GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Do We Prefer Consensual Advice – Even When It is Detrimental to Our Judgment Quality?

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades

"Doctor rerum naturalium"

der Georg-August-Universität Göttingen

im Promotionsprogramm Biologie

der Georg-August University School of Science (GAUSS)

vorgelegt von

Stella Katherina Wanzel,

geb. in Heidelberg

Göttingen, 2017

Betreuungsausschuss

Prof. Dr. Stefan Schulz-Hardt, Abteilung Wirtschafts- und Sozialpsychologie, Georg-Elias-Müller-Institut für Psychologie

Prof. Dr. Michael Waldmann, Abteilung Kognitionswissenschaft und Entscheidungspsychologie, Georg-Elias-Müller-Institut für Psychologie

Mitglieder der Prüfungskommission

Referent: Prof. Dr. Stefan Schulz-Hardt, Abteilung Wirtschafts- und Sozialpsychologie, Georg-Elias-Müller-Institut für Psychologie

Korreferent: Prof. Dr. Michael Waldmann, Abteilung Kognitionswissenschaft und Entscheidungspsychologie, Georg-Elias-Müller-Institut für Psychologie

Weitere Mitglieder der Prüfungskommission:

Prof. Dr. Margarete Boos, Abteilung Sozial- und Kommunikationspsychologie, Georg-Elias-Müller-Institut für Psychologie

Prof. Dr. York Hagmayer, Abteilung Kognitionswissenschaft und Entscheidungspsychologie, Georg-Elias-Müller-Institut für Psychologie

Prof. Dr. Uwe Mattler, Abteilung Kognitionswissenschaften, Georg-Elias-Müller-Institut für Psychologie

Prof. Dr. Annekathrin Schacht, Abteilung Affektive Neurowissenschaft und Psychophysiologie, Georg-Elias-Müller-Institut für Psychologie

Tag der mündlichen Prüfung: 11.12.2017

Danksagung

Viele Menschen haben mich während des gesamten Prozesses meines Dissertationsprojekts begleitet und den Abschluss dieses Projekts durch ihre fachliche und emotionale Unterstützung überhaupt erst möglich gemacht. Im Folgenden möchte ich diesen Menschen danken.

An erster Stelle gilt dabei mein Dank meinem Doktorvater Stefan Schulz-Hardt, der es mir ermöglicht hat, dieses Dissertationsprojekt umzusetzen. Seine konstruktiven und hilfreichen Ratschläge und Anregungen in jeder Phase meiner Promotion waren stets eine wertvolle Bereicherung. Seine ausgeprägte fachliche Kompetenz und seine strukturierte, immer wissenschaftlich präzise Arbeitsweise haben mich von Anfang an beeindruckt.

Danken möchte ich außerdem Michael Waldmann dafür, dass er sich bereit erklärt hat, mein Dissertationsprojekt als Zweitgutachter zu begleiten. Weiterhin bedanke ich mich bei Margarete Boos, York Hagmeyer, Uwe Mattler und Annekathrin Schacht dafür, dass sie sich die Zeit genommen haben, in meinem Prüfungskomitee mitzuwirken.

Thomas Schultze-Gerlach danke ich dafür, dass er jederzeit bereit war, mich zu beraten und mich mit seinem ausgeprägten statistischen Wissen dabei zu unterstützen, auch komplexere Probleme zu lösen. Christian Treffenstädt möchte ich dafür danken, dass er mir bei der technischen Umsetzung meiner Experimente eine große Hilfe war.

Annika Giersiepen gilt mein Dank, da sie mir in den letzten Jahren sowohl bei fachlichen Fragen immer geduldig weitergeholfen hat, als auch immer bereit war, mich zu den teilweise notwendigen Ablenkungen in Form von Kaffee-/Tee-/Kuchen-Trips zu „Maria“ zu begleiten.

Von ganzem Herzen möchte ich mich außerdem bei meinen Eltern bedanken. Vom Beginn des Projekts an bis zu dessen Abschluss haben sie mich stets bedingungslos unterstützt, mich aufgemuntert und immer an mich geglaubt. Zu jeder Zeit konnte ich mich darauf verlassen, dass ich bei ihnen auf offene Ohren stoße und sie jede meiner Entscheidungen respektieren.

Nicht zuletzt gilt mein Dank meinem Freund Hendrik, der mich in den vergangenen drei Jahren in jeglicher Hinsicht unterstützt und mir zur Seite gestanden, und, um es mit den Worten Grönemeyers zu sagen, jeden Verdruss ins Gegenteil verkehrt hat. Seine unendliche Geduld und sein Zuspruch haben Vieles leichter gemacht. Ihm und meinen Eltern ist diese Arbeit gewidmet.

Göttingen, im Oktober 2017 *Stella Wanzel*

Table of Contents

1 Introduction ... 1

1.1 The Judge-Advisor-System ... 1

1.2 Motivational Aspects of Advice Seeking and Taking ... 2

1.3 Combining Independent versus Dependent opinions ... 3

1.4 Why Individuals May Favor Dependent Opinions ... 5

2 Summary of Manuscript 1: Disentangling the Effects of Advisor Consensus and Advice Proximity ... 9

3 Summary of Manuscript 2: Do Individuals Have a General Preference for Dependent Advice? – Pitting Advisor Accuracy Against Advisor Consensus ... 14

4 General Discussion ... 19

5 Conclusion ... 23

6 References ... 24

Appendix A: Manuscript 1

Disentangling the Effects of Advisor Consensus and Advice Proximity

Appendix B: Manuscript 2

Do Individuals Have a General Preference for Dependent Advice? – Pitting Advisor Accuracy Against Advisor Consensus

Appendix C: Supplementary Study Materials & Additional Results Manuscript 2

Appendix D: Curriculum Vitae

1 Introduction

1.1 The Judge-Advisor-System

In the real world, we often encounter situations in which we seek the advice of other individuals instead of making a judgment or decision by ourselves. Importantly, as Heath and Gonzalez (1995) point out, most of our relevant decisions are made after we have consulted (at least) one other person. For instance, we may solicit opinions from others when we are planning to make a big investment, when we have to decide if we want to undergo a risky surgery, and in many more situations. Hence, advice seeking is prevalent in many domains, such as matters of taste, medical decisions, etc. Soliciting opinions from others, however, is also involved when making numeric judgments, such as stock or weather forecasts, or judgments regarding legal matters. In our daily lives, we may seek advice regarding numeric judgments when we want to sell something on, for example, the online marketplace Ebay and need to determine how much money we can obtain for our item. Recognizing the importance of this aspect of social life, a large amount of research has directed its focus on the social context of judgment and decision making in the last twenty years, primarily investigating advice giving and advice taking (e.g., Dalal & Bonaccio, 2010; Gino, 2008; Harvey & Fischer, 1997; Yaniv, 1997; Yaniv, Choshen-Hillel, & Milyavsky, 2011, to name only a few). The Judge-Advisor System (JAS; Sniezek and Buckley (1995)), which usually involves the role of one judge and at least one advisor, is a helpful tool to study the social context of advice giving. In a prototypical JAS, the participant – generally the judge – works on qualitative choice or quantitative judgment tasks. He or she makes an initial decision or judgment and is aided by one or more advisors in his or her final choice or judgment. Numerous studies have employed the JAS, often to investigate the amount of advice utilization individuals demonstrate after receiving the opinions of others (for a review, see Bonaccio & Dalal, 2006). Advice utilization describes how much the judge adjusts his or her final estimate in the direction of the advice. Although receiving advice has proven to be beneficial for the judge's accuracy (e.g., Harvey & Fischer, 1997; Yaniv, 2004a; 2004b), individuals often discount it to a high extent (Harvey & Fischer, 1997; Yaniv & Kleinberger, 2000). To study the amount of advice utilization in numerical tasks, Harvey and Fischer (1997) developed a formula which defines advice utilization as a ratio between the adjustment of the judge's initial estimate to its final estimate and the adjustment of the judge's initial estimate to the advice:

$$\text{advice taking} = \frac{\text{judge final estimate} - \text{judge initial estimate}}{\text{advisor recommendation} - \text{judge initial estimate}}$$

With the help of this formula, we can study the amount of advice taking as well as possible influences on weighting of the advice that might help to explain individuals' advice taking behavior.

1.2 Motivational Aspects of Advice Seeking and Taking

Informational gain. Why do individuals seek and take advice? The motives are manifold. Next to social reasons, like distributing the responsibility for a decision or judgment across several individuals (Harvey & Fischer, 1997), the uncertainty arising from missing information and the lack of clear directions on how to proceed in real-life judgment and decision tasks make seeking out advice particularly important. Thus, filling in the missing pieces of information is usually one motivation behind consulting others during judgment and decision making (Budescu & Rantilla, 2000; Valley, White, Neale, & Bazerman, 1992).

Accuracy gains. A major motive, which is central to my research question, is to improve the accuracy of one's own judgment. As already noted above, it has been widely shown that heeding advice produces accuracy gains (e.g. Harvey & Fischer, 1997; Johnson, Budescu, & Wallsten, 2001; Yaniv, 1997). This improvement is especially large when combining the opinions of more than one advisor (cf. Budescu & Yu, 2007). This is because aggregation reduces random errors – at least, when the opinions are independent (Herzog & Hertwig, 2009).

1.3 Combining Independent versus Dependent opinions

Importantly, combining multiple opinions entails one particular hazard, however: these opinions might not be independent; in fact, in many instances, they are interdependent to some extent. In this dissertation, interdependence is used to describe the amount of positive correlation of the biases among a set of opinions. Hence, interdependence can be defined by the following statement: knowing the opinion of Person A helps individuals to guess the opinion of Person B (see also Yaniv, Choshen-Hillel, & Milyavsky, 2009). This inter-correlation can stem from various sources: individuals may use the same methods to interpret the information at hand, they may influence each other, or they may share the same information. All these situations eventually result in correlated opinions. Interdependence does not necessarily have detrimental effects on the accuracy of the judge's estimates, as individuals influencing each other through discussion may actually result in more accurate judgments (Schultze, Mojzisch, & Schulz-Hardt, 2012). However, in several instances interdependence between advisory opinions does indeed affect judgments negatively. To understand this problem, we first have to take a look at an estimate itself and the parts it consists of. According to Yaniv (2004b), a quantitative estimate consists of the true value, an unsystematic error, and a systematic bias, that is, the idiosyncratic under- or overestimation of the true value. For dependent opinions, the idiosyncratic biases are correlated. As a result, these opinions exhibit higher levels of consensus than independent opinions, because they co-vary regarding the true value as well as part of their idiosyncratic biases. Furthermore, when the biases are positively correlated, interdependence is detrimental to the advisor's (and, if the advice is taken into account, also the judge's) accuracy. The reason is that, contrary to independent opinions for which unsystematic errors and systematic biases cancel each other out during aggregation, only unsystematic errors are reduced when aggregating dependent opinions. Consequently, dependent opinions are (on average) less accurate than independent opinions.

This characteristic of interdependent opinions should not prove problematic as long as individuals seek out the more accurate advice and take it into account more. Therefore, it is important to determine whether or not individuals actually do identify the type of advice that is more accurate, especially when it goes along with lower levels of consensus among the advisory opinions. How do individuals deal with interdependence among opinions? Do they understand the possible implications of interdependence, and how do they determine what kind of opinions (dependent or independent) to take into account? In fact, research suggests that individuals tend to rely on consensus as a signal for accuracy (e.g., Chaiken & Stangor, 1987). This heuristic may serve them well as long as consensus is indeed indicative of accuracy (which can often be the case); however, when higher agreement stems from correlated biases, consensus is misleading. Furthermore, individuals apparently have difficulties understanding the concept of correlated errors and do not necessarily take meta-information about the correlation of errors into account (e.g., Maines, 1996). In the following, I will address relevant findings regarding the before-mentioned topics.

1.4 Why Individuals May Favor Dependent Opinions

Preference for consensus information. In a rather unpredictable world that is full of uncertainties, paying attention to consensus information may serve as way to gain some stability. When individuals confirm each other's opinions, this can imply that these opinions are correct, that the individuals have common knowledge of the truth. In fact, according to Chaiken and Stangor (1987), individuals seem to be guided by a *consensus implies correctness heuristic*. More precisely, when judging how valid a message is, individuals agree more with messages that other individuals agree with, without investing a lot of cognitive effort (cf. p. 599). Lopes, Vala, and Garcia-Marques (2007) indeed found that individuals attributed higher validity to the opinions of a group who displayed higher consensus than a comparison group. In this context, validity was measured via an index of, on the one hand, certainty attributed to the group members regarding their ideas and, on the other hand, the ascribed credibility of those ideas.

In the domain of judgments, Sniezek and Kolzow (1994; cited in Savadori, van Swol, & Sniezek, 2001) demonstrated in an unpublished study that individuals display reduced levels of confidence when they experience disagreement between the opinions of others, suggesting that individuals' confidence levels would benefit from the agreement. In addition, Sniezek and Buckley (1995) were able to demonstrate that conflict between advisors' opinions in a JAS does not increase confidence of the judge, whereas the absence of conflict bolsters judges' confidence. These findings also imply that individuals follow the heuristic that consensus implies validity. Since, in many cases, agreement among opinions may indeed be indicative of accuracy, using consensus as a cue is not always ill-advised. However, as soon as the consensus is only illusory or spurious (see Yaniv et al. 2009) in that it may, for example, stem from correlated biases, it cannot be deemed as valid, thus making it an inadequate cue for accuracy. As a matter of fact, individuals are indeed more confident after receiving agreeing opinions, even when the agreement results from correlated biases (Budescu and Rantilla, 2000).

Bearing the findings cited above in mind, I assume that, in a JAS, individuals show a preference for dependent over independent opinions because, on average, they tend to agree more (i.e., have higher consensus) than the latter. In the next section, I address the question of whether individuals have general knowledge about correlated errors and how they deal with this information.

Difficulties in understanding correlated errors. Several studies suggest that individuals do not fully understand the concept of correlated errors or, at least, that they apparently do not know how to use information regarding the correlation of errors correctly. For example, Larrick and Soll (2006) found that participants underestimate the value of averaging opinions. Averaging has been widely shown to reduce judgment errors (e.g., Ariely et al., 2000; Johnson et al., 2001). However, Larrick and Soll showed that participants misjudged that averaging the estimates of multiple judges outperforms (in terms of accuracy) the estimate of the average judge. In fact, in their first experiment, the authors presented participants with performance summaries of two judges and asked them to estimate how well a certain set of strategies (for using the forecasts of the two judges) would perform. Among these strategies were averaging the forecasts of the two judges, picking the forecasts of one judge over the other or deciding which to choose based on the judges' confidence levels. The authors found that 57% of the participants rated averaging as no better than the performance of the average judge. This misappreciation was reduced when the bracketing rate (i.e., the bracketing of the true value between the two judgments) was increased. Since bracketing rates are lower in the case of shared biases or positively correlated random errors (see Larrick & Soll, 2006) the finding suggests that individuals may have difficulties understanding the implications of averaging independent compared to dependent opinions, namely reducing random as well as systematic errors instead of reducing only random errors. While Larrick and Soll’s results are merely suggestive of individuals’ difficulty to fully comprehend the concept of correlated errors, a study by Maines (1996) provides additional insights. To be exact, participants were presented with past sales forecasts of three analysts, the respective actual sales and the signed forecast errors. Furthermore, Maines gave her participants precise information regarding the interdependence of forecasters by providing them with the correlation between the forecast errors. Participants either knew that all pairwise correlations between the forecast errors were close to zero (independent condition) or that two of the three pairwise correlations were close to zero, but one was .97 (dependent condition). However, the results showed that participants’ combined estimates were insensitive to differences in interdependence of the forecasters. Furthermore, Maines states that participants’ written comments revealed that half of those who mentioned interdependence either did not use this information correctly, or they did not know how to use it.

Interestingly, Soll (1999) found that individuals basically understand how biases can be reduced but, unfortunately, they perceive a trade-off where there is none. In fact, they correctly assume that aggregating dependent sources only reduces bias, but also falsely assume that aggregating independent sources only reduces unsystematic errors. Therefore, they might prefer one source over the other depending on their idea of which kind of error contributes more to total error (cf. p.323). In sum, these studies suggest that it is not a given that individuals know how to deal with one specific characteristic of dependent opinions – correlated errors. This may make it particularly difficult for them to correctly asses the value of consensus as a cue for validity, especially when this consensus stems from correlated biases.

As stated above, the cited studies suggest that individuals may have a preference for agreement among opinions and, furthermore, may not realize whether this agreement arises from correlated errors or, at least, may not know how to use this information. I believe that, in combination, these factors may lead participants to favor dependent over independent opinions, even when the former are less accurate. Only a few studies have touched this issue by demonstrating that individuals display higher confidence levels when presented with interdependent opinions (e.g., Budescu & Yu, 2007; Kahneman & Tversky, 1973). So far, no study has measured participants' preference for dependent opinions directly by investigating which type of advice they would choose when asked to do so, and which one they would weight more strongly. The studies described in the next sections were developed to close this research gap. In my opinion, addressing this issue is important because, as explained above, soliciting advice is an everyday occurrence, which makes it especially essential to find out whether individuals use good strategies – and, if they do not, at long last to aid them in that matter. In the first manuscript, we investigated whether and, if so, what kind of influence advisor interdependence has on individuals' advice taking behavior. Since, in the only study which has investigated advisor interdependence in a JAS (Yaniv, et al., 2009) the effects of social validation and advisor interdependence were confounded, we separated the effects of advisor consensus and proximity of the advice to the judge's own estimate. By doing so, we were able to investigate whether there was a unique effect of interdependence on participants' weighting and adjustment rates. With our second manuscript, we deepened our understanding of individuals' preference for interdependence by exploring their (advice) choosing and weighting behavior in a context where a preference for dependent advice is beneficial to the participants' own judgment accuracy as well as in a context where choosing dependent advice and weighting it more strongly is detrimental to their own judgment.

2 Summary of Manuscript 1: Disentangling the Effects of Advisor Consensus and Advice Proximity

Wanzel, S. K., Schultze, T., & Schulz-Hardt (2017), Journal of Experimental Psychology: Learning, Memory, and Cognition, 43, 1669-1675.

In our first experiment, we pursued two main goals. First, we wanted to remove a confound we detected in a study by Yaniv et al. (2009). To my knowledge, this is the only study to investigate dependent versus independent advice in the context of the JAS. The authors found that individuals were less accurate after receiving dependent (compared to independent) advice but, interestingly, also more confident in their judgments. Furthermore, they revised their opinion more often in the independent condition. Yaniv and colleagues concluded that their participants placed too much weight on the spurious consensus that accompanies dependent advice while disregarding the informational value of independent advice. This conclusion may be premature, however. The authors manipulated interdependence by drawing three pieces of advice that were close to the judge's initial estimate from a pool of 100 estimates (the three opinions were 1st, 7th and 15th closest to the initial estimate). As we explain in the manuscript, by doing this they not only manipulated interdependence but also the amount of social validation participants experienced. Social validation means that individuals receive support for their opinions (e.g., Mojzisch, Schulz-Hardt, Kerschreiter, Brodbeck, & Frey, 2008; see also Schultz-Hardt, Frey, Lüthgens, & Moscovici, 2000) and constitutes an alternative explanation for the dissociation of accuracy and confidence found by Yaniv and colleagues. For instance, Schultze, Rakotoarisoa and Schulz-Hardt (2015) showed that individuals display higher levels of confidence and adjust their judgments less frequently when the advice is largely similar to their own estimate. In the study by Yaniv and colleagues, dependent advice was close to the judge's initial estimate, while the randomly drawn independent advice was more distant from it. Therefore, we aimed to investigate if advice proximity rather than interdependence caused the effects found by Yaniv and colleagues or whether both factors are at play in explaining the effects.

