

Rebholz, Tobias R.; Biella, Marco; Hütter, Mandy

Article — Published Version

Mixed-effects regression weights for advice taking and related phenomena of information sampling and utilization

Journal of Behavioral Decision Making

Provided in Cooperation with:

John Wiley & Sons

Suggested Citation: Rebholz, Tobias R.; Biella, Marco; Hütter, Mandy (2024) : Mixed-effects regression weights for advice taking and related phenomena of information sampling and utilization, Journal of Behavioral Decision Making, ISSN 1099-0771, Wiley, Hoboken, NJ, Vol. 37, Iss. 2,
<https://doi.org/10.1002/bdm.2369>

This Version is available at:

<https://hdl.handle.net/10419/294002>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<http://creativecommons.org/licenses/by/4.0/>

RESEARCH ARTICLE

Mixed-effects regression weights for advice taking and related phenomena of information sampling and utilization

Tobias R. Rebholz¹  | Marco Biella^{1,2}  | Mandy Hütter¹ 

¹Psychology Department, University of Tübingen, Tübingen, Germany

²Faculty of Business and Economics, University of Basel, Basel, Switzerland

Correspondence

Tobias R. Rebholz, Psychology Department, University of Tübingen, Tübingen, Germany.
Email: tobias.rebholz@uni-tuebingen.de

Funding information

Deutsche Forschungsgemeinschaft (DFG), Grant/Award Number: 2277

Abstract

Advice taking and related research is dominated by deterministic weighting indices, specifically ratio-of-differences-based formulas for investigating informational influence. Their arithmetic is intuitively simple, but they pose several measurement problems and restrict research to a particular paradigmatic approach. As a solution, we propose to specify how strongly peoples' judgments are influenced by externally provided evidence by fitting corresponding mixed-effects regression models. Our approach explicitly distinguishes between endogenous components, such as updated beliefs, and exogenous components, such as independent initial judgments and advice. Crucially, mixed-effects regression coefficients of various exogenous sources of information also reflect individual weighting but are based on a conceptually consistent representation of the endogenous judgment process. The formal derivation of the proposed weighting measures is accompanied by a detailed elaboration on their most important technical and statistical subtleties. We use this modeling approach to revisit empirical findings from several paradigms investigating algorithm aversion, sequential collaboration, and advice taking. In summary, we replicate and extend the original finding of algorithm appreciation and initially demonstrate a lack of evidence for both systematic order effects in sequential collaboration as well as differential weighting of multiple pieces of advice. In addition to opening new avenues for innovative research, appropriate modeling of information sampling and utilization has the potential to increase the reproducibility and replicability of behavioral science. Furthermore, the proposed method is relevant beyond advice taking, as mixed-effects regression weights can also inform research on related cognitive phenomena such as multidimensional belief updating, anchoring effects, hindsight bias, or attitude change.

KEYWORDS

advice taking, belief updating, judge–advisor system, multilevel modeling, weight of advice

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Authors. *Journal of Behavioral Decision Making* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

New information technologies and social networks make a wide variety of opinions and advice easily accessible across different contexts. Therefore, assessing how much people are affected by informational influences is gaining importance in the social sciences. However, it is an ongoing debate how much people make use of others' opinions, and a plethora of different approaches exists to investigate belief updating and judgment formation in light of external evidence. In this paper, we highlight critical aspects of common variants of the established modeling approach that is based on calculating descriptive weighting indices. More importantly, as a solution to the most crucial limitations of the state of the art, we propose to alternatively implement a temporally consistent mixed-effects regression (MER)¹ model of belief updating.

Psychologists, economists, and other social scientists often rely on experiments to generate insights with respect to peoples' advice taking behavior. In the dyadic judge–advisor system (JAS) as introduced by Sniezek and Buckley (1995), the participant is asked to judge stimulus items with the help of passively presented or actively sampled pieces of external information from one or multiple advisors. In most experiments, participants initially judge the same items free of any external influences. It is assumed that the shift (from initial to final) judgment indicates the amount of advice which was taken by that person. Specifically, a discrepancy between initial beliefs and advice may indicate some degree of initial bias or error, both of which can be compensated for by integrating external evidence into one's initial judgment (Deutsch & Gerard, 1955; Soll & Larrick, 2009). As a classic instance of capitalizing on the wisdom of crowds (e.g., Galton, 1907; Surowiecki, 2005), the relative size of judgmental shift is accordingly called weight of advice (WOA) and commonly of central interest to most advice taking researchers.

In anchoring paradigms, external pieces of information are also integrated into one's judgment, for instance, by insufficiently adjusting away from unrelated numbers (Epley & Gilovich, 2006; Tversky & Kahneman, 1974; see also Furnham & Boo, 2011, for a review). In contrast to WOA, the integration of information is referred to as “anchoring effect” because it is considered inappropriate. The same reasoning about (faulty) integration of external evidence applies to “hindsight biases” in memory research (Hoffrage et al., 2000), especially in Hawkins and Hastie's (1990) sense of systematically biased re-judgments in light of outcome knowledge. Essentially, there is no consensus in the literature on how to measure advice weighting (Bonaccio & Dalal, 2006) or anchoring effects (e.g., Turner & Schley, 2016, Footnote 3; see also Jacowitz & Kahneman, 1995). We will focus on advice taking here but argue that most claims, formulas, and findings can be transferred to other cognitive phenomena with similar structure such as anchoring, hindsight, or persuasion (e.g., Bochner & Insko, 1966; see also Yaniv, 2004; Yaniv & Milyavsky, 2007) due to conceptual and paradigmatic similarities.

Harvey and Fischer's (1997) advice taking index, as formally introduced below, has dominated the recent literature (e.g., Hütter & Ache, 2016; Schultze et al., 2015; Soll & Larrick, 2009; see Bailey et al., 2022, for a review). In economics, random variables are divided into “endogenous” and “exogenous,” with the former variables being determined by the model, such as modeling the process of updating judgments in response to receiving advice and the latter variables being measured or manipulated (Chiang, 1984, p. 8; Varian, 1992, p. 202). Accordingly, the advice taking index reflects how much people may have adjusted their final judgments, that is, the endogenous components, in units of the distance between the advice and their independent initial judgments, which constitute the two exogenous components. Although initial judgments also emerge from an “internal sampling” and judgment process (e.g., Fiedler & Kutzner, 2015; Henriksson et al., 2010; Stewart et al., 2006), they are treated as exogenous variables because they represent the outcome of a self-contained process during the advice taking phase.

We suspect that it is the simplicity of this established ratio-of-differences (ROD) formula paired with its capability to capture inter- and intra-individual differences that is responsible for its popularity. Capturing individual differences refers to explicitly accounting for variability between and within individuals. That is, to compute separate advice taking scores that describe participants' behavior on each experimental trial (Baayen et al., 2008; Bauer, 2011). Consequently, such metrics allow the measurement of idiosyncrasies of persons, items, and other clustering variables (e.g., multiple advisors) in specific situations (see also Kämmer et al., 2023). On the downside, however, and in contrast to the alternative modeling approach proposed here, they also require or impose critical paradigmatic peculiarities. For instance, the ROD-WOA cannot be used to measure informational influences if participants' prior beliefs are unobserved or unobservable (see Section 3.2). Also, if participants are granted access to multiple pieces of advice, the ROD-WOA is limited in explanatory power, as demonstrated in Section 4.2.

The same reasoning about popularity due to simplicity and capturing idiosyncrasies applies to other established criteria from traditional ROD-type modeling, such as the “anchoring index” of Jacowitz and Kahneman (1995) or the “hindsight bias index” of Hell et al. (1988). Crucially, however, intermixing endogenous and exogenous components to construct ROD scores implies a number of conceptual and measurement problems (Cronbach, 1943; Cronbach & Furby, 1970; Edwards, 1995; Firebaugh & Gibbs, 1985). For instance, ratios are often not normally distributed (see Soll & Larrick, 2009, for the characteristic W-shaped distribution of ROD-WOA) and imply a linear relationship between the numerator and denominator (but see Schultze et al., 2015, for evidence of a linear relationship between absolute judgment shift and advice distance). Moreover, difference scores suffer from outcome ambiguity when the separate effects of independent variables on each score component are reduced to a single coefficient (Edwards, 1995).

Regression-based methods, by contrast, are consistent with the recommendation to use endogenous components as criteria in an analysis that controls for exogenous components (Cronbach &

¹The terms multilevel, mixed-effects, and hierarchical modeling or regression all refer to the same statistical procedure in which coefficients comprise fixed and random, that is, mixed components.

Furby, 1970; but see Allison, 1990). Back in the 1980s and 1990s, some lines of research indeed assessed advice utilization by regressing final judgments simultaneously on all sources of information—advice in Brunswikian advice taking research (e.g., Brehmer & Hagafors, 1986; Harvey et al., 2000) plus independent initial judgments in the forecasting literature (e.g., Lim & O'Connor, 1995). The major limitation of regression-based approaches as available back then was their aggregate data analysis scheme (Bonaccio & Dalal, 2006). MER modeling instead allows to simultaneously account for participant and stimulus item variation (Baayen et al., 2008; Raudenbush & Bryk, 2002). Thus, the major limitation of regression-based analyses according to Bonaccio and Dalal (2006) can be resolved by explicit consideration of the multilevel structure that most experimental advice taking data inherits from repeated measures designs. In other words, as demonstrated formally and for several empirical examples below, the crossing of participants and stimulus items allows inferences about each individual trial of an advice taking experiment even for (mixed-effects) regression modeling of the data. Additional advantages of regression analyses include, but are not limited to, the easy handling of unbalanced data, missing observations, or arbitrary types of predictor and response variables, where the latter is the main reason to not distinguish between discrete choice and quantitative judgment here.

Our goal is to extend the toolbox for quantifying advice weighting by proposing a method which is technically more advanced than state-of-the-art modeling as assessed by the above-mentioned criteria. We will show that, by accounting for the (temporal) dependency of final judgments on exogenous sources of information, our approach will enable researchers to more flexibly measure the psychological construct “advice taking.” Nevertheless, the proposed method is not restricted to situations of people taking advice but can also be applied to other information acquisition phenomena such as anchoring effects, hindsight bias, or attitude change. First, we will formally establish the intended data analysis approach and elaborate on its most important technical and statistical subtleties. Then, corresponding insights into established empirical phenomena will be provided before the article concludes with a critical discussion of limitations and merits of the proposed modeling approach.

2 | MODELING OF INFORMATIONAL INFLUENCES IN ADVICE TAKING

It is common practice in advice taking research to rely on a formula involving ratios of differences of judgments to quantify advice weighting. The following formula as introduced by Harvey and Fischer (1997) measures how strongly people adapt their initial judgment toward advice:

$$\omega_{A,ij} = \frac{F_{ij} - I_{ij}}{A_{ij} - I_{ij}}, \quad (1)$$

where I_{ij} and F_{ij} indicate the initial and final judgments of a participant $i = 1, \dots, N$ about a given stimulus item $j = 1, \dots, M$, and A_{ij} the advice

received. This formula identifies the relative amount of judgmental shift from initial to final estimation that can be attributed to a single piece of new evidence which was passively encountered or actively acquired. Thus, $\omega_{A,ij} = 1$ indicates complete adoption, $\omega_{A,ij} = 0$ entire disregard, and everything in between a corresponding weighting of advice with $\omega_{A,ij} \in (0,1)$.² Exhibiting conceptual resemblance to Jacowitz and Kahneman's (1995) anchoring indices, it accordingly provides a “readily interpretable” but merely descriptive measure of advice weighting.

By definition, the traditional index implies a cognitive processing scheme according to which the final judgment is inherently defined as a weighted linear combination of all available sources of information—external (i.e., advice A_{ij}) and internal (i.e., initial estimates I_{ij}). This becomes apparent by rearranging the ROD formula as follows:

$$F_{ij} = I_{ij} + \omega_{A,ij}(A_{ij} - I_{ij}) = \omega_{A,ij}A_{ij} + (1 - \omega_{A,ij})I_{ij}, \quad (2)$$

that is, to account for the endogenous formation of final judgments (Hoffman, 1960). Mathematically, “choosing” the advisor (self) is simply a special case of “averaging” where the weight has an extreme value of one (zero; Soll & Larrick, 2009). In that sense, averaging is just another term for a specific type of cognitive processing for which the weighting policy amounts to (adaptively) compromise between two (or more; see below) exogenous sources of information. Notably, there is ample normative (e.g., Clemen, 1989; Mannes, 2009) and empirical (e.g., Anderson, 1981; Budescu & Rantilla, 2000; Slovic & Lichtenstein, 1971) evidence for simple averaging.

2.1 | A MER model of advice taking

By accounting for overall error $\varepsilon \sim N(0, \sigma^2)$ in Equation (2), a regression-based correspondence of Harvey and Fischer's (1997) ROD-WOA can be derived from the resulting regression model:

$$F_{ij} = \omega_{A,ij}A_{ij} + (1 - \omega_{A,ij})I_{ij} + \varepsilon_{ij}, \quad (3)$$

where the coefficient (or often “weight”) $\omega_{A,ij}$ measures the effect of advice on final judgment. Ordinary estimation techniques, however, do not enable estimating separate weights of individual pieces of advice $\hat{\omega}_{A,ij}$ or the self $1 - \hat{\omega}_{A,ij}$, respectively (Bauer, 2011). In repeated measures designs, multiple observations are usually available per participant and item. Therefore, the residuals of the coefficient can be disentangled from the overall error term ε such that individual regression coefficients of the form

$$\omega_{A,ij} = \beta_A + \alpha_{A,i}^S + \alpha_{A,j}^T \quad (4)$$

²Theoretically, $\omega_{A,ij} \notin [0,1]$ is possible if participants shift in the opposite direction than advised. Accordingly, Önkal et al. (2009) interpret Equation (1) merely as a “positional measure” indicating the location of final judgments relative to advice and initial judgments (i.e., closer to one or the other).

become admissible (Baayen et al., 2008; Bates et al., 2015; Brown et al., 2018; Raudenbush & Bryk, 2002). In this equation, β_A denotes the fixed effect of advice on the final judgment and $\alpha^q \sim N(0, \tau_{A,q}^2)$ the random effects of participants S and stimulus items T , respectively, with $\tau_{A,S}^2$ and $\tau_{A,T}^2$ mutually independent. In summary, the multi-level regression model specified in Equations (3) and (4) essentially consists of a regression of participants' observed judgments as the dependent and independent variables at level 1, henceforth the "judgment level," and a regression of these coefficients as the latent or unobserved dependent variable at level 2, henceforth the "weighting level." Consequently, the proposed MER-WOA simply corresponds to the estimated regression coefficients.³

Formally, the proposed model assumes that participants and items, "although unique in many ways, have certain common characteristics that may be accounted for in the modeling process" (Afshartous & de Leeuw, 2005, p. 111). For instance, different values of $\alpha_{A,j}^S$ capture that people differ in their selection of advice taking strategies from a disjunctive set of strategies including full adoption, disregard, and equal weighting (Soll & Larrick, 2009). However, idiosyncratic characteristics of stimulus items may also influence participants' strategy selection or weighting. Specifically, participants may be more knowledgeable about judging a particular item j due to their experience with the underlying judgment domain, and thus need or accept less help, resulting in a relatively small value of $\alpha_{A,j}^T$ for that item. For instance, in Experiment 2 of Ache et al. (2020), it was easier for participants to estimate the airline distance between native than non-native city pairs. Indeed, task difficulty (Gino & Moore, 2007) and level of expertise (Harvey & Fischer, 1997; Sniezek & Buckley, 1995) were found to reliably affect advice weighting in respectively opposite directions.

