Advanced Statistical Modeling of Ecological Constraints in Information Sampling and Utilization

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Tobias Robert Rebholz

aus Sigmaringen

Tübingen

2023

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	26.04.2023
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatterin:	Prof. Dr. Mandy Hütter
2. Berichterstatter:	Prof. Dr. Arndt Bröder
3. Berichterstatter:	PD Dr. Thomas Schultze-Gerlach

For my family

Contents

Summary

Information sampling and utilization are ubiquitous in daily life. Accordingly, both processes are affected by a variety of environmental factors. This dissertation is primarily concerned with ecological constraints that are implemented by the social context. In particular, people often consider the opinions and beliefs of others in their judgments and decisions.

Research on advice taking and related cognitive phenomena such as anchoring, hindsight, or attitude change traditionally relies on ratio-of-differences-type formulas to determine informational influences. In this dissertation, two alternative modeling frameworks are presented for specifying how strongly peoples' judgments are influenced by external information. In contrast to the traditional approach, the proposed methods are consistent with the dependency of endogenous judgments (i.e., potentially updated beliefs) on exogenous sources of information (e.g., advice, base rates, anchors). Corresponding statistical modeling has the advantage of avoiding critical measurement problems of the traditional approach and is shown to enable new substantive research. A Bayesian account provides the opportunity to test for adaptive strategy selection in sequential advice seeking by explicitly distinguishing Thurstonian and Brunswikian sampling. Moreover, mixed-effects regression of final judgment on any exogenous sources of information resolves further paradigmatic peculiarities of the classic experimental procedure. For instance, the traditional modeling approach requires independent initial judgments as well as observable intermediate judgments, or presupposes equal weighting of sequentially sampled advice, respectively.

Empirical investigations of advice expectation and sequential advice seeking highlight two particularly relevant and novel ecological constraints of social information acquisition. First, traditional modeling reveals a positive effect of advice expectation on weighting for a trial-by-trial contrast of low versus high expectation to receive advice. The proposed regression-based approach validates this finding by means of process-consistent statistical modeling. Second, final judgment correspondence is taken as evidence for Bayesian advice taking in sampling extensions of the classic experimental paradigm. Indeed, empirical mixed-effects regression weights of sequentially sampled advice are moderately to strongly correlated with Bayesian weights constituting the normative benchmark. Moreover, both more advanced modeling approaches provide first evidence for nonlinear serial weighting of sequentially sampled advice.

In summary, the process-consistent statistical modeling proposed in this dissertation facilitates and extends substantive research on important ecological constraints of (social) information acquisition, such as the expectation of external influences and the sequential sampling of information.

Contributions

This dissertation is based on three manuscripts, one of which has been published and two of which have been submitted for publication. Copies of the three manuscripts can be found in Appendix D. The following list of manuscripts is accompanied by statements on the individual contributions of all contributing authors and details about the publication status.

Manuscript I

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

Author contributions:

Author	Scientific Ideas	Data Generation	Analysis & Interpretation	Paper Writing
Rebholz, T. R.	80%	95%	75%	80%
Hütter, M.	20%	5%	25%	20%

Status in publication process: Published

Manuscript II

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

Author contributions:

Author	Scientific Ideas	Data Generation	Analysis & Interpretation	Paper Writing
Rebholz, T. R.	40%	-	70%	70%
Hütter, M.	30%	-	10%	20%
Voss, A.	30%	-	20%	10%

Status in publication process: Submitted for publication

Manuscript III

Rebholz, T. R., Biella, M., & Hütter, M. (2023). *Mixed-effects regression weights (of advice): Process-consistent modeling of information sampling and utilization.* PsyArXiv. https://doi.org/10.31234/osf.io/x36az

Author contributions:

Author	Scientific Ideas	Data Generation	Analysis & Interpretation	Paper Writing
Rebholz, T. R.	90%	-	90%	85%
Biella, M.	5%	-	5%	5%
Hütter, M.	5%	-	5%	10%

Status in publication process: Submitted for publication

1 Introduction

In daily life, people base minor and major judgments and decisions on information that is acquired from the environment. For instance, when deciding whether to go to work by bus or bicycle on a cloudy morning, at least some of the following factors probably come into play: the weather forecast and bus schedule, both in the morning and for the way back home, personal perceptions of humidity, the technical condition of the bicycle, and so on. Additionally, one might also observe the behavior of other people such as neighbors or ask family members for their opinions. For instance, although one would normally risk getting wet on the bicycle under these circumstances, one might eventually decide to follow the partner's advice and take the bus. Regardless of what the final decision is, it reflects the contributions of environmental information on peoples' judgments.

Historically, multiple lines of research have investigated information acquisition (i.e., sampling and utilization) from partly overlapping and partly diverging perspectives (see Slovic & Lichtenstein, 1971, for a review). The Bayesian literature on belief updating covered normative aspects of evidence accumulation (e.g., W. Edwards, 1962; W. Edwards et al., 1963). For instance, participants' probabilistic inferences were compared to the normative solutions of vignettes according to Bayes' rule (Kahneman & Tversky, 1972). Similarly, research on forecast combinations (e.g., Lim & O'Connor, 1995) and "wisdom of crowds" (e.g., Galton, 1907; Surowiecki, 2005) focused on the accuracy of aggregated information. Other investigations of information integration tested mathematical models against empirical data to gain insights into how people actually utilize information (e.g., Hoffman, 1960). In his Information Integration Theory, Anderson (1971, 1981) focused on different mathematical rules or algebraic operations, respectively, as applied by different persons to integrate information from multiple sources. Instead of focusing only on the judge, the so-called "lens model" integrated environmental factors (i.e., context) of the judgment situation (Brunswik, 1952, 1956). All these aspects are also relevant for advice taking, which is primarily concerned with the social aspects of information sampling and integration (see Bonaccio & Dalal, 2006; Kämmer et al., 2023; Rader et al., 2017, for reviews).