In order to separate distance from consensus effects, we replicated the experiment by Yaniv and colleagues and added a third condition. In this third condition, the three pieces of advice were dependent, but far from the judge's initial estimate. This was achieved by calculating an interval of one standard deviation of all estimates in the advisor pool and then drawing the three pieces that were 1^{st}, 7^{th} and 15^{th} closest to either limit of this interval. As in the study by Yaniv and colleagues, we explicitly informed participants in each condition about how the advice was sampled. Additionally, the experimental procedure was the same as in the original study where participants made 30 calorie judgments for various foods in a JAS and were presented with the three different conditions in random order in a within-subject manner.

Regarding accuracy, we found the same result pattern as Yaniv and colleagues. Participants achieved lower accuracy gains after receiving dependent advice – independent of its proximity to the judge's initial estimate. Consequently, the authors correctly attributed their accuracy results to the level of advisor consensus. More interesting were the outcomes for confidence and adjustment. We found that, while participants indeed displayed higher confidence gains after receiving dependent and simultaneously close advice than after receiving independent advice, they displayed the lowest (and at the same time not significant) confidence gains in the condition where advice was dependent but also far from the judge' first estimate. This leads us to two conclusions. First, since, when comparing the three conditions, we found the lowest confidence gains in the one condition where the advice was far from the judge's own estimate and at the same time more consistent (than independent advice), it probably was advice proximity rather than interdependence that caused the confidence findings in the original study. This is in line with research demonstrating that agreement of others with one's own opinion makes individuals more confident, as well as with research showing that disagreement results in lower confidence (Festinger, Gerard, Hymovitch, Kelley, and Raven, 1952; Schultze et al., 2015). Second, agreement among advisors does not necessarily make judges more confident, at least when it does not simultaneously support the judge's own opinion.

Regarding adjustment rates and weighting, we found that participants changed their estimates most frequently in the dependent/far condition (compared to the other two conditions) and also weighted the advice most in this condition. They changed their estimates least frequently and weighted the advice the least in the dependent/close condition. These findings also suggest that advice proximity, not interdependence, explains why participants in the original study adjusted their estimates more in the independent condition. However, consensus also plays a role in explaining our weighting effects. The fact that, considering all three conditions, participants weighted the advice most in the dependent/far and least in the dependent/close condition shows that it is particularly likely for participants to adjust their estimates when the advice not only seems to confirm their own judgment but is also more consistent (i.e., has higher consensus). This result underlines that individuals overweight the value of consensus as a cue for accuracy. Therefore, we can add to Chaiken and Stangor's (1987) theory by providing evidence that individuals rely on consensus cues even when there are precise hints that the consensus results from poor sampling of the advice.

Nevertheless, the result that the advice was weighted the least in the dependent/close condition renders us optimistic insofar as individuals apparently realize that there is no benefit in weighting opinions that already support one's own. Of course, this could merely be due to our measure of advice taking. As it stands, when the judge's own initial opinion is already similar to the advice, there is not much room for adjustments. However, our finding is also in line with research showing that individuals acknowledge the value of opinions or information that is different from one's own (Gonzalez, 1994; Van Swol, 2009; Van Swol & Ludutsky, 2007). In light of individuals' motivation to receive alternative information (Heath & Gonzalez, 1995) or information they do not already possess themselves (Budescu & Rantilla, 2000; Valley et al. 1992), the higher weighting of dependent/far compared to dependent/close advice makes sense.

Taking all results into account, an important implication is that individuals do not sufficiently consider the meta-information they receive. Although they were explicitly told how the advice was sampled, they were apparently unable or not willing to take this information sufficiently into account. This finding is also in line with results of Maines (1996) who found that individuals did not know how to make use of the information that the errors of their advisors were correlated. Therefore, it is crucial to aid individuals in developing strategies to overcome this challenge.

In sum, with our first manuscript we were able to show that individuals have a preference for dependent advice by weighting it more and adjusting their judgments more often, at least when the advice is far from their own judgment. It is important to bear in mind, however, that the advice's interdependence was manipulated as a function of its proximity to the judge's own estimate and was not a result of realistic circumstances. Therefore, as in the original study by Yaniv and colleagues, the operationalization of interdependence was considerably artificial. This raises the question of whether individuals might be better able to detect interdependence and correctly take its implications into account when interdependence arises more realistically, for example, from advisors influencing each other through discussion. Furthermore, since we only investigated participants' advice taking behavior we do not know what happens when they have the opportunity to decide which type of advice (dependent or independent) they want to receive, before making their own judgments. These issues will be addressed in our second manuscript.

3 Summary of Manuscript 2: Do Individuals Have a General Preference for Dependent Advice? – Pitting Advisor Accuracy Against Advisor Consensus

With the present manuscript, we pursued several goals. First, we wanted to investigate individuals' preference for dependent versus independent advice in more realistic settings. As stated above, the operationalization of interdependence in our first study was somewhat artificial. In the studies reported in Manuscript 2, we used two different operationalizations of interdependent advice. As described before, among others, interdependence may arise from individuals influencing each other or from individuals sharing information. Therefore, in two of our studies, dependent advisors – who were participants in a pre-test designed to generate the advice – discussed their estimates before making an individual judgment, while independent advisors worked alone. Through the discussion, individuals automatically influence each other, resulting in correlated opinions. In the three other studies in this manuscript, we used a different, but also realistic manipulation of dependence among advisors, namely the amount of information the advisors shared. In order to do this, we employed a paradigm developed by Schultze (2015) in which advice is simulated by the computer. In this paradigm, participants take the role of meteorologists and estimate the precipitation amounts of cities in Asia. They are assisted by advisors who either share two of three weather stations aiding them in their estimates (dependent advisors) or those who have access to three unique weather stations (independent advisors). Sharing information in this way should also make the advice more interdependent.

By manipulating interdependence using the methods described above, we also realized another very important goal, namely to investigate individuals' preference for dependent versus independent advice both in a setting where dependent advice is more accurate and in a setting where it is less accurate than independent advice. This way, we were able to determine whether participants act rationally by only preferring dependent advice when it proves beneficial for their own outcome or if they have a general preference for dependent advice, independent of its accuracy. In fact, letting advisors discuss their estimates should not only make their estimates interdependent and consensual, but should also produce more accurate estimates. The reason is that, through discussion, individuals can exchange information not known to all members prior to discussion and also have the possibility to correct each other (e.g., Schultze, Mojzisch, & Schulz-Hardt, 2012). On the other hand, when advisors have overlapping information, as was the case in the scenario with shared versus unshared weather stations, this should result in less accurate judgments, since advisors in this case also share their biases. We hypothesized that individuals would display a preference for dependent advice, irrespective of the setting, since they would primarily focus on consensus cues, and these consensus cues were present in both types of dependent advice.

Third, we wanted to investigate individuals' preference for dependent versus independent advice in these two settings via two different measures. First, we were interested in finding out which type of advice participants would choose if they were given the opportunity to do so. Therefore, in both contexts, we presented participants with explicit descriptions of how the advice was sampled, and then asked them to actively choose one set of advisors before working on the task. Second, as in Manuscript 1, we measured participants' amount of weighting of dependent and independent advice. We believe that, if participants are distracted by consensus cues, a preference for dependent advice should also be demonstrated by weighting it more strongly, especially since in this type of experiment individuals are not only presented with the descriptions of advice sampling but also see the advice and, thus, can detect its level of consensus.

As a fourth goal, we wanted to collect initial evidence that would help us to resolve the question of why individuals prefer dependent advice. To this end, one of our choosing studies (with less accurate dependent advice) was designed to investigate whether individuals are indeed distracted by consensus information. We presented participants with example values of both types of advice before they made a decision. Half of the participants were also shown the true value, thereby giving them the possibility to compare the advice regarding its accuracy. In our last study with the same scenario, on the other hand, participants did not see any example values and could, therefore, not detect the advice's similarity. A change in participants' preference from dependent to independent advice in these two studies would thus suggest that they indeed have a tendency to rely on consensus cues (as long as they are retrievable from the data) when making their decisions.

Surprisingly, the results only partly confirmed our expectations. Participants indeed consistently preferred dependent advice in our weighting studies, by weighting it more strongly independent of its level of accuracy. We are given a different picture, however, when we look at the studies, where participants were given the opportunity to choose between dependent and independent advice. In these studies, participants in fact always chose the more accurate advice. This was the case even when participants were presented with example values of dependent and independent advice. The fact that participants prefer dependent advice throughout the weighting studies suggests that individuals might be more prone to consensus cues when this type of measurement is used compared to when preference is operationalized through active choosing. Why is that? It should be noted that the measurement of preference is structurally very different in these two types of studies. In fact, our participants chose the advice more or less directly after having read the descriptions about how the advice was sampled. Even when they saw example values of the advice there was only one screen interposed (with all example values being presented at once) between the presentation of meta-information regarding advice sampling and the decision being made. In the weighting studies on the other hand, participants are occupied with the very specific group of advisors in each trial while the advice sampling information fades into the background. This way, it may have been easier for respondents in the choosing studies to take the sampling information into account, being less distracted by the actual advice.

Moreover, a critical issue that distinguishes the decision tasks from the weighting tasks is the fact that in the former, participants know from the beginning that there are two types of advice and how they were generated, respectively. In contrast, in the weighting studies, participants learn only after the first fifteen trials with one type of advice that there will be another type of advice for the next fifteen trials and are then informed how it was generated. Therefore, when choosing advice, participants instantly have the possibility to compare the two types of advice. As a result, they may focus on other aspects compared to the participants who are not aware from the beginning, that there are two different types of advice. Having in mind that their task was to choose between the two types of advice participants in the choosing studies may have analyzed more carefully on what basis the advisors actually formed their advice.

Regarding the judges' accuracy the results confirmed our expectations. Participants benefitted more from dependent advice when the advisors had discussed their estimates and they benefitted more from independent advice when dependent advisors had overlapping information.

In sum, consensus seems to be a salient cue in the weighting studies misleading participants to weight interdependent advice more strongly independent of its accuracy. On the other hand, in the choosing studies the meta-information regarding the advice's interdependence is apparently more salient, aiding participants in making a decision which is beneficial to their own accuracy.

Our studies add to previous findings that individuals show higher confidence after being presented with dependent compared to independent opinions (e.g., Budescu & Rantilla, 2000; Kahneman & Tversky, 1973) by directly investigating participants' preference in terms of choosing one type of advice over the other, as well as by providing a second, somewhat indirect preference measure in terms of how strongly participants weight dependent vs. independent advice. With our studies, we were able to qualify previous findings by showing that a preference for dependent advice is not universal. Participants clearly preferred dependent advice when weighting it, but, in each scenario, they chose the more accurate advice regardless of its interdependence.

4 General Discussion

In a world in which individuals rarely have access to all the relevant information for a decision to be made, they seek the advice of others to fill in the missing pieces. As a result of limited knowledge (i.e., missing information concerning the advisors' reputation or expertise), we may resort to using heuristics in order to determine which type of advice to use. One such heuristic could be to rely on consensus as a cue for validity (see Chaiken & Stangor, 1987). With the two manuscripts that I summarized above, we intend to make important contributions to the literature concerning advice taking by showing that individuals indeed seem to rely on consensus cues by favoring dependent (and more consensual) over independent (disagreeing) advice, even when the former is less accurate than the latter – but only in situations where they do not directly compare independent with dependent advice. Furthermore, our research is the first to investigate the choice and weighting of dependent versus independent advice in a JAS. Although there is one study by Yaniv and colleagues (2009) in which they investigated a preference for dependent advice via confidence gains and adjustment rates in the JAS, we were able to demonstrate in our first manuscript that the lion's share of the effects in the original study was due to social validation rather than consensus. This finding is crucial insofar as we could show that agreement among advisors – one typical characteristic of interdependence – does not necessarily increase the judge's confidence. In fact, our participants displayed the lowest confidence gains when the advisors agreed without simultaneously supporting the opinion of the judge. This is important as previous research has shown that individuals are more confident when they find agreement among opinions (Sniezek & Buckley, 1995). Our research qualifies this aspect by showing that, at least in the JAS, the level of agreement with the judge's own opinion has to be added to the equation.

The series of experiments described in our second set of experiments further qualifies the understanding of our findings above by not only showing that the effect of weighting also holds true when interdependence is operationalized in an ecologically valid way; but by furthermore revealing that participants in fact do not have an overall preference for dependent advice: they only prefer dependent over independent advice when weighting it, and, in contrast, display a preference for the more accurate advice when being able to choose between the two types of advice.

This means that individuals are not generally prone to use consensus as a cue for accuracy. When they are aware that there are two different types of advice and they have the opportunity to compare them, they also take information regarding the advice's interdependence into account. This is important as it shows that decision making in these kinds of scenarios is guided by common sense. Apparently, it is harder for individuals to ignore consensus cues when they are confronted with only one type of advice at a time and preference is measured in a more indirect manner. Since this preference for dependent advice in the weighting studies was displayed irrespective of the advice's accuracy the signal that consensus sends is in this case indeed stronger than the effect of meta-information concerning accuracy.

Limitations and directions for future research. As described above, participants in the weighting studies did not know from the beginning that there would be two different types of advice and only later learned, how the advisors differed in forming the advice – the two advisors teams were never put in direct juxtaposition. Therefore, the varying degree of interdependence might have been more salient in the choosing studies, making it more obvious for participants to take this information into account. Therefore, in future studies respondents in the weighting studies could be made aware from the beginning there would be two advisors teams who differed in the way they formed the advice.

Furthermore, from our results it is not perfectly clear if individuals lack a general understanding of correlated errors. On the one hand, our findings regarding weighting in Manuscripts 1 and 2 suggest that this is indeed the case. If participants had understood that the biases of dependent opinions are positively correlated, they should not have weighted dependent advice more strongly. This is in line with previous findings by Soll and Larrick (2009) and Minson and Mueller (2012). For one, Soll and Larrick showed that participants react to differences in competence of advisors but not to differences in bracketing rates. This suggests that they were not sensitive to the existence of shared biases (this is because bracketing rates are low in the case of shared biases and higher for opposing biases; see Larrick & Soll, 2006). Furthermore, Minson and Mueller (2012) demonstrated that participants making estimates through discussion in dyads put less weight on judgments of their peers compared to participants working alone. This effect was mediated by higher confidence levels displayed by individuals working in dyads. At the same time, dyads were not more accurate in their final judgments than individuals who had worked alone. This suggests that participants fail to realize that their confidence in their judgments is not the result of two independent estimates. Therefore, in the same way, our participants might not have recognized that the consensus displayed by the advisors was not a result of independent estimates.

In contrast to Maines (1996), we did not provide participants with explicit information regarding the correlation of errors. Rather, they had to draw the conclusion that the advisor consensus they perceived in the data was a result of correlated errors. Thus, it is possible that they had difficulties in seeing a connection between the consensus they saw and the meta-information on advice interdependence they had received at the beginning of each study. The fact that our participants chose the more accurate team of advisors based on the descriptions regarding advice sampling also speaks for this interpretation. Perhaps individuals do generally understand the concept of correlated errors, but, in scenarios where consensus is perhaps more salient than the information on correlated errors, find it challenging to relate this knowledge to the evident advisor consensus (and its source). Just from the fact that participants weighted dependent advice more we cannot definitively tell whether or not they lack a general understanding of correlated errors. Future research could therefore ask participants at the end of the experiment why, in their opinion, the advice was consensual or not and whether they thought that the level of consensus had an effect on their behavior.

Finally, we have to keep in mind that we only investigated conditions in which the consensus of dependent opinions was higher than that of independent opinions. As I explained in the introduction, such consensus is a typical characteristic of dependent advice. However, generally speaking it is possible that, incidentally, independent opinions are as consensual as dependent opinions. In these cases, independent opinions might indeed be weighted more strongly than dependent opinions. This would indicate that individuals value consensus more when it comes from independent sources. Therefore, future research could manipulate interdependence and consensus orthogonally by pairing low and high levels of interdependence with low and high levels of consensus.

5 Conclusion

It is a well-established fact that individuals gain most from combining several independent opinions when making their judgments (Budescu & Yu, 2007; Johnson, Budescu, & Wallsten, 2001). Our research contributes to this finding by demonstrating in a JAS that individuals often do not realize these accuracy gains, because they are distracted by the advice's consensus. We confirmed this result using different scenarios and different measures, thereby showing the robustness of the effect. Our findings underline the importance of training individuals in focusing on the advice's actual accuracy without being misguided by consensus cues. Nevertheless, we could also show that individuals do in fact choose the more accurate advice, irrespective of its interdependence, when they are given the opportunity to compare the two advisors teams directly. Therefore, individuals apparently do act rationally when given all relevant information beforehand.

6 References

Ariely, D., Au, W. T., Bender, R. H., Budescu, D. V., Dietz, C. B., Gu, H., Wallsten, T. S., & Zauberman, G. (2000). The effects of averaging subjective probability estimates between and within judges. Journal of Experimental Psychology: Applied, 6, 130–147.

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,* 127-151.

Budescu, D. V., & Rantilla, A. K. (2000). Confidence in aggregation of expert opinions. *Acta Psychologica, 104,* 371-398.

Budescu, D. V., & Yu, H.-T. (2007). Aggregation of opinions based on correlated cues and advisors. *Journal of Behavioral Decision Making, 20,* 153-177.

Chaiken, S., & Stangor, C. (1987). Attitudes and attitude change. *Annual Review of Psychology, 38,* 575-630.

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

Festinger, L., Gerard, H. B., Hymovitch, B., Kelley, H. H., & Raven, B. (1952). The influence process in the presence of extreme deviates. *Human Relations, 5,* 327-346.

Gino, F. (2008). Do we listen to advice just because we paid for it? The impact of advice cost on its use. *Organizational Behavior and Human Decision Processes, 107,* 234-245.

Gonzalez, R. (1994). When words speak louder than actions: Another's evaluations can appear more diagnostic than their decisions. *Organizational Behavior and Human Decision Processes,* 58, 214– 245.

Harvey, N., & Fischer, I. (1997). Taking advice: Accepting help, improving judgment, and sharing responsibility. *Organizational Behavior and Human Decision Processes, 70,* 117-133.

Heath, C., & Gonzalez, R. (1995). Interaction with others increases decision confidence but not decision quality: Evidence against information collection views of interactive decision making. *Organizational Behavior and Human Decision Processes, 61*, 305–326.

Herzog, S. M., & Hertwig, R. (2009). The wisdom of many in one mind: Improving individual judgments with dialectical bootstrapping. *Psychological Science, 20,* 231-237.

Johnson, T. R., Budescu, D. V., & Wallsten, T. S. (2001). Averaging probability judgments: Monte Carlo analyses of asymptotic diagnostic value. *Journal of Behavioral Decision Making, 14*, 123-140.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*, 237-251.

Larrick, R. P., & Soll, J. B. (2006). Intuitions about combining opinions: Misappreciation of the averaging principle. *Management Science, 52,* 111-127.

Lopes, D., Vala, J., &Garcia-Marques, L. (2007). Social validation of everyday knowledge: Heterogeneity and consensus functionality. *Group Dynamics: Theory, Research, and Practice*, *11,* 223-239.

Maines, L. A. (1996). An experimental examination of subjective forecast combination. *International Journal of Forecasting, 12*, 223–234.

Minson, J. A., & Mueller, J. S. (2012). The cost of collaboration: Why joint decision making exacerbates rejection of outside information. *Psychological Science*, *23,* 219-224.

Mojzisch, A., Schulz-Hardt, S., Kerschreiter, R., Brodbeck, F. C., & Frey, D. (2008). Social validation in group decision-making: Differential effects on the decisional impact of preference-consistent and preference-inconsistent information. *Journal of Experimental Social Psychology, 44*, 1477-1490.