2.2 | First empirical application

Level of expertise was also found to matter in research on algorithmic advice taking. Logg et al. (2019) provide initial evidence for "algorithm appreciation," that is, people integrating algorithmic advice more than quantitatively equivalent advice provided by human advisors. In contrast, previous findings report a reluctance to integrate advice from (erroneous) algorithms as compared to (erroneous) humans, which was accordingly termed "algorithm aversion" (Dietvorst et al., 2015; see Burton et al., 2020; Jussupow et al., 2020; Mahmud et al., 2022, for reviews). In order to reassess the existence of algorithm appreciation by means of the proposed modeling procedure, we conducted a reanalysis of data from Experiment 4 of Logg et al. (2019). Our aim is to demonstrate the approach in a rather standard, simplistic JAS situation before highlighting more flexible and powerful modeling solutions

in the remaining sections.⁴ A significance level of 5% was used throughout for statistical testing in all three empirical applications. Furthermore, reproducible analysis scripts and supplementary material containing plots of model diagnostics are publicly available online (<https://osf.io/6gmhs>).

In Experiment 4, Logg et al. (2019) investigated the influence of judges' expertise (laypersons recruited via Amazon's Mechanical Turk vs. national security experts) and the source of advice (human vs. algorithm) on advice weighting. All groups received quantitatively equivalent advice for four different quantitative judgment/forecasting tasks related to weight estimation, business events, and geopolitical incidents (Mellers et al., 2014). Consistent with their other experiments, participants in this study placed more weight on algorithmic than human advice indicating algorithm appreciation. However, the national security experts discounted all advice sources more strongly than lay participants. This result is consistent with the findings mentioned above that experts across various fields tend to be more resistant to advice (Harvey & Fischer, 1997; Sniezek & Buckley, 1995).

2.2.1 | Methods

For reanalyzing the data of Experiment 4 of Logg et al. (2019) by means of the proposed modeling approach, we extended the most basic model from Equations (3) and (4) only at the weighting level, such that

$$F_{ij} = \omega_{A,ij} A_{ij} + (1 - \omega_{A,ij}) I_{ij} + \varepsilon_{ij}, \quad (5)$$

$$\omega_{A,ij} = \beta_A + \alpha_{A,i}^S + \alpha_{A,j}^T + \beta_{A \times Fam} Fam_i + \beta_{A \times Exp} Exp_i + \beta_{A \times Src} Src_i + \beta_{A \times Exp \times Src} Exp_i Src_i, \quad (6)$$

can be used to examine the effects of judges' expertise, the source of advice, and their interaction on weighting from estimates of the corresponding fixed effects $\hat{\beta}_{A \times Exp}$, $\hat{\beta}_{A \times Src}$, and $\hat{\beta}_{A \times Exp \times Src}$, respectively. As in the original study, the set of additional explanatory variables included participants' familiarity with the word algorithm Fam_i , measured on a 7-point scale (1 = *not at all certain* to 7 = *extremely certain*), as z-standardized control variable at the weighting level. The variable $Exp_i \in \{-0.5, 0.5\}$ denotes the expertise of judge i , contrast-coded with -0.5 for laypersons and 0.5 for national security experts, and $Src_i \in \{-0.5, 0.5\}$ corresponds to the source condition, also contrast-coded with -0.5 for human advice and 0.5 for algorithmic advice. Contrast-coding of the treatment variables and z-standardization of the control variable were implemented for the random intercepts $\alpha_{A,i}^S$ and $\alpha_{A,j}^T$ at the weighting level, which correspond to random slopes at the judgment level (see below), to capture participant- and item-wise

³Formally, the multilevel regression-based estimator of WOA is defined as $\hat{\omega}_{A,ij} = \hat{\beta}_A + \hat{\alpha}_{A,i}^S + \hat{\alpha}_{A,j}^T$, which denotes (empirical Bayes) estimated MER coefficients of the judgment model. Accordingly, participant- and item-wise idiosyncrasies of mean advice weighting $\hat{\beta}_A$ are captured by the conditional modes of the random effects $\hat{\alpha}_{A,i}^S$ and $\hat{\alpha}_{A,j}^T$. As such, the individual weight estimates $\hat{\omega}_{A,ij}$ are treated as observed draws of random variables at the respective grouping level.

⁴Additional applications to data from our own lab include the validation of a positive effect of a trial-by-trial contrast in advice expectation on weighting and the estimation of separate weights of sequentially sampled advice (Rebholz, 2023, Chapters 3 and 4, respectively), as well as the investigation of various strategies for belief updating based on ambiguous scientific evidence (Schreiner et al., 2023).

TABLE 1 Multilevel model according to Equations (5) and (6) for Experiment 4 of Logg et al. (2019).

	Estimate	95% CI	SE	t	df	p
β_A	0.3240***	[0.2452, 0.4027]	0.0402	8.07	5.80	<.001
$\beta_{A \times Fam}$	-0.0608***	[-0.0935, -0.0281]	0.0167	-3.65	345.58	<.001
$\beta_{A \times Exp}$	-0.2536***	[-0.3391, -0.1680]	0.0436	-5.82	351.21	<.001
$\beta_{A \times Src}$	0.0820	[-0.0029, 0.1668]	0.0433	1.90	348.86	.058
$\beta_{A \times Exp \times Src}$	-0.2330**	[-0.4035, -0.0625]	0.0869	-2.68	348.50	.007
$\tau_{A,S}$	0.2763	[0.2528, 0.3012]				
$\tau_{A,T}$	0.0667	[0.0088, 0.1186]				
σ	7.2278	[6.9149, 7.5431]				
ICC	0.67					
R_{marg}^2	0.55					
R_{cond}^2	0.86					

Note: Two-sided * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$. Wald 95% CI for fixed and bootstrap 95% CI (with 1000 iterations) for random effects are shown. Sample sizes of $N = 343$ participants S and $M = 4$ stimulus items T resulted in a total number of 1372 observations.

random deviations in both experimental conditions and for an average level of familiarity (Judd et al., 2017).

The model equation as fitted in the statistical software, in our case R Version 4.3.1 (R Core Team, 2023) using the packages lme4 (Bates et al., 2015) and lmerTest (Kuznetsova et al., 2017), can be reconstructed by plugging Equation (6) into Equation (5):

$$F_{ij} = I_{ij} + \beta_A (A_{ij} - I_{ij}) + \alpha_{A_i}^S (A_{ij} - I_{ij}) + \alpha_{A_i}^T (A_{ij} - I_{ij}) + \beta_{A \times Fam} Fam_i (A_{ij} - I_{ij}) + \beta_{A \times Exp} Exp_i (A_{ij} - I_{ij}) + \beta_{A \times Src} Src_i (A_{ij} - I_{ij}) + \beta_{A \times Exp \times Src} Exp_i Src_i (A_{ij} - I_{ij}) + \varepsilon_{ij}. \quad (7)$$

Accordingly, we first computed the advice distance as a separate data column and then used these values as a predictor at the judgment level of the model. Moreover, the judgment-level intercept was fixed to zero, and the initial judgments I_{ij} were specified as offsets to estimate this sum-to-one constrained model version. The integrated model equation shows that the fixed control and treatment effects at the weighting level are simply modeled as interaction terms at the judgment level of the proposed model. More generally, some factors are included only as interaction terms with no main effects at the judgment level, which also applies to all subsequent reanalyses that include additional explanatory variables with effects only at the weighting level. Therefore, for the sake of brevity, the models are no longer written out explicitly in the following.

2.2.2 | Results

We fitted the model from Equations (5) and (6), or Equation (7), respectively, to the data of Experiment 4 of Logg et al. (2019). There was a significant effect of the control variable on MER-WOA indicating a significantly reduced weighting of advice by participants who were more familiar with the word algorithm (Table 1). For an average level of familiarity, the negative effect of expertise on advice

weighting could be reproduced using the proposed multilevel modeling framework. Experts weighted advice from any source significantly less than laypersons as indicated by $\hat{\beta}_{A \times Exp}$. In contrast, the original evidence for algorithm appreciation could not be reproduced as the coefficient estimate $\hat{\beta}_{A \times Src}$ was not statistically significant. For one-tailed hypothesis testing, however, participants collapsed across expertise indeed weighted algorithmic advice significantly more strongly than quantitatively equivalent human advice. Moreover, the coefficient estimate $\hat{\beta}_{A \times Exp \times Src}$ provided evidence that laypersons and experts differed in their treatment of algorithmic versus human sources of advice.

The sign and relative magnitude of the significant interaction term $\hat{\beta}_{A \times Exp \times Src}$ suggest that the effect of source on weighting may also differ qualitatively between experts and laypersons. This impression was reinforced by plots of the weighting distributions in Figure 1. Whereas the distributions of MER-WOA for laypersons were more left-skewed for algorithmic advice and more right-skewed for human advice, there were no such striking visual differences for experts according to both MER- and ROD-WOA. In order to test our hypothesis, we fitted two additional multilevel regression models with dummy-coded expertise (i.e., adding 0.5 to Exp_i) and reverse dummy-coded expertise (i.e., subtracting 0.5 from Exp_i and multiplying the result by -1), respectively.⁵ Overall, laypersons indeed weighted algorithmic advice significantly more strongly than human advice (see Table A1). The positive coefficient estimate $\hat{\beta}_{A \times Src}$ in this model with dummy-coded expertise provided evidence for algorithm appreciation by laypersons. Descriptively, the same effect even pointed in the opposite direction for experts as indicated by the negative coefficient estimate $\hat{\beta}_{A \times Src}$ in the reverse dummy-coded expertise model (see Table A2). In essence,

⁵In general, contrast-coding should be preferred over dummy-coding for the estimation of MER-WOA with crossed clustering instances (Judd et al., 2017, Footnote 1). The reason is that dummy-coding implies that any random effects—in our case, these are the participant- and item-wise random intercepts at the weighting level—capture randomness only for the reference level, that is, the group coded as zero according to the corresponding explanatory variable. Therefore, we fit two models with the two expertise samples as the respective reference groups.

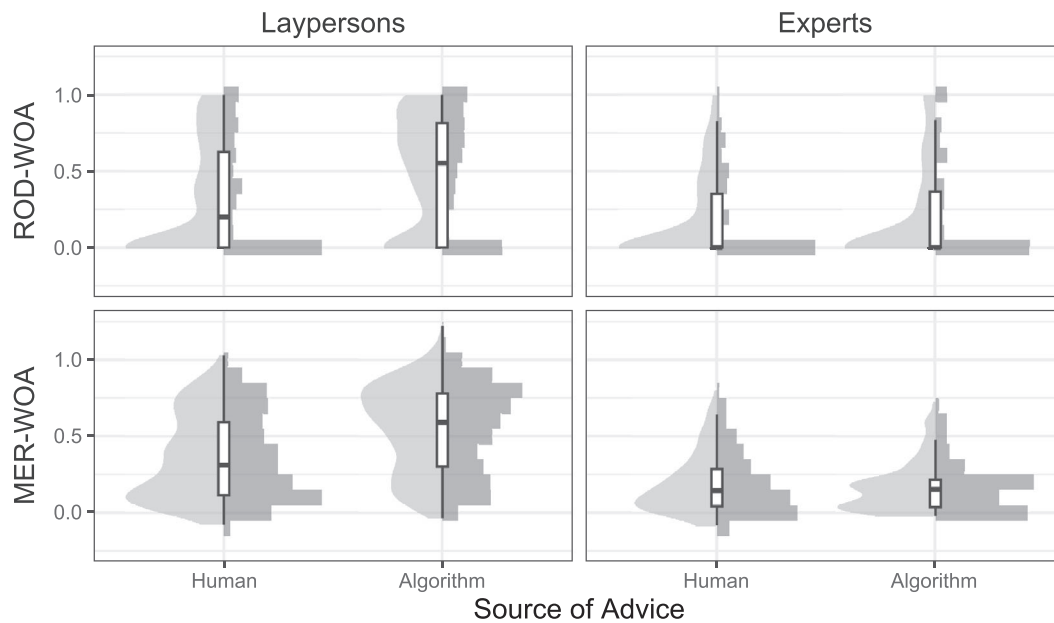


FIGURE 1 Distributions and box plots of the descriptive ratio-of-differences (ROD; top) and the estimated mixed-effects regression (MER; bottom) weight of advice (WOA) per source condition and the expertise of participants in Experiment 4 of Logg et al. (2019). Note: Distributions are represented as half-violin plots (i.e., smoothed densities) on the left and histograms on the right. Summary statistics are presented as box plots and do not include outliers.

the proposed modeling approach enabled us to extend the original findings by providing evidence that experts did not discriminate between the sources of advice as much as laypersons.

2.2.3 | Discussion

In summary, our results using the proposed analysis method are not entirely consistent with the original findings reported in Logg et al. (2019) and extend them in crucial aspects. Based on the original analysis of variance of ROD-WOA, there was significant evidence for algorithm appreciation. In our reanalysis, the difference between the average weighting of quantitatively equivalent algorithmic and human advice is significant only for one-tailed testing of treatment effects on MER-WOA. However, both modeling approaches consistently provide evidence that experts weight advice significantly less strongly than laypersons.