Advice taking is ubiquitous in judgment situations that involve social contexts such as in the introductory example about deciding to go by bus or bike to work. Essentially, by seeking advice, the informational basis about a specific matter is extended by the opinions of others. In the traditional experimental paradigm, the Judge-Advisor

System (JAS), participants receive the judgments of other participants as advice to potentially revise their own initial judgments (Sniezek & Buckley, 1995). In extensions of this classic version of the JAS, different aspects and boundary conditions of advice taking as a form of social information acquisition were investigated. For instance, people are generally too uncritical about the sources from which they receive information. This "metacognitive myopia" was shown to render people overly susceptible to misleading advice (Fiedler et al., 2019). Moreover, people seek more advice when they are less confident about their own judgment (Gibbons et al., 2003). In the introductory example, the reason for asking family members for advice could be a lapse in memory about the weather forecast from yesterday's newscast. From a Thurstonian perspective, the (perceived) insufficiency of one's own informational basis (i.e., the internal sample) for making an independent final judgment or decision is reflected in low initial confidence (Koriat, 2012a, 2012b; Koriat et al., 1980; see also Hütter & Fiedler, 2019). According to the informational asymmetry account of advice taking, people are privy to their own thoughts (Yaniv, 2004a, 2004b). Hence, sampling extensions of the original paradigm posit that people can overcome this informational asymmetry by more extensive external sampling (Ache, 2017; Hütter & Ache, 2016). Nevertheless, people are also often reluctant to seek advice, for instance, because they fear appearing incompetent (Brooks et al., 2015). In general, cognitive idiosyncrasies such as confidence, knowledge, memory, and metacognitive processes additionally reflect and/or are affected by social factors of the information ecology (Kämmer et al., 2023; Rader et al., 2017).

1.1 Ecological Constraints of Social Information Acquisition

For Bayesian updating, "conservatism" refers to the weighting of evidence that does not reflect its true diagnosticity (Slovic & Lichtenstein, 1971). In contrast, research on social information acquisition adopts a cognitive perspective to contextualize corresponding behavior. Underweighting of advice relative to a normative benchmark is called "egocentric discounting" and describes peoples' general tendency to prefer own judgments over others' judgments (Harvey & Fischer, 1997; Yaniv & Kleinberger, 2000). Initially, differences in perceived expertise—that is, cognitive processes—were offered as an explanation for this phenomenon (but see Soll & Larrick, 2009; and Section 1.2). In general, however, peoples' advice taking behavior is not only influenced by cognitive idiosyncrasies but also the information ecology (i.e., "extra-psychic" factors; Fiedler & Kutzner, 2015). For instance, objective differences in expertise constrain the relative judgment quality (e.g., Larrick & Feiler, 2015), and often diverge from perceived differences in expertise (Rader et al., 2017; see also Bednarik & Schultze, 2015). Crucially, advisory judgments of relatively higher quality were found to be weighted more strongly (Yaniv & Kleinberger, 2000; but see Schultze et al., 2017, for the independence of anchoring-based advice integration from advice quality). Throughout this dissertation, environmental influences on the sampling and utilization of external information are thus referred to as “ecological constraints” of information acquisition.

In addition to differences in objective expertise, many other ecological constraints are implemented by the social context in which many judgments and decisions are made. Additional examples include, but are not limited to, the congruency of desired and actual weighting (with or without communication; Ache et al., 2020), and, in a similar vein, to anchor advisors by including one’s own judgment in the request for advice (Reif et al., 2022). Digital transformation and corresponding information technology (e.g., social and professional network platforms) foster the convenience of requesting and/or accessing advice in a variety of contexts. Methodologically, however, studying advice taking under laboratory conditions restricts the social nature of judge-advisor interactions (Kämmer et al., 2023; Rader et al., 2017; but see Minson et al., 2011; Van Swol, 2011, for exceptions). Accordingly, this dissertation focuses on two rather novel ecological constraints of social information acquisition that are particularly relevant for JASs in digitalized societies. First, digital transformation reduces the effort and cost of acquiring advice, which should imply high expectations of external influences on one’s judgments and decisions. Second, spatially distant and largely anonymous interactions foster the abundance of potential advisors to be sampled via digital information technology (Schulz & Roessler, 2012).

Under real-world circumstances, the interactivity of most advice taking situations offers plenty of opportunities for acquiring advice (Ache, 2017). However, it is not as easy to recruit advisors in some situations as in others (see also Gibbons et al., 2003). For instance, there may be a lack of access to the right (social) network, or the judgment is about a particularly sensitive matter such as political orientation (Schulz & Roessler, 2012). Cognitively, provisional initial judgments and ongoing mental tasks, as featured by expecting external influences on one’s judgment in the future, should trigger relatively more assimilative processing of expected than unexpected advice (cf. Alexopoulos et al., 2012). Indeed, in the experiments reported in first manuscript, we found that expected advice is taken significantly more than unexpected advice for a trial-by-trial contrast of low versus high expectations to receive advice (Rebholz & Hütter, 2022). That is, expecting external influences can affect peoples’ processing of a single piece of advice when it is eventually received at a later point in time.

