOS ONE*, *8*(11), e78433. https://doi.org/10.1371/journal.pone.0078433.

Norenzayan, A., Choi, I., & Nisbett, R. E. (2002). Cultural similarities and differences in social inference: Evidence from behavioral predictions and lay theories of behavior. *Personality and Social Psychology Bulletin*, *28*(1), 109–120. https://doi.org/10.1177/0146167202281010.

OECD. (2023). *PISA 2022 results (volume I): The state of learning and equity in education*. OECD. https://doi.org/10.1787/53f23881-en.

Önkal, D., Goodwin, P., Thomson, M., Gönül, S., & Pollock, A. (2009). The relative influence of advice from human experts and statistical methods on forecast adjustments. *Journal of Behavioral Decision Making*, *22*(4), 390–409. https://doi.org/10.1002/bdm.637.

Prahl, A. & Van Swol, L. (2017). Understanding algorithm aversion: When is advice from automation discounted? *Journal of Forecasting*, *36*(6), 691–702. https://doi.org/10.1002/for.2464.

Rakoczy, H., Ehrling, C., Harris, P. L., & Schultze, T. (2015). Young children heed advice selectively. *Journal of Experimental Child Psychology*, *138*, 71–87. https://doi.org/10.1016/j.jecp.2015.04.007.

Schermerhorn, J. R. & Harris Bond, M. (1997). Cross-cultural leadership dynamics in collectivism and high power distance settings. *Leadership & Organization Development Journal*, *18*(4), 187–193. https://doi.org/10.1108/01437739710182287.

Schmidt, F. L. & Hunter, J. E. (2015). *Methods of meta-analysis: Correcting error and bias in research findings*. SAGE Publications, Ltd. https://doi.org/10.4135/9781483398105.

https://doi.org/10.1017/jdm.2024.43 Published online by Cambridge University Press

Schultze, T., Gerlach, T. M., & Rittich, J. C. (2018). Some people heed advice less than others: Agency (but not communion) predicts advice taking. *Journal of Behavioral Decision Making*, *31*(3), 430–445.

Schultze, T., Mojzisch, A., & Schulz-Hardt, S. (2017). On the inability to ignore useless advice. *Experimental Psychology*, *64*(3), 170–183. https://doi.org/10.1027/1618-3169/a000361.

Schultze, T., Rakotoarisoa, A.-F., & Stefan, S.-H. (2015). Effects of distance between initial estimates and advice on advice utilization. *Judgment and Decision Making*, *10*(2), 144–171. https://doi.org/10.1017/S1930297500003922.

See, K. E., Morrison, E. W., Rothman, N. B., & Soll, J. B. (2011). The detrimental effects of power on confidence, advice taking, and accuracy. *Organizational Behavior and Human Decision Processes*, *116*(2), 272–285. https://doi.org/10.1016/j.obhdp.2011.07.006.

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. https://doi.org/10.1006/obhd.1995.1040.

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*, 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.

Tavakoli, M. & Tavakoli, A. (2010). A cross-cultural study of advice and social pressure. *Procedia – Social and Behavioral Sciences*, *5*, 1533–1539. https://doi.org/10.1016/j.sbspro.2010.07.321.

Tinghu, K., Li, W., Peiling, X., & Qian, P. (2018). Taking advice for vocational decisions: Regulatory fit effects. *Journal of Pacific Rim Psychology*, *12*, e7. https://doi.org/10.1017/prp.2017.12.

Treffenstaedt, Christian & Wiemann, Paul. (2018). *Alfred—A library for rapid experiment development* (version v0.2b5) [Computer software]. Zenodo. https://doi.org/10.5281/ZENODO.1437220.

Tzini, K. & Jain, K. (2018). The role of anticipated regret in advice taking. *Journal of Behavioral Decision Making*, *31*(1), 74–86. https://doi.org/10.1002/bdm.2048.

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 Swol, L. M. (2009). The effects of confidence and advisor motives on advice utilization. *Communication Research*, *36*(6), 857–873.

Wang, X. & Du, X. (2018). Why does advice discounting occur? The combined roles of confidence and trust. *Frontiers in Psychology*, 9. https://www.frontiersin.org/articles/10.3389/fpsyg.2018.02381

Wright, G. N., Phillips, L. D., Whalley, P. C., Choo, G. T., Ng, K.-O., Tan, I., & Wisudha, A. (1978). Cultural differences in probabilistic thinking. *Journal of Cross-Cultural Psychology*, *9*(3), 285–299. https://doi.org/10.1177/002202217893002.

Yaniv, I. (2004). Receiving other people's advice: Influence and benefit. *Organizational Behavior and Human Decision Processes*, *93*(1), 1–13. https://doi.org/10.1016/j.obhdp.2003.08.002.

Yaniv, I. & Choshen-Hillel, S. (2012). Exploiting the wisdom of others to make better decisions: Suspending judgment reduces egocentrism and increases accuracy. *Journal of Behavioral Decision Making*, *25*(5), 427–434. https://doi.org/10.1002/bdm.740.

Yaniv, I. & Kleinberger, E. (2000). Advice taking in decision making: Egocentric discounting and reputation formation. *Organizational Behavior and Human Decision Processes*, *83*(2), 260–281. https://doi.org/10.1006/obhd.2000.2909.

Yaniv, I. & Milyavsky, M. (2007). Using advice from multiple sources to revise and improve judgments. *Organizational Behavior and Human Decision Processes*, *103*(1), 104–120. https://doi.org/10.1016/j.obhdp.2006.05.006.

Yates, J. F. & de Oliveira, S. (2016). Culture and decision making. *Organizational Behavior and Human Decision Processes*, *136*, 106–118. https://doi.org/10.1016/j.obhdp.2016.05.003.

Yates, J. F., Lee, J.-W., Shinotsuka, H., Patalano, A. L., & Sieck, W. R. (1998). Cross-cultural variations in probability judgment accuracy: Beyond general knowledge overconfidence? *Organizational Behavior and Human Decision Processes*, *74*(2), 89–117. https://doi.org/10.1006/obhd.1998.2771.

You, S., Yang, C. L., & Li, X. (2022). Algorithmic versus human advice: Does presenting prediction performance matter for algorithm appreciation? *Journal of Management Information Systems*, *39*(2), 336–365. https://doi.org/10.1080/07421222.2022.2063553.

Cite this article: Schultze, T. and Chen, Z. (2025). The development of a task to study advice taking across nations and its application in a China-Germany comparison. *Judgment and Decision Making*, e8. https://doi.org/10.1017/jdm.2024.43