So far, we have primarily demonstrated *how* it is possible to account for endogenous judgment formation and individual differences in advice taking using the proposed multilevel modeling framework. A more crucial question is *why* the results of our reanalysis differ from the original analysis with respect to the two main findings. Whereas the p value for the main effect of the source of advice is substantially smaller here than in the original analysis, the opposite is true for the p value of the interaction between source and expertise. In fact, the significant p value of the interaction term being substantially lower in the reanalysis as compared to the original analysis suggests critical qualitative differences between expert and lay augmented judgment and decision-making. From the model with reverse dummy-coded expertise, we can infer that laypersons do indeed

weight algorithmic advice significantly more than human advice, whereas experts do not discriminate as strongly between the two sources of advice. These observations suggest that modeling the data from more classical JAS experiments in a framework that reflects the endogenous judgment process provides stronger evidence for the focal effect predicted by Logg et al. (2019, Experiment 4).

There are many methodological and theoretical considerations that underlie the practical relevance of the proposed modeling approach. For instance, sum-to-one constrained weighted averaging at the judgment level in Equation (3) as well as linearity at the weighting level in Equation (4) eventually appear to be overly restrictive. Nevertheless, both constraints were also implemented in the MER-based reanalysis to promote comparability with the original ROD-based analyses of Logg et al. (2019). Therefore, in the following sections, we will discuss two of the most valuable extensions of the multilevel modeling framework, which resolve at least some of these restrictions. Ultimately, we aim to illustrate *why* it is worth considering more flexible methods for analyzing advice taking behavior by providing additional empirical examples that convincingly demonstrate the practical applicability and substantiate the merits of the proposed modeling framework.

3 | BEYOND RELATIVITY TO THE SELF

Harvey and Fischer (1997) called it “perverse” advice taking when a final judgment does not lie strictly in between the initial judgment and the advice. For Equation (1), negative weights result from eventually shifting in the opposite direction than advised. Similarly, WOA greater than one indicates overshooting the advice. Researchers often either

distort perverse advice taking behavior by taking a ratio-of-absolute-differences approach with weighting specified as

$$\hat{\omega}_{A,ij} = \frac{|F_{ij} - I_{ij}|}{|A_{ij} - I_{ij}|}, \quad (8)$$

(e.g., Gino, 2008; Yaniv, 2004), or actively discard it altogether by winsorizing negative values to zero and values greater than one to one, or alternatively truncating them (e.g., Gino & Schweitzer, 2008; Gino et al., 2009; Schultze et al., 2015; Soll & Larrick, 2009). Any approach suffers from potentially undefined or ambiguous values. Essentially, the same observation can even be interpreted in opposite ways depending on the pre-processing applied. For instance, consider a shift from 100 to 90 when the advice was 110. This behavior would be truncated from the data and classified as an outlier, considered as non-weighting according to the winsorizing approach, or full weighting as measured by the ratio-of-absolute-differences formula. Therefore, Bonaccio and Dalal (2006) recommended analyzing the data twice, with and without “problematic” WOA values, to ensure invariant conclusions (e.g., Himmelstein & Budescu, 2022; Hütter & Ache, 2016).

Even if the results do not change, the central dependent variable is incomparable across advice taking studies applying different data pre-processing techniques, which thus renders advice taking research prone to replicability issues. Alternatively, a less restrictive interpretation of the original index is that it merely captures relative positioning. In other words, WOA merely measures to which of the two sources of information final judgment is relatively closer (Önkal et al., 2009). Accordingly, $\omega_{A,ij} \notin [0,1]$ “reflect deliberate behavior rather than being noise or a nuisance” (Soll et al., 2022), let alone not being “well defined” (Gino, 2008; Yaniv & Kleinberger, 2000) or representing “errors” (Hou & Jung, 2021). Indeed, there can be good psychological reasons for “pushing away” from advice, for instance, if it encourages additional search (Rader et al., 2015) or raises suspicion. In contrast, more technical reasons include erroneous judgment responses, such as missing a trailing 0 and entering 100 instead of 1000, but also revisions based on the recognition of such errors and other inconsistencies in initial judgments, regardless of the advice received. Additionally, cross-cultural differences may induce systematic variations in advice taking, for instance, the increased willingness to fully accept advice in collectivistic societies (Mercier et al., 2012). Whereas such systematic effects can be taken into account by including observed sample characteristics as additional predictors in the model, extreme distortions due to inattentiveness or technical errors can be minimized by excluding judgment outliers from the data. In general, however, we argue against (arbitrary) pre-processing or the exclusion of “odd” observations in terms of weighting.

3.1 | Unconstrained regression modeling

The regression-based approach is much more flexible regarding the definition and interpretation of weights. Instead of restrictively

conceptualizing advice taking as the weight of the advisor *relative to the self*, the sum-to-one constraint can be abandoned in favor of separate weights for each individual source of information. The system of Equations (3) and (4) can be rewritten as

$$F_{ij} = \omega_{A,ij}A_{ij} + \omega_{I,ij}I_{ij} + \varepsilon_{ij}, \quad (9)$$

$$\omega_{p,ij} = \beta_p + \alpha_{p,i}^S + \alpha_{p,j}^T, \quad (10)$$

where $p \in \{A, I\}$ indicates the weights of advice and the self with fixed effects β_p and random effects $\alpha^a \sim N_{||p||}(\mathbf{0}_{2 \times 1}, \Sigma_a)$, $q \in \{S, T\}$, where $\mathbf{0}_{2 \times 1}$ is the zero vector and the covariance matrices Σ_a of participants $q=S$ and stimulus items $q=T$ being mutually independent by assumption.

The coefficient regressions of Equation (10) capture the partial effects (i.e., the value of the respective other source of information held constant) of advice $\omega_{A,ij}$ and initial judgment $\omega_{I,ij}$ on final judgment, respectively. Thus, the approach still allows comparing the WOA to the weight of the initial judgment. Moreover, sum-to-one constraining can be restored as follows:

$$\tilde{\omega}_{A,ij} = \frac{\hat{\omega}_{A,ij}}{\hat{\omega}_{A,ij} + \hat{\omega}_{I,ij}}. \quad (11)$$

Applying the divide-by-total principle post hoc (cf. Harvey et al., 2000) restores the original scaling and thus also the intuitive interpretability of relative weights $\tilde{\omega}_{p,ij} \forall p \in \{A, I\}$. Essentially, however, it is neither necessary to apply potentially problematic data pre- or post-processing approaches nor to analyze the data twice. More important, and to de facto move *beyond* the relativity to the self by estimating WOAs from unconstrained regressions, it is possible to consider alternative formulations of the judgment formation model in Equation (9).

Partial effects $\omega_{p,ij}$ are particularly relevant under the following three circumstances. First, more than the traditional two sources of information in the JAS are available. For instance, additional cues to expertise or accuracy (e.g., Budescu et al., 2003; Mannes et al., 2014; Soll & Larrick, 2009; Yaniv & Kleinberger, 2000), multiple advisors (e.g., Brehmer & Hagafors, 1986; Harvey et al., 2000; Hütter & Ache, 2016; see also below), “automated advice” from algorithms additional to traditional human judgments (e.g., Logg et al., 2019; Prahll & van Swol, 2017), and so on. Second, advice and initial judgments are non-orthogonal. For instance, multiple regression with partial effects would be more appropriate than traditional analysis approaches for experimentally manipulated advice distance (e.g., Schultze et al., 2015; Rebholz & Hütter, 2022). Moreover, non-orthogonality is highly ecological as judges often anchor advisors by including their own judgments in their requests for advice (Reif et al., 2022). Third, no initial judgments are recorded (e.g., Brehmer & Hagafors, 1986; Harvey et al., 2000; Mayer & Heck, 2022). In these situations, the ROD-calculus cannot be applied, which is the main focus of the following second empirical application.

3.2 | Second empirical application

A blind spot of advice taking research is informational influence without prior formulation of independent judgments. Although it is not explicitly framed as such, the “sequential collaboration” experiments of Mayer and Heck (2022) are procedurally equivalent to the traditional JAS. In their paradigm, participants received the estimate of a previous participant to answer general knowledge questions, such as “How tall is the Eiffel Tower?” in Experiments 1 and 2, or to locate cities on maps in Experiment 3. The most important difference from classical studies on advice taking is that no independent initial judgments were required. Consequently, although the provision of previous participants' estimates essentially constitutes an advice taking situation in the traditional sense, informational influence could not be assessed by indices of relative positioning such as ROD-WOA.

The original study was limited to a performance perspective, providing evidence for increasing judgment accuracy along sequential collaboration chains. Although the number of previous contributors was unknown to the participants, positioning effects on informational influences would provide valuable additional insight into whether more accurate advice is weighted more strongly (e.g., Yaniv & Kleinberger, 2000; but see Fiedler et al., 2019; Schultze et al., 2017). Thus, by reanalyzing the data from Mayer and Heck (2022, Experiments 1 and 2) using the proposed regression-based approach, we address this important blind spot in research on advice taking without prior formulation of independent judgments.

3.2.1 | Methods

Essentially, separate weights for individual sources of information can be calculated by means of MER even when advice is the only source of information observed. If participants make unobserved or unobservable initial judgments, these can be estimated as random intercepts of participants and items (see Section 3.2.3 for more details). In the regression model fitted here, that is,

$$F_{ij} = \omega_{A,ij} A_{ij} + \varepsilon_{ij}, \quad (12)$$

$$\omega_{A,ij} = \beta_A + \alpha_{A,i}^S + \alpha_{A,j}^T + \beta_{A \times C} c_i, \quad (13)$$

potentially formed initial judgments are part of the residuals ε_{ij} . The coefficient $\omega_{A,ij}$ now captures the partial effect of advice on the final judgment. Put differently, weighting was operationalized as the change in F in response to a unit change in A .⁶ Accordingly, whereas $\omega_{A,ij} = 0$ indicates unresponsiveness or no advice taking as usual, the

interpretation of $\omega_{A,ij} = 1$ is slightly different: The final judgment exactly keeping pace with changes in A does not necessarily imply the complete adoption of the advice, or in other words, maintaining a previous participants' judgment. As usual, $\omega_{A,ij} \in (0,1)$ indicates less than full but more than no responsiveness or weighting, and $\omega_{A,ij} \notin [0,1]$ captures effects akin to pushing away from or being over-responsive to advice.

The potential temporal dynamics of informational influences are the most interesting comparison for advice taking in sequential chains. Whereas Mayer and Heck (2022) found that the change probability decreased for later collaborations and that participants changed their judgments less in absolute terms relative to the true value, the partial effect estimates may tell a different story. Therefore, we also included a linear trend of participants' centered chain position $c_i = 2, \dots, C$ (with $C = 4$ in Experiment 1 and $C = 6$ in Experiment 2) at the weighting level of the model in Equation (13). Observations with $c_i = 1$ were excluded as no advice was provided to participants at position 1.

3.2.2 | Results

To extend the original research beyond the investigation of accuracy gains to informational influences on general knowledge questions via sequential collaboration, we estimated the model from Equations (12) and (13) for data from Experiments 1 and 2 of Mayer and Heck (2022). In Experiment 1, the mean responsiveness/weighting was slightly smaller than in Experiment 2 as indicated by $\hat{\beta}_A$ (Tables 2 and 3, respectively). Overall, earlier participants exerted relatively more informational influence on their successors in the second experiment. In other words, sequential collaboration was slightly less pronounced in Experiment 1 than Experiment 2. In classical research on advice taking, the distribution of weights often follows a characteristic W-shaped form, with modes at no and full weighting of advice as well as for equal weights averaging (Soll & Larrick, 2009). In contrast, the distributions of partial effects were left-skewed with modes at full responsiveness to judgments of previous participants, that is, $\omega_{A,ij} = 1$ at most chain positions in both experiments (Figure 2). This could be a consequence of the possibility to opt out of changing a previous participant's judgment. In general, participants were quite responsive to advice in most trials and almost never completely unresponsive. Nevertheless, the average responsiveness to advice was significantly less than one as indicated by the 95% CIs of $\hat{\beta}_A$.

In both experiments, there was no evidence for a fixed positioning effect on MER-WOA. As participants were unaware of their position in the chain, it is not surprising that the effect of chain position on weighting $\hat{\beta}_{A \times C}$ was not significantly different from zero (Tables 2 and 3). Nevertheless, the positioning effects in both experiments pointed descriptively in the same direction, indicating increasing informational influence along the sequential collaboration chains (see also Figure 2). According to the distributions of trial-wise partial effects plotted separately for each chain position, non-weighting is almost

⁶To promote model convergence, we applied item-wise z-standardization to the raw judgments instead of their distances from the true values as in the original study. Consequently, unit changes were measured in standard deviations, and the intercept could be set to zero in Equation (12).

TABLE 2 Multilevel model according to Equations (12) and (13) for Experiment 1 of Mayer and Heck (2022).

	Estimate	95% CI	SE	t	df	p
β_A	0.8104***	[0.7512, 0.8697]	0.0302	26.82	149.44	<.001
$\beta_{A \times c}$	0.0520	[-0.0107, 0.1147]	0.0320	1.63	111.51	.104
$\tau_{A,S}$	0.2607	[0.2228, 0.2989]				
$\tau_{A,T}$	0.1225	[0.0953, 0.1488]				
σ	0.5650	[0.5556, 0.5748]				
ICC	0.20					
$R^2_{\text{marg.}}$	0.62					
$R^2_{\text{cond.}}$	0.69					

Note: Two-sided * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$. Wald 95% CI for fixed and bootstrap 95% CI (with 1000 iterations) for random effects are shown. Sample sizes of $N = 111$ participants S and $M = 65$ stimulus items T resulted in a total number of 6621 observations after excluding 594 outliers.

exclusively restricted to initial or earlier interactions. In contrast, the variances of the distributions for later sequential collaborations seem to stabilize at a relatively lower level than observed for earlier interactions.