In the experiments providing evidence for a positive effect of advice expectation on weighting, we manipulated high versus low expectations to receive advice via instructions (Rebholz & Hütter, 2022). Specifically, advice provision was implemented as a probabilistic within-participants factor that informed participants about either high chances (i.e., 80% probability) or low chances (i.e., 20% probability) to receive advice on the following trial. In contrast, expectation effects on weighting were suppressed by sequencing multiple different judgments into initial versus final estimation blocks, which supports the notion of ongoing mental tasks as reason for expectation effects on weighting (cf. Alexopoulos et al., 2012). We also provided evidence for the independence of this null effect from the extremeness of expectations, that is, inducing high expectations by communicating an 80% chance of receiving advice as opposed to a guaranteed provision of advice. Hence, the experimental procedure highlights two important ecological constraints of information integration in social contexts: volatile chances to receive advice or being able to recruit advisors and the sequential conclusion of multiple judgment tasks. Nevertheless, the ecological validity of our dichotomous probabilistic manipulations is restricted by peoples' construal of probability as a non-linear dimension of "hypothetical" distance (Kahneman & Tversky, 1979; Trope & Liberman, 2010). Moreover, receiving a piece of advice in the first place might change the expectations about *additional* external influences on ongoing mental tasks. In general, believing it is easy (difficult) to recruit advisors, which might change over time, should be reflected in higher (lower) expectations to receive advice. That is, advice expectation is inextricably linked to the sampling ecology of a certain judgment situation or task.

Traditionally, most JAS research prevents active information seeking (see Bailey et al., 2022, Table 1). In the original paradigm, only one piece of advice was provided by default (Sniezek & Buckley, 1995). In advice taking research related to wisdom of crowds, by contrast, multiple pieces of advice were available but provided simultaneously (e.g., Adjodah et al., 2021; Budescu & Yu, 2006; Molleman et al., 2020; Yaniv & Milyavsky, 2007). In other applications such as those reported in the first manuscript, participants received at most one piece of advice per trial, if any (Rebholz & Hütter, 2022; see also Schrah et al., 2006). All of these applications have in common that advice was passively presented without an opportunity for participants to actively control the sampling of advice.[1] In contrast, sampling approaches to advice taking allow multiple pieces of advice to be sampled sequentially (e.g., Ache, 2017; Hütter & Ache, 2016). People often consider more than one piece of external information for making final judgments or decisions. For instance, the partner's recommendation to ride the bicycle in the introductory example may be based on admittedly only remembering the weather forecast for the morning, which does not help much for the way back home and is thus insufficient. Similarly, successfully acquired advice that contradicts one's opinions is ascribed a lower quality and thus deemed rather unsatisfying (Minson et al., 2011; Pronin et al., 2004). So why not put extra effort into seeking additional advice in both cases? In other words, consulting additional advisors to test whether the first piece of advice merely constitutes an outlier instead of representing (potentially unsatisfying) consensus (see also Rader et al., 2015; Schrah et al., 2006).

One reason against additional sampling is that seeking advice from multiple advisors can also have negative interpersonal consequences such as reduced relational

[1] One exception is the study of Schrah et al. (2006), in which participants could freely determine the exact timing of receiving a single piece of advice during information search. Technically, they also had the option to forego advice completely. However, all participants opted for being provided with advice on all trials.

closeness or less willingness to provide advice in future interactions (Blunden et al., 2019; Feng & MacGeorge, 2006; Feng & Magen, 2016; see also Biella & Hütter, 2023, for a related sampling perspective on the formation of trust impressions). Nevertheless, if allowed to freely sample as many pieces of advice as desired (up to a certain threshold), which is more ecological in digitalized societies, most participants indeed consider more than just one piece of external input (Hütter & Ache, 2016). Moreover, adding sampling costs yielded significantly smaller average advice sample sizes as compared to free sampling (Ache, 2017). Specifically, the experimental procedure implemented additional waiting time before a new piece of advice could be sampled in the costly sampling condition. Due to sunk cost fallacy (Arkes & Blumer, 1985), costly advice is weighted more strongly than free advice (Ache, 2017; Gino, 2008; see also Sniezek et al., 2004). More generally, the trade-off between benefits (e.g., additional viewpoints and shared responsibility) and costs (e.g., the time needed for sampling or negative interpersonal consequences) of receiving additional external information influences advice seeking (Bonaccio & Dalal, 2006; Kämmer et al., 2023). Accordingly, focusing on single advice taking situations in large parts of the literature constitutes another important ecological constraint for generating insights with respect to real-world behavior. Instead, allowing participants to actively sample advice boosts the ecological validity of corresponding experimental research.

The focus is on advice taking in this dissertation, but many aspects related to ecological constraints of information acquisition generalize to non-social, external sources of information. For instance, expertise of another social agent can be construed as the competence of any source of information (e.g., the prediction accuracy of algorithmic market or weather forecasts). Indeed, many aspects of the genuinely social process of taking judgments from others into consideration also applies to interactions of humans with artificial intelligence that acts as advisor or recommender system (e.g., Hütter & Fiedler, 2019; Logg et al., 2019). As will be argued in the following section, statistically appropriate modeling of corresponding judgment processes is essential for the investigation of any kind of information ecology and thus constitutes the central objective of this dissertation on a methodological level.

1.2 Objective

Initially, differences in perceived expertise were provided as substantive explanation for preferring own judgments over judgments of other people (Harvey & Fischer, 1997; Yaniv & Kleinberger, 2000). However, later it was shown that this finding could mainly be attributed to the aggregate analysis schemes applied in this research (Soll & Larrick, 2009). Advice weighting typically features a W-shaped distribution with three modes of different height representing three distinct advice taking strategies: In most cases, people do not shift away from their own initial judgment (i.e., no advice taking), followed by choosing advisory judgments (i.e., full advice taking) or simple (i.e., unweighted) averaging of the two. Consequently, the overall mean of this tri-modal distribution indicates less than equal weighting of both judgments. In this spirit, the aim of this dissertation is to extend and improve substantive research on ecological constraints in advice taking and related fields by instantiating an integrative modeling framework that explicitly takes idiosyncrasies in peoples' strategy selections and cognitive processes for (externally influenced) judgment formation into consideration.