Savadori, L., Van Swol, L. M., & Sniezek, J. A. (2001). Information sampling and confidence within groups and judge advisor systems. *Communication Research*, *28*, 737-771.

Schulz-Hardt, S., Frey, D., Lüthgens, C., & Moscovici, S. (2000). Biased information search in group decision making. *Journal of Personality and Social Psychology, 78*, 655–669.

Schultze, T. (2015, November). *A comparison of actual and optimal advice taking in quantitative judgments*. Poster session presented at the 36th annual conference of the Society for Judgment and Decision making, Chicago.

Schultze, T., Mojzisch, A., & Schulz-Hardt, S. (2012). Why groups perform better than individuals at quantitative judgment tasks: Group-to-individual transfer as an alternative to differential weighting. *Organizational Behavior and Human Decision Processes*, *118*, 24-36.

Schultze, T., Rakotoarisoa, A., & Schulz-Hardt, S. (2015). Effects of distance between initial estimates and advice on advice utilization. *Judgment and Decision Making*, *10*, 144-171.

Sniezek, J. A., & Buckley, T. (1995). Cueing and cognitive conflict in Judge-Advisor decision making. *Organizational Behavior and Human Decision Processes, 62*, 159–174.

Sniezek, J. A., & Kolzow, K. (1994). *From disagreement and uncertainty to consensus and* *confidence in group judgment*. Unpublished manuscript, University of Illinois at Urbana-Champaign.

Soll, J. B. (1999). Intuitive theories of information: Beliefs about the value of redundancy. *Cognitive Psychology, 38*, 317–346.

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.

Valley, K. L., White, S. B., Neale, M. A., & Bazerman, M. H. (1992). Agents as information brokers: The effects of information disclosure on negotiated outcomes. *Organizational Behavior and Human Decision Processes*, *51,* 220-236.

Van Swol, L. M. (2009). Factors affecting decision makers' preference for unshared information. *Group Dynamics: Theory, Research, and Practice, 13,* 31-45.

Van Swol, L. M., & Ludutsky, C. (2007). Tell me something I don't know: Decision makers' preference for advisors with unshared information. *Communication Research, 34,* 297–312.

Yaniv, I. (1997). Weighting and trimming: heuristics for aggregating judgments under uncertainty. *Organizational Behavior and Human Decision Processes, 69,* 237-249.

Yaniv, I. (2004a). Receiving other people's advice: Influence and benefit. *Organizational Behavior and Human Decision Processes*, *93,* 1-13.

Yaniv, I. (2004b). The benefit of additional opinions. *Current Directions in Psychological Science*, *13,* 75–78.

Yaniv, I., & Kleinberger, E. (2000). Advice taking in decision making: Egocentric discounting and reputation formation. *Organizational Behavior and Human Decision Processes*, *83*, 260–281

Yaniv, I., Choshen-Hillel, S., & Milyavsky, M. (2009). Spurious consensus and opinion revision: Why might people be more confident in their less accurate judgments? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 558-563.

Yaniv, I., Choshen-Hillel, S., & Milyavsky, M. (2011). Receiving advice on matters of taste: Similarity, majority influence and taste discrimination. *Organizational Behavior and Human Decision Processes*, *115,* 111-120.

Manuscript 1

Disentangling the Effects of Advisor Consensus and Advice Proximity

Disentangling the Effects of Advisor Consensus and Advice Proximity

Stella Katherina Wanzel, Thomas Schultze, and Stefan Schulz-Hardt

University of Goettingen

Reproduced from Wanzel, S. K., Schultze, T., & Schulz-Hardt, Stefan (2017). Disentangling the effects of advisor consensus and advice proximity. Journal of Experimental Psychology: Learning, Memory, and Cognition, 43, 1669-1675

Abstract

When advice comes from interdependent sources (e.g., from advisors who use the same database), less information should be gained as compared to independent advice. On the other hand, since individuals strive for consistency, they should be more confident in consistent compared to conflicting advice, and interdependent advice should be more consistent than independent advice. In a study investigating the differential effects of interdependent vs. independent advice on a judge's accuracy and confidence (Yaniv, Choshen-Hillel & Milyavsky, 2009), advice interdependence was confounded with another variable, namely closeness of the advice to the judge's estimate: Interdependent advice was not only more consistent than independent advice but also closer to the judge's first estimate. The present study aimed at disentangling the effects of consensus and closeness of the advice by adding a third experimental condition in which interdependent (and, hence, consistent) advice was far from the judge's own estimate. We found that, as suggested by Yaniv et al., accuracy gains were indeed a consequence of advisor interdependence. However, in contrast to Yaniv et al.'s conclusions, confidence in the correctness of one's estimates was mostly a function of the advice's proximity to the participants' initial estimations, thereby indicating a social validation effect.

Keywords: advice, proximity, consensus, interdependent, independent

Disentangling the Effects of Advisor Consensus and Advice Proximity

We frequently receive advice, either because we actively seek it, or because others offer it unsolicitedly. Sometimes, we might feel that we need advice from several advisors before making a decision. For example, when making predictions regarding the development of a stock index – in order to know whether and how much to invest in a certain index fund – a person might consult more than one analyst. Another example is environmental legislation. Before deciding whether to regulate carbon dioxide emissions more strictly, a government might consult several experts on climate development. Typically, individuals seek more than one opinion in order to prevent following one piece of bad advice, or to receive sufficient confirmation that they are doing the right thing. Therefore, decision makers will hope for consensus among advisors, that is, the second (and third) advisor should support the first one. If advice is consistent, it is taken as accurate (Kahneman & Tversky, 1973).

Advisor consensus and accuracy

Advice from multiple advisors may not always be equally beneficial, however. An important variable that should affect the value that multiple pieces of advice can have for judgmental accuracy is whether the different advisors are *independent* or *interdependent*. If the financial analysts in the first example are freelancing, intuition holds that their advice may be independent, because the analysts do not influence each other, and they may use different sources of information. If, however, the financial analysts work for the same company (thus sharing a common background and applying similar methods), we should expect their advice to be interdependent. That is, the advice is – to some extent – redundant.

If a person aims to make accurate judgments, multiple pieces of advice are particularly valuable if they are independent. As Yaniv (2004) points out, a quantity estimate consists of the true value, an idiosyncratic bias (i.e., an individual's tendency to under- or overestimate target values), and a random error. The random errors of two individuals are (by definition) independent, but this is not necessarily true for their idiosyncratic biases. Interdependence of opinions means that the idiosyncratic biases are (positively) correlated. This has two consequences: On the one hand, interdependent advice must be – on average – more similar to each other than independent advice. The reason is that for independent opinions the only source of covariation (and thus, similarity) is the true value. In contrast, interdependent opinions co-vary because they share the true value and (part of) the idiosyncratic biases. Thus, interdependent opinions will appear more similar to one another on average – thereby suggesting greater consensus regarding the true value. Second, this greater consensus usually does not go along with greater accuracy – quite the contrary: In the case of independent opinions, aggregation will lead to cancellation of unsystematic (or random) error and idiosyncratic biases, since these biases are mutually independent. In interdependent opinions, only unsystematic errors are reduced in the same fashion, but this is not so for idiosyncratic biases. In essence, the higher the correlation of idiosyncratic biases (i.e., the greater the interdependence of the opinions), the less potential there is for cancellation. Therefore, assuming equal unsystematic errors, independent opinions are, on average, superior to interdependent opinions in terms of accuracy (see also Hogarth, 1978, for a formal analysis).

The question is if decision-makers can recognize the value of independent advice, or whether a desire for advisor consensus pushes them towards interdependent advice. So far, only one published study has addressed this question: Yaniv, Choshen-Hillel, and Milyavsky (2009) investigated how advice from interdependent vs. independent sources affected opinion revision. Their idea was that individuals pay more attention to consensus cues (like the consistency among the advisors) than to interdependence cues (e.g., how the advice was sampled). This should lead to higher confidence due to consistent opinions, but simultaneously less accurate judgments when participants are dealing with interdependent information. According to Yaniv et al., people tend to overlook that consensus can be spurious, defined as “a set of consistent opinions produced by interdependent sources” (p. 561). In contrast, only consensus arising from independent opinions is a valid indicator of the correctness of this advice (cf. 559).

To investigate their hypotheses, Yaniv and colleagues (2009) let participants perform a series of quantitative estimation tasks in a one-factorial within-subjects design. In each trial, participants first estimated the caloric value of a specific food item as accurately as possible. Then, they received advice in the form of the estimates that three participants of a previous study had made for the same food item. Afterwards, participants gave a final – and possibly revised – estimate. Participants also indicated their confidence in the accuracy of their initial and their final estimate. The type of advice (independent or interdependent) was the independent variable. Independent advice was drawn randomly from a pool of 100 previously collected estimates. Interdependent advice, on the other hand, was generated by drawing those three estimates that were closest, 7th closest and 15th closest to participants’ initial estimates. This constraint ensured a high degree of advisor consensus. Depending on the condition participants were either informed that the estimates were selected from those that were closest to their own initial judgment, or that they were randomly drawn.

Confirming their hypotheses, Yaniv et al. (2009) found that participants’ judgments were less accurate when they received interdependent compared to randomly sampled (i.e., independent) advice. Furthermore, participants revised their judgments less often and displayed higher levels of confidence in the interdependent condition. According to the authors, these findings indicate a dissociation of accuracy and confidence, resulting from participants’ preference for advisor consensus and their failure to realize that the underlying dependence caused this spurious consensus.

Advisor consensus or social validation?

Although the authors' interpretation is plausible, their experimental manipulation leaves room for an alternative interpretation: by drawing advice that was close to the participants' own opinion in the interdependent condition, Yaniv et al. (2009) not only manipulated advisor consensus, but also the distance of the advice to the judge's initial estimate. Because of this confound, the findings could just as well be the result of the *proximity* of the interdependent advice to participants' initial estimates. This possibility suggests an alternative explanation for the higher levels of confidence along with lower levels of accuracy in the independent vs. the independent condition, namely social validation. Social validation means that a person's opinion is reinforced by others signaling that they hold similar beliefs or have come to similar conclusions (e.g., Mojzisch, Schulz-Hardt, Kerschreiter, Brodbeck, & Frey, 2008). Schultze, Rakotoarisoa, and Schulz-Hardt (2015) demonstrated that social validation also occurs in the judge advisor system. They found that advice that was more similar to the judge's own initial opinion led to higher levels of confidence and, at the same time, to less frequent adjustments of the initial estimates.

Thus, it seems reasonable to conclude that the high levels of confidence found by Yaniv et al. (2009) are not necessarily due to the high consistency among the advice, but may rather have (at least partially) resulted from the judge being socially validated by the three advisors. Interestingly, one particular result in the Yaniv et al. (2009) study itself already hints at the possibility of social validation effects: participants were more than twice as likely to retain their initial estimate after receiving interdependent advice (which was close to their own initial estimate), as compared to independent (and far) advice (65% vs. 30%). It seems that they were more convinced that their first estimates were accurate and needed no further adjustment when presented with advice similar to these estimates, which is in line with the social validation hypothesis.

The present study aimed to disentangle the effects of the two confounded[1] variables on judges' willingness to heed the advice, their accuracy gains due to taking advice, and increases in confidence. To this end, we replicated the original Yaniv et al. (2009) study and added a third experimental condition in which advisors' estimates were interdependent, but at the same time, notably different from the judge's initial opinion.[2] If the interpretation by Yaniv and colleagues holds true, judges' behavior in the two interdependent advice conditions should differ from that in the independent advisor condition in the same way as in the original study, but the two interdependent conditions should not be different from each other. If, in contrast, participants' behavior in the new condition (interdependent advice dissimilar to the participants' estimations) differs from the original interdependence condition, and more closely resembles behavior in the independence condition, the effects observed by Yaniv and colleagues (2009) were misattributed and, in fact, driven by advice distance rather than interdependence.

[1] Interestingly, Yaniv et al. (2009) were aware of their experimental conditions differing in two ways (proximity and advisor consensus), and they also acknowledged that individuals tend to give more weight to opinions that agreed with their own. However, eventually they decided to interpret the results as an effect of advisor consensus alone, while disregarding proximity to the judge's initial estimate.

[2] An even stricter test of our idea would be to implement a fully orthogonal 2x2-design in order to completely disentangle the effects of the two variables. However, the one particular cell that this experimental design would have in comparison to our design (namely a condition with independent advice close to the judge's initial estimate) is practically impossible to realize, as drawing advice that is close to the judge's estimate automatically means that it is interdependent (in terms of the definition of interdependence in the original study).

Method

Participants and design. Eighty[3] participants were recruited via the Online Recruitment System for Economic Experiments (ORSEE, Greiner, 2015). One person was excluded because she correctly guessed the purpose of the study, leaving 79 cases for further analyses. Fifty-two participants were female (66%), mean age was 24.89 years (*SD* = 5.79), and all participants were university students. As in the original study, we used a within-subjects design with the method of advice sampling (independent vs. dependent/far vs. dependent/close) being the independent variable.

Procedure. Our study was similar to that of Yaniv and colleagues (2009) in as many aspects as possible to ensure the comparability of the results. Participants worked on individual computers. They were informed that the experiment had two phases, in each of which they had to estimate the calorie content of various foods. We incentivized participants in the same ways as Yaniv et al.: On the one hand, participants learned that they would receive a bonus of 2 Cents for each estimate that fell in the range of +/- 12% of the true value. On the other hand, they were informed that they could gain an extra bonus of 40 Cents by betting on their answers on 15 out of 30 trials. They received the bonus if their judgment fell in an interval of +/- 12% of the correct answer. In each trial, participants were told how often they had bet so far, and how many remaining bets they had. In the first phase, participants were asked to make initial estimates for the caloric value per 100g/ml of 30 different foods (such as zucchini or butter). As in the original study, we also asked them to provide

[3] A post-hoc power analysis (G*Power; Faul, Erdfelder, Lang, & Buchner, 2007) revealed a test power of .999 for medium effects (f = .25) when assuming a correlation of r = .50 among the repeated measures. For small to medium-sized effects (f = .18), power was .95. When testing the null-hypothesis with an alpha-level of α = .25 in order to avoid committing a type-II-error, we achieved a test-power of .80 for small effects (f = .10).

confidence ratings for their estimates on a scale from 0% (not confident at all) to 100% (completely confident).

In phase two, participants were presented with the same 30 food items. They also received the estimates of three advisors and saw their own estimate from phase one. Phase two involved the three advice-sampling conditions (independent condition, dependent/far condition and dependent/close condition), and each participant passed through ten trials per condition. For each participant, the 30 food items were presented in the same sequence; each condition was presented in 10 trials; the order of conditions, however, was fully randomized for each participant in order to prevent confounding the experimental conditions with specific food items. In each trial, respondents saw a header, accurately informing them about how the three pieces of advice were sampled before they made their final judgments. The independent and the dependent/close condition were identical to those in Yaniv et al. study, whereas the new dependent/far condition was designed in order to disentangle consensus and proximity.

In the independent condition, the header stated that the three advisory estimates were selected randomly from a pool of 112 estimates, which had been made by participants in a previous study. In the dependent/close condition, it said that the advice came from persons who had given estimates close to the participant's opinion. Furthermore, they were informed that the advice was sorted according to its proximity to the judge's original estimate. Similar to the Yaniv et al. study, the computer picked the values that were closest, 7th closest, and 15th closest to the participants' initial estimation. In the dependent/far condition, we first computed an interval of one standard deviation (based on all estimates in the advisor pool for this specific trial). We then selected the piece of advice that was closest to either limit of this interval. This estimate then served as the anchor for drawing the three pieces of advice that were presented to the participant. Similar to the dependent/close condition, all estimates were sorted according to their proximity to this fourth participant's estimate with the selected pieces of advice having positions 1, 7, and 15. The header in the dependent/far condition informed participants that the advice were estimates close to the estimate of a fourth, allegedly randomly selected, participant (therefore, the three pieces of advice were also close to each other). After receiving the advice, participants provided a final, possibly revised, estimate and stated their level of confidence.

Results

In all analyses, one trial out of 30 had to be excluded for 7 persons in the independent condition, since they received advice from only two advisors due to technical problems.
In addition to frequentist tests, we also report Bayes Factors (BF) obtained using the free software JASP (2016). We used the default JZS priors for our analyses of variance (ANOVAs) and *t* tests (see Morey and Rouder, 2011; Rouder, Morey, Speckman, & Province, 2012; Rouder, Speckman, Sun, Morey, & Iverson, 2009). We report the BF_{10} for the Bayesian tests. The BF_{10} is an odds ratio stating how likely H1 is relative to H0 given the observed data. In line with Jeffrey (1961), we interpret a BF_{10} of 3 or greater as positive evidence for the alternative hypothesis, and a BF_{10} smaller then 1/3 as positive evidence in favor of the Null. In some analyses, we tested for differences in the change in certain variables (e.g., changes in accuracy or changes in confidence) using two-factorial ANOVAs with one factor representing time (pre-advice vs. post-advice). In those cases, we only report the BF_{10} for the interaction effect, as this is the effect of interest.[4]

---

[4] The interpretation of main or interaction effects in two-factorial ANOVAs in JASP does not correspond well to the frequentist significance test. The BF_{10} of an effect represents the relative likelihood of a model containing this effect and all subordinate effects to the null-model containing only an intercept. Since the effect of interest in the two-factorial ANOVAs we report is the interaction of time and experimental condition, we chose

Manipulation checks

Advisor Consensus. First, we investigated the degree of advisor consensus in the three conditions by assessing the mean coefficient of variation (CoV). A repeated measures ANOVA showed a main effect for condition, $F(1.73$[5]$, 134.60) = 1309.00$, $p < .001$, $\eta_p^2 = .94$, $BF_{10} = 2.90 \times 10^{128}$. Paired *t*-tests revealed that, as expected, the CoV was higher in the independent condition (M=.61, SD = .09) than in the dependent/far (M = .17, SD = .05), $t(78) = 37.31$, $p < .001$, $d = 4.35$, $BF_{10} = 8.36 \times 10^{47}$ or the dependent/close condition (M =.13, SD = .04), $t(78) = 43.89$, $p < .001$, $d = 5.32$, $BF_{10} = 1.27 \times 10^{53}$. The difference between the two dependent conditions was also significant, $t(78) = 5.01$, $p < .001$, $d = 0.57$, $BF_{10} = 4.74 \times 10^{3}$; compared to the independent condition, however, the CoV was small in both dependent conditions.

Advice Proximity. As a measure of advice proximity, we calculated the absolute difference between judges' initial estimates and the mean of the advice for each trial, and we then averaged these absolute differences for each condition. A repeated measures ANOVA on advice proximity revealed an effect of condition, $F(1.58, 123.48) = 429.17$, $p < .001$, $\eta_p^2 = .85$, $BF_{10} = 4.64 \times 10^{74}$. As expected, proximity was higher in the dependent/close condition (M = 18.37, SD = 19.72) than in the independent condition (M = 113.19, SD = 46.88), $t(78) = 19.38$, $p < .001$, $d = 2.55$, $BF_{10} = 1.94 \times 10^{28}$, or the dependent/far condition (M = 161.23, SD = 27.76), $t(78) = 38.13$, $p < .001$, $d = 4.36$, $BF_{10} = 4.06 \times 10^{48}$. The difference between the

---

to report the BF_{10} for the a one-factorial ANOVA on the difference scores. A one-factorial test on the difference scores is equivalent to testing the interaction of time and condition, thus providing the desired BF_{10}.

[5] Whenever the assumption of sphericity was violated, we used the Greenhouse-Geisser correction and report the corrected fractional degrees of freedom.

independent and the dependent/far condition was also significant[5], $t(78) = -8.02$, $p < .001$, $d = -0.093$, $BF_{10} = 1.01 \times 10^9$. In summary, the manipulations of interdependence (consensus) and proximity were successful.