3.2.3 | Discussion

Independent initial judgments are irrelevant from an accuracy perspective as applied in the original study. However, they are key to traditional investigations of informational influence by means of calculating ROD-based weighting indices. While the original study found that accuracy increased with the number of collaborations, the proposed regression-based approach shows that this was not due to changes in informational influence along the chains. This observation contradicts findings that advice of (seemingly objective) higher quality, that is, for later positions in the scenario at hand, is weighted relatively more strongly (Yaniv & Kleinberger, 2000).⁷ Accordingly, the new evidence suggests that the positive effect of advice quality on weighting may be driven by implementing a reference point for advice quality assessment in the traditional paradigm. Specifically, the distance of the advice from the independent initial judgment that serves as the benchmark is considered informative of the quality of the advice (Zaleskiewicz & Gasiorowska, 2023; see also Rebholz et al., 2023). In contrast, this reference point was not available in the experiments of Mayer and Heck (2022), which focused on the effectiveness of sequential collaborations.

Although not recorded, this does not preclude the possibility that participants may have formed independent judgments. Accordingly, their *unobserved* priors being represented in the overall error term ϵ may eventually imply omitted variable bias for the model as specified in Equation (12) (see also Raudenbush & Bryk, 2002, Chapter 9; Snijders & Bosker, 2012, Chapter 14). The formation of independent

judgments without explicit request, and thus the possibility for the experimenter to record them, would imply that the reported effect of advice on the final judgment includes the effects of such missing variables that are correlated with the advice presented. Indeed, the influence of advice on judgment turned out to be surprisingly strong, at least as compared to traditional advice taking research (see Bailey et al., 2022, for a meta-analysis on ROD-WOA). Notably, this observation could only be made by estimating unrestricted MER-WOAs.

By including random intercepts per participant and item in Equation (12) (but see Footnote 6), one could additionally obtain estimates of latent initial judgments, or control for their potential formation, respectively. On the one hand, however, the simultaneous presentation of stimulus items and collaborators' judgments constituted a strong procedural barrier to the formation of independent beliefs in Mayer and Heck (2022). Additionally, it implies mental off-loading in the sense that participants' effort to think through their task is reduced because someone else has already done it for them, which positively affects advice taking (Sniezek & Buckley, 1995). On the other hand, omitted variable bias can also not be ruled out completely if initial judgments were incorporated into the model. Some of the reasons for this have already been mentioned above, such as the lack of data on other potentially correlated cues, for instance, advisor expertise or advice quality. In general, the inevitable temporal distance between initial and final judgments in the original paradigm could itself cause discrepancies in participants' judgments. For instance, advisees have limited working memory capacity (e.g., Cowan, 2010), or they may deliberately change their minds for other unobservable reasons, such as continuing the internal sampling process after having already submitted their initial judgment (Juslin & Olsson, 1997; Thurstone, 1927).

The simultaneous confrontation with the judgment task together with the collaborators' judgment in Mayer and Heck (2022) also carries the risk of having triggered an anchoring and adjustment process in the classical sense (see also Minson & Mueller, 2012; Schultze et al., 2019; Tversky & Kahneman, 1974). Consequently, our findings provide evidence for participants having (insufficiently) adjusted away from earlier anchors of lower quality as much as from later anchors of higher quality (but see Hütter & Fiedler, 2019; Röseler et al., 2023;

⁷Although judgments were not independent, increasing advice quality is probably a consequence of the increasing crowd wisdom with more contributors (Hogarth, 1978; but see Davis-Stober et al., 2014). This might explain why advice non-taking and the larger variance in weighting were particularly prominent features of relatively early inter-individual interactions (see Figure 2).

	Estimate	95% CI	SE	t	df	p
β_A	0.8250***	[0.7764, 0.8736]	0.0248	33.29	207.16	<.001
$\beta_{A \times c}$	0.0165	[-0.0095, 0.0424]	0.0133	1.24	263.44	.214
$\tau_{A,S}$	0.2825	[0.2553, 0.3098]				
$\tau_{A,T}$	0.1305	[0.1048, 0.1566]				
σ	0.5703	[0.5638, 0.5768]				
ICC	0.23					
$R^2_{\text{marg.}}$	0.61					
$R^2_{\text{cond.}}$	0.70					

TABLE 3 Multilevel model according to Equations (12) and (13) for Experiment 2 of Mayer and Heck (2022).

Note: Two-sided $^*p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$. Wald 95% CI for fixed and bootstrap 95% CI (with 1000 iterations) for random effects are shown. Sample sizes of $N = 254$ participants S and $M = 65$ stimulus items T resulted in a total number of 15,076 observations after excluding 1434 outliers.

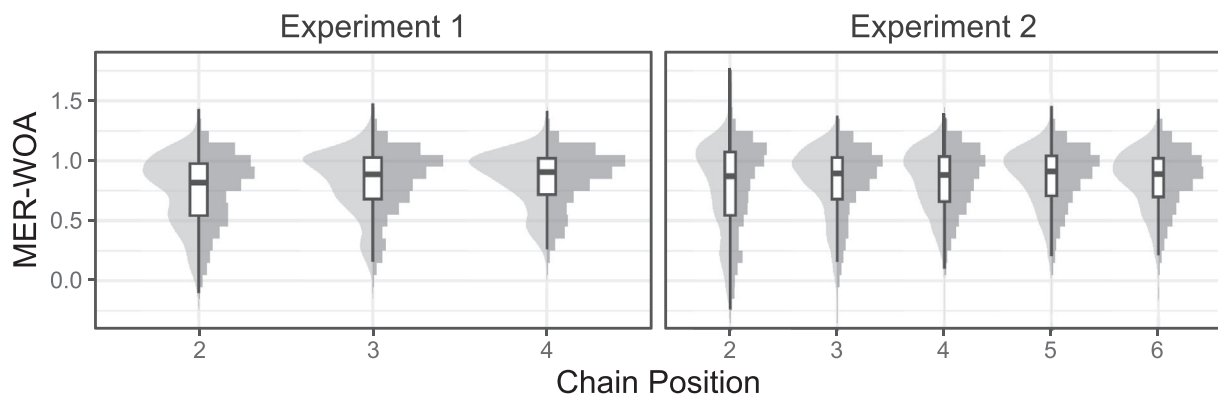


FIGURE 2 Distributions and box plots of the mixed-effects regression weight of advice (MER-WOA) per chain position in Experiments 1 and 2 of Mayer and Heck (2022). Note: Distributions are represented as half-violin plots (i.e., smoothed densities) on the left and histograms on the right. Summary statistics are presented as box plots and do not include outliers.

Schultze et al., 2017; Schweickart et al., 2021). In the traditional paradigm, however, independent initial judgments also constitute anchors in a sense, from which participants tend to adjust away as they consider the external evidence provided by their advisors (Bonaccio & Dalal, 2006; Harvey & Fischer, 1997; Lim & O'Connor, 1995). The possibility of omitting the initial judgment phase thus reduces the risk of self-anchoring in the JAS. As informational influences can still be estimated using the proposed analytical method, MER-WOA also helps to avoid confounding anchoring effects by resolving this critical paradigmatic peculiarity of advice taking research.

4 | SEPARATE WEIGHTS OF MULTIPLE PIECES OF ADVICE

In extensions of the traditional paradigm to multiple advice taking, participants receive advice from more than one single advisor. In sampling approaches, for instance, participants have the opportunity to freely and sequentially sample additional advice. In two of three

experiments of Hütter and Ache (2016), participants were allowed to sample up to 20 pieces of advice about the caloric content of dishes (e.g., fish pasta) before stating their final, possibly revised estimates. Due to the lack of more advanced techniques, the factually sequential taking of multiple pieces of advice was modeled as the taking of the mean of all advisory judgments within a trial:

$$\omega_{\bar{A}_{ij}} = \frac{F_{ij} - I_{ij}}{\bar{A}_{ij} - I_{ij}}, \quad (14)$$

where $\bar{A}_{ij} = \frac{1}{K_{ij}} \sum_{k=1}^{K_{ij}} A_{ijk}$, for $k = 1, \dots, K_{ij}$ sampled pieces of advice per trial ij . Plugging \bar{A}_{ij} into Equation (14) and rearranging the result to account for endogenous judgment formation yields:

$$F_{ij} = \frac{\omega_{\bar{A}_{ij}}}{K_{ij}} A_{ij1} + \dots + \frac{\omega_{\bar{A}_{ij}}}{K_{ij}} A_{ijK_{ij}} + (1 - \omega_{\bar{A}_{ij}}) I_{ij}. \quad (15)$$

Thus, rearrangement reveals that conceiving of advice as an unweighted linear combination of all sampled advisory judgments imposes an equal weighting constraint, that is, $\omega_{A_{ijk}} = \omega_{A_{ijl}}$, on each

individual piece of advice $k, l = 1, \dots, K_{ij}$.⁸ Put differently, $\omega_{A_{ij}}$ indicates the *total* WOA which is defined as the *sum* of all single, equally weighted pieces of advice that are sampled on a specific trial. In contrast, there are many empirical examples that provide evidence against equal weighting.

Similar to the influence of sequentially sampled anchors, there may be ordering effects (e.g., primacy or recency) in the weighting of sequentially sampled advice (Hogarth & Einhorn, 1992). Moreover, due to differences in perceived expertise, advisors are often egocentrically discounted or, in other words, weighted less strongly than the self (Harvey & Fischer, 1997; Yaniv & Kleinberger, 2000). The same likely holds true for perceived or actual differences in expertise among a set of distinct advisors (Brehmer & Hagafors, 1986; Harvey & Harries, 2004). Further counterexamples include the stronger egocentric discounting of relatively more as compared to less distant advice (Yaniv & Milyavsky, 2007), and that judgment updating was found to be proportional to multiple judges' consensus (Schultze et al., 2022). In general, presupposing equal weighting of multiple advisors conceals potentially differential weighting with respect to, for instance, expertise or sampling position. Essentially, the latter is not limited to applications that only involve sequential sampling of multiple pieces of advice. Instead, the notion of sampling positions also applies to scenarios in which attention is sequentially allocated to different pieces of advice presented simultaneously. Therefore, our criticism of assuming equal effects of multiple social influences also applies to the Behavioral Experiment of Molleman et al. (2020), in which participants received advice from multiple peers simultaneously.

4.1 | Multiple (advisors) regression modeling

In MER, each sampled piece of advice may be construed as additional predictor on the level of trials. Treating the advice coefficients of Equation (15) as free parameters, separate weights of individual pieces of advice can be estimated by fitting the following model:

$$F_{ij} = \sum_{k=1}^{K_{ij}} \omega_{A_{jk}} A_{ijk} + \left(1 - \sum_{k=1}^{K_{ij}} \omega_{A_{jk}}\right) I_{ij} + \varepsilon_{ij}, \quad (16)$$

$$\omega_{A_{ijk}} = \beta_{A_k} + \alpha_{A_k, i}^S + \alpha_{A_k, j}^T, \quad (17)$$

where $\omega_{A_{jk}}$ denotes the sum-to-one constrained weights of the k th piece of advice A_k encountered or sampled during the i th trial. In the past, utilization of multiple pieces of advice per judgment was indeed assessed by regressing final judgments simultaneously on all sources of information (e.g., Brehmer & Hagafors, 1986; Harvey et al., 2000; Lim & O'Connor, 1995).

Multiple advice taking scenarios provide an additional substantive reason against commonly used outlier criteria for advice weighting measures. Imagine a participant in the experiment of Molleman et al. (2020) who made an initial judgment of 100 and received advice of 80, 120, and 140. In general, it would be rather surprising to observe a final judgment that is smaller than 100 given the consensus of the majority of judges that the true value should be equal to or larger than 100. For instance, equal weighting suggests a final judgment of 110. Essentially, looking at the weighting of the first piece of advice in isolation, we would conclude having observed a “perverse” judgmental shift in terms of a ROD-WOA of -0.50 , because this hypothetical advisee has shifted away from the first piece of advice of 80. Accordingly, given a certain sequence of advice, parts of which may also be unobserved in the real world, negative weights (e.g., of advice on the opposite side of the majority consensus) indeed constitute quite reasonable behavior in terms of deliberate linear combinations of multiple judgments. This example once again nicely illustrates why it is so important to be able to estimate separate weights for each individual piece of advice, as will be demonstrated in the following third empirical application.

4.2 | Third empirical application

Molleman et al. (2020) conducted an online experiment to investigate how people integrate disparate social information from multiple, simultaneously presented advisors. In a perceptual judgment task that has been shown to have a high test-retest reliability, participants recruited via Amazon's Mechanical Turk and rewarded for accuracy were asked to estimate the number of animals shown for 6 s in 20 images of herds with 50 to 100 animals in total (Molleman et al., 2019). After making independent initial judgments on visual sliders ranging from 1 to 150, participants were provided with their own and three advisors' judgments from a pretest with no social information to incorporate into their final judgments. While keeping constant the distance between the sample mean of the distributions of social information and participants' initial judgment, the variance (low vs. high) and in the high-variance condition also the skew (not skewed vs. skewed) of the distributions were systematically manipulated within participants. Crucially, the advisors' judgments always pointed in the same and correct direction. Advice weighting, or what the authors originally referred to as “social information use,” was measured as shifting toward the sample mean of a particular distribution, that is, implicitly assuming equal weighting of all three sources of information.

According to the original results, participants egocentrically discounted advice in all treatment conditions. However, the sample mean of advice was weighted more strongly in the condition with low variance and no skew than in all conditions with high variance. Moreover, the direction of skew in the high-variance conditions affected participants' advice taking behavior as follows: For a cluster of two advisors that were relatively close to participants' initial judgment, they shifted more toward the sample mean of advice compared to a cluster of two advisors that were relatively far from their initial

⁸Conversely, this modeling procedure assumes that participants build the mean of all advice values and take this summary value weighted by the total ROD-WOA (cf. Hogarth & Einhorn, 1992, eq. 5). To the best of our knowledge, there are no published findings about the relevance of this implicit assumption.