The established data analysis approach as applied in the first manuscript is to rely on descriptive weighting indices for quantifying informational influence (Rebholz & Hütter, 2022). As this dissertation is the product of research conducted in the research training group "Statistical Modeling in Psychology" (SMiP), it integrates research on statistical techniques, model families, and application fields in line with the agenda of SMiP. The focus is on advanced statistical techniques from two different model families that facilitate and extend our understanding of human behavior in application fields related to information sampling and utilization. In the second manuscript, two substantively distinct Bayesian updating strategies are formulated to enable the investigation of strategy selection in sequential advice seeking (Rebholz, Hütter, & Voss, 2023). We found that Bayesian-type compromising is the most frequently applied strategy as compared to choosing internal or external judgments that were sequentially sampled. In the third manuscript, a frequentist model is proposed that relies on a similar conceptual understanding of evidence-based information integration in social contexts (Rebholz, Biella, & Hütter, 2023). Specifically, mixed-effects regression of final judgment on exogenous sources of information provided initial evidence for temporally invariable informational influence in sequential collaboration chains (Mayer & Heck, 2022) as well as for recency effects in Experiment 2 of Hütter and Ache (2016)—at least when accounting for nonlinear effects of distance on the weighting of sequentially sampled advice (cf. Moussaïd et al., 2013; Schultze et al., 2015). In contrast to the traditional approach, the proposed methods are consistent with the judgment formation process in advice taking and related paradigms such as anchoring, hindsight, or attitude change.

The following chapter contains a brief recap of the traditional and more advanced modeling frameworks as applied and/or developed in each individual manuscript. In additional empirical applications, cross-comparisons of different modeling approaches will provide a more integrative perspective on the research included in this dissertation. In Chapter 3, the positive effect of advice expectation on weighting as indicated by traditional modeling will be verified by process-consistent regression analyses. The normativity of sequential advice seeking will be reassessed in the second empirical application as presented in Chapter 4. There are moderate to strong correlations between empirical mixed-effects regression and Bayesian weights of sequentially sampled advice, depending on model specification and the temporal perspective of the updating process. Finally, in an integrative General Discussion, I will outline consequences of the insights derived in this dissertation as well as directions for future substantive research on expectation and sampling as important ecological constraints of information acquisition in social contexts.

2 Statistical Modeling of Information Sampling and Utilization

Historically, advice taking and related research (e.g., on anchoring effects, hindsight bias, or attitude change) used deterministic formulas to specify empirically how much information from external sources was integrated into own judgments (Bonaccio & Dalal, 2006; see also Turner & Schley, 2016). Depending on a certain reference point or initial status, respectively, Bayes' theorem provides a normative account of responsiveness to new evidence. Alternatively, individual amounts of integration can be estimated from a mixed-effects regression (MER) model that is consistent with the endogenous formation of externally influenced judgments. In a brief recap of the methodological and statistical concepts involved in each of those modeling procedures, I will tentatively describe their most substantial merits and limitations. A more detailed discussion in the light of new empirical evidence regarding expectations and sampling as ecological constraints of (social) information acquisition is postponed to the General Discussion.

2.1 Deterministic Weighting Indices

In traditional analyses of advice taking behavior, such as those conducted in the first manuscript (Rebholz & Hütter, 2022), the index of Harvey and Fischer (1997) is used to quantify informational influences. The weight of advice (WOA) index is specified as

$$\omega_{A,ijk} = \frac{E_{ijk} - E_{ij0}}{A_{ijk} - E_{ij0}}, \tag{2.1}$$

where E_{ij0} denotes the initial judgment of participants $i = 1, \ldots, N$ for stimulus items $j = 1, \ldots, M$ (e.g., the caloric content of food; Hütter & Ache, 2016; Schultze et al., 2015; Yaniv et al., 2009). Endogenous, potentially revised judgments after having received advice A_{ijk} are denoted as E_{ijk}. Accordingly, Equation 2.1 applies a deterministic ratio-of-differences (ROD) arithmetic to measure the degree of integrating other peoples' judgments into one's own judgment. Consequently, $\omega_{A,ijk} = 0$ indicates no advice taking (i.e., $E_{ijk} = E_{ij0}$) whereas $\omega_{A,ijk} = 1$ corresponds to complete integration of advice (i.e., $E_{ijk} = A_{ijk}$). Every other value of $\omega_{A,ijk}$ denotes corresponding weighted linear combinations of the two exogenous sources of information, that is, advice A and the initial judgment E_0.

Descriptive ROD-WOA requires a two-stage procedure for deriving inferential conclusions about information integration. To investigate ecological constraints such as the consequences of advice expectation as induced by specific advice probabilities on weighting, $\omega_{A,ijk}$ is taken as the dependent variable in statistical optimization such as multilevel modeling on stage two (Rebholz & Hütter, 2022). However, (mixed-effects) regression of ROD-WOA on explanatory variables such as experimental conditions or control variables was only recently established in empirical practice (e.g., Ache et al., 2020; Minson & Mueller, 2012; Schultze et al., 2015). In contrast, Information Integration Theory mostly relied upon factorial designs and thus analysis of variance for statistical testing (Anderson, 1971, 1981; see also Slovic & Lichtenstein, 1971). Indeed, I suspect one of the main reasons for the traditional popularity of ROD-WOA is that it enables the use of analysis of variance as an analytical tool for advice weighting investigations. This comes at the cost of measurement problems like conceptual or outcome ambiguity of difference scores (J. R. Edwards, 1994, 1995).[2]

Additionally, building ratios of difference scores has critical limiting properties. Advice that is very close to initial judgments of participants makes the index in Equation 2.1 converge to infinity. Consequently, corresponding weighting is usually classified as an outlier and excluded from the analysis. In empirical practice, computational issues are reduced by alternatively taking absolute differences or truncating outliers to the respective boundaries of the $[0, 1]$ interval. However, these two approaches can yield undefined or ambiguous values for confirmatory and largely confirmatory advice (i.e., dividing judgmental shift by advice distance scores close to zero in Equation 2.1) as well as for shifting away from advice (Rebholz, Biella, & Hütter, 2023; see also Rader et al., 2015). More importantly, calculation of judgmental shift and advice distance in the traditional index requires prior and posterior judgments to be both observable and observed. In research on sequential collaboration, for instance, prior judgments are typically not observed (Mayer & Heck, 2022). The observation of intermediate posterior judgments is particularly relevant for integrating multiple, sequentially sampled pieces of external information. Put differently, Equation 2.1 is defined only for $k = K = 1$, that is, single advice taking (but see Ache, 2017; Hütter & Ache, 2016, for mean advice taking). Alternatively, Bayes' theorem can be used to infer normative latent intermediate judgments, as will be shown in the following recap of the modeling approach developed in the second manuscript (Rebholz, Hütter, & Voss, 2023).