Main Analyses.

Accuracy. Table 1 depicts the mean absolute errors of the initial and final estimates, as well as the improvements (in percent) for each condition. A 2 (Phase 1 vs. 2) × 3 (Condition) analysis of variance with repeated measures on the mean absolute errors revealed a main effect of phase, $F(1, 78) = 25.59$, $p < .001$, $\eta_p^2 = .25$, no overall effect of condition, $F(2, 156) = 1.45$, $p = .24$, $\eta_p^2 = .02$, and a significant interaction $F(1.77, 148.97) = 5.66$, $p = .006$, $\eta^2 = .07$, $BF_{10} = 7.54$, indicating that accuracy gains differed between conditions. We analyzed the interaction more closely by investigating the accuracy gains in the three conditions. To this end, we calculated difference scores between the mean absolute errors of participants' initial and final estimates for each condition. Testing the difference scores against zero revealed significant gains in the independent condition ($M = 16.77$), $t(78) = 6.23$, $p < .001$, $d = 0.32$, $BF_{10} = 5.43 \times 10^5$, and in the dependent close condition ($M = 4.64$), $t(78) = 3.05$, $p = .003$, $d = 0.09$, $BF_{10} = 8.80$. In the dependent/far condition, accuracy also increased significantly ($M = 8.52$), $t(78) = 2.15$, $p = .034$[6], $d = 0.16$, $BF_{10} = 1.09$, but since the BF_{10} did not exceed 3, we interpret this latter accuracy gain with caution. The results are in line with previous research showing that individuals benefit from advice (e.g., Soll & Larrick, 2009; Yaniv, 2004). However, the magnitude of the accuracy gains differed between the conditions:

[5] Since advice in the independent condition was randomly drawn, it can be close to the initial estimate by chance. However, this is impossible in the dependent/far condition.

[6] One person in the dependent/far condition had a particularly high accuracy gain of 114 points. When we exclude this person from the analysis, the gain in accuracy for this condition (M = 7.16) no longer reaches significance, $t(77) = 1.90$, $p = .061$, $d = 0.43$, $BF_{10} = .69$.

the independent condition yielded significantly larger gains than the dependent/close condition, $t(78) = 4.08$, $p < .001$, $d = -0.62$, $BF_{10} = 188.34$. This result replicates the finding by Yaniv and colleagues (2009). When comparing our new dependent/far condition with the other two, accuracy gains were also greater in the independent than in the dependent/far condition, $t(78) = 2.00$, $p = .049$, $d = -0.27$, $BF_{10} = 0.81$. Technically, while the *p*-value supports the H1, the BF_{10} tends toward the null and is in a range that Bayesian conventions consider inconclusive (Jeffreys, 1961). There was no significant difference between the two dependent conditions, $t(78) = 1.01$, $p = .316$, $d = -0.15$, $BF_{10} = 0.20$. Here, the BF_{10} shows positive evidence of the Null, suggesting that accuracy gains did not differ substantially in the two dependent conditions. In sum, the pattern is not completely unequivocal but, overall, the independent condition seems to be different from the two dependent conditions which, in turn, do not seem to differ reliably from each other. Hence, our findings suggest that, more or less, Yaniv et al. correctly attributed the accuracy findings to consensus. However, this was to be expected, because social validation, as outlined, should mainly affect confidence-related variables.

Frequency of revision. For each condition, we calculated the frequency of revision as the percentage of trials in which participants' initial and final estimates differed (see also Table 1). A repeated measures ANOVA revealed a significant effect of condition, $F(1.65, 128.60) = 92.79$, $p < .001$, $\eta_p^2 = .54$, $BF_{10} = 2.07\times10^{24}$. Participants changed their initial judgments less often in the dependent/close condition (44%) than in the dependent/far (82%), $t(78) = 11.29$, $p < .001$, $d = -1.28$, $BF_{10} = 1.24\times10^{15}$ or the independent condition (71%), $t(78) = 9.26$, $p < .001$, $d = -1.05$, $BF_{10} = 2.11\times10^{11}$. Furthermore, participants changed their initial judgments less often in the independent than in the dependent/ far condition, $t(78) = -4.86$, $p < .001$, $d = -0.55$, $BF_{10} = 2.73\times10^{3}$. Note that the values for the independent and the dependent/close condition are very similar to those found by Yaniv and colleagues (2009), which speaks for the comparability of our results to theirs. However, the fact that we observed the highest adjustment rates in the dependent/far condition suggests that reduced willingness to adjust in the dependent/close condition is due to advice proximity rather than interdependence.

Weighting of advice. We analyzed an additional variable not included in the analyses by Yaniv et al. (2009), namely the weight of advice, calculated as the difference between the final and initial estimates, divided by the difference between the mean of the three advisors and the initial estimate. This measure corrects for the systematic difference in advice distance between the three experimental conditions. It is structurally similar to the advice taking measure Harvey and Fischer (1997) developed for the case of single advisors. Weights of advice usually range from 0 to 1. In line with previous studies, we truncated weighting scores greater than 1 to a value of 1, and negative scores to a value of 0 (Schultze et al., 2015; Soll & Larrick, 2009). A repeated measures ANOVA revealed a significant effect of condition, F (1.72, 134.20) = 31.77, $p < .001$, $\eta_p^2 = .29$, $BF_{10} = 1.55 \times 10^9$. Paired t-tests showed greater weights in the dependent/far condition ($M = 0.43$) than in the independent condition ($M = 0.32$), $t(78) = -5.31$, $p < .001$, $d = 0.63$, $BF_{10} = 9785.36$, or the dependent/close condition ($M = 0.24$), $t(78) = 6.71$, $p < .001$, $d = 0.79$, $BF_{10} = 2.57 \times 10^6$. Furthermore, weighting was greater in the independent than in the dependent/close condition, $t(78) = 3.62$, $p = .001$, $d = 0.41$, $BF_{10} = 40.84$. This pattern supports the findings for revision rate, indicating that differential weighting of advice is a function of advice proximity rather than advisor consensus.

Confidence and betting. We analyzed confidence ratings in a 2 (Phase 1 vs. 2) × 3 (Condition) within-subjects ANOVA. This analysis revealed main effects of phase, $F(1, 78) = 19.48$, $p < .001$, $\eta^2 = .20$, and condition, $F(1.74, 135.49) = 16.66$, $p < .001$, $\eta_p^2 = .18$, both of which were qualified by an interaction effect, $F(1.76, 136.99) = 28.36$, $p < .001$, $\eta_p^2 = .18$,

$BF_{10} = 2.65 \times 10^8$. Paired *t*-tests revealed that confidence was higher after receiving advice (as compared to before) in the independent condition ($M_{final} = 51.18$ vs. $M_{initial} = 46.71$), $t(78) = -3.80$, $p < .001$, $d = -0.24$, $BF_{10} = 75.16$, and in the dependent/close condition ($M_{final} = 57.10$ vs. $M_{initial} = 47.80$), $t(78) = -6.61$, $p < .001$, $d = -0.49$, $BF_{10} = 2.64 \times 10^6$. However, in the dependent/far condition, confidence did not change significantly ($M_{final} = 49.03$ vs. $M_{initial} = 47.60$), $t(78) = -1.10$, $p = .274$, $d = -0.08$, $BF_{10} = 0.22$. Note that the BF_{10} for this comparison shows positive evidence in favor of the Null. Additional *t*-tests with confidence gains (i.e., the difference between final and initial confidence) as the dependent variable revealed greater gains in the dependent/close condition ($M = 9.30$) than in the independent condition ($M = 4.47$), $t(78) = -5.62$, $p < .001$, $d = 0.65$, $BF_{10} = 4.89 \times 10^4$. This finding parallels the corresponding result in the Yaniv et al. (2009) study. In the new dependent/far condition, we found even smaller confidence gains ($M = 1.43$) than in the independent ($M = 4.47$), $t(78) = -2.85$, $p = .006$, $d = -0.32$, $BF_{10} = 5.15$, or the dependent/close condition ($M = 9.30$), $t(78) = -6.53$, $p < .001$, $d = -0.74$, $BF_{10} = 1.85 \times 10^6$. The strong difference between the dependent/close and the dependent/far condition shows that, contrary to what Yaniv et al. assumed, it is not the interdependence (the consensus) among advisors that drives the confidence findings – rather, it is the proximity of the advice to the participants' initial estimates.

Similar to Yaniv and colleagues (2009), we also analyzed participants' betting behavior. A χ^2-test for equal distribution of frequencies showed a significant difference between conditions, $\chi^2(2) = 95.16$, $p < .001$, $BF_{10} = 4.04 \times 10^{18}$. Supporting the results for the subjective confidence ratings, and in line with the original study, participants bet more often in the dependent/close (69%) condition than in the independent condition (50%), $\chi^2(1) = 52.02$, $p < .001$, $BF_{10} = 1.56 \times 10^{18}$. When adding the dependent/far condition, we found that participants bet less in this condition (44%) than in the dependent/close condition (69%),

$\chi^2(1) = 88.29$, $p < .001$, $BF_{10} = 4.04\times10^{18}$. They also bet less in the dependent/far condition than in the independent condition, but this difference was relatively small at 6 percentage points, $\chi^2(1) = 4.83$, $p = .028$, $BF_{10} = 0.74$. Also, contrary to the p-value, the BF_{10} tends toward the Null, but suggest the results are inconclusive. Consistent with our reasoning above, we chose the frequentist interpretation but – due to the weak evidence – refrain from drawing strong conclusions from the observed difference. Nevertheless, because the dependent/close condition clearly differs from the other two ones, the results suggest that – once more – proximity is the main driving force here.

Table 1. Overview of main results (Means with Standard Deviations in parentheses)

	Method of sampling advice		
Measure	Independent	Dependent/far	Dependent/close
Accuracy			
initial error	**119.50** (53.28)	**110.37** (55.35)	**119.51** (54.34)
final error	**102.72** (45.08)	**101.85** (43.30)	**114.86** (50.43)
improvement (difference scores)	**16.77** (23.94)	**8.52** (35.19)	**4.64** (13.51)
% improvement	**14**	**8**	**4**
Confidence			
initial confidence	**46.71** (18.43)	**47.60** (18.15)	**47.80** (18.57)
final confidence	**51.18** (18.95)	**49.04** (18.73)	**57.10** (19.27)
confidence gain	**4.47** (10.47)	**1.43** (11.60)	**9.30** (12.50)
Rate of betting (%)	**51** (22.14)	**46** (23.50)	**70** (22.09)
Revision Process			
rate of changing initial estimates (%)	**71** (22.73)	**82** (23.56)	**44** (26.78)
Weighting	**.33** (.37)	**.44** (.28)	**.12** (.54)

Discussion

In their study on interdependent versus dependent advice, Yaniv et al. (2009) showed a dissociation between accuracy and confidence. Compared to independent advice, receiving interdependent advice led to less accurate judgments but, paradoxically, also to increased confidence in the accuracy of those judgments. In their experiment, Yaniv et al. focused on advisor consensus, resulting from interdependence among advisors, to explain this dissociation. However, since interdependent advice in their study was not only more consistent, but also closer to the judge's first estimate, one cannot firmly draw the conclusion that advisor consensus is responsible for the effects. It is possible that these effects are caused by the proximity of the advice to the participants' initial estimates. In the present study, we conducted an extended replication of the Yaniv et al. study. By introducing a third condition where advice had similarly high consensus as in the interdependent condition of the original study, but was far from the participants' initial estimates, we disentangled consensus and distance.

First of all, it is important to note that we were able to replicate the main findings of the Yaniv et al. (2009) study. Compared to the independent condition, participants were less accurate after receiving interdependent and close advice, while simultaneously displaying higher levels of confidence and betting more often on their final estimates. Furthermore, they revised their initial estimates less frequently. However, we reach a different conclusion than Yaniv et al. when we take the results of the new dependent/far condition into account. The fact that we found the lowest (and non-significant) confidence gains in the dependent/far condition suggests that proximity, not consensus, is the driving factor here, since otherwise we would have expected roughly the same confidence gains in the two interdependent conditions. Apparently, when the dependent advice is far from one's own estimate (and the distance was greatest in the dependent/far condition), thus not validating the judge's opinion, this results in lower (or even no) confidence gains. This finding is in line with the result of Schultze et al. (2015) that proximity serves as a source of social validation. These conclusions are further supported by participants' reduced willingness to bet on their final estimates in the independent or the dependent/far condition, as compared to the dependent/close condition.

For revision rates, we observed an interesting pattern: Participants revised their estimates most frequently in the dependent/far condition, and least frequently in the dependent/close condition. This is in line with the idea that individuals use consensus as a cue for validity (see Kahneman & Tversky, 1979). When advice is far from one's own opinion, but the advisors are consistent, this results in opinion change. When advice is consistent and close, however, this encourages individuals to keep their opinion. These findings suggest that consensus is always used, but the rule of conduct it signals differs.

In sum, the results so far suggest that greater confidence as well as the lower revision rates after receiving dependent advice in the original study were most likely effects of advice proximity, since our two conditions with dependent advice differed on both confidence measures in the way a social validation account would suggest. Regarding our newly added measure of advice weighting, the results also speak to proximity as the driving force. Supporting the findings for revision rates, participants weighted the advice most in the dependent/far condition and least in the dependent/close condition. Thus, the probability that advice is weighted is higher when it does not resemble individuals' own opinion. The pattern is completely in line with an effect of proximity. However, we can not rule out that proximity in combination with consensus affects our measure in a particular way: when advice is not only far from one's own estimate but also more consistent, this may appear as an especially good reason to adjust to the advice.

With respect to accuracy gains, our results draw a somewhat different picture. While accuracy gains were lower in the two interdependent conditions than in the independent condition, the two interdependent conditions did not differ. In accordance with its lower informational value, interdependent advice leads to less accurate final estimates as compared to independent advice. Descriptively, participants' accuracy improved somewhat more in the dependent/far than in the dependent/close condition, and this pattern makes sense, given that participants retained their initial estimates less often in the latter condition and, accordingly, had less opportunity to improve on their initial estimates. Thus, our data suggest that Yaniv et al. (2009) correctly attributed the lower accuracy gains to the interdependence of advice.

In sum, our results suggest that, while consensus may play a role in explaining some of the findings by Yaniv and colleagues (2009), it is certainly not the only driving force. Quite the contrary: For the majority of variables measured in the Yaniv et al. study (namely confidence, revision rates, and betting), and also for our newly added weighting variable, we showed that proximity rather than consensus drives the effects. Hence, the dissociation between confidence and accuracy Yaniv et al. reported is due to a combination of both proximity and consensus effects, with the former contributing somewhat more than the latter.

Limitations and directions for future research. As an extended replication, our study shares most of the limitations of the original study. First, we only investigated opinions on matters of fact; thus, the results may differ for matters of taste. Second, the operationalization of interdependence was somewhat artificial, in that it was based solely on statistical calculations. Such a procedure is unlikely to map onto everyday life when individuals usually solicit advice. Outside the laboratory, interdependence can arise from advisors belonging to the same group or using shared information. For example, doctors who work in the same hospital or medical center may influence each other when they discuss certain cases or they may underlie a common practice and, therefore, may draw similar conclusions even when they come from different fields. If interdependence arises from realistic circumstances, it might be easier to detect, since individuals are more familiar with these situations. They might intuitively conclude that they do not necessarily receive new information when the advisors all belong to the same institution and pursue the same agenda, or when they have the same data sources. Therefore, an interesting avenue for further research would be to investigate whether the findings still hold true when an ecologically valid operationalization of advisor interdependence is used.

Conclusion. In conclusion, we can state that proximity and consensus are both at play in creating the effects found by Yaniv and colleagues (2009). While social validation can explain why individuals are guided by the closeness of the advice to their own prior opinion, it is rather unclear why it is so difficult for judges to understand the implications of interdependence. Our data suggest that in the absence of proximity (that is, when the advice is far from one's opinion), individuals pay unwarranted attention to (spurious) consensus. Compared to the independent condition, individuals changed their final judgments more often and weighted the advice – which was in fact also far from one's opinion, but more consistent (albeit less beneficial) – more strongly. Individuals apparently feel that there is safety in consensus, even though this consensus is the result of poor sampling methods. Since individuals have a significant disadvantage when they only pay attention to consensus cues, it would be wise to investigate possibilities to train individuals how to use meta-information (like the consensus cues and sampling information) correctly.

Furthermore, our findings tell us that it is also important to take into account the advice's proximity to the advice seeker's own opinion when studying the effects of interdependent vs. independent advice. In real-life situations, advice might be close to the judge's own opinion when the latter also comes from the same institution as the advisors (e.g., a co-worker) or uses the same database, etc. As we have seen, proximity influences individuals' confidence in their own judgements after receiving interdependent vs. independent advice, as well as the extent to which they use the advice. Thus, in such cases it does not suffice to consider only consensus when explaining differential effects in interdependence vs. independence conditions.

References

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association*, *1*, 114-125.

Hogarth, R. M. (1978). A note on aggregating opinions., *Organizational Behavior and Human Performance 21*, 40-46.

JASP Team (2016). JASP (Version 0.8.0.0)[Computer software].

Jeffreys, H. (1961). *Theory of probability.* Oxford, England: Oxford University Press.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80,* 237-251.

Mojzisch, A., Schulz-Hardt, S., Kerschreiter, R., Brodbeck, F. C., & Frey, D. (2008). Social validation in group decision-making: Differential effects on the decisional impact of preference-consistent and preference-inconsistent information. *Journal of Experimental Social Psychology, 44*, 1477-1490.

Morey, R. D., & Rouder, J. N. (2011). Bayes factor approaches for testing interval null hypotheses. *Psychological Methods, 16*, 406-419.

Schultze, T., Rakotoarisoa, A., Schulz-Hardt, S. (2015). Effects of distance between initial estimates and advice on advice utilization. *Judgment and Decision Making, 10*, 144-171.

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.

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology, 56*, 356-374.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t-Tests for accepting or rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*, 225-237.

Yaniv, I. (2004). Receiving other people's advice: Influence and benefit. *Organizational Behavior and Human Decision Processes, 93,* 1-13.

Yaniv, I., Choshen-Hillel, S., & Milyavsky, M. (2009). Spurious consensus and opinion revision: Why might people be more confident in their less accurate judgments? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 558-563.

Manuscript 2

Do Individuals Have a General Preference for Dependent Advice? – Pitting Advisor Accuracy Against Advisor Consensus

Do Individuals Have a General Preference for Dependent Advice? – Pitting Advisor Accuracy Against Advisor Consensus

Stella Katherina Wanzel, Thomas Schultze, and Stefan Schulz-Hardt

University of Goettingen

Abstract

Advice stemming from dependent sources is usually less accurate than advice provided by independent sources, while simultaneously exhibiting greater consensus. Previous literature has shown that individuals display higher levels of confidence after having received information from dependent as compared to independent sources. In our current experiments, we address a more general question, namely whether or not individuals express a preference for dependent advice by choosing it over independent advice, and by weighting it more strongly. We test for this preference both in a situation when dependent advice is more accurate (as a result of group discussion) as well as in a situation when it less accurate (as a result of the advisors using partially identical information sources) than independent advice. In a series of altogether six studies, we show that, when being given the opportunity to choose between independent and dependent advice, participants predominantly select the type of advice that is more accurate in this particular situation. However, when being provided with both types of advice, they generally weight the dependent advice stronger than the independent advice, even if the latter is more accurate. The role of consensus in explaining our findings is discussed.