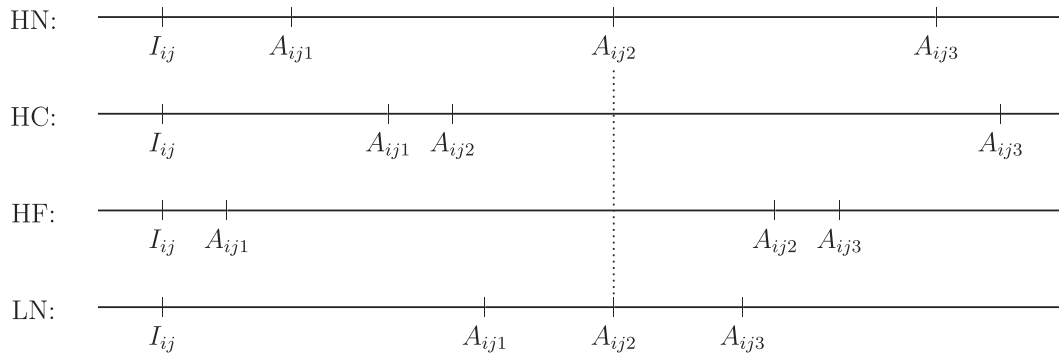


FIGURE 3 Distributions of social information across the experimental conditions implemented in Molleman et al. (2020). Note: For the experimental conditions, we have HN, high variance, no skew; HC, high variance, cluster close to I_{ij} ; HF, high variance, cluster far from I_{ij} ; and LN, low variance, no skew. Across the four conditions, the distance between the sample mean (dotted line) of advice A_{ijk} , $k = 1, \dots, 3$, and participants' initial judgment I_{ij} was held constant. Adapted from Molleman et al. (2020, fig. 1e). CC BY 4.0.

judgment. These results highlight that variance and skew in social information distributions significantly influence participants' utilization of advice at the aggregate level. However, participants were also shown to consistently exhibit inter-individual differences in their selections of advice weighting strategies across conditions (cf. Rebholz et al., 2023). Also note that on trials in the control condition, which mimicked the four distributions of the experimental conditions but are excluded as participants only acted as the observers of four judgments of peers and did not see the stimulus items, their own judgments closely matched the sample mean of all observed peer judgments. Moreover, more weight was assigned to peer judgments that were closer to each other as analyzed descriptively and by means of a cognitive model that emulated Bayesian updating and implemented different proximity-based weighting mechanisms (see Molleman et al., 2020, figs. 2 and 3, respectively).

4.2.1 | Methods

For reanalyzing the experimental data of Molleman et al. (2020) by means of the proposed modeling approach, we extended the multiple advisors model from Equations (16) and (17) for $K_{ij} = K = 3$ as follows:

$$F_{ij} = \omega_{A_{ij1}} A_{ij1} + \omega_{A_{ij2}} A_{ij2} + \omega_{A_{ij3}} A_{ij3} + (1 - \omega_{A_{ij1}} - \omega_{A_{ij2}} - \omega_{A_{ij3}}) I_{ij} + \varepsilon_{ij}. \quad (18)$$

$$\omega_{A_{ijk}} = \beta_{A_k} + \alpha_{A_k, i}^S + \alpha_{A_k, j}^T + \beta_{A_k \times HC} HC_{ij} + \beta_{A_k \times HF} HF_{ij} + \beta_{A_k \times LN} LN_{ij}. \quad (19)$$

Accordingly, the estimates $\hat{\omega}_{A_{ij1}}$, $\hat{\omega}_{A_{ij2}}$, and $\hat{\omega}_{A_{ij3}}$ measure the separate weights assigned to all three advisors A_{ij1} , A_{ij2} , and A_{ij3} , respectively. The advice was sorted by proximity to the participants' initial judgment I_{ij} , so that $\hat{\beta}_{A_1}$ measures the mean weight of the closest advisor's judgment, whereas $\hat{\beta}_{A_3}$ measures the mean weight of the most distant advisor's judgment. Similar to the random effects structure of the unconstrained regression model in Section 3, we have mutually independent random effects $\alpha^q \sim N_K(\mathbf{0}_{K \times 1}, \Sigma_q)$, $q \in \{S, T\}$, per advisor. Furthermore, the correlation coefficients between random effects are

denoted as $\rho_{A_k A_{k'}, q} \forall k, k' = 1, \dots, K$. Compared to the reference level with high variance and no skew (HN), $\hat{\beta}_{A_k \times HC}$ and $\hat{\beta}_{A_k \times HF}$ measure the fixed trial-wise treatment effects on weighting the advice of clusters of two advisors far from I_{ij} versus close to I_{ij} , respectively, and $\hat{\beta}_{A_k \times LN}$ measures the effect of relatively lower variance. This is achieved by implementing the contrast-coding scheme from Table 4 for the predictors representing the different distributions of social information as exemplified in Figure 3.

The same animals were used multiple times in the experiment of Molleman et al. (2020). Moreover, herds of crickets of size 58 occurred twice during the experimental trials. Therefore, the item index $j = 1, \dots, M$ was created by concatenating the name of the animal, the actual size of the herd, and the number of times participants had already encountered a herd of the same animal and size. This procedure resulted in 20 unique item identifiers, one per experimental trial.

4.2.2 | Results

On average over all treatment conditions, each individual piece of advice was weighted positively in the linear combination of all available judgments as indicated by the significantly positive coefficient estimates $\hat{\beta}_{A_1}$, $\hat{\beta}_{A_2}$, and $\hat{\beta}_{A_3}$ (Table 5). More importantly, all three confidence intervals overlap with each other. As such, there was no evidence for significant differences in the mean weighting of these three individual pieces of advice. Descriptively, however, advice of intermediate distance, A_2 , was on average weighted somewhat more strongly than relatively close advice, A_1 , which again was weighted somewhat more strongly than relatively distant advice, A_3 . This finding is in line with empirical evidence for an inverse-U-shaped relation between advice distance and weighting, where advice of intermediate distance is weighted the most both according to ROD-WOA in the traditional paradigm (Moussaïd et al., 2013; Schultze et al., 2015), as well as according to separate MER-WOAs of sequentially sampled advice (Rebholz, 2023, Chapter 4).

TABLE 4 Contrast-coding scheme of the experimental conditions implemented in Molleman et al. (2020) for the estimation of MER-WOA according to Equations (18) and (19).

Variance	Skew	HC _{ij}	HF _{ij}	LN _{ij}
High (H)	No (N)	-.25	-.25	-.25
	Cluster close to I _{ij} (C)	.75	-.25	-.25
	Cluster far from I _{ij} (F)	-.25	.75	-.25
Low (L)	No (N)	-.25	-.25	.75

Only one treatment effect in the model from Table 5 reached statistical significance. For relatively lower variance in the distribution of social information, the first and closest piece of advice was weighted significantly more strongly than for relatively higher variance. As A₁ was significantly closer to I in terms of absolute distances on HN- (M = 3.93, 95% CI [3.67, 4.20], SD = 2.93) than LN-trials (M = 9.93, 95% CI [9.59, 10.26], SD = 3.72), $t(899.33) = 27.60, p < .001$ (see also Molleman et al., 2020, fig. 1e), this was again in line with an inverse-U-shaped relation between advice distance and weighting. All other treatment effects on separate advice weighting did not reach statistical significance. Descriptively, however, there was a consistent pattern that all three treatment effects positively affected the weighting of the first advisor, whereas there were mixed treatment effects on the weighting of the other two advisors that were more distant to participants' initial judgment.

Compared to the inter-individual differences in advice weighting $\hat{\tau}_{A_k,S}$, there was notably less variability for different items as measured by $\hat{\tau}_{A_k,T}$. Also in comparison to other judgment tasks, such as product carbon footprint or quantity estimation (Rebholz & Hütter, 2022; see also Rebholz, 2023, Chapter 3, for MER-WOAs), the differences between items were rather negligible. In contrast to the item-wise random effects, we indeed found significant differences between some of the participant-wise random effects. As indicated by the non-overlapping bootstrap 95% CIs, there was significantly less variation in participants' weighting of the third and most distant advisor as compared to the two advisors that were relatively closer to participants' initial judgment. This finding, along with the one above that the most distant advisor received the lowest descriptive weight on average, accounted for the notable difference in the weighting distributions between the same two groups of advisors (Figure 4). Across all four treatment conditions, the distributions of MER-WOA for the third advisor were centered much closer to zero and were much narrower than for the first two advisors, for whom the distributions were more similar.

4.2.3 | Discussion

Significance testing for treatment effects on the *separate* MER weights of multiple pieces of advice is one of the key advantages over the more established aggregate ROD-based modeling procedure as applied in the original study. In essence, there is no statistical evidence

for the differential weighting of multiple individual pieces of advice in the Behavioral Experiment of Molleman et al. (2020). Descriptively, however, the typical inverse-U-shaped relationship between advice distance and weighting can also be observed in their data. Together with findings from the literature on (conformity) effects of consensus on advice weighting (e.g., Fiedler et al., 2019; Larson et al., 2020; Minson & Mueller, 2012; Schultze et al., 2019; see also Asch, 1946; Deutsch & Gerard, 1955; Larrick et al., 2012; Schultze et al., 2013), this result can also explain the descriptive patterns as observed for the treatment effects.

In the HC condition, participants form a consensus cluster together with the first and second advisor, whereas advice A₃ represents the minority opinion (see Figure 3). Accordingly, it is not surprising that the supporting majority of judgments is weighted relatively more strongly as compared to the minority (Larson et al., 2020). In contrast, in the HF condition, there are opposing opinion clusters of equal size. Thus, egocentric discounting implies that the single supporting judgment A₁ is weighted relatively more strongly, whereas the cluster containing two opposing judgments is discounted (Schultze et al., 2013; Yaniv & Milyavsky, 2007). Although the sign of $\hat{\beta}_{A_3 \times HF}$ is positive, its relative size supports this interpretation (Table 5). Finally, there is another possible explanation other than the inverse-U-shaped distance-weighting relationship for why the first advisor is weighted significantly more strongly in the LN condition than in the HN condition, despite being closer to I in the latter. Related to the salience of consensus in multiple advice taking scenarios that is absent in the traditional JAS paradigm, whereas the judge is part of the largest crowd (i.e., narrowest consensus region) on HN-trials, participants may overvalue the single advisor who is closest to them in a crowd of advisors that constitutes a relatively distant consensus region on LN-trials.

Participants' initial judgments consistently marked extreme values of the judgment distributions in the sense that all advisors were on the same side of it. On the one hand, this constitutes an important limitation of the ecological validity of the paradigm implemented in Molleman et al. (2020). On the other hand, treatment effects might become more pronounced when the initial judgments can also lie within the range of advice, such as in Yaniv and Milyavsky (2007) or Schultze et al. (2022). Qualitatively different constellations in terms of the coverage and location of initial judgments can be included as additional predictors with fixed effects on the MER-WOA.

From a technical point of view, the risk of linear dependencies ("redundancy;" Soll & Larrick, 2009; Yaniv et al., 2009) between the advice A_{ijk} and A_{ijk'}}, $\forall k \neq k'$, or I_{ij} naturally increases in K. Accordingly, the reliability of weights estimated from a regular, non-hierarchical regression model with fixed order effects implemented as separate coefficients is likely to suffer from multicollinearity for large K (see also Hoffman, 1960). In contrast, the proposed multilevel modeling framework implements separate coefficient regressions at the weighting level. And because multilevel modeling is mathematically equivalent to ridge regression, it provides multicollinearity stable shrinkage estimates of WOA by design (Baayen & Linke, 2020; Brown et al., 2018; Helwig, 2017; see Section 5 for more details). The downside of

	Estimate	95% CI	SE	t	df	p
β_{A_1}	0.1201*	[0.0130, 0.2273]	0.0546	2.20	615.23	.028
β_{A_2}	0.1420*	[0.0273, 0.2566]	0.0584	2.43	1111.05	.015
β_{A_3}	0.0954*	[0.0125, 0.1784]	0.0423	2.26	528.66	.024
$\beta_{A_1 \times HC}$	0.1118	[-0.1543, 0.3779]	0.1357	0.82	891.38	.410
$\beta_{A_2 \times HC}$	0.1215	[-0.1395, 0.3825]	0.1331	0.91	1209.38	.362
$\beta_{A_3 \times HC}$	-0.0055	[-0.0754, 0.0644]	0.0356	-0.15	28.19	.877
$\beta_{A_1 \times HF}$	0.0980	[-0.1403, 0.3363]	0.1215	0.81	386.09	.420
$\beta_{A_2 \times HF}$	-0.0564	[-0.3864, 0.2737]	0.1683	-0.34	600.90	.738
$\beta_{A_3 \times HF}$	0.0024	[-0.2800, 0.2848]	0.1440	0.02	442.20	.987
$\beta_{A_1 \times LN}$	0.2621*	[0.0394, 0.4849]	0.1136	2.31	667.22	.021
$\beta_{A_2 \times LN}$	-0.1415	[-0.3458, 0.0629]	0.1042	-1.36	832.45	.175
$\beta_{A_3 \times LN}$	0.0790	[-0.0802, 0.2382]	0.0812	0.97	504.58	.331
$\tau_{A_1,S}$	0.1700	[0.1218, 0.2152]				
$\tau_{A_2,S}$	0.1225	[0.0915, 0.1562]				
$\tau_{A_3,S}$	0.0665	[0.0417, 0.0885]				
$\tau_{A_1,T}$	0.0146	[0.0000, 0.0468]				
$\tau_{A_2,T}$	0.0127	[0.0033, 0.0424]				
$\tau_{A_3,T}$	0.0179	[0.0034, 0.0391]				
$\rho_{A_1 A_2,S}$	0.4375	[0.0853, 0.7595]				
$\rho_{A_1 A_3,S}$	-0.1465	[-0.4964, 0.3309]				
$\rho_{A_2 A_3,S}$	0.3484	[-0.0922, 0.9314]				
$\rho_{A_1 A_2,T}$	0.9693	[-0.9860, 0.9988]				
$\rho_{A_1 A_3,T}$	-0.9881	[-0.9995, 0.5733]				
$\rho_{A_2 A_3,T}$	-0.9956	[-1.0000, 0.8181]				
σ	2.6637	[2.5678, 2.7506]				
R^2_{marg}	0.73					

TABLE 5 Multilevel model according to Equations (18) and (19) for Molleman et al. (2020).

Note: Two-sided * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$. Wald 95% CI for fixed and bootstrap 95% CI (with 1000 iterations) for random effects are shown. Sample sizes of $N = 95$ participants S and $M = 20$ stimulus items T resulted in a total number of 1900 observations.

this property, however, is that the estimates for MER-WOAs are biased toward the sample means of the crossed clustering variables, typically participants i and stimulus items j .⁹ A more thorough discussion of the technical properties of MER-WOAs is provided in the following section.