2.2 Normative Bayesian Updating

To model sequential advice seeking in the second manuscript, we applied the following perspective of Morris (1974) on "expert use:"

[2]Conceptual ambiguity is due to the implicit equal weighting concealing the relative variance contributions of individual difference score components. Outcome ambiguity describes the confounding of difference score components by reducing individual effects of separate independent variables to a single coefficient.

> Conceptually, consulting an expert is like performing an experiment where the observed data is a function (probability distribution) rather than a number. Just as the results of an experiment are a priori unknown to an experimenter, the quantity $[A_k]$ is uncertain to the decision maker prior to receiving advice. (p. 1235)

Hence, we explicitly distinguished between external ("Brunswikian") and internal ("Thurstonian") sources of information (Juslin & Olsson, 1997). Advice takers' judgment formation regarding the unknown truth was modeled as a Bayesian updating process (Rebholz, Hütter, & Voss, 2023). The notion of Thurstonian sampling implies that initial judgments represent summary statistics of an internal sampling process (Sniezek & Buckley, 1995; Thurstone, 1927; see also Fiedler & Kutzner, 2015; Henriksson et al., 2010; Stewart et al., 2006). Therefore, we specified participant i's prior judgment E_{ij0} about the true value θ of stimulus item j as

$$\theta \sim N(E_{ij0}, C_{ij0}), \tag{2.2}$$

that is, as the center of normally distributed initial beliefs.[3] Uncertainty C_{ij0} results from a random or quasi-random internal sampling process (Fiedler & Juslin, 2006; Juslin & Olsson, 1997; but see Herzog & Hertwig, 2014; Rauhut & Lorenz, 2011; Soll & Klayman, 2004). In contrast, sampling advice corresponds to observing additional external evidence with likelihood

$$A_{ijk} \sim N\Big(\theta, \tau^{-2}\Big), \tag{2.3}$$

where the advice precision τ^{-2} is typically unknown to the judge. Advice that is centered at the true value of an item is justified for participants—albeit imperfectly—appreciating the wisdom of crowds (Larrick & Soll, 2006; Mannes, 2009). But what are reasonable assumptions for participants' beliefs about the validity or variability of advice, respectively? In Rebholz, Hütter, and Voss (2023), we implemented two different solutions to account for unknown advice precision τ^{-2}.

The hierarchical Bayesian account from the second manuscript builds on assuming exchangeability of the judgment and inverse confidence parameters. The corresponding prior follows a joint normal-inverse-χ^2 distribution (see Rebholz, Hütter, & Voss, 2023, Equation 12) rather than the distributions specified in Equations 2.2 and 2.3. Such a prior specification implies judgment updating that is mathematically equivalent to simple cumulative averaging of internal and external judgments (Gelman et al., 2013). Alternatively, the sequential Bayesian account implements first-stage updating of participants' intuitions about advice precision

$$\tau^2 \sim \text{Inv-}\chi^2\Big(L_{ij0}, T^2_{ij0}\Big), \tag{2.4}$$

[3] Conjugate normal distributions involve convenient analytical solutions and were appropriate in many empirical examples (e.g., Adjodah et al., 2021; Moussaïd et al., 2013; Soll & Klayman, 2004). Nevertheless, normal beliefs can be replaced by any other distributional assumptions to match different judgment tasks and imply corresponding updating specifications.

where L_{ij0} denotes the internal sample size and $T^2_{ij0} = C^{-1}_{ij0}$ captures similar expectations about external and internal sampling. This process separation effectively renders corresponding Bayesian updating truly sequential, that is, non-invariant with respect to the sequence in which advice is sampled. Therefore, I will mainly focus on the sequential Bayesian account here. More details about technical and substantive differences between the hierarchical and sequential Bayesian accounts can be found in the second manuscript (Rebholz, Hütter, & Voss, 2023).

Formally, the sequential Bayesian account implies belief updating in response to sequentially sampled advice based on the following rules:

$$L_{ijk} = L_{ij(k-1)} + 1, \tag{2.5}$$

$$\hat{T}^2_{ijk} = \frac{L_{ij(k-1)}}{L_{ijk}} \hat{T}^2_{ij(k-1)} + \frac{1}{L_{ijk}} (A_{ijk} - \hat{E}_{ij(k-1)})^2, \tag{2.6}$$

$$\hat{C}_{ijk} = \hat{C}_{ij(k-1)} + \hat{T}^{-2}_{ijk}, \tag{2.7}$$

$$\hat{E}_{ijk} = \frac{\hat{C}_{ij(k-1)}}{\hat{C}_{ijk}} \hat{E}_{ij(k-1)} + \frac{\hat{T}^{-2}_{ijk}}{\hat{C}_{ijk}} A_{ijk}, \tag{2.8}$$