Keywords: advice taking, accuracy, consensus, dependent advice, independent advice

Do Individuals Have a General Preference for Dependent Advice? – Pitting Advisor Accuracy Against Advisor Consensus

Individuals often solicit advice from others in order to make the best possible judgments and decisions. For instance, if we want to take out disability insurance, we may consult several insurance brokers in order to determine the amount of money to invest. In addition, when we want to sell a car, we may seek out experts to help us estimate a reasonable price. Research on such advice taking usually employs the *judge advisor system* (JAS; see Sniezek & Buckley, 1995; Sniezek & Van Swol, 2001) as a research paradigm. In the JAS, one individual (the judge) works on a series of judgment trials, during which she or he receives advice from one or more other individuals (the advisors) before making a final estimate (for a review see Bonaccio & Dalal, 2006). One common finding is that advice improves judgments and decisions (e.g. Sniezek & Buckley, 1995; Soll & Larrick, 2009; Yaniv, 2004a). These accuracy gains increase even further when the judge integrates the opinions of multiple advisors rather than consulting only one advisor. Several studies demonstrated that the accuracy of the average of advisors' opinions increases monotonically with the number of advisors – as long as the estimates are not completely interdependent (e.g., Ariely et al., 2000; Johnson, Budescu, & Wallsten, 2001; Wallsten & Diederich, 2001). The potential for accuracy gains should be highest, on average, when the advice is independent, and should decrease as the interdependence of advice increases[7] (i.e. when one can guess person B's opinion by knowing the opinion of person A, thereby gaining no additional information; see Yaniv, Choshen-Hillel, & Milyavsky, 2009). Interdependence of opinions can stem from various sources: the advisors may be socialized in the same environment and, therefore, use

[7]To be precise, the best judgment can theoretically be made when the biases of the advisors are perfectly negatively correlated. In this case, the bracketing rate would be 100% for moderate unsystematic errors. However, such a case should hardly ever occur in practice.

the same techniques to interpret the data, they may have access to the same data sources, or they may know each other and discuss the issue at hand, thereby influencing each other.

Why does independent advice have a higher potential for accuracy gains than dependent advice has? As Yaniv (2004b) points out, an estimate consists of a true value, a random error, and an idiosyncratic bias (i.e. a systematic tendency to over- or underestimate the true value). Random errors are generally independent, whereas the idiosyncratic biases of two or more individuals can be positively correlated. If this is the case, these opinions are dependent. As a consequence, dependent opinions will, on average, be more similar than independent opinions, since the latter only co-vary regarding their true value. This means that dependent opinions may create the impression that there is greater consent regarding the true value. However, the positive correlation of biases not only implies that dependent opinions are more similar; it also – on average – results in lower accuracy. This is because, when dealing with independent opinions, both random error and idiosyncratic bias can be reduced by aggregation. Since for dependent advice the biases are positively correlated, only random error is reduced by aggregating the opinions. As a result, assuming (roughly) equal unsystematic errors, the aggregates of dependent advice tend to be less accurate than those of independent advice (see also Hogarth, 1978, for a formal analysis). Yet, the question is, do individuals take this (explicitly or implicitly) into account? In other words, do they systematically choose independent over dependent advice, and do they weight the former more strongly than the latter? Previous research raises reasonable doubts about this, as we shall detail next.

Preference for dependent vs. independent advice. In several studies outside the judge-advisor-paradigm, participants' confidence in their final estimate served as an indicator of preference for a certain type of information. In a study by Kahneman and Tversky (1973), for example, participants were asked to predict grade point averages from two highly correlated tests (creative thinking and symbolic ability) as well as from two uncorrelated tests (mental flexibility and systematic reasoning). They were explicitly told that the tests were equally predictive of college performance and, furthermore, that one pair of tests was correlated, and that the other was not. Although making predictions based on two correlated tests should reduce accuracy, participants were more confident when predicting from the correlated tests compared to the uncorrelated tests. Similarly, Budescu and Rantilla (2000; see also Budescu & Yu, 2007) showed that participants were more confident when presented with correlated opinions. More precisely, participants were provided with a decision scenario where each expert had an equal amount of cues (e.g. six cues for each expert) and, amongst other things, were also given information about the structural overlap of their expert's cues. That is, participants knew whether each expert had unique information (independent experts) or whether experts had shared cues (dependent experts), and they could infer that the gross information content was greater for the independent experts. Participants were much more confident when experts were dependent; in fact, the highest confidence levels were found when the advisors were completely redundant in terms of the cues they had seen and, furthermore, were completely consenting.

The above-mentioned studies did not measure preference for dependent versus independent advice directly. Soll (1999) went a step further by asking participants to choose between two sources of information, one of which was redundant to the information they already possessed, while the other was not. Furthermore, he manipulated whether measurement error in the scenario was high or low. Soll found that 61% of the participants preferred the redundant source, but only when measurement error was high[8].

[8] Soll assumed that individuals believe that there is a tradeoff between reducing measurement error, on the one hand, and reducing bias, on the other hand. Specifically, according to Soll, individuals think that using redundant tests/sources only reduces measurement error (which is true), and that using independent tests only reduces bias (which is false).

So far, only one study compared participants' reactions to dependent versus independent advice in the context of the JAS (Yaniv et al., 2009). Similar to the studies mentioned before, respondents in this study were more confident after receiving dependent advice compared to independent advice. Furthermore, they revised their initial judgments more frequently after receiving dependent advice, indicating a tendency to adjust their judgments in the direction of the advice. However, as recently demonstrated in an extended replication of this study (Wanzel, Schultze, & Schulz-Hardt, 2017), in the Yaniv et al. (2009) study interdependence of advice, on the one hand, and proximity of the advice to the judge's initial estimate, on the other hand, were confounded, and most of the effects described in the original paper seem to be driven by proximity.

In conclusion, the majority of the cited studies hint at the possibility that individuals might have a general preference for dependent advice, but they do not provide direct evidence for this idea. In the present paper, we aim to close this research gap by investigating individuals' preference for dependent vs. independent advice. Before we describe our studies in detail, however, we briefly address another important question, namely: is a preference for dependent advice necessarily a dysfunctional strategy? Everything else being equal, independent advice should lead to higher accuracy gains (as outlined above). However, in many real-life situations, all else is not equal and, hence, under certain circumstances individuals might benefit more from dependent advice, because the very same factors that introduce correlation among advisors' errors may also enhance their accuracy. A well-known example for this is a situation where advisors discuss their opinions before making a judgment. For instance, several financial analysts may discuss their ratings before submitting individual recommendations regarding credit risk. Here, discussion creates interdependence of opinions, because the advisors influence each other (their post-discussion opinions correlate more strongly than their uninfluenced pre-discussion opinions). At the same time, discussion often leads to improved accuracy of group members' post-discussion opinions, because the group members can exchange task-relevant information and correct each other's misconceptions (Schultze, Mojzisch, & Schulz-Hardt, 2012; Stern, Schultze & Schulz-Hardt, 2017). This natural confound is beneficial for the judge, and a preference for dependent advice would, therefore, often be the objectively correct strategy (regarding accuracy gains) when interdependence of opinions stems from advisors' joint discussion of the problem at hand.

In order to determine whether individuals differentiate between dependent and independent advice in a functional way, we conducted a series of six studies. In these studies, we investigated individuals' preference for independent vs. dependent advice both in a situation where dependent advice is more accurate than independent advice (Studies 1 and 2), as well as in a context where the opposite is true (Studies 3a, b and 4a, b). In Studies 1 and 2, dependent advisors discussed their estimates whereas independent advisors worked alone. In the discussion, the advisors can exchange the relevant information and correct each other's misconceptions. This should result in advice that is both more correlated and more accurate than that of a comparable number of independent advisors. In contrast, in Studies 3a, 3b, 4a and 4b , we used a computer-simulated scenario where dependent advisors shared part of their information, while independent advisors each held unique information. As described above, the overlap of information between advisors and the resulting lower amount of total information available to advisors leads to lower accuracy of dependent compared to independent advice. In Studies 1, 3a and 3b, participants could actively choose between two advisor teams (dependent vs. independent). In Studies 2, 4a and 4b, we used a different measure of advice preference, namely how much participants actually weighted dependent versus independent advice. Based on the aforementioned theoretical considerations, we formulated the following hypotheses.

Hypothesis 1: Judges prefer dependent advice by choosing it more often and weighting it more strongly, independent of whether dependent advice is more accurate or less accurate than independent advice.

Hypothesis 2a: Judges benefit more from dependent (compared to independent) advice if interdependence results from group discussion (dependent advice is more accurate than independent advice).

Hypothesis 2b: Judges benefit less from dependent (compared to independent) advice if interdependence results from informational overlap (dependent advice is less accurate than independent advice).

Study 1 – Choice of advice when interdependent advice is more accurate

Study 1 was designed to investigate participants' preference for dependent versus independent advice in a context where a preference for dependent advice would be beneficial for the judge. We measured preference by asking participants to choose between the two types of advice.

Method

Participants and design. Eighty-three participants were contacted via the Online Recruitment System for Economic Experiments (ORSEE, Greiner, 2015), via the social network Facebook, and via an online job forum for students (https://www.uni-goettingen.de/de/29718.html). Participants' age ranged from 19 to 62[9] years (*M* = 25.80, *SD* = 6.70), 27 of them were male, 54 female, and one person did not specify his or her gender. Twenty-three of those participants had indicated that they had already participated in a similar experiment (e.g., one of our earlier experiments on advice taking). In the present as

[9] Excluding those three individuals who were outliers regarding their age (50, 52 and 62 years) did not affect the result pattern.

well as in the following studies, we retained participants who had participated in a similar experiment in our analyses, as excluding these individuals did not change our results. The primary dependent variable was participants' choice of dependent versus independent advice.

Materials. We conducted a pre-test in order to generate a pool of dependent and independent advice which participants would receive in Studies 1 and 2. An ecologically valid manipulation of interdependence was realized by having advisors discuss their estimates in the dependent condition. This way, the individuals necessarily influenced each other, leading to dependent estimates.

For this pretest, 188 participants were contacted via ORSEE, Facebook, and the online job forum for students. Their task was to estimate the caloric value per 100 g/ml of 30 different foods on paper and pencil questionnaires. In the dependent condition (N = 93), participants first discussed each caloric value in groups of three. After the discussion, each group member made an individual estimate in private. In the independent condition, each participant (N = 95) worked individually on the 30 trials without discussing the judgments with other participants, and we later formed nominal groups of three independent advisors. We had to exclude three groups from this initial sample. One group had a member with extremely disruptive behavior[10], in the second group, only two of the three participants had shown up, and in the third group, the mean absolute percentage error (MAPE) deviated three standard deviations from the mean (the latter is a common criterion for excluding participants from analyses; cf. van Selst & Jolicoeur, 1994). Thus, 180 participants (118 female, 61 male, one person did not specify his or her gender; M_{age}= 23.84, SD_{age} = 4.25), were left for our analyses.

[10] This person was being loud and did not join the discussion, resulting in the other two group members working individually on the task.

We ran several analyses on the pre-test data. First, we tested whether the manipulation of interdependence of advice had worked as intended. To this end, we calculated the percentage errors per trial for each (nominal) group member. We then computed the inter-correlations of these errors per group (i.e. three bivariate correlations). Finally, we calculated the mean of these three correlations, tested the correlations for significance in the two conditions and afterwards compared the Fischer-Z-transformed correlations between dependent and independent (nominal) groups with a two-sample *t*-test. The average correlation was significant for dependent advice, $r(28) = .87$, $p < .001$, as well as for independent advice, $r(28) = .48$, $p = .007$. As predicted, the correlation of error was greater in the dependent advice condition compared to the independent advice condition ($r = .87$ vs. $r = .48$), $t(43.89)$[11] $= 7.41$, $p < .001$. We also measured the consensus in the groups by calculating the coefficient of variation (CoV). A *t*-test for independent samples revealed that, as expected, the CoV was lower in the dependent condition ($M = 0.21$, $SD = 0.10$) than in the independent condition ($M = 0.57$, $SD = 0.13$), $t(58)= -11.90$, $p < .001$, $d = -3.10$. Thus, dependent opinions showed greater consensus than independent opinions. Finally, we tested whether dependent advice was more accurate by comparing the MAPE scores of the advisor groups between conditions. A *t*-test for independent samples confirmed that dependent advice was more accurate ($M = 63.90$, $SD = 28.38$) than independent advice ($M = 88.46$, $SD = 50.85$), $t(58) = -2.31$, $p = .024$, $d = -0.62$. As we anticipated, creating interdependence by giving participants the opportunity to discuss their estimates increased both the interdependence of the advice and its accuracy.

[11] Throughout the manuscript, for *t*-tests fractional degrees of freedom are reported in the case of unequal variances.

In the main experiment, participants worked on the same calorie judgments as the advisors in the pre-test, this time on the computer. In contrast to the pre-test, the experiment was divided into two phases, each consisting of fifteen trials.

Procedure. Upon arriving in the laboratory, participants were each seated in front of individual computers. Participants were thanked for their willingness to participate and were informed that their participation was voluntary, and that they could quit the experiment at any time. Furthermore, it was explained to them that their data would be treated anonymously. They learned that, across 30 trials, they would be asked to estimate the caloric value of various foods, and to indicate how confident they were in their judgment. Additionally, they were informed that, on each trial, they would receive the estimates of three advisors who had worked on the same trials in a previous experiment. Participants learned that they would have the opportunity to choose between two different advisor trios, and that they would receive advice from the chosen trio for the next fifteen trials. Furthermore, they would receive a bonus of up to 3€ for particularly accurate judgments. Then, participants saw an example of the task in order to familiarize themselves with the task structure. Afterwards, they received detailed information about how the two advisor trios made their judgments: in trio A (the independent advisors), the advisors did not communicate while making their judgments and worked completely on their own. Trio B (the dependent advisors), on the other hand, discussed the target values and, after that discussion, each of the three advisors gave his or her final individual estimate (each of the 30 trials followed this discuss-then-estimate procedure). For better comprehension, participants also saw the respective instructions the advisors had received in the pre-test. Finally, the respondents were asked to choose one of the advisor trios.

After this choice, the first phase comprising 15 trials began. The order of the two phases was counterbalanced across the two conditions such that one group saw the trials of the other group's second phase as the first one. On each trial, participants were first asked to give an initial estimate for the caloric value of a specific food item (e.g. bananas) and to indicate their confidence in this estimate on a scale ranging from 0% to 100%, with intervals of 10%. Then, they were presented with the chosen advice in form of the estimates of three different advisors, who were drawn randomly from the respective condition of the pre-test pool (they received advice from the same set of advisors in all of the fifteen trials). Participants also saw their own initial estimate and were asked to make a final judgment along with a rating of their post-advice confidence.

After fifteen trials, participants received feedback regarding each of the trials. Specifically, for each trial they were presented with a table containing the true caloric value, their own final estimate, and the three pieces of advice they had received, so they could assess the accuracy of their advisors. Afterwards, they had to choose another advisor trio (dependent vs. independent advisors) for the second phase of the experiment. Participants knew that even if they chose the same type of advisors, they would not keep their chosen trio from the first phase. Instead, a new advisor trio would be chosen randomly from the pre-test data. This way, we could explore whether the same pattern (as for the first choice) would emerge when participants already had experience with the advice. Respondents were informed once more that they would receive advice from the chosen trio on the next fifteen trials. After finishing the second set of trials, they answered a suspicion check and indicated whether they had already participated in a similar experiment involving the estimation of the caloric value of various foods in the past. Finally, participants wrote down their rationale for choosing the type of advice in the first and the second phase.[12] Furthermore, they stated their

[12] In the present as well as in the following studies, we measured several variables, some of which were intended for further exploratory analyses. Throughout the manuscript, we report only the results for those variables that are central to our research questions, but provide all measured variables in the methods sections for greater transparency.

beliefs about how the advisors' judgments were affected by the discussion as compared to working alone. Participants were then paid and debriefed. The payment consisted of a show-up fee of 4€ and a bonus which depended on the accuracy of the final trials. Specifically, they received an additional 3€ when their average MAPE score was between 0% and 50%, 2€ when the MAPE fell into the range of >50% and 75%, 1€ when it fell into the range of >75% and 100%, and no bonus when the MAPE exceeded 100%.

Results and discussion

Main analyses. First, we explored whether there were any effects of the order of phases. Our analyses showed no significant effects for either phase, all $\chi s^2 < 0.80$, all $ps >$ .370. In order to test whether participants preferred dependent advice, we ran a two-tailed binomial test on the first decision. As hypothesized, the test revealed a preference for dependent over independent advice above chance level. Fifty-five (66%) of the participants chose dependent advice, while twenty-eight (34%) of the participants chose independent advice, $p = .005$. Since dependent advice in this experiment was more accurate than independent advice, participants made a normatively correct first decision, based on the information they had received.

Exploratory analyses[13]. We also examined participants' second choice in order to investigate if we would find the same overall pattern after they had actually seen the advice. Since dependent advice was more accurate, one might expect that only participants who had chosen independent advice for the first phase changed their opinion afterwards (at least, when they recognized the advice's accuracy on the basis of the true value, presented to them in form of the feedback table after Phase 1). In fact, the same overall preference emerged, but the difference was no longer statistically significant. In their second decision, 46 (56%)

[13] The results for accuracy gains, confidence gains and weighting of the advice for Studies 1, 3a and 3b are presented in appendices C through E.

participants chose dependent advice, and 37 (44%) chose independent advice, *p* = .32. Descriptively, 10 of the 28 (36%) participants who selected independent advice in the first phase chose dependent advice in the second. Of the 54 participants who chose dependent advice in the first phase, 18 (33%) chose independent advice in the second. To check whether participants changed their opinion based on the advice's accuracy, we first calculated a variable indicating whether participants changed their opinion or not. We then ran a *t*-test with the MAPE of the advice as the dependent variable and change of opinion as grouping variable. Descriptively, those who changed their opinion had received slightly less accurate advice than those who chose the same type of advice in both decisions (M = 77.08, SD = 62.53 versus M = 71.95, SD = 40.11), but this difference in accuracy was far from statistical significance, $t(81)$= -.45, p = .65, d = .11. Therefore, the advice's accuracy cannot explain why some participants changed their opinion. Instead, a possible explanation for their behavior might be variety seeking, meaning that participants may have changed their preferences for the sake of trying out something new (e.g., Givon, 1984).

In sum, we have seen that, as predicted, participants prefer dependent advice by choosing it more frequently than independent advice. It is important to note that while we could also have investigated the weight of advice in Study 1, any differences we might have found would have been difficult to interpret, because participants self-selected themselves to the different types of advice they received. Therefore, we would have been unable to tell whether a possible difference in weighting was actually due to differences in the interdependence of advice, or whether differences might have been masked by personal differences between the two types of participants. For this reason, our next study addresses the question whether the preference for dependent advice can also be found in participants' weighting behavior.

Study 2 – Advice taking when interdependent advice is more accurate

Based on what we found in Study 1, in Study 2 we wanted to explore whether respondents would still display the same preference for interdependence when actually weighting the advice. The important difference is that in this second study participants were able to compare dependent and independent advice since they received both types of advice. We measured advice weighting using the advice taking score (AT, Harvey and Fischer, 1997):

$$AT = \frac{final\ estimate - initial\ estimate}{advice - initial\ estimate}$$

The AT score represents the percent weight of advice when making the final judgment. In Study 2, we avoided self-selection: instead of choosing one set of advisors, all participants received both dependent and independent advice, and we measured how much they weighted each type of advice.

Method

Participants and design. The experiment employed a 2 (advice interdependence: independent vs. dependent advice) x 2 (order of presentation: dependent advice first vs. independent advice first) mixed design with advice interdependence as a within-subject factor and order as a between-subject factor. Advice was drawn from the pool generated in the pre-test of Study 1.

Sixty-seven participants[14] were contacted via ORSEE, Facebook, and the online job forum for students. Participants ranged in age from 19 to 62 years[15] (*M* = 25.55, *SD* = 6.38), 26 of them male and 41 female.