⁹Consider a hypothetical single advice taking experiment with only one item per participant, and thus for simplicity, no item-wise random effects $\alpha_{A_j}^T$. If only one participant i neglects the advice, and all other participants $i \neq i' \forall i, i' = 1, \dots, N$ assign a non-negative weight to the advice, then a good model of participants' behavior would predict the lowest conditional mode across all participant-wise random effects $\alpha_{A_j}^S$ for this specific participant's random effect $\alpha_{A_j}^S$ (see also Footnote 3). Nevertheless, it would be unreasonable to expect $\hat{\alpha}_{A_j}^S = -\hat{\beta}_A$, implying $\hat{\omega}_{A_j} = \hat{\beta}_A + \hat{\alpha}_{A_j}^S = \hat{\beta}_A - \hat{\beta}_A = 0$, due to the shrinkage property of the proposed MER-WOA. Instead, incorporating information from the whole sample's behavior in this hypothetical study, it is more likely to observe $\hat{\omega}_{A_j} > 0$, but ceteris paribus also $\hat{\omega}_{A_j} > \hat{\omega}_{A_{j'}} \forall i \neq i'$, such that at least the relative rankings of the different weighting strategies are preserved.

5 | GENERAL DISCUSSION

We proposed MER-WOAs to analyze advice taking behavior and data related to other information acquisition phenomena with similar structure (e.g., anchoring effects, hindsight bias, or attitude change). The method is based on exploiting or explicitly modeling the multilevel structure of most data collected in these paradigms. The multilevel data structure results from the experimental crossing of grouping factors, often at least participants and stimulus items, rarely also advisors. As a consequence, and in contrast to state-of-the-art ROD modeling (Bailey et al., 2022), the proposed framework is consistent with the endogenous formation of judgments based on exogenous sources of information such as advice. It also has many technical merits, some of which are discussed in more detail below.

In several reanalyses of empirical data, the more advanced modeling approach has been demonstrated to be practically applicable and easy to implement. More importantly, the proposed method has been shown to provide more clear-cut as well as novel and extended

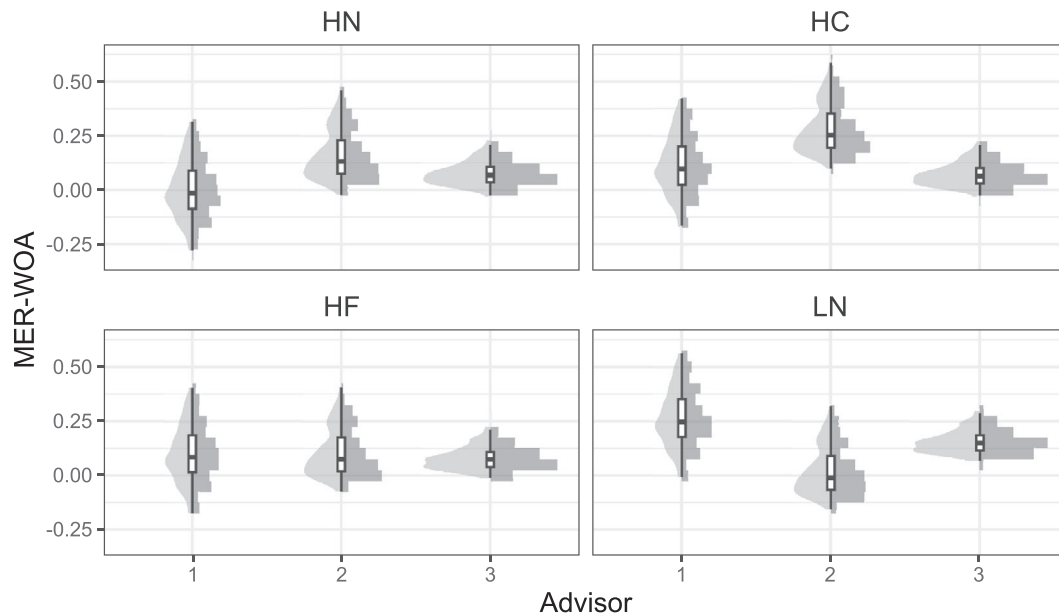


FIGURE 4 Distributions and box plots of the mixed-effects regression weight of advice (MER-WOA) per experimental condition and advisor in Molleman et al. (2020). Note: Distributions are represented as half-violin plots (i.e., smoothed densities) on the left and histograms on the right. Summary statistics are presented as box plots and do not include outliers. For the experimental conditions, we have HN, high variance, no skew; HC, high variance, cluster close to I_{ij} ; HF, high variance, cluster far from I_{ij} ; and LN, low variance, no skew.

evidence not only for more traditional advice taking paradigms, such as the one used by Logg et al. (2019, Experiment 4). In particular, the blind spot of informational influences without recording initial judgments, as in Mayer and Heck (2022), could be resolved for the first time. Also, in scenarios where participants are given access to multiple advisors, we provide initial insights into the differential weighting of individual pieces of advice. In fact, our results do not provide evidence for statistically significant differences in participants' separate weighting of individual pieces of advice, and thus support the more simplistic total weighting approximation of Molleman et al. (2020).

In methodological research, “more advanced” is often connoted with “more complex” modeling. In fact, instead of calculating individual weights per trial using simple arithmetic, a statistical optimization procedure is required to estimate mean weights and predict individual deviations thereof, which is relatively more complex in practice. However, at least parsimony within the multilevel regression framework can be controlled by means of model specification. For instance, modeling advisors as a third, crossed clustering instance would reduce the problem of estimating separate mixed-effects coefficient regressions for multiple advisors to estimating only one additional variance term at the weighting level. Although this can help reduce the statistical complexity of the estimation problem, the interpretation of the estimated weights is tied to their specification as random effects (see also Rebholz, 2023, Chapter 4).

To take advantage of all the benefits of MER-WOA that are presented in this paper, it seems worthwhile to sacrifice parsimony in favor of more complex model building and simultaneous statistical optimization. The prevailing demand for inferential conclusions anyway challenges the existence of factual differences in (model) complexity between our

framework and the more established modeling approach. The reason is that in the latter, researchers have to rely on a two-step procedure in which purely descriptive ROD-WOAs from the first step are used in the second step as dependent variables in statistical test procedures, such as multilevel modeling (e.g., Ache et al., 2020; Hütter & Fiedler, 2019; Minson & Mueller, 2012; Rebholz & Hütter, 2022; Schultze et al., 2015). Therefore, these and other important merits and limitations of the two methods will be critically evaluated in the following.

5.1 | The blind spot of (extremely) close advice

Reported evidence suggests that close advice is not taken in terms of judgmental shift but increased confidence instead (Hütter & Ache, 2016; Moussaïd et al., 2013; Schultze et al., 2015; see Soll et al., 2022, for a more detailed discussion). Everything else being equal, ROD-WOA converges to infinity for closer advice as the denominator approaches zero for smaller distances between advice and initial judgment (see Equation 1). Therefore, the probability to be classified as outlier and excluded from the analyses or truncated to one is relatively high for close advice. Accordingly, conclusions from corresponding evidence do not reflect the most relevant cases with very high or even undefined (ROD) weights of extremely close advice. The proposed MER approach, by contrast, provides well-defined—in terms of person- and item-specific deviations from mean weighting tendencies—estimates of weights for these situations as well. For the limiting case in which advice and initial judgment are the same, we can still find informative partial effects as long as there is sufficient variance in advice distance over all trials. From a substantive point of

view, MER-WOAs thus also carry the potential to clarify empirical findings about (extremely) close advice. In other words, research on judgmental shifts are enabled even in advice distance regions that are highly sensitive in terms of tiny distance changes implying huge judgmental shifts as measured by ROD-WOA.

For complete consideration of information from all experimental trials that were conducted, a healthy modeling strategy should be resilient to data pre-processing artifacts. The outlier sensitivity of ROD-WOA, caused by its specific arithmetic as described above, often makes exclusions or truncation unavoidable in established empirical practice (e.g., Hütter & Ache, 2016; Schultze et al., 2015). By contrast, shrunken MER-WOAs do not require the implementation of often complex and ambiguous (see also the example in Section 3) outlier specifications. As data pre-processing implies researcher degrees of freedom, the proposed method can thus also be seen as a tool to counteract the reproducibility crisis. This is achieved in addition to enabling extensions of existing empirical findings or developing new substantive research domains.

5.2 | Admissible interpretation of (shrunken) regression coefficients as “weights”

On balance, Bonaccio and Dalal (2006) favor regression modeling, more specifically, dominance weights of advice.¹⁰ Historically, regression analysis of advice taking data relied on “utilization indices rather than beta weights” (Harvey et al., 2000, p. 258) for two reasons. First, ordinary regression coefficients do not reflect individual differences (Bauer, 2011; Bonaccio & Dalal, 2006). This is no longer a limitation of MER-WOAs from the proposed multilevel modeling framework. Second, regression coefficients are problematic indicators of variable importance in case of multicollinearity—especially for regressing on the judgments of multiple, interrelated judges (Harvey et al., 2000; Hoffman, 1960). Thus, a limitation of the new approach is its proposal of potentially poor measures of information utilization. On the one hand, however, ROD-WOA as the most common utilization index in advice taking research also effectively conceptualizes importance in terms of beta weights as became obvious from rearranging the formula (see Equation 2). Accordingly, the proposed MER approach relies on a well-established convention. On the other hand, rearrangement additionally established a crucial qualification of this limitation. In the most parsimonious case, sum-to-one constrained MER-WOAs can be estimated free of multicollinearity based on a multilevel model with advice distance as the only predictor (see Equation 3).

Multilevel modeling is mathematically equivalent to ridge regression (Baayen & Linke, 2020; Brown et al., 2018). In classical ridge regression, coefficient stability is achieved by adding a penalization term of the squared coefficients to the optimization problem

(Helwig, 2017). Optimally balancing the bias-variance trade-off reduces the expected mean squared error of the resulting shrinkage estimates. Consequently, penalization is the key to stabilizing coefficients also from more complex regression models with many predictors, such as additional control variables. Put differently, MER-WOAs (i.e., shrinkage estimates) are generally stable against multicollinearity by design. Accordingly, the proposed regression-based measures of advice weighting are appropriate also for testing theories that involve additional predictors.

5.3 | Model comparisons and nonlinear advice weighting

Additional predictors can be added at any level of a multilevel regression model. Modeling that explicitly considers endogenous judgment formation thus allows for the consideration of additional predictors at any of the grouping levels or on the level of trials, which may but do not necessarily have random effects on judgment too. By simply adding such variables (e.g., additional cues to expertise) to the regression model at the respective level, researchers can investigate the weighting behavior more holistically. In general, any number and type of main or (cross-level) interaction effects of additional predictors on judgment or weighting can be assessed. An accordingly extended system of Equations (3) and (4) thus also allows to test moderation effects on the WOA or to compare different weighting policies.

Often people combine different sources of information in a nonlinear way (e.g., Ganzach, 1995, for clinical judgment). Actually, ROD-WOA usually features an inverse-U-shaped relation with advice distance in many empirical studies: Close and more distant advice is typically taken less than advice of “intermediate” distance (Hütter & Ache, 2016; Moussaïd et al., 2013; Rebholz, 2023; Schultze et al., 2015). As the regression framework allows to incorporate any number and type of additional predictors at any level, researchers no longer have to assume *linear* processing of advice for calculating the weights. To model nonlinear weighting as requested by Bailey et al. (2022), the set of additional predictors can include higher order terms such as polynomials (e.g., Schreiner et al., 2023) or nonlinear transformations of advice (distance). Base expansions provide an even more flexible solution in the “generalized additive modeling” framework (e.g., Baayen & Linke, 2020). In essence, as already requested by Slovic and Lichtenstein (1971), the proposed multilevel modeling approach allows statistical tests of alternative theoretical accounts of external information utilization.

5.4 | On the duality of advice utilization and beyond

More recently, advice taking research focused on dimensions of advice utilization beyond judgmental shift (e.g., Soll et al., 2022; Yaniv et al., 2009). Essentially, Moussaïd et al. (2013) were the first to find that the utilization dimension depends on the distance of advice.

¹⁰General dominance weights of advice capture “the average percentage increase in criterion variance explained ... when the focal predictor [i.e., advice] is added to models containing all possible subsets of the other predictors” (Bonaccio & Dalal, 2006, p. 142). As investigations of treatment effects would thus require comparing dominance weights between different regressions models, Önkal et al. (2009) favored ROD-WOAs instead.

Instead of causing judgmental shifts from initial to final estimation, the confirmatory value of close advice is taken by participants in terms of enhanced confidence (as measured on 6-point Likert scales) in their initial estimates. This so-called “duality of advice utilization” (Schultze et al., 2015, p. 170) is assumed to be responsible for the characteristic inverse-U-shaped relation of advice weighting and distance. In a regression framework, variables from different scales can be included side by side in the same statistical model, at both the independent and dependent levels. For instance, although ordinal confidence and numerical judgments were measured on different scales in the example above, MER-confidence weights of advice can be derived for confidence as a second dependent variable regressed on advice.

Extensions to multivariate multilevel regression, that is, multiple response variables modeled collectively in one larger regression model, are equally straightforward (Snijders & Bosker, 2012, Chapter 16). Such a model can capture advice taking simultaneously on multiple dimensions, for instance, judgment shift *plus* confidence change. In contrast to the “influence of advice” measure proposed by Soll et al. (2022), the duality of advice taking is reflected by dimensional interdependencies of the coefficient regressions. In addition, judgment accuracy can further enrich the corresponding multivariate multilevel regression modeling and hypothesis testing on a third, wisdom of crowds-related dimension, and so on. Notably, multivariate methods are especially powerful if the set of dependent variables is strongly correlated, such as judgment and confidence as described above. Accordingly, a valid and powerful test of multidimensional advice taking naturally requires joint hypothesis testing of the multidimensional weights of advice and thus ultimately a regression-based procedure.