where $\hat{T}^2_{ij(k-1)} = C^{-1}_{ij0}$, $\hat{C}_{ij(k-1)} = C_{ij0}$, and $\hat{E}_{ij(k-1)} = E_{ij0}$ for $k = 1$. That is, step-wise updating of previous beliefs based on new evidence takes the relative uncertainties associated with both sources of information into consideration. On most trials in Experiment 5 of Ache (2017), we indeed found a rather good correspondence of predicted final judgment, $\hat{E}_{ijK_{ij}}$, and actual final judgment, $E_{ijK_{ij}}$, where K_{ij} denotes the realized advice sample sizes (see also Chapter 4). More importantly, the updating processes as specified in Equations 2.6 to 2.8 formalizes normative laws from which reference points for actual behavior can be derived that are more comprehensive than only in terms of final belief formation. As will be shown in the next section, empirical weights of sequentially sampled advice can be estimated by the multilevel modeling approach as proposed in the third manuscript (Rebholz, Biella, & Hütter, 2023). Comparing those MER-weights of exogenous sources of information to the normative benchmark additionally enables investigations of strategy selection on the *weighting* level. In terms of a more fine-grained model comparison, close correspondence of individual weights of sequentially sampled advice is more informative about the goodness of fit of a particular strategy than accurate outcome predictions only.

2.3 Partial Mixed-Effects

Bayesian updating and the traditional ROD-based analysis of advice taking rely on a shared definition of judgment formation. Specifically, rearranging Equation 2.1 to account for endogenous posteriors specifies belief updating as a sum-to-one constrained weighted linear combination of priors and external evidence. Put differently, the Bayesian estimate of WOA according to Equations 2.7 and 2.8 is

$$\hat{\omega}^{*}_{A,ijk} = \frac{\hat{T}^{-2}_{ijk}}{\hat{C}_{ij(k-1)} + \hat{T}^{-2}_{ijk}}, \tag{2.9}$$

which corresponds to a measure of relative uncertainty. Hence, comparison of Bayesian and ROD-WOA for $k = 1$ provides insights with respect to the normativity of participants' single advice taking behavior. Alternatively, a mixed-effects regression model that is consistent with the endogenous process of final judgment formation as derived from Bayes' theorem can be specified as

$$E_{ijk} = (1 - \omega_{A_k,ij})E_{ij(k-1)} + \omega_{A_k,ij}A_{ijk} + \varepsilon_{ij}, \tag{2.10}$$

where the residuals of the coefficients can be disentangled from overall error $\varepsilon_{ij} \sim N(0, \sigma^2)$ by multilevel modeling (Bates et al., 2015; Brown et al., 2018; Raudenbush & Bryk, 2002). Corresponding MER-weights (of advice) are formally defined as

$$\omega_{p,ij} = \beta_p + \alpha^S_{p,i} + \alpha^T_{p,j}, \tag{2.11}$$

where $p = A_k$ (Rebholz, Biella, & Hütter, 2023). The conceptual specification of individual weighting hence boils down to participant- and item-wise random deviations $\alpha^q_p \sim N\left(0, \tau^2_{p,q}\right)$, $q \in \{S, T\}$, from mean weighting, β_p. In general, the variance terms of the crossed random effects, $\tau^2_{p,S}$, $\tau^2_{p,T}$, are mutually independent by assumption.

MER-weights can be estimated for various ecological constraints of information acquisition by corresponding extensions of the basic regression model (Rebholz, Biella, & Hütter, 2023). For instance, sum-to-one constraining in Equation 2.10 can be abandoned in favor of partial mixed-effects estimation. From Equation 2.11 for $p = E_0$, individual weights of initial judgment, $\omega_{E_0,ij}$, can be estimated in addition to individual weights of advice, $\omega_{A_k,ij}$. This unconstrained model enables the quantification of informational influences in situations without formulation of independent initial judgments (e.g., in sequential collaboration; Mayer & Heck, 2022). In addition, more than just two sources of information can be included. Accordingly, the regression model can also be extended by fixed or random order effects for deriving empirical weights of sequentially sampled advice (see Chapter 4). Although multiple judgments of different agents are generally not independent of each other in typical advice taking scenarios (e.g., Harvey et al., 2000; see also Hoffman, 1960), MER-weights are stable against multicollinearity by design (Baayen & Linke, 2020; Brown et al., 2018). In summary, multiple paradigmatic peculiarities of the JAS can be resolved by process-consistent statistical modeling. In the following two empirical applications, this is demonstrated by further elaborating on expectation and sampling as ecological constraints of (social) information acquisition.

3 Validation of Expectation Effects on Advice Taking

In the first manuscript, we investigated the influence of expectation to receive a single piece of advice in the classic JAS paradigm for a total of $N = 2019$ participants (Rebholz & Hütter, 2022). For instance, participants estimated the carbon footprint of selected products and received a single piece of advice based on estimates from a pretest of the material. In Experiments 3 and 4, a trial-by-trial contrast of low (i.e., $P(\text{advice}) = 0.20$) versus high (i.e., $P(\text{advice}) = 0.80$) probability to receive advice was implemented. According to the original evidence, unexpected advice was significantly less taken than expected advice, but only in the two experiments that use a within-participants manipulation of expectation (see also Figure 3.1, top panel). However, this evidence was derived from ROD-WOAs, which have some undesirable properties as discussed in Section 2.1. Resolving the conceptual inconsistency of the ROD formula with respect to the underlying judgment process has the potential to avoid corresponding measurement problems (Rebholz, Biella, & Hütter, 2023; see also Footnote 2). Therefore, MER-weights of (un)expected advice are used here to validate the expectation effect on advice taking by means of process-consistent statistical modeling (see Section 2.3). A significance level of 5% is used for statistical testing throughout. Reproducible analysis scripts for both empirical applications are publicly available online (https://osf.io/pva5b).