---

[14] One person had no yoking partner because of an odd-numbered sample. Excluding this person did not affect the result pattern, therefore this person was retained in our data.

[15] Excluding three individuals who were outliers regarding their age (44, 45 and 62 years, respectively) did not affect the pattern of results.

Procedure. The procedure of Study 2 was identical to that of Study 1, with the following exceptions: First, before working on the actual trials, participants were presented with ten practice trials, each containing a calorie estimate and a confidence judgment on a scale from 0% to 100%, with intervals of 10%. Then, they provided answers to four questions regarding these trials: they should indicate on an 11-point Likert-scale ranging from 0 (not at all) to 10 (absolutely) how accurate they thought their judgments were. This was followed by indicating which rank they thought they would reach among 100 persons. Next, participants were asked on an 11-point Likert-scale ranging from 0 (none) to 10 (all of them) how many of their ten practice judgments would deviate from the true value by no more than ten percent. Finally, they indicated whether they had previously elaborated on the caloric value of foods more closely, for example, in the context of a fitness program.

The second exception was that, instead of choosing the type of advice, all participants received dependent advice on half of the trials and independent advice on the other half. The order of advice type was counterbalanced. For a completely balanced design, we yoked participants between the two order conditions. Each participant who received dependent advice first was paired with another participant who received independent advice first. The set of advisors was identical for the two yoked respondents, that is, one participant received the same advice in the second phase that the other participant had received in the first phase (and vice versa). At the beginning of each phase, participants learned how their advisors had made their judgments. As a check of whether participants correctly understood and recognized the instructions, respondents were then asked to describe how the advisors had made their judgments.

Third, after each phase, respondents were asked to answer three questions on 11-point Likert-scales. These questions were designed to investigate participants' understanding of the potential consequences of interdependence/independence. First, they were asked to indicate on a scale from 0 (not at all) to 10 (very much) whether they felt that they had received new information through the advisors' estimates. Second, in order to test whether participants had read and understood the instructions, we asked them to state how strongly, in their opinion, the estimates of the advisor trio were influenced by each other on a Likert scale ranging from 0 (not at all) to 10 (very much). Third, they were asked on a scale form 0% (not at all accurate) to 100% (completely accurate) how accurate the three pieces of advice were on average. They were also asked to provide reasons for their answers to the three above-mentioned questions, using open-response formats.

Results and discussion

Instruction check. To explore whether participants had experienced dependent and independent advice differently, we had asked them to indicate to what extent the three estimates were influenced by each other. We calculated a 2 (advice interdependence: independent vs. dependent advice) x 2 (order of presentation: dependent advice first vs. independent advice first) mixed ANOVA with advice interdependence as the within-subject variable. We found a main effect of advice interdependence, $F(1, 65) = 76.83$, $p < .001$, $\eta_p^2 = .54$, a main effect of order of presentation, $F(1, 65) = 14.93$, $p < .001$, $\eta_p^2 = .19$, and an interaction effect, $F(1, 65) = 6.68$, $p = .012$, $\eta_p^2 = .11$. Regarding the main effect of advice interdependence, we found that participants correctly rated dependent estimates as being more influenced by each other ($M = 6.34$, $SD = 2.20$) compared to independent estimates ($M = 3.48$, $SD = 2.61$). Also, participants thought that the estimates were more influenced by each other when independent advice was presented first ($M = 5.75$, $SD = 2.20$) compared to when dependent advice was presented first ($M = 4.05$, $SD = 2.95$). Investigating the interaction further, we found that dependent pieces of advice were always rated as being more influenced by each other than independent advice; this difference was stronger, however, when dependent advice was presented first (*M* = 5.91, *SD* = 2.37; *M* = 2.18, *SD* = 2.21), $F(1, 32) = 79.50$, $p < .001$, $\eta_p^2 = .71$, compared to independent advice being presented first (*M* = 6.76, *SD* = 1.93; *M* = 4.74, *SD* = 2.35), $F(1, 33) = 16.22$, $p < .001$, $\eta_p^2 = .33$. Thus, in principle, participants correctly understood and were aware of the manipulation of advice dependence.

Main analyses

Weighting of advice. We first computed the mean AT scores as a measure of advice taking for each trial. In line with previous studies (e.g., Gino, Brooks, & Schweitzer, 2012; Schultze, Rakotoarisoa, & Schulz-Hardt, 2015; Soll & Larrick, 2009), we truncated AT scores at 0 and 1. We then computed mean AT scores for each participant and phase. A 2 (advice interdependence: independent vs. dependent advice) x 2 (order of presentation: dependent advice first vs. independent advice first) mixed ANOVA on the mean AT scores revealed a main effect of advice interdependence, $F(1, 65) = 80.11$, $p < .001$, $\eta_p^2 = .55$. As predicted, participants gave more weight to the advice in the dependent condition (*M* = 0.41, *SD* = 0.21) than in the independent condition (*M* = 0.22, *SD* = 0.14). The main effect of order of presentation and the interaction effect were not significant, $F(1, 65) = 1.00$, $p = .32$, $\eta_p^2 = .02$, and $F(1, 65) = 0.55$, $p = .46$, $\eta_p^2 = .01$. Table 1 in Appendix A depicts the mean values and standard deviations for all of our main dependent variables.

To explore whether the effect is robust, we used a second advice taking measure. To this end, we calculated the frequency of revision, that is, the percentage of trials in which participants gave final estimates different from their initial estimates. For this measure, a 2 (advice interdependence: independent vs. dependent advice) x 2 (order of presentation: dependent advice first vs. independent advice first) mixed ANOVA revealed a main effect of advice interdependence, $F(1, 65) = 47.42$, $p < .001$, $\eta_p^2 = .42$. In line with the results for the AT scores, participants changed their estimates more often in the dependent condition (68%)

than in the independent condition (53%), again indicating that they used dependent advice more. The main effect of order of presentation and the interaction effect were not significant, $F(1, 65) = 0.75$, $p = .39$, $\eta_p^2 = .01$, and $F(1, 65) = 3.64$, $p = .06$, $\eta_p^2 = .05$. Hence, the results of Study 2 consistently show that, as hypothesized, participants weight dependent advice more strongly and change their initial opinion more frequently after receiving dependent advice as compared to independent advice. Table 1 depicts the mean values and standard deviations for all of our main dependent variables.

Table 1. Overview of main results of Study 2 (Means with Standard Deviations in parentheses).

	Method of Sampling	
Measure	Independent	Dependent
Accuracy		
initial MAPE	**82.06** (46.28)	**83.82** (43.10)
final MAPE	**71.59** (41.79)	**62.75** (33.42)
improvement (difference scores)	**10.46** (17.22)	**21.07** (29.90)
Weighting	**.22** (.14)	**.41** (.21)
Revision Process		
rate of changing initial estimates (%)	**53** (26)	**68** (24)
Confidence		
initial confidence	**43.91** (22.83)	**42.57** (22.31)
final confidence	**48.00** (23.40)	**47.44** (23.88)
confidence gain	**4.09** (6.87)	**4.88** (9.41)

Accuracy gains. We had also hypothesized that, in the situation implemented in Study 2, participants would benefit more from dependent as compared to independent advice. To test this hypothesis, we first explored whether there were any accuracy gains at all in the two conditions by running two *t*-tests against zero on the difference scores between the initial and the final MAPE (positive values of this measure indicate a reduction in MAPE scores, that is, an increase in accuracy). We then compared the two conditions regarding their accuracy gains by calculating a 2 (advice interdependence: independent vs. dependent advice) x 2 (order of presentation: dependent advice first vs. independent advice first) repeated measures ANOVA. We found significant accuracy gains in the dependent (M = 21.07, SD = 29.90) as well as in the independent condition (M = 10.46, SD = 17.22), $t(66)$ = 5.77, p < .001, d = 1.42, and $t(66)$ = 4.98, p < .001, d = 1.23. Furthermore, the ANOVA revealed a significant main effect of advice interdependence, $F(1, 65)$ = 8.25, p = .006, η_p^2 = .11: Participants' accuracy gains were greater in the dependent condition compared to the independent condition. Therefore, we can confirm Hypothesis 2a. The main effect of order of presentation and the interaction effect were not significant, all Fs < 1.73, all ps > .19, all η_p^2 < .026.

An interesting question that arises is whether participants in the dependent advice condition benefitted more from the advice because they weighted it more, or because it was more accurate, or both. To explore this question, we first calculated a difference score between the accuracy gains (the difference score between the MAPEs of the initial and final estimates) in the dependent condition and those in the independent condition. Positive scores indicate higher accuracy gains in the dependent condition. We then entered this variable as the criterion in a regression analysis. Furthermore, we calculated the difference score between the AT values in the dependent and the independent condition (positive values indicate greater weight placed on the advice in the dependent condition) as well as a difference score between the MAPEs for dependent and independent advice (positive values indicate lower accuracy/greater errors in the dependent condition). We mean-centered those two variables and entered them as well with their interaction term as predictors in the regression analysis. The regression revealed that both the difference in the MAPEs of the advice, b = -0.11, t(63) = -2.50, p = .015, as well as the difference in the AT values, b = 57.46, t(63) = 2.77, p = .007, significantly predicted participants accuracy gains. The interaction was not significant, b = -0.32, t(63) = -1.23, p = .22.

Exploratory analyses

Perceived and objective advice accuracy. Yaniv and Kleinberger (2000) demonstrated that individuals are able to identify (in)accurate advice even in the absence of performance feedback or a-priori information about the expertise of the advisors. Since participants saw the advice before they decided whether to heed it, it is conceivable that the preference for dependent advice in our study is, at least to some extent, a preference for accurate advice. In fact, participants correctly rated dependent advice to be more accurate than independent advice (M = 57.01, SD = 19.10 vs. M = 43.73, SD = 18.60), F(1, 65) = 28.64, p < .001, η_p^2 = .31, indicating that they were able to recognize which type of advice was more accurate. The effect of order was also significant, with participants rating the advice as more accurate when independent advice was presented first compared to when dependent advice was presented first (M = 54.26, SD = 19.50 vs. M = 46.36, SD = 19.70), F(1, 65) = 4.45, p = .039, η_p^2 = .06. The interaction effect was not significant, F(1, 65) = 2.88, p = .095, η_p^2 = .04. To test whether our findings for weighting were (at least in part) due to a preference for accurate advice, we repeated the analysis on the AT scores, this time controlling for advice accuracy (measured as the MAPE of the advice). To this end, we analyzed AT scores in a linear mixed model, with type of advice and order of presentation as independent variables and advice accuracy as a covariate. A random intercept model was used. The effect of condition was still significant, F(1, 64) = 24.70, p < .001. The main effect of the covariate was also significant, F(1, 64) = 5.81, p = .019. There were no other significant effects, all *Fs* < 0.55, all *ps* > .46. Hence, even when the accuracy of the advice was held constant, participants still preferred dependent advice.

Confidence gains. To possibly add to the understanding of our main results, we also investigated respondents' gains in confidence. We first calculated confidence shifts as the difference between the confidence in the final estimate and confidence in the initial estimate (positive values indicate an increase in confidence). One-sample *t*-tests revealed that confidence increased in the dependent (M = 4.88, SD = 9.41) as well as in the independent condition, (M = 4.09, SD = 6.87), $t(66) = 4.24$, $p < .001$, $d = 1.04$ and $t(66) = 4.88$, $p < .001$, $d = 1.20$. A 2 (advice interdependence: independent vs. dependent advice) x 2 (order of presentation: dependent advice first vs. independent advice first) repeated measures ANOVA on the confidence shifts explored differences between the two conditions. We found no significant effects, all *Fs* < 0.96, all *ps* > .33, all $\eta_p^2 < .015$. Thus, participants actually benefitted from weighting the dependent advice more, and they also perceived dependent advice as more accurate, but the stronger increase in accuracy when receiving dependent advice was not reflected in participants' confidence ratings.

Study 2 shows that, just like in Study 1, participants have a clear preference for dependent advice. Since participants also gained more from dependent advice, a preference for the latter is indeed favorable in the very situation that we tested in these two studies. Furthermore, although participants correctly believed dependent advice to be more accurate, the preference for dependent advice remained even after controlling for the accuracy of the advice. Thus, participants' perceptions of advice quality are no sufficient explanation for their preference for dependent advice. Instead, individuals might actually prefer dependent advice because of its consensus and, thus, might display this preference even when dependent advice is less accurate than independent advice. We investigated this possibility in our next studies by exploring participants' choice and weighting behavior when dependent advice is less accurate. Before describing the actual studies, we first explain the paradigm used in both of these studies.

Studies 3a and 3b – Choice of advice when interdependent advice is less accurate

Task

In Studies 3 (a, b) and 4 (a, b), we investigated the effects of interdependence versus independence on choosing and weighting advice in a situation where interdependence is detrimental to the advice's accuracy. To this end, we employed a task developed by Schultze (2015), which was slightly adapted for the purposes of the present research. In this task, participants estimate the annual precipitation at several locations in Asia unknown to them, and they do so based on a set of independent weather stations, each of which provides one measurement. In our studies, participants received data from three weather stations on each trial to make their initial estimates. Furthermore, they received the estimates of four advisors for their final judgments. Participants knew that the advisors were simulated by the computer, based on the typical behavior of real participants. Using simulated advisors, we were able to ensure experimental control and fine-tuning of the parameters we manipulated. Similar to the judge, each advisor had access to three weather stations using the same technology (i.e., the reliability of their measurements was identical to the reliability of the judges' measurements). Dependent advisors shared two of the three stations. Therefore, the overall information base was smaller in the case of dependent advisors (eight measurements) as compared to independent advisors (twelve measurements). Second, the errors of dependent advisors should be positively correlated, whereas those of independent advisor should be uncorrelated. Accordingly, an aggregate of the four pieces of advice should be (on average) more accurate in the independent condition than in the dependent condition.

Purpose of Study 3a

Study 3a was designed to investigate participants' choice of dependent versus independent advice within a scenario where interdependence has a detrimental effect on the advice's accuracy. If they still choose dependent advice more often than independent advice this would suggest that participants have a tendency to be guided by cues suggesting interdependence rather than by accuracy cues. If, on the other hand, they choose independent advice more often, this indicates that they are able to make a functional choice.

Method

Participants and design. Fifty-nine participants (25 male, 34 female) were contacted via the same recruiting channels as in the other studies. Their age ranged from 18 to 36 years (*M* = 22.90, *SD* = 3.52). The primary dependent variable was participants' first choice of dependent versus independent advice.

Procedure. The procedure of Study 3a was similar to that of Study 1. Therefore, we report only the differences in the instructions. Participants were first asked to imagine they were meteorologists and had to judge the annual precipitation amount at several locations in Asia in millimeters per square meter, based on data from three weather stations. They learned that each weather station's measures were unbiased but prone to unsystematic measurement error. Furthermore, each weather station produced an independent measurement of the true value. Data entry was limited to values from 1 to 5,000. Participants were also informed that they would receive four pieces of advice by computer-generated advisors who held similar positions as meteorologists, and that they could choose between two advisor quartets. They also learned that each advisor used the same technology that participants used, that is, the average quality of the measurements was identical for judge and advisors. Afterwards, they were informed that in quartet A all advisors had access to three different weather stations (i.e., an advisor's weather stations were all different from both the judge's and the other advisors' weather stations), whereas advisors in quartet B shared two of the three weather stations (i.e., each advisor had one unique and two shared measurements, but the three measurements were still different from those of the judge). All measurements were drawn from a normal distribution centered on the true values which were retrieved from the database of the German Weather Service (DWD), with a standard deviation of one third of the true value. Drawings were truncated at values of 10 on the low end and at 5,000 on the high end, in order to prevent negative or excessively small or large values. For each advisor, we computed advice by averaging over the three measurements (note that the mean of the measurements is the best unbiased linear estimate for the true values in this experiment).

After having seen the descriptions of the advice, participants chose a type of advice and proceeded with the fifteen trials of estimating precipitation amounts. Following the fifteen trials during which they made estimates of precipitation amounts, participants had the opportunity to choose again a team of advisors which they would keep for the next fifteen trials.

Participants received a show-up fee of 5€. In addition, they could earn up to 3€ as a bonus depending on the accuracy of their final estimates. Specifically, they received 3€ when their MAPE was lying between 0% and 6%, 2€ when the MAPE fell into the range of > 6% and 12%, 1€ when it fell into the range of >12% and 18%, and nothing when the MAPE exceeded 18%.

Results and discussion

Control measures. Since we were primarily interested in participants' first choice, only the results of Phase 1 are relevant for a strict test of our hypothesis. However, for exploratory purposes we also analyzed the second phase and report the corresponding results.

First, we calculated the inter-correlations of the percentage errors per group and compared these correlations between dependent and independent groups of advisors. As predicted, advisors' errors were significantly correlated in the dependent, $r = .65$, $p = .009$, but not in the independent condition, $r = .02$, $p = .944$. The same was true for the second phase $r =.72$, $p = .002$ and $r = -.01$, $p = .97$. Furthermore, the correlation of errors was significantly higher for dependent advice than for independent advice in the first as well as in the second phase, $t(24.80)= -13.28$, $p < .001$, and $t(19.98) = -17.90$, $p < .001$. To find out which type of advice was more accurate, we compared the MAPE scores of dependent and independent teams of advisors. A two-tailed *t*-test revealed that, as expected, independent advice was more accurate ($M = 7.62$, $SD = 1.67$) than dependent advice ($M = 13.30$, $SD = 3.30$), $t(24.11) = -7.25$, $p < .001$, $d = -2.43$. The same was true for Phase 2 ($M = 7.66$, $SD = 1.41$; $M= 13.92$, $SD = 2.45$), $t(20.44) = -9.90$, $p < .001$, $d = -3.55$. Furthermore, in Phase 1, the CoV was greater for independent advice ($M = 0.17$, $SD = 0.02$) than for dependent advice ($M = 0.11$, $SD = 0.01$), $t(57) = 15.10$, $p < .001$, $d = 4.15$, suggesting greater consensus for the latter. We found the same pattern for Phase 2 ($M = 0.18$, $SD = 0.02$; $M = 0.10$, $SD = 0.01$), $t(57) = 13.41$, $p < .001$, $d = 3.86$. Thus, the advice type manipulation was successful.

Main analyses. As in Study 1, we first investigated whether the order of phases had any effect on participants' choices. We found no significant effects of order for the first nor the second choice, all $\chi^2(1) < .42$, all $p > .519$. To find out if participants had a preference for dependent or independent advice, we ran a two-tailed binomial test. This test revealed that the observed proportion of the first decision for dependent versus independent advice was different from chance. Specifically, 39 (66%) of the participants chose independent advice, while only 20 (34%) of the participants chose dependent advice, $p = .018$.

As in Study 1, we also examined participants' second choice after they had seen the actual advice. This time, even more participants chose independent advice, namely 42 (71%), while only 17 participants (29%) chose dependent advice, *p* = .002. Descriptively, only 8 of the 20 participants (= 40%) who chose dependent advice in the first phase chose it again in the second phase. Thirty of the 39 participants (= 77%) who chose independent advice in the first phase chose it again in the second[16]. In other words, participants exhibited a substantial preference for independent (i.e. the more accurate type of) advice. This means that participants were able to choose the more accurate advice.