For the sake of brevity, we refrain from discussing other valuable extensions such as the following two examples in detail. For instance, multilevel quantile regression could be used to examine statistics of distributions of weights other than their means. Furthermore, the ROD formula is extremely limited in explanatory power for choices among a set of discrete, qualitatively different alternatives (Bailey et al., 2022). For binary choice, ROD-WOA merely describes “matching” in terms of accepting or disregarding advice (Snizek & Buckley, 1995). Instead, implementing appropriate link functions such as logit or probit for modeling qualitative decisions in a generalized multilevel regression or linear mixed modeling framework allows for more informative weighting parameters, such as choice probabilities or odds ratios.

5.5 | Toward a process-consistent modeling of information sampling and utilization

Unfortunately, multilevel regression models of advice taking will not be a panacea after all. For instance, the estimation of mixed-effects generally requires a relatively large number of observations. Beside being increasingly common in experimental research practice, however, it is nowadays at least possible to estimate the practical constrain (i.e., number of

subjects and items required) of an experiment that is analyzed by means of multilevel modeling (Brybaert & Stevens, 2018; Green & MacLeod, 2016). More importantly, the traditional modeling of advice taking is responsible for a variety of limitations (see Bonaccio & Dalal, 2006, for a review). For instance, there may be sources of information other than a single piece of advice and, if any (see Section 3), one's independent initial judgment. If multiple pieces of external evidence are presented simultaneously (e.g., Molleman et al., 2020), the calculation of the established weighting index of Harvey and Fischer (1997) for each individual source of information is infeasible. Previous regression-based approaches have already taken into account the endogenous formation of judgments (e.g., Brehmer & Hagafors, 1986; Harvey et al., 2000; Lim & O'Connor, 1995). Put differently, we propose to model endogenous variables such as updated beliefs as—potentially nonlinearly—weighted combinations of exogenous sources of information. However, unlike these previous analytical methods, which are limited to aggregated inferences, the proposed multilevel regression framework still enables researchers to estimate individual weights for each experimental trial.

5.6 | Conclusion

In summary, the main strength of the proposed MER weights (of advice) is that they enable research on individual differences in the sampling and utilization of information from a variety of sources, on multiple dimensions, and beyond social contexts. New information technologies and online networks make it increasingly convenient to access multiple opinions and advisors. Therefore, more flexible tools for examining informational influences in these situations are invaluable.

ACKNOWLEDGMENTS

This research was funded by the Deutsche Forschungsgemeinschaft (DFG), grant 2277, Research Training Group “Statistical Modeling in Psychology” (SMiP). Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

We have no known conflicts of interest to disclose.

DATA AVAILABILITY STATEMENT

Reproducible analysis scripts are available in Open Science Framework at <https://osf.io/6gmhs/>. The data that support the findings of this study were derived from the following resources available in the public domain: Logg et al. (2019, Experiment 4), <https://osf.io/y2thn/>; Mayer and Heck (2022, Experiments 1 and 2), <https://osf.io/96nsk/>; and Molleman et al. (2020, Behavioral Experiment), <https://osf.io/dasn2/>.

ORCID

Tobias R. Rebholz  <https://orcid.org/0000-0001-5436-0253>

Marco Biella  <https://orcid.org/0000-0002-8039-0170>

Mandy Hütter  <https://orcid.org/0000-0002-0952-3831>

REFERENCES

- Ache, F., Rader, C. A., & Hütter, M. (2020). Advisors want their advice to be used—But not too much: An interpersonal perspective on advice taking. *Journal of Experimental Social Psychology*, 89, 103979. <https://doi.org/10.1016/j.jesp.2020.103979>
- Afshartous, D., & de Leeuw, J. (2005). Prediction in multilevel models. *Journal of Educational and Behavioral Statistics*, 30(2), 109–139. <https://doi.org/10.3102/10769986030002109>
- Allison, P. D. (1990). Change scores as dependent variables in regression analysis. *Sociological Methodology*, 20, 93–114. <https://doi.org/10.2307/271083>
- Anderson, N. H. (1981). *Foundations of information integration theory*. Academic Press.
- Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal Psychology*, 41, 258–290. <https://doi.org/10.1037/h0055756>
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Baayen, R. H., & Linke, M. (2020). Generalized additive mixed models. In Paquot, M., & Gries, S. T. (Eds.), *A practical handbook of corpus linguistics*. Springer International Publishing, pp. 563–591.
- Bailey, P. E., Leon, T., Ebner, N. C., Moustafa, A. A., & Weidemann, G. (2022). A meta-analysis of the weight of advice in decision-making. *Current Psychology*, 42, 24516–24541. <https://doi.org/10.1007/s12144-022-03573-2>
- Bates, D. M., Mächler, M., Bolker, B. M., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bauer, D. J. (2011). Evaluating individual differences in psychological processes. *Current Directions in Psychological Science*, 20(2), 115–118. <https://doi.org/10.1177/0963721411402670>
- Bochner, S., & Insko, C. A. (1966). Communicator discrepancy, source credibility, and opinion change. *Journal of Personality and Social Psychology*, 4(6), 614–621. <https://doi.org/10.1037/h0021192>
- Bonaccio, S., & Dalal, R. S. (2006). Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational Behavior and Human Decision Processes*, 101(2), 127–151. <https://doi.org/10.1016/j.obhdp.2006.07.001>
- 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(2), 181–195. [https://doi.org/10.1016/0749-5978\(86\)90015-4](https://doi.org/10.1016/0749-5978(86)90015-4)
- Brown, L. D., Mukherjee, G., & Weinstein, A. (2018). Empirical Bayes estimates for a two-way cross-classified model. *The Annals of Statistics*, 46(4), 1693–1720. <https://doi.org/10.1214/17-AOS1599>
- Brybaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, 1(1), 1–20. <https://doi.org/10.5334/joc.10>
- Budescu, D. V., & Rantilla, A. K. (2000). Confidence in aggregation of expert opinions. *Acta Psychologica*, 104(3), 371–398. [https://doi.org/10.1016/S0001-6918\(00\)00037-8](https://doi.org/10.1016/S0001-6918(00)00037-8)
- Budescu, D. V., Rantilla, A. K., Yu, H.-T., & Karelitz, T. M. (2003). The effects of asymmetry among advisors on the aggregation of their opinions. *Organizational Behavior and Human Decision Processes*, 90(1), 178–194. [https://doi.org/10.1016/S0749-5978\(02\)00516-2](https://doi.org/10.1016/S0749-5978(02)00516-2)
- Burton, J. W., Stein, M.-K., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making*, 33(2), 220–239. <https://doi.org/10.1002/bdm.2155>
- Chiang, A. C. (1984). *Fundamental methods of mathematical economics* (3rd ed.). McGraw-Hill.
- Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, 5(4), 559–583. [https://doi.org/10.1016/0169-2070\(89\)90012-5](https://doi.org/10.1016/0169-2070(89)90012-5)
- Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science*, 19(1), 51–57. <https://doi.org/10.1177/0963721409359277>
- Cronbach, L. J. (1943). Note on the reliability of ratio scores. *Educational and Psychological Measurement*, 3(1), 67. <https://doi.org/10.1177/001316444300300106>
- Cronbach, L. J., & Furby, L. (1970). How we should measure “change”—Or should we? *Psychological Bulletin*, 74(1), 68–80. <https://doi.org/10.1037/h0029382>
- Davis-Stober, C. P., Budescu, D. V., Dana, J., & Broomell, S. B. (2014). When is a crowd wise? *Decision*, 1(2), 79–101. <https://doi.org/10.1037/dec0000004>
- Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgement. *Journal of Abnormal Psychology*, 51(3), 629–636. <https://doi.org/10.1037/h0046408>
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 114–126. <https://doi.org/10.1037/xge0000033>
- Edwards, J. R. (1995). Alternatives to difference scores as dependent variables in the study of congruence in organizational research. *Organizational Behavior and Human Decision Processes*, 64(3), 307–324. <https://doi.org/10.1006/obhd.1995.1108>
- Epley, N., & Gilovich, T. (2006). The anchoring-and-adjustment heuristic: Why the adjustments are insufficient. *Psychological Science*, 17(4), 311–318. <https://doi.org/10.1111/j.1467-9280.2006.01704.x>
- Fiedler, K., Hütter, M., Schott, M., & Kutzner, F. (2019). Metacognitive myopia and the overutilization of misleading advice. *Journal of Behavioral Decision Making*, 32(3), 317–333. <https://doi.org/10.1002/bdm.2109>
- Fiedler, K., & Kutzner, F. (2015). Information sampling and reasoning biases. In Keren, G., & Wu, G. (Eds.), *The Wiley Blackwell handbook of judgment and decision making*. John Wiley & Sons, pp. 380–403.
- Firebaugh, G., & Gibbs, J. P. (1985). User's guide to ratio variables. *American Sociological Review*, 50(5), 713–722. <https://doi.org/10.2307/2095384>
- Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics*, 40(1), 35–42. <https://doi.org/10.1016/j.socsec.2010.10.008>
- Galton, F. (1907). Vox populi. *Nature*, 75(1949), 450–451. <https://doi.org/10.1038/075450a0>
- Ganzach, Y. (1995). Nonlinear models of clinical judgment: Meehl's data revisited. *Psychological Bulletin*, 118(3), 422–429. <https://doi.org/10.1037/0033-2909.118.3.422>
- 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(2), 234–245. <https://doi.org/10.1016/j.obhdp.2008.03.001>
- Gino, F., & Moore, D. A. (2007). Effects of task difficulty on use of advice. *Journal of Behavioral Decision Making*, 20(1), 21–35. <https://doi.org/10.1002/bdm.539>
- Gino, F., & Schweitzer, M. E. (2008). Blinded by anger or feeling the love: How emotions influence advice taking. *The Journal of Applied Psychology*, 93(5), 1165–1173. <https://doi.org/10.1037/0021-9010.93.5.1165>
- Gino, F., Shang, J., & Croson, R. (2009). The impact of information from similar or different advisors on judgment. *Organizational Behavior and Human Decision Processes*, 108(2), 287–302. <https://doi.org/10.1016/j.obhdp.2008.08.002>
- Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, 7(4), 493–498. <https://doi.org/10.1111/2041-210X.12504>
- Harvey, N., & Fischer, I. (1997). Taking advice: Accepting help, improving judgment, and sharing responsibility. *Organizational Behavior and*

- Human Decision Processes*, 70(2), 117–133. <https://doi.org/10.1006/obhd.1997.2697>
- Harvey, N., & Harries, C. (2004). Effects of judges' forecasting on their later combination of forecasts for the same outcomes. *International Journal of Forecasting*, 20(3), 391–409. <https://doi.org/10.1016/j.ijforecast.2003.09.012>
- Harvey, N., Harries, C., & Fischer, I. (2000). Using advice and assessing its quality. *Organizational Behavior and Human Decision Processes*, 81(2), 252–273. <https://doi.org/10.1006/obhd.1999.2874>
- Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin*, 107(3), 311–327. <https://doi.org/10.1037/0033-2909.107.3.311>
- Hell, W., Gigerenzer, G., Gauggel, S., Mall, M., & Müller, M. (1988). Hindsight bias: An interaction of automatic and motivational factors?. *Memory & Cognition*, 16(6), 533–538. <https://doi.org/10.3758/BF03197054>
- Helwig, N. E. (2017). Adding bias to reduce variance in psychological results: A tutorial on penalized regression. *The Quantitative Methods for Psychology*, 13(1), 1–19. <https://doi.org/10.20982/tqmp.13.1.p001>
- Henriksson, M. P., Elwin, E., & Juslin, P. (2010). What is coded into memory in the absence of outcome feedback? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(1), 1–16. <https://doi.org/10.1037/a0017893>
- Himmelstein, M., & Budescu, D. V. (2022). Preference for human or algorithmic forecasting advice does not predict if and how it is used. *Journal of Behavioral Decision Making*, 36(1), e2285. <https://doi.org/10.1002/bdm.2285>
- Hoffman, P. J. (1960). The paramorphic representation of clinical judgment. *Psychological Bulletin*, 57(2), 116–131. <https://doi.org/10.1037/h0047807>
- Hoffrage, U., Hertwig, R., & Gigerenzer, G. (2000). Hindsight bias: A by-product of knowledge updating? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(3), 566–581. <https://doi.org/10.1037/0278-7393.26.3.566>
- Hogarth, R. M. (1978). A note on aggregating opinions. *Organizational Behavior and Human Performance*, 21(1), 40–46. [https://doi.org/10.1016/0030-5073\(78\)90037-5](https://doi.org/10.1016/0030-5073(78)90037-5)
- Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology*, 24(1), 1–55. [https://doi.org/10.1016/0010-0285\(92\)90002-J](https://doi.org/10.1016/0010-0285(92)90002-J)
- Hou, Y. T.-Y., & Jung, M. F. (2021). Who is the expert? Reconciling algorithm aversion and algorithm appreciation in AI-supported decision making. *Proceedings of the ACM on Human-Computer Interaction*, 5-(CSCW2), 1–25. <https://doi.org/10.1145/3479864>
- Hütter, M., & Ache, F. (2016). Seeking advice: A sampling approach to advice taking. *Judgment and Decision Making*, 11(4), 401–415.
- Hütter, M., & Fiedler, K. (2019). Advice taking under uncertainty: The impact of genuine advice versus arbitrary anchors on judgment. *Journal of Experimental Social Psychology*, 85, 103829. <https://doi.org/10.1016/j.jesp.2019.103829>
- Jacowitz, K. E., & Kahneman, D. (1995). Measures of anchoring in estimation tasks. *Personality and Social Psychology Bulletin*, 21(11), 1161–1166. <https://doi.org/10.1177/01461672952111004>
- Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology*, 68, 601–625. <https://doi.org/10.1146/annurev-psych-122414-033702>
- Juslin, P., & Olsson, H. (1997). Thurstonian and Brunswikian origins of uncertainty in judgment: A sampling model of confidence in sensory discrimination. *Psychological Review*, 104(2), 344–366. <https://doi.org/10.1037/0033-295X.104.2.344>
- Jussupow, E., Benbasat, I., & Heinzl, A. (2020). Why are we averse towards algorithms? A comprehensive literature review on algorithm aversion. In Rowe, F. (Ed.), *Proceedings of the 28th European Conference on Information Systems (ECIS)*. AISel, pp. RP 168. https://aisel.aisnet.org/ecis2020_rp/168
- Kämmer, J. E., Choshen-Hillel, S., Müller-Trede, J., Black, S. L., & Weibler, J. (2023). A systematic review of empirical studies on advice-based decisions in behavioral and organizational research. *Decision*, 10(2), 107–137. <https://doi.org/10.1037/dec0000199>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Larrick, R. P., Mannes, A. E., & Soll, J. B. (2012). The social psychology of the wisdom of crowds. In Krueger, J. I. (Ed.), *Social judgment and decision making*. Psychology Press, pp. 227–242.
- Larson, J. R., Tindale, R. S., & Yoon, Y.-J. (2020). Advice taking by groups: The effects of consensus seeking and member opinion differences. *Group Processes & Intergroup Relations*, 23(7), 921–942. <https://doi.org/10.1177/1368430219871349>
- Lim, J. S., & O'Connor, M. (1995). Judgemental adjustment of initial forecasts: Its effectiveness and biases. *Journal of Behavioral Decision Making*, 8(3), 149–168. <https://doi.org/10.1002/bdm.3960080302>
- Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 151, 90–103. <https://doi.org/10.1016/j.obhdp.2018.12.005>
- Mahmud, H., Islam, A. K. M. N., Ahmed, S. I., & Smolander, K. (2022). What influences algorithmic decision-making? A systematic literature review on algorithm aversion. *Technological Forecasting and Social Change*, 175, 121390. <https://doi.org/10.1016/j.techfore.2021.121390>
- Mannes, A. E. (2009). Are we wise about the wisdom of crowds? The use of group judgments in belief revision. *Management Science*, 55(8), 1267–1279. <https://doi.org/10.1287/mnsc.1090.1031>
- Mannes, A. E., Soll, J. B., & Larrick, R. P. (2014). The wisdom of select crowds. *Journal of Personality and Social Psychology*, 107(2), 276–299. <https://doi.org/10.1037/a0036677>
- Mayer, M., & Heck, D. W. (2022). Sequential collaboration: The accuracy of dependent, incremental judgments. *Decision*. Advance online publication. <https://doi.org/10.1037/dec0000193>
- Mellers, B., Ungar, L., Baron, J., Ramos, J., Gurcay, B., Fincher, K., Scott, S. E., Moore, D., Atanasov, P., Swift, S. A., Murray, T., Stone, E., & Tetlock, P. E. (2014). Psychological strategies for winning a geopolitical forecasting tournament. *Psychological Science*, 25(5), 1106–1115. <https://doi.org/10.1177/0956797614524255>
- Mercier, H., Yama, H., Kawasaki, Y., Adachi, K., & van der Henst, J.-B. (2012). Is the use of averaging in advice taking modulated by culture? *Journal of Cognition and Culture*, 12(1), 1–16. <https://doi.org/10.1163/156853712X633893>
- Minson, J. A., & Mueller, J. S. (2012). The cost of collaboration: Why joint decision making exacerbates rejection of outside information. *Psychological Science*, 23(3), 219–224. <https://doi.org/10.1177/0956797611429132>
- Molleman, L., Kurvers, R. H. J. M., & van den Bos, W. (2019). Unleashing the BEAST: A brief measure of human social information use. *Evolution and Human Behavior*, 40(5), 492–499. <https://doi.org/10.1016/j.evolhumbehav.2019.06.005>
- Molleman, L., Tump, A. N., Gradassi, A., Herzog, S., Jayles, B., Kurvers, R. H. J. M., & van den Bos, W. (2020). Strategies for integrating disparate social information. *Proceedings of the Royal Society B: Biological Sciences*, 287(1939), 20202413. <https://doi.org/10.1098/rspb.2020.2413>
- Moussaïd, M., Kämmer, J. E., Analytis, P. P., & Neth, H. (2013). Social influence and the collective dynamics of opinion formation. *PLoS ONE*, 8(11), 1–8. <https://doi.org/10.1371/journal.pone.0078433>
- Önköl, D., Goodwin, P., Thomson, M., Gönül, S., & Pollock, A. (2009). The relative influence of advice from human experts and statistical methods on forecast adjustments. *Journal of Behavioral Decision Making*, 22(4), 390–409. <https://doi.org/10.1002/bdm.637>