3.1 Method

In the original analyses, we applied second-stage multilevel modeling for significance testing of ROD-WOA. That is, fitting regression models on Harvey and Fischer's (1997) weighting index with random intercepts of participants and stimulus items, and fixed effects of expectation condition. Instead, treatment effects can be incorporated directly on the weighting level of the process-consistent modeling framework as proposed in the third manuscript (Rebholz, Biella, & Hütter, 2023). Expectation condition was included as fixed treatment effect on weighting by specifying MER-WOA as

$$\omega_{A,ij} = \beta_A + \alpha^S_{A,i} + \alpha^T_{A,j} + \beta_{A\times C} C_{ij}, \tag{3.1}$$

where advice expectation was contrast-coded as

$$C_{ij} = \begin{cases} \text{-0.50}, & \text{for } P(\text{advice}) = 0.20 \\ 0.50, & \text{for } P(\text{advice}) = 0.80. \end{cases} \tag{3.2}$$

Consequently, $\beta_{A \times C}$ captures the fixed effect of commonly high expectations to receive advice on weighting in traditional JAS-type studies. In between-participants designs, that is, Experiments 1, 2, and 5 of Rebholz and Hütter (2022), expectation condition was implemented as $C_{ij} = C_{ij'} \ \forall j, j' = 1, \ldots, M$. Moreover, judgments were log-transformed to account for positive skew before being included in the judgment level model as specified in Equation 2.10 for $k = 1$. Expectation condition could alternatively be included as third crossed random effect (see Rebholz, Biella, & Hütter, 2023, for technical details). However, the practical recommendation for a minimum of five factor levels per clustering instance (Bolker, 2015; see also Oberpriller et al., 2022) is not met for advice expectation implemented as a binary treatment condition. Moreover, significance testing of group differences as fixed effect more closely resembles the original analyses.

3.2 Results

Across all experiments, average advice taking according to MER-WOA (see β_A in Table 3.1) was in a similar range as measured by ROD-WOA (multiplied by 100 for the measurement in percent; Rebholz & Hütter, 2022, Table 1). However, there is substantially less disregard of advice (i.e., $\hat{\omega}_{A,ij} = 0$) as measured by random deviations from mean advice weighting in MER-WOA (see Figure 3.1, top vs. bottom panel). In general, the characteristic W-shaped distribution of ROD-WOA with three modes at no weighting, equal weights averaging, and complete adoption (Soll & Larrick, 2009) is not reproduced for MER-WOA. Substantively, such a distributional pattern is rather unlikely to be observed for the latter as most participants do not apply one specific strategy across all trials of an experiment (cf. Rebholz, Hütter, & Voss, 2023, for sequential advice seeking). Instead, using information from the whole sample and determining (conditional modes of) random deviations thereof represents a more holistic approach (Baayen et al., 2008; Bates et al., 2015; Brown et al., 2018; Raudenbush & Bryk, 2002). Similarly, narrower distributions are a result of the shrinkage property of MER-WOA (see Rebholz, Biella, & Hütter, 2023, for a discussion of the implications of shrinkage in the context of advice weighting). Nevertheless, the two weighting measures are moderately to strongly correlated across all five experiments (Figure 3.2).

Replicating the original results, the fixed treatment effects of contrast-coded expectation on MER-WOA, $\hat{\beta}_{A \times C}$, were significantly positive only for within-participants manipulations of expectation in Experiments 3 and 4 (Table 3.1). According to second-stage multilevel modeling of ROD-WOA, the original expectation effect was slightly smaller in Experiment 4 than in Experiment 3 ($\hat{\beta}_1 = 2.29$ vs. $\hat{\beta}_1 = 4.62$; Rebholz &

Table 3.1: Fixed Effects of Multilevel Models of Final Judgment According to Equation 2.10 (for $k = 1$) With Fixed Treatment Effects of Contrast-Coded Advice Expectation Condition (C) as Specified in Equation 3.1 for Experiments 1 to 5 of Rebholz and Hütter (2022)

	β_A		$\beta_{A \times C}$	
	Estimate	*SE*	*Estimate*	*SE*
Experiment 1	0.3550 ***	0.0179	-0.0471	0.0331
Experiment 2	0.3338 ***	0.0230	0.0173	0.0257
Experiment 3	0.3334 ***	0.0275	0.0374 *	0.0158
Experiment 4	0.2346 ***	0.0138	0.0273 **	0.0084
Experiment 5	0.5908 ***	0.0289	-0.0193	0.0210

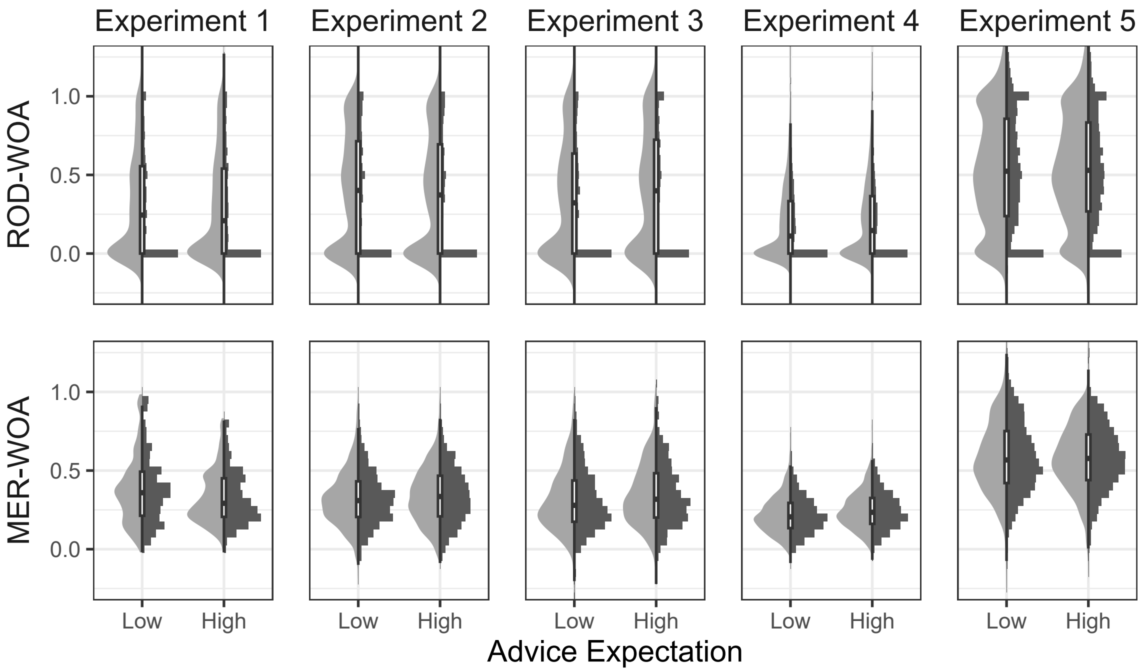
Note. $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$. The full models can be found in Appendix A, Tables A.1 to A.5.