Purpose of Study 3b

The advice sampling procedure was more abstract and unfamiliar to our participants in Study 3a compared to Studies 1and 2. Therefore, to make sure that the results still hold true when the preconditions are less complex, we presented participants with a table displaying example values of two randomly selected potential advisor teams (one with overlapping and one with non-overlapping information) across 15 trials before they chose the independent or the dependent advice. The participants made no judgments in these trials; they just read the two series of advisor values that were provided for each of the trials. Half of the participants also received the true value during these trials. We did so for the following reason: From the example values, advisor consensus is instantly apparent, whereas participants can only infer the accuracy of the advice when they have sufficient task-relevant knowledge, and this might give a certain advantage to the dependent advice. For accuracy to become apparent, the true value has to be provided. However, providing the true value might make accuracy extremely salient and, hence, might almost put a demand effect on participants to choose the more accurate advice, which might then give an advantage to the

[16] While participants could technically observe the advice's consensus in the first phase, there was still no possibility for them to compare the two types of advice regarding their level of consensus as they only had seen the chosen advice.

independent advice. To counterbalance these possibilities, we implemented them as two different experimental conditions.

Method

Participants and design. Sixty-one participants (24 male, 36 female, one participant did not indicate her or his gender) were contacted via ORSEE, Facebook, and the online job forum for students. Their age ranged from 18 to 65 years[17] (M = 25.89, SD = 6.43). The primary dependent variable was participants' first choice of dependent versus independent advice.

Procedure. The procedure of Study 3b was similar to that of Study 3a. Therefore, we report only the differences in the instructions. After learning how the two advisor quartets (dependent and independent advisors) made their judgments, participants were then presented with a table with exemplary values for fifteen trials, to help them in getting an idea of the advice. That is, before choosing between independent and dependent advice, participants saw example pieces of advice from both types of advisor quartets. This way, participants could compare the advice regarding, for example, its consistency. The example values were drawn randomly for each participant. It was explained to participants that the chosen advisor quartet would not be identical to one of the two examples shown to them; instead, their advisor quartet would be randomly drawn after they had decided for one of the two advice options. We manipulated between participants whether the exemplary table also contained the true values to explore whether participants rely on accuracy cues rather than cues about advisor consensus when choosing a team of advisors. After having seen the descriptions of the advice as well as the table, participants chose a type of advice and proceeded with the fifteen trials

[17] Excluding two individuals who were outliers regarding their age (40 and 65 years, respectively) produced the same result pattern.

of estimating precipitation amounts. Since in this study we were primarily interested in finding out which type of advice participants preferred for their first choice and how this choice was influenced by the additional information contained in the table, there was no second decision and no second phase.

Results and discussion

Manipulation check. As in the previous studies, we first calculated the inter-correlations of the percentage errors per advisor quartet and compared these correlations between dependent and independent groups of advisors. As predicted, advisors' errors were highly correlated in the dependent advice sets, $r = .657$, $p = .008$, whereas errors were uncorrelated in the independent advice sets, $r = -.017$, $p = .952$. The difference between the two types of advice was significant, $t(27.45) = -31.70$, $p < .001$, $d = -11.57$. Furthermore, the CoV was greater for independent advice ($M = 0.18$, $SD = 0.01$) than for dependent advice ($M = 0.11$, $SD = 0.004$), $t(19.54) = 25.96$, $p < .001$, $d = 9.48$, suggesting greater consensus for the latter. To find out which type of advice was more accurate, we computed the average of the four pieces of advice for each advisor quartet and computed MAPE scores based on these aggregates. We found that, as expected, independent advice was more accurate ($M = 7.47$, $SD = 0.80$) than dependent advice ($M = 13.14$, $SD = 1.12$), $t(25.46) = -15.95$, $p < .001$, $d = -5.83$. Thus, the advice type manipulation was successful.

Main analyses. A two-tailed binomial test showed that the observed distribution of participants' choices for dependent versus independent advice was different from an equal distribution. As predicted, considerably more participants chose independent advice over idependent advice (45 vs. 16, or 74% vs. 26%), $p < .001$. Hence, participants preferred the type of advice that was more accurate. This preference was independent of whether or not the participants were shown the true values in the introductory examples, $\chi^2(1) = 0.66$, $p = .42$.

Descriptively, 20 out of 29 participants (69%) chose independent advice when the true value was displayed. When the true value was not displayed, 25 out of 32 participants (78%) chose independent advice.

In sum, the results of studies 3a and 3b clearly show that, when being given the choice between independent and dependent advice in a situation where independent advice is more accurate, participants predominantly selected the independent advice. Hence, when also taking the findings from Study 1 into account, it becomes clear that participants do not exhibit a general preference for dependent advice in their advice selection behavior, but rather pick the type of advice that is more accurate in the particular situation. Our next (and final) aim was to investigate whether the same functional preference is also present in participants' advice weighting.

Studies 4a and 4b – Advice taking when interdependent advice is less accurate

In Study 2, we had found that, when being given both independent and dependent advice, participants weight dependent advice more strongly in a situation where the dependent advice is more accurate. However, does this weighting behavior change if, as it is usually the case, dependent advice is *less* accurate than independent advice? We address this question in the fourth study. Study 4 is divided into Parts A and B because, in reaction to a methodological issue, we ran the same study twice. The specific reason for this doubling is explained in the "procedure" section.

Study 4a

Method

Participants and design. For Study 4a, 63[18] participants were contacted via ORSEE and through the before-mentioned recruiting channels. They ranged in age in age from 16 to 49[19] years (*M* = 23.21, *SD* = 4.62). Twenty-six of them were male and 37 female. The experiment employed a 2 (advice interdependence: independent vs. dependent advice) x 2 (order: dependent advice first vs. independent advice first) mixed design. Advice interdependence was manipulated within participants, and order was randomized between participants.

Procedure. The procedure of Study 4a was similar to that of Study 2, but the task (estimating precipitation amounts), the manipulation of advice dependence, and the payment scheme were the same as in Study 3. The only difference was that, after each phase, participants were asked how similar they thought the pieces of advice were. This question was designed to measure the subjective (i.e. perceived) consensus of the advice.

During our analyses, we detected a programming error due to which the estimate of the third advisor was missing in the data file, while the advice of the fourth advisor was saved twice. This programming error did not affect the study itself, because participants in both conditions were displayed the correct values, that is, they saw the four different estimates. Technically, this loss of data has consequences for all analyses that require computations based on the individual pieces of advice. Specifically, this concerns the manipulation checks and the analysis of advice weighting. However, the missing piece of advice is more of a nuisance rather than a fundamental problem for the following reason: In all of the above-

[18] Five persons had no yoking partner. Excluding these individuals did not affect our result pattern, therefore we retained all participants in our analyses.

[19] Excluding the 16- and the 49 year old (as they were outliers with respect to their age) did not affect the result pattern.

mentioned analyses, we can compute the relevant measures using the three correctly saved pieces of advice. For example, we can estimate advisor consensus based on the similarity of those three pieces of advice that were correctly saved in the data file. Naturally, the respective estimates will slightly deviate from the actual values we would have obtained if we had used all four pieces of advice. However, these differences are unsystematic. As such, they do not introduce systematic biases, but rather add to the error variance in our statistical tests. Since this works against our hypotheses, we decided to retain the study and report its results based on the three pieces of advice. However, to fully ensure that our effects are valid, we ran the experiment again and report this experiment later as Study 4b.

Results and discussion

Manipulation checks. First, we checked whether the correlation of errors in Phase 1[20] differed significantly from zero in the two conditions. As predicted, errors were significantly correlated in the dependent, $r = .67$, $p = .006$, but not in the independent condition, $r = -.01$, $p = .97$. Next, we investigated if, as intended, the errors of dependent advisors had a higher inter-correlation than those of independent advisors. As expected, the correlation of errors was significantly higher for dependent advice than for independent advice, $t(62) = -33.92$, $p < .001$, in Study 4a. Regarding the advice's accuracy, a paired *t*-test showed that, as intended, independent advice was indeed more accurate ($M = 8.72$, $SD = 1.96$) than dependent advice ($M = 13.24$, $SD = 2.40$), $t(62) = -13.68$, $p < .001$, $d = -0.83$. Moreover, a paired *t*-test with the CoV as the dependent variable revealed that variability was greater for independent advice ($M = 0.17$, $SD = .02$) than for dependent advice ($M = 0.10$, $SD = .01$), $t(62) = 22.44$, $p < .001$, $d = 0.94$, indicating higher consensus for dependent advice. To analyze participants' ratings of the similarity of the advice provided by the advisors, we calculated a 2 (advice

[20] As participants were yoked to each other, the advice in the second phase was the same the yoked partner had received in the first phase, meaning that the result is the same. Therefore, it is not included in this analysis.

interdependence: independent vs. dependent advice) x 2 (order of presentation: dependent advice first vs. independent advice first) ANOVA and found a main effect of advice interdependence. Since dependent advice had lower variation, it should also be rated as more similar by our participants. Indeed, they accurately perceived information contained in dependent advice as being more similar (M = 6.59, SD = 2.01) as compared to information contained in independent advice (M = 5.49; SD = 1.96), $F(1, 61) = 10.19$, $p = .002$, $\eta_p^2 = .14$. The main effects of order and the interaction effects were not significant, all $Fs < 0.55$, all $ps > .464$, all $\eta_p^2 < .02$. Hence, our experimental manipulation was successful, and dependent advice was also perceived as having greater consensus.

Main analyses

Weighting of advice. In order to test whether participants weighted dependent advice more strongly than independent advice, we calculated mean AT scores per participant and phase and analyzed them in a 2 (advice interdependence: independent vs. dependent advice) x 2 (order of presentation: dependent advice first vs. independent advice first) repeated measures ANOVA. As predicted, participants weighted dependent advice more (M = 0.55; SD = 0.21) than independent advice (M = 0.44; SD = 0.19), $F(1, 61) = 21.74$, $p < .001$, $\eta_p^2 = .26$, although the former was – on average – less accurate. The main effect of order of presentation and the interaction effect were not significant, all $Fs < 0.60$, all $ps > .444$, all $\eta_p^2 = .02$.

Similar to Study 2, we also investigated the frequency of revision as an alternative measure of advice weighting. A 2 (order of presentation) x 2 (advice interdependence) repeated measures ANOVA revealed a main effect of advice interdependence, $F(1, 61) = 9.85$, $p = .003$, $\eta_p^2 = .14$ and no significant order or interaction effects, all $Fs < 0.16$, all $p > .700$, all $\eta_p^2 < .01$. In line with the findings for the AT scores, participants changed their estimates more often in the dependent condition than in the independent condition (80% vs.

72%), thereby displaying a preference for dependent advice. Table 2 depicts the values of the main dependent variables.

Table 2. Overview of main results of Study 4a (Means with Standard Deviations in parentheses).

	Method of Sampling	
Measure	Independent	Dependent
Accuracy		
initial MAPE	**19.37** (8.51)	**18.82** (8.37)
final MAPE	**10.34** (3.07)	**12.57** (3.25)
improvement (difference scores)	**9.03** (7.74)	**6.25** (8.48)
Weighting	**.44** (.19)	**.55** (.21)
Revision Process		
rate of changing initial estimates (%)	**72** (23)	**80** (21)
Confidence		
initial confidence	**42.17** (25.82)	**42.81** (23.88)
final confidence	**47.38** (25.60)	**47.02** (24.74)
confidence gain	**5.21** (10.81)	**4.20** (7.48)

Accuracy gains. As in Study 2, we also investigated whether participants actually benefitted from the advice. Three persons were excluded from the analyses on accuracy gains, as their difference scores deviated more than three standard deviations from the mean. First, we computed accuracy gains in the same fashion as in Study 2. One-sample *t*-tests revealed significant accuracy gains in the independent (M = 9.03, SD = 7.74) as well as in the dependent condition (M = 6.25, SD = 8.48), $t(62) = 9.26$, $p < .001$, $d = 2.35$, and $t(62) = 5.85$, $p < .001$, $d = 1.49$. Although the means indicate somewhat larger accuracy gains in the independent as compared to the dependent condition, the corresponding main effect in a 2 (advice interdependence: independent vs. dependent advice) x 2 (order of presentation: dependent advice first vs. independent advice first) mixed ANOVA failed to reach conventional levels of significance, $F(1, 61) = 3.62$, p 0 .062, $\eta_p^2 = .056$ (the other effects were also insignificant, all $Fs < 0.76$, all $ps > .38$, all $\eta_p^2 < .013$. Hence, although the independent advice was clearly more accurate, participants' underweighting of this advice (relative to the dependent advice) prevented them from benefitting more from it. Table 2 depicts the values of the main dependent variables.

This reasoning is further underlined if we, similar to Study 2, analyze to what extent relatively higher accuracy gains from independent as compared to dependent advice depend on the amount that the two types of advice are taken into account. To this end, as Study 2, we first calculated a difference score between the accuracy gains in the independent versus the dependent condition, such that positive scores indicate higher gains in the independent condition. This variable was entered as the criterion in a regression analysis. We then calculated the difference score between the AT values (positive scores indicate higher weighting in the dependent condition) as well as the difference score between the MAPEs of dependent and independent advice (positive scores indicate lower accuracy in the dependent condition). We mean-centered these variables for their interaction term and entered them together with the latter in the regression analysis. We found that only the difference score between the AT values was a significant predictor of participants' accuracy gains[21], b = -19.36, $t(59) = -2.25$, $p = .029$, indicating that, as could be expected, individuals achieved higher accuracy gains when they weighted dependent advice less. Or, to put the same result differently, the more the participants weighted the dependent advice relative to the independent advice, the less they benefitted from the superior accuracy of the latter.

[21] The results for the difference score of the advice's MAPEs and for the interaction term were b = 0.30, t(59) = 0.05, $p = .96$, and b = 2.11, t(59) = 0.69, $p = .49$, respectively.

.***Exploratory analyses***

Confidence gains. As in Study 2, we also investigated whether participants displayed gains in confidence after receiving advice. One sample *t*-tests on the confidence shifts revealed significant confidence gains both in the independent (M = 5.21, SD = 10.81) as well as in the dependent condition (M = 4.20, SD = 7.48), $t(62) = 3.82$, $p < .001$, $d = 0.97$, and $t(62) = 4.46$, $p < .001$, $d = 1.13$. A 2 (advice interdependence: independent vs. dependent advice) x 2 (order of presentation: dependent advice first vs. independent advice first) mixed ANOVA on the confidence shifts revealed no significant effects, all $Fs < 0.65$, $ps > .424$, all $\eta_p^2 < .02$. In other words, although participants benefitted more from independent advice, this did not result in higher confidence gains. We address these findings in the general discussion.

Perceived accuracy. Did our participants recognize that independent advice was more accurate than dependent advice? To explore this question, we calculated a 2 (advice interdependence: independent vs. dependent advice) x 2 (order of presentation: dependent advice first vs. independent advice first) mixed ANOVA on the accuracy ratings. We found no significant effects, all $Fs < 0.75$, all $ps > .392$, all $\eta_p^2 < .02$. In other words, participants gave roughly similar ratings in the dependent (M = 58.78, SD = 18.27) as well as in the independent condition (M = 60.00, SD = 19.00). This could explain why there was no difference in confidence gains between the two conditions. It is important to note, however, that participants were not directly asked to compare the advice. This may have masked whether or not participants detected differences in accuracy.

In conclusion, Study 4a adds to our previous findings by demonstrating that participants weight dependent advice more even when it is less accurate. However, as already mentioned, the data of Study 4a were affected by an unsystematic data loss introducing some noise into the analyses. Therefore, we ran Study 4b as a full replication of Study 4a, this time without any technical errors.

Study 4b

Method

Participants and design. In Study 4b, we had 110[22] participants, ranging in age from 19 to 33 years (M = 23.46, SD = 3.06). Forty of them were male and 69 female. One person did not specify his or her gender.

Procedure. The procedure was exactly the same as in Study 4a. This time, the values of all four advisors were correctly saved.

Results and discussion

Manipulation checks. Analogous to Study 4a, we first tested the correlations of errors in each condition against zero. As predicted, advisors' errors were significantly correlated in the dependent condition r = .69, p = .004, but not in the independent condition r = .01, p = .97. The inter-correlation was higher for dependent compared to independent advice, t(109)= -48.75, p < .001. Regarding advice accuracy, a paired t-test showed that, as predicted, independent advice was more accurate (M = 7.71, SD = 1.38) than dependent advice (M = 13.27, SD = 2.18), t(109) = -24.12, p < .001, d = -2.06. Moreover, the variability of the advice was greater in the independent condition (M = 0.17, SD = 0.03) than in the dependent condition (M = 0.10, SD = 0.01), t(109) = 26.76, p < .001, d = 2.40.

Similar to Study 4a, we calculated a 2 (advice interdependence: independent vs. dependent advice) x 2 (order of presentation: dependent advice first vs. independent advice first) ANOVA to analyze participants' ratings of the similarity of the advice provided by the advisors. The main effect of advice interdependence was significant, F(1, 108) = 10.95, p = .001, η_p^2 = .09. Participants correctly perceived information contained in dependent advice as being more similar (M = 6.14, SD = 2.07) as compared to independent advice (M = 5.35; SD

[22] Excluding 33 participants without yoking-partner did not change the results. Therefore, all participants were retained in our analyses.

= 1.88). The main effects of order and the interaction effect were not significant, all *Fs* < 1.56, all *ps* > .22, all η_p^2 < .015. In sum, our manipulation was successful.

Main analyses

Weighting of advice. A 2 (advice interdependence: independent vs. dependent advice) x 2 (order of presentation: dependent advice first vs. independent advice first) repeated measures ANOVA on the mean AT scores per participant and phase revealed a main effect of advice interdependence, $F(1, 108) = 7.58$, $p = .008$, $\eta_p^2 = .06$. Participants weighted dependent advice more ($M = 0.49$; $SD = 0.20$) than independent advice ($M = 0.44$; $SD = 0.20$), thereby replicating the effect we found in Study 4a. This time, the main effect of phase was also significant, $F(1, 108) = 4.50$, $p = .036$, $\eta_p^2 = .04$, indicating that AT scores were higher overall if the dependent advice was presented first (M = .54, SD = 0.18), as compared to when the independent advice was presented first (M = .44, SD = 0.20). The interaction was not significant, $F(1, 108) = 1.09$, $p = .299$, $\eta_p^2 = .01$.

Also parallel to Study 4a, we investigated the frequency of revision as an alternative measure of advice weighting. A 2 (advice interdependence: independent vs. dependent advice) x 2 (order of presentation: dependent advice first vs. independent advice first) repeated measures ANOVA revealed that participants changed their estimates more often in the dependent than in the independent condition (80% vs. 75%), $F(1, 108) = 7.93$, $p = .006$, $\eta_p^2 = .07$. The main effect of order was also significant, $F(1, 108) = 9.63$, $p = .002$, $\eta_p^2 = .08$. Participants changed their estimates more often when receiving dependent advice first compared to receiving independent first (83% vs. 73%). However, this finding is independent of the effect of advice interdependence, since the interaction effect was not significant, $F(1, 108) = 1.98$, $p = .16$, $\eta_p^2 = .02$. Therefore, it has no implication for the interpretation of the effect of advice interdependence. In sum, we were able to replicate the effects of advice

Table 3. Overview of main results of Study 4b (Means with Standard Deviations in parentheses).

	Method of Sampling	
Measure	Independent	Dependent
Accuracy		
initial error	**17.23** (3.91)	**18.38** (8.83)
final error	**11.00** (3.76)	**13.08** (7.05)
improvement (difference scores)	**6.23** (3.57)	**5.29** (10.35)
Weighting	**.44** (.20)	**.49** (.20)
Revision Process		
rate of changing initial estimates (%)	**75** (20)	**80** (20)
Confidence		
initial confidence	**42.50** (20.29)	**41.98** 19.08)
final confidence	**45.96** (20.10)	**45.60** (18.35)
confidence gain	**3.45** (8.90)	**3.63** (7.52)

interdependence on our main dependent variables. Table 3 depicts the values of the main dependent variables.