- Prahl, A., & van Swol, L. (2017). Understanding algorithm aversion: When is advice from automation discounted? *Journal of Forecasting*, 36(6), 691–702. <https://doi.org/10.1002/for.2464>
- R Core Team (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rader, C. A., Soll, J. B., & Larrick, R. P. (2015). Pushing away from representative advice: Advice taking, anchoring, and adjustment. *Organizational Behavior and Human Decision Processes*, 130, 26–43. <https://doi.org/10.1016/j.obhdp.2015.05.004>
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.), Advanced Quantitative Techniques in the Social Sciences Series, Vol. 1. Sage Publications.
- Rebholz, T. R. (2023). *Advanced statistical modeling of ecological constraints in information sampling and utilization*. (Ph.D. Thesis), University of Tübingen. <https://doi.org/10.15496/publikation-83023>
- Rebholz, T. R., & Hütter, M. (2022). The advice less taken: The consequences of receiving unexpected advice. *Judgment and Decision Making*, 17(4), 816–848. <https://doi.org/10.1017/S1930297500008950>
- Rebholz, T. R., Hütter, M., & Voss, A. (2023). *Bayesian advice taking: Adaptive strategy selection in sequential advice seeking*. PsyArXiv. <https://doi.org/10.31234/osf.io/y8x92>
- Reif, J., Larrick, R. P., & Soll, J. B. (2022, November 10–13). *Anchoring the advisor: Do advice-seekers induce cognitive biases in their advisors?* [Poster], Society for Judgment and Decision Making (SJD) Annual Meeting, San Diego, CA. <https://sjdm.org/presentations/2022-Poster-Reif-Jessica-advice-anchoring-influence~.pdf>
- Röseler, L., Incerti, L., Rebholz, T. R., Seida, C., & Papenmeier, F. (2023). *Falsifying the insufficient adjustment model: No evidence for unidirectional adjustment from anchors*. PsyArXiv. <https://doi.org/10.31234/osf.io/jztk2>
- Schreiner, M. R., Quevedo Pütter, J., & Rebholz, T. R. (2023). *Time for an update: Belief updating based on ambiguous scientific evidence* [Manuscript submitted for publication]. Department of Psychology, University of Mannheim.
- Schultze, T., Mojzisch, A., & Schulz-Hardt, S. (2013). Groups weight outside information less than individuals do because they should: Response to Minson and Mueller (2012). *Psychological Science*, 24(7), 1371–1372. <https://doi.org/10.1177/0956797612472206>
- Schultze, T., Mojzisch, A., & Schulz-Hardt, S. (2017). On the inability to ignore useless advice: A case for anchoring in the judge-advisor-system. *Experimental Psychology*, 64(3), 170–183. <https://doi.org/10.1027/1618-3169/a000361>
- Schultze, T., Mojzisch, A., & Schulz-Hardt, S. (2019). Why dyads heed advice less than individuals do. *Judgment and Decision Making*, 14(3), 349–363. <https://doi.org/10.1017/S1930297500004381>
- Schultze, T., Rakotoarisoa, A.-F., & Schulz-Hardt, S. (2015). Effects of distance between initial estimates and advice on advice utilization. *Judgment and Decision Making*, 10(2), 144–171. <https://doi.org/10.1017/S1930297500003922>
- Schultze, T., Treffenstädt, C., & Schulz-Hardt, S. (2022). How people use information about the number and distribution of judgments when tapping into the wisdom of the crowds. *Decision*. Advance online publication. <https://doi.org/10.1037/dec0000194>
- Schweickart, O., Tam, C., & Brown, N. R. (2021). When “bad” is good: How evaluative judgments eliminate the standard anchoring effect. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, 75(1), 56–63. <https://doi.org/10.1037/cep0000209>
- Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance*, 6(6), 649–744. [https://doi.org/10.1016/0030-5073\(71\)90033-X](https://doi.org/10.1016/0030-5073(71)90033-X)
- Sniezek, J. A., & Buckley, T. (1995). Cueing and cognitive conflict in judge-advisor decision making. *Organizational Behavior and Human Decision Processes*, 62(2), 159–174. <https://doi.org/10.1006/obhd.1995.1040>
- Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Sage Publications.
- Soll, J. B., & Larrick, R. P. (2009). Strategies for revising judgment: How (and how well) people use others' opinions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(3), 780–805. <https://doi.org/10.1037/a0015145>
- Soll, J. B., Palley, A. B., & Rader, C. A. (2022). The bad thing about good advice: Understanding when and how advice exacerbates overconfidence. *Management Science*, 68(4), 2949–2969. <https://doi.org/10.1287/mnsc.2021.3987>
- Stewart, N., Chater, N., & Brown, G. D. A. (2006). Decision by sampling. *Cognitive Psychology*, 53(1), 1–26. <https://doi.org/10.1016/j.cogpsych.2005.10.003>
- Surowiecki, J. (2005). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations*. Anchor Books.
- Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, 34(4), 273–286. <https://doi.org/10.1037/h0070288>
- Turner, B. M., & Schley, D. R. (2016). The anchor integration model: A descriptive model of anchoring effects. *Cognitive Psychology*, 90, 1–47. <https://doi.org/10.1016/j.cogpsych.2016.07.003>
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131. <https://doi.org/10.1126/science.185.4157.1124>
- Varian, H. R. (1992). *Microeconomic analysis* (3rd ed.). Norton.
- Yaniv, I. (2004). Receiving other people's advice: Influence and benefit. *Organizational Behavior and Human Decision Processes*, 93(1), 1–13. <https://doi.org/10.1016/j.obhdp.2003.08.002>
- Yaniv, I., Choshen-Hillel, S., & 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(2), 558–563. <https://doi.org/10.1037/a0014589>
- Yaniv, I., & Kleinberger, E. (2000). Advice taking in decision making: Ego-centric discounting and reputation formation. *Organizational Behavior and Human Decision Processes*, 83(2), 260–281. <https://doi.org/10.1006/obhd.2000.2909>
- Yaniv, I., & Milyavsky, M. (2007). Using advice from multiple sources to revise and improve judgments. *Organizational Behavior and Human Decision Processes*, 103(1), 104–120. <https://doi.org/10.1016/j.obhdp.2006.05.006>
- Zaleskiewicz, T., & Gasiorowska, A. (2023). Taking financial advice: Going beyond making good decisions. In Altman, M. (Ed.), *Handbook of research methods in behavioural economics*. Edward Elgar Publishing, pp. 226–241.

AUTHOR BIOGRAPHIES

Tobias R. Rebholz is a postdoctoral researcher at the University of Tübingen, Germany, where he also received his PhD in the transregional research training group Statistical Modeling in Psychology (SMiP). His research interests include quantitative methods, social cognition, and judgment and decision-making, especially in the context of information sampling and utilization.

Marco Biella is currently working at the Faculty of Business and Economics at the University of Basel, Switzerland, after spending three years in Tübingen and Heidelberg, Germany. He investigates social interaction and decision-making. Under the lenses provided by the Sampling Approaches, his current research

focuses on how the information ecology shapes socially embedded decision-making.

Mandy Hütter is a full professor of Social Cognition and Decision Sciences at the University of Tübingen, Germany. She received her PhD from the University of Freiburg, Germany. Her research interests include information sampling and utilization in judgment and decision-making as well as attitude acquisition and change.

How to cite this article: Rebholz, T. R., Biella, M., & Hütter, M. (2024). Mixed-effects regression weights for advice taking and related phenomena of information sampling and utilization. *Journal of Behavioral Decision Making*, 37(2), e2369. <https://doi.org/10.1002/bdm.2369>

APPENDIX A: ADDITIONAL RESULTS FOR THE REANALYSIS OF LOGG ET AL. (2019)

	Estimate	95% CI	SE	t	df	p
β_A	0.4507***	[0.3757, 0.5258]	0.0383	11.78	4.78	<.001
$\beta_{A \times Fam}$	-0.0608***	[-0.0935, -0.0281]	0.0167	-3.65	345.58	<.001
$\beta_{A \times Exp}$	-0.2536***	[-0.3391, -0.1680]	0.0436	-5.82	351.21	<.001
$\beta_{A \times Src}$	0.1985***	[0.1278, 0.2692]	0.0361	5.51	337.43	<.001
$\beta_{A \times Exp \times Src}$	-0.2330**	[-0.4035, -0.0625]	0.0869	-2.68	348.50	.007
$\tau_{A,S}$	0.2763	[0.2528, 0.3012]				
$\tau_{A,T}$	0.0667	[0.0088, 0.1186]				
σ	7.2278	[6.9149, 7.5431]				
ICC	0.67					
$R^2_{marg.}$	0.55					
$R^2_{cond.}$	0.86					

Note: Two-sided * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$. Wald 95% CI for fixed and bootstrap 95% CI (with 1000 iterations) for random effects are shown. Sample sizes of $N = 343$ participants S and $M = 4$ stimulus items T resulted in a total number of 1372 observations.

	Estimate	95% CI	SE	t	df	p
β_A	0.1972***	[0.0950, 0.2993]	0.0521	3.79	16.18	<.001
$\beta_{A \times Fam}$	-0.0608***	[-0.0935, -0.0281]	0.0167	-3.65	345.58	<.001
$\beta_{A \times Exp}$	0.2536***	[0.1680, 0.3391]	0.0436	5.82	351.21	<.001
$\beta_{A \times Src}$	-0.0345	[-0.1892, 0.1202]	0.0789	-0.44	351.09	.662
$\beta_{A \times Exp \times Src}$	0.2330**	[0.0625, 0.4035]	0.0869	2.68	348.50	.007
$\tau_{A,S}$	0.2763	[0.2528, 0.3012]				
$\tau_{A,T}$	0.0667	[0.0088, 0.1186]				
σ	7.2278	[6.9149, 7.5431]				
ICC	0.67					
$R^2_{marg.}$	0.55					
$R^2_{cond.}$	0.86					

Note: Two-sided * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$. Wald 95% CI for fixed and bootstrap 95% CI (with 1000 iterations) for random effects are shown. Sample sizes of $N = 343$ participants S and $M = 4$ stimulus items T resulted in a total number of 1372 observations.

TABLE A1 Multilevel model according to Equations (5) and (6) for dummy-coded expertise in Experiment 4 of Logg et al. (2019).

TABLE A2 Multilevel model according to Equations (5) and (6) for reverse dummy-coded expertise in Experiment 4 of Logg et al. (2019).