Accuracy gains. For our accuracy gain measures, two persons were excluded as their difference scores between the MAPEs for the initial and the final judgment deviated more than three standard deviations from the mean. One-sample t-tests on the difference scores showed significant accuracy gains in the independent (M = 6.23, SD = 3.57) as well as in the dependent condition (M = 5.29, SD = 10.35), $t(109)$ = 18.28, p < .001, d = 3.50, and $t(109)$ = 5.36, p < .001, d = 1.03. As in Study 4a, on a descriptive level the accuracy gains were somewhat higher in the independent than in the dependent condition, but neither this effect nor any other effect in a 2 (advice interdependence: independent vs. dependent advice) x 2 (order of presentation: dependent advice first vs. independent advice first) repeated measures

ANOVA on the difference scores became significicant, all *F*s < 3.55, all *p*s > .05, all $\eta_p^2 <$ .04. In contrast to Study 4a, differences in accuracy gains between the independent and the dependent condition were not significantly predicted by differences in AT scores, *b* = -7.09, *t*(106) = -1.28, *p* = .20. In line with Study 4a, these differences were (also) not significantly related to the differences in advice accuracy, , *b* = 0.38, *t*(106) = 0.85, *p* = .40, or to the interaction of advice accuracy and AT scores, *b* = -0.99, *t*(106) = -0.42, *p* = .68. Table 3 depicts the means and standard deviations of the main dependent variables.

Exploratory analyses

Confidence gains. As in Study 4a, we ran one-sample *t*-tests on the confidence shifts. Participants had significant confidence gains in the independent (*M* = 3.45, *SD* = 8.90), *t*(109) = 4.07, *p* < .001, *d* = 0.78, as well as in the dependent condition (*M* = 3.63, *SD* = 7.52), *t*(109) = 5.06, *p* < .001, *d* = 0.97. A 2 (advice interdependence: independent vs. dependent advice) x 2 (order of presentation: dependent advice first vs. independent advice first) mixed ANOVA on the confidence shifts revealed no significant differences between conditions, all *F*s < 0.91, *p*s > .34, both $\eta_p^2 < .01$.

Perceived accuracy. We calculated a 2 (advice interdependence: independent vs. dependent advice) x 2 (order of presentation: dependent advice first vs. independent advice first) mixed ANOVAs. The main effect of advice interdependence was not significant, *F*(1,108) = 0.00, *p* = .995, $\eta_p^2 = .00$. The main effect of order and the interaction effect, however, were significant, *F*(1, 108) = 6.11, *p* = .015, $\eta_p^2 = .05$, and *F*(1, 108) = 9.50, *p* = .002, $\eta_p^2 = .08$. Advice was overall rated as more accurate when dependent advice was presented first (*M* = 61.19, *SD* = 17.73 vs. *M* = 53.60, *SD* = 18.84)). Investigating the interaction, we found that participants in the dependent condition perceived advice as more accurate when dependent advice was presented first (*M* = 64.05, *SD* = 15.31) compared to when independent advice was presented first (*M* = 50.74, *SD* = 18.23), *F*(1, 108) = 15.59, *p* <

.001, η_p^2 = .13. There was no difference in the perceived accuracy in the independent condition (*M* = 58.33, *SD* = 19.62 vs. *M* = 56.47, *SD* = 19.14), *F*(1, 108) = 0.24, *p* = .62, η_p^2 = .002. This means that, if participants perceived a difference in the accuracy of dependent and independent advice at all, they falsely attributed higher accuracy to dependent advice.

In conclusion, the findings of both Study 4a and Study 4b showed that individuals put more weight on dependent advice even when it is less accurate than independent advice. In both studies, participants did not realize that independent advice was more accurate than dependent advice, although the instructions clearly indicated that dependent advice was characterized by partially redundant information and, as a consequence, independent advice was based on more information than dependent advice was. This failure may explain why participants did not take the advice's accuracy into account. In fact, participants may have been guided by the more obvious characteristics of the advice, like its level of consensus.

General discussion

The present paper contributes to the decision-making and advice-taking literature by investigating individuals' preference for dependent versus independent advice in two different contexts, and with two different measures. More precisely, across four studies with a total of six experiments, we examined which type of advice participants choose and which they weight more strongly, both in a context where dependent advice is more accurate (Studies 1 and 2), as well as in a context in which independent advice is more accurate (Studies 3, and 4). We found that individuals prefer dependent over independent advice in the weighting studies (Studies 2 and 4) while they prefer independent over dependent advice in the choosing studies (Studies 1 and 3). Therefore, our first hypothesis was only partially confirmed as our participants did not display a general preference for dependent advice by weighting it more strongly *and* choosing it more often. Hypothesis 2a and 2b were confirmed, however, as we were able to find the hypothesized effects regarding accuracy gains. That is, individuals benefitted more from dependent advice as compared to independent advice when interdependence resulted from group discussion, and they benefitted less from dependent advice when interdependence resulted from informational overlap.

Why is it that our first hypothesis was only partially confirmed? In fact, our findings in the weighting studies are in line with previous studies showing that individuals rely less on their meta-knowledge about the advice (i.e., how it was sampled) when they also observe its level of consensus (e.g. Yaniv et al., 2009). One explanation is that consensus, for example, is easier to process than accuracy. That is, in order to compare the respective levels of accuracy, one has to calculate the errors (of the advice) first, and one then needs to know whether these errors are small or large, depending on the context. Consensus, on the other hand, is more or less immediately obvious. Furthermore, individuals apparently do not completely understand the consequences of correlated errors (i.e., that consensus may stem from correlated errors) or, at least, they do not know how to use this information correctly. Maines (1996), for example, showed that her participants' combined estimates were insensitive to differences in the interdependence of forecasts. Moreover, Soll and Larrick (2009) demonstrated that participants react to differences in competence, but not to the existence of shared biases. This deficiency may impede the comparison between the accuracy levels of independent and dependent advice even more when consensus cues are present. Thus, participants may have difficulties in judging the advice's accuracy – this is supported by the findings of Study 4a and b, where participants gave similar accuracy ratings to independent and dependent advice.[23] Our findings are also in line with the *consensus implies correctness heuristic* (Chaiken and Stangor, 1987), which basically says that individuals use consensus as an indication for accuracy. In combination with their difficulty to understand or detect correlated errors, this means that they may believe that consensus generally stems from shared information on the true value, which, as we have shown, can be misleading.

But why, then, were our participants perfectly capable of making a functional decision in our choosing studies by always choosing the more accurate advice irrespective of its level of consensus? On the one hand, choosing between two types of advice and weighting it is structurally very different. In fact, in Studies 1 and 3 participants made their choice directly after being informed how the advisors. That is, when making their choice, they probably still had in mind how the respective advisors made their judgments. Even in Study 3b, where respondents were presented with example values of the advice before making their choice, there was only one intermediate screen involved, between reading the descriptions and making the choice. In contrast, in the weighting studies the sampling information becomes less salient from trial to trial. Thus, participants may have been more distracted by the more obvious features of the advice, like its level of consensus. Therefore, the findings in our choosing studies suggest, that participants do in fact have a basic understanding of the effects of (spurious) consensus, but it may depend on the specific terms if they live up to their potential.

Furthermore, in the choosing studies participants are aware from the very beginning that they can choose between two types of advice and how these two types were generated, respectively. This is essentially different from the weighting studies where participants only learn after the first fifteen trials with one team of advisors that there will be another team for

[23] In Study 2, participants correctly rated dependent advice as more accurate. This does not contradict the argument that individuals have difficulties with the concept of correlated errors as in this specific case, dependent advice was more accurate.

the next fifteen trials and that this team arrived at their judgments in a different manner. This means that participants in the choosing studies had the possibility to compare these two types of advice more closely and may therefore have focused on different aspects than participants in the weighting studies.

Interestingly, across our studies, we found no differences regarding participants' confidence gains between the independent and the dependent condition. One explanation for this finding might be that participants perceived no difference in the advice's accuracy (Study 4a and Study 4b). When the two types of advice are perceived as similarly accurate, this should also lead to similar gains in confidence. On the other hand, it is also possible that agreement among advisors does not necessarily lead to higher confidence gains. In fact, a study by Wanzel and colleagues (2017) found that dependent advice only led to higher confidence gains when it simultaneously confirmed the judge's opinion. Therefore, a difference in confidence gains after having received dependent versus independent advice might also be a matter of varying proximity of the advice to the judge's own estimate. The next section addresses limitations of our studies and gives some directions for future research.

Limitations and directions for future research. As outlined above, participants in the choosing studies, before making their decision, know that there are two teams of advisors and that the way, how the advice was generated in these teams, differed. Obviously, this fact is inherent to this type of task. However, in future studies, participants in the weighting studies could also be informed in the beginning that they will be presented with two types of advice which was generated in two different ways. This way, participants in these studies also have the possibility to compare the two advisor teams from the start.

Furthermore, as described above, one possible explanation as to why individuals prefer dependent advice in the weighting studies could be its higher consensus. Participants may have perceived more agreement among dependent than independent advice and thus were drawn to it. In our weighting studies, we did not include a proper measure of potential mediators such as the perceived unambiguousness of the advice (i.e., how clearly it pointed in one direction in terms of the true value). This was done to prevent any demand effects. However, future studies would certainly benefit from adding mediator measures that capture individuals' perception of advisors' unambiguity. This could be done, for example, by asking participants how unambiguous they perceived the advice to be. Ideally, these measurements should take place before participants submit their final estimates in each trial, in order to minimize the possibility of reverse causation.

Moreover, our participants were presented with the individual advisor opinions. By this means, advisors' consensus is comparatively easy to detect while, as mentioned before, the advice's accuracy is not as obvious. Therefore, as a possible intervention to aid individuals in not being misdirected by the advice's consensus, participants could also be presented with aggregated rather than individual pieces of advice. For numeric advice, this means that individuals would only see the mean of several pieces of advice and could no longer determine the level of agreement between advisors. Another possibility would be to present them not only with the mean but also with dispersion information.

Finally, we only used quantitative judgment tasks, which is the default in the JAS. This means, however, that we cannot make any claims about individuals' preference for dependent or independent advice in other types of tasks. Therefore, future research might investigate if participants' overall preference for dependent advice can also be found in other tasks, for example decisions between several alternatives. Individuals might look for different qualities in their advisors across different tasks, resulting in different preferences as well. For instance, when there is greater uncertainty regarding the true value (as is often the case regarding the tasks used in the JAS, where participants often have to estimate distances, heights, calories, etc.), individuals might be more prone to look for consensus in their advisors, compared to tasks where uncertainty might be lower (as in knowledge questions with few alternatives).

Conclusion. With this paper, we contribute to the advice taking literature by showing that individuals prefer dependent over independent advice only when weighting the advice. When they have the possibility to choose between two advisor teams, individuals choose the more accurate advisors independent of its level of consensus. The fact that individuals across all scenarios choose the more accurate advice suggests that they have the right intuition about which type of advice would be better, but that this intuition becomes overruled when a more subtle measure of preference is used. Therefore, it is important to develop strategies to help individuals to avoid solely focusing on advisors' consensus in these situations. As a closing remark, it is important to note that in Study 2 as well as in Study 4 participants benefited from the advice in both conditions. This finding is in line with the advice taking literature stating that integrating advice in general improves judgment, even when it is not one hundred percent accurate (e.g., Brehmer & Hagafors, 1986;Yaniv, 2004a; 2004b). Therefore, seeking out independent advice is, in many situations, the better option with respect to accuracy gains, but even taking dependent advice might be better than not taking any advice at all, at least when the judge has no prior knowledge regarding the judgment issue.

References

Ariely, D., Au, W. T., Bender, R. H., Budescu, D. V., Dietz, C. B., Gu, H., Wallsten, T. S., & Zauberman, G. (2000). The effects of averaging subjective probability estimates between and within judges. Journal of Experimental Psychology: Applied, 6, 130–147.

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,* 127-151.

Brehmer, B., & Hagafors, R. (1986). Use of experts in complex decision making: A paradigm for the study of staff work. *Organizational Behavior and Human Decision Processes, 38,* 181-195.

Budescu, D. V., & Rantilla, A. K. (2000). Confidence in aggregation of expert opinions. *Acta Psychologica, 104*, 371-398.

Budescu, D. V., & Yu, H.-T. (2007). Aggregation of opinions based on correlated cues and advisors. *Journal of Behavioral Decision Making, 20,* 153-177.

Chaiken, S., & Stangor, C. (1987). Attitudes and attitude change. *Annual Review of Psychology, 38,* 575-630.

Gino, F., Brooks, A. W., & Schweitzer, M. E. (2012). Anxiety, advice, and the ability to discern: Feeling anxious motivates individuals to seek and use advice. *Journal of Personality and Social Psychology, 102,* 497-512.

Givon, M. (1984). Variety seeking through brand switching. *Marketing Science, 3*, 1-22.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association, 1*, 114-125.

Harvey, N., & Fischer, I. (1997). Taking advice: Accepting help, improving judgment, and sharing responsibility. *Organizational Behavior and Human Decision Processes, 70,* 117-133.

Hogarth, R. M. (1978). A note on aggregating opinions., *Organizational Behavior and Human Performance 21*, 40-46.

Johnson, T. R., Budescu, D. V., & Wallsten, T. S. (2001). Averaging probability judgments: Monte Carlo analyses of asymptotic diagnostic value. *Journal of Behavioral Decision Making, 14*, 123-140.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*, 237-251.

Maines, L. A. (1996). An experimental examination of subjective forecast combination. *International Journal of Forecasting*, *12*, 223–234.

Schultze, T. (2015, November). *A comparison of actual and optimal advice taking in quantitative judgments*. Poster session presented at the 36th annual conference of the Society for Judgment and Decision making, Chicago.

Schultze, T., Mojzisch, A., & Schulz-Hardt, S. (2012). Why groups perform better than individuals at quantitative judgment tasks: Group-to-individual transfer as an alternative to differential weighting. *Organizational Behavior and Human Decision Processes*, *118*, 24-36.

Schultze, T., Rakotoarisoa, A., & Schulz-Hardt, S. (2015). Effects of distance between initial estimates and advice on advice utilization. *Judgment and Decision Making, 10,* 144-171.

Sniezek, J. A., & Buckley, T. (1995). Cueing and cognitive conflict in Judge-Advisor decision making. *Organizational Behavior and Human Decision Processes*, *62*, 159–174.

Sniezek, J. A., & Van Swol (2001). Trust and expertise in a Judge Advisor System.

Organizational Behavior and Human Decision Processes, *84*, 288-307.

Soll, J. B. (1999). Intuitive theories of information: Beliefs about the value of redundancy. *Cognitive Psychology, 38*, 317–346.

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.

Stern, A., Schultze, T., & Schulz-Hardt, S. (2017). How Much Group is Necessary? Group-To-Individual Transfer in Estimation Tasks. *Collabra: Psychology,* 3: 16, 1-17.

Van Selst, M., & Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 47A*, 631-650.

Wallsten, T. S., & Diederich, A. (2001).Understanding pooled subjective probability estimates. *Mathematical Social Sciences*, *18*, 1–18.

Wanzel, S. K., Schultze, T., & Schulz-Hardt, S. (2017). Disentangling the effects of advisor consensus and advice proximity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*, 1669-1675.

Yaniv, I. (2004a). Receiving other people's advice: Influence and benefit. *Organizational Behavior and Human Decision Processes, 93*, 1-13.

Yaniv, I. (2004b). The benefit of additional opinions. *Current Directions in Psychological Science*, *13*, 75–78.

Yaniv, I., & Kleinberger, E. (2000). Advice taking in decision making: Egocentric discounting and reputation formation. *Organizational Behavior and Human Decision Processes*, *83*, 260–281.

Yaniv, I., Choshen-Hillel, S., & Milyavsky, M. (2009). Spurious consensus and opinion revision: Why might people be more confident in their less accurate judgments? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 558-563.

Supplementary Study Materials &

Additional Results Manuscript 2

Supplementary Study Materials

Upon publication all study materials will be accessible under the following link https://osf.io/f7wj9/. The material will be found in the folder “Stella Wanzel Dissertation”.

Additional Results Manuscript 2

Additional Results of Study 1

Accuracy gains in Phase 1 and Phase 2

Phase 1:

Descriptive statistics:

Independent: *M* = 19.45, *SD* = 23.35

Dependent: *M* = 13.74, *SD* = 20.69

t-test: *t*(80) = 1.13, *p* = .26, *d* = 0.26

Phase 2:

Descriptive statistics:

Independent: *M* = 6.44, *SD* = 20.25

Dependent: *M* = 13.79, *SD* = 28.53

t-test: *t*(79.32) = -1.36, *p* = .177, *d* = -0.29

Confidence gains in Phase 1 and Phase 2

Phase 1:

Descriptive statistics:

Independent: *M* = 5.45, *SD* = 8.15

Dependent: *M* = 7.77, *SD* = 12.26

t-test: *t*(80) = -0.90, *p* = .371, *d* = -0.21

Phase 2:

Descriptive statistics:

Independent: *M* = 4.37, *SD* = 7.33

Dependent: *M* = 5.86, *SD* = 10.30

t-test: *t*(80) = -0.73, *p* = .467, *d* = -0.16

Weighting of Advice (AT) in Phase 1 and Phase 2

Phase 1:

Descriptive statistics:

Independent: M = 0.31, SD = 0.20

Dependent: M = 0.45, SD = 0.21

t-test: $t(80) = -2.94$, $p = .004$, $d = -0.69$

Phase 2:

Descriptive statistics:

Independent: M = 0.26, SD = 0.17

Dependent: M = 0.44, SD = 0.18

t-test: $t(80) = -4.69$, $p < .001$, $d = -1.04$

Additional Results of Study 3a

Accuracy gains in Phase 1 and Phase 2

Phase 1:

Descriptive statistics:

Independent: M = 7.67, SD = 3.57

Dependent: M = 6.00, SD = 3.73

t-test: $t(57) = 1.67$, $p = .101$, $d = 0.46$

Phase 2:

Descriptive statistics:

Independent: M = 7.98, SD = 14.61

Dependent: M = 2.71, SD = 2.45

t-test: $t(46.34) = 2.26$, $p = .029$, $d = 0.42$

Confidence gains in Phase 1 and Phase 2

Phase 1:

Descriptive statistics:

Independent: M = 3.52, SD = 9.67

Dependent: M = 0.97, SD = 8.61

t-test: $t(57) = 1.00$, $p = .324$, $d = .027$

Phase 2:

Descriptive statistics:

Independent: M = 3.56, SD = 8.45

Dependent: M = -3.18, SD = 8.49

t-test: $t(57) = 2.77$, $p = .008$, $d = 0.80$

Weighting of Advice (AT) in Phase 1 and Phase 2

Phase 1:

Descriptive statistics:

Independent: M = 0.49, SD = 0.18

Dependent: M = 0.46, SD = 0.23

t-test: $t(57) = 0.33$, $p = .741$, $d = 0.09$

Phase 2:

Descriptive statistics:

Independent: M = 0.50, SD = 0.17

Dependent: M = 0.47, SD = 0.28

t-test: $t(20.88) = 0.44$, $p = .664$, $d = 0.16$

Additional Results of Study 3b

Accuracy gains

Descriptive statistics:

Independent: M = 8.11, SD = 4.55

Dependent: M = 4.21, SD = 7.16

t-test: $t(59) = 2.03$, $p = .047$, $d = 0.59$

Confidence gains

Descriptive statistics:

Independent: M = 6.25, SD = 10.54

Dependent: M = 7.83, SD = 10.77

t-test: $t(59) = -0.41$, $p = .685$, $d = -0.12$

Weighting of Advice (AT)

Descriptive statistics:

Independent: M = .52, SD = 0.15

Dependent: M = .58, SD = 0.19

t-test: $t(59) = -1.12$, $p = .269$, $d = -0.35$

Curriculum Vitae

Stella Wanzel

Curriculum Vitae

NICHT IN DER ONLINEVERSION VERFÜGBAR