Hütter, 2022, Table 1). For Experiment 4, the average reduction in weighting of unexpected advice was slightly higher than originally reported. In contrast, for Experiment 3 the effect of expectation on MER-WOA was almost one percentage point smaller than according to the original evidence. As a result, the absolute difference between significant expectation effects in within-participants designs was less pronounced with about one percentage point according to MER-WOA as compared to more than two percentage points according to ROD-WOA. Consequently, differences in task difficulty and knowledge requirements were probably not as large as originally suggested for product carbon footprint and quantity estimation as judgment tasks in Experiments 3 and 4, respectively.

3.3 Discussion

The hypothesis of central interest in the first manuscript concerns the positive effect of advice expectation on weighting. According to the original evidence, expected advice is weighted significantly more strongly than unexpected advice only for a trial-by-trial contrast of low versus high expectation (Rebholz & Hütter, 2022). However, if expectation does not change across trials, there is no evidence for differences in weighting of expected and unexpected advice. Similar statistical conclusions are derived for expectation effects on MER-WOA. Hence, process-consistent statistical modeling verifies the implications of high advice expectation on weighting, specifically, ongoing mental tasks featuring assimilative processing (cf. Alexopoulos et al., 2012). However, the fixed treatment effects on MER-WOA in Experiments 3 and 4 are slightly smaller than on ROD-WOA for which outliers were excluded as preregistered. Potentially ambiguous outlier criteria (see Section 2.1) are necessary to stabilize and sometimes even enable (e.g., ROD-WOA converges to infinity for extremely close advice) statistical optimization with ROD-type weighting indices as dependent variables (Rebholz, Biella,

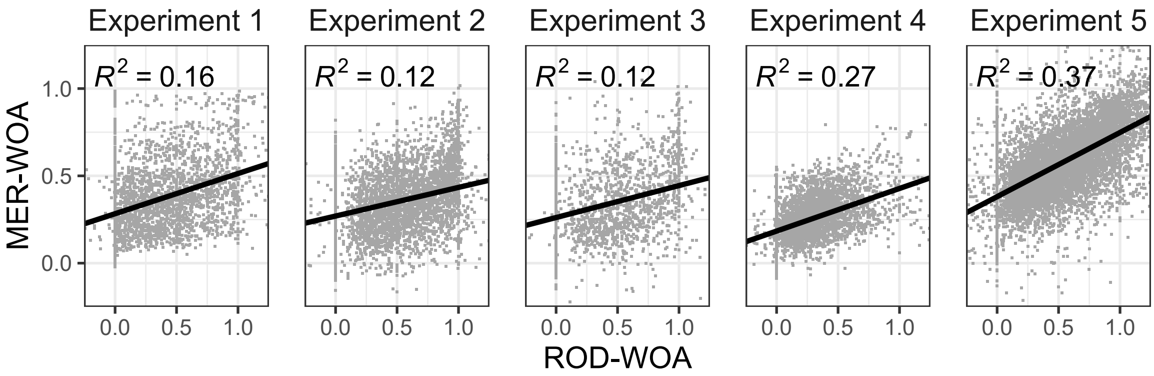
Figure 3.1: Distributions of Ratio-of-Differences (ROD; Top) and Mixed-Effects Regression (MER; Bottom) Weight of Advice (WOA) With Fixed Treatment Effects of Contrast-Coded Advice Expectation Condition in all Trials of Experiments 1 to 5 of Rebholz and Hütter (2022)



& Hütter, 2023). In contrast, the proposed regression-based approach does not require outlier criteria for either weighting or (finite) judgment. Thus, MER-WOA captures a broader range of "deliberate behavior" (Soll et al., 2022), which might exhibit reduced or no expectation effects. For instance, anchoring- or quality-related processes that lead participants to push away from advice (Rader et al., 2015) should not be affected by a more or less assimilative mindset. In summary, inducing high expectations about the opportunity to revise their initial judgment in the light of advice by informing participants in advance about the study procedure (e.g., Fiedler et al., 2019; Sniezek & Buckley, 1995; Soll & Larrick, 2009) indeed constitutes an important ecological constraint in JAS-type experiments.

In all but the last experiment, also weighting as measured by MER-WOA indicated egocentric discounting in both expectation conditions (Figure 3.1). Statistical testing of expectation effects in specific conditions can be conducted in multilevel models with dummy-coded treatment conditions. For (reverse) dummy-coded C_{ij}, β_A from Equation 3.1 measures average advice weighting in the low (high) expectation condition. The corresponding 95% CI enables significance testing for specific values other than zero. For instance, 0.50 represents the equal weighting or egocentric discounting threshold for single advice taking. Indeed, there is significant evidence against egocentric discounting in both expectation conditions of Experiment 5 (Appendix A, Tables

Figure 3.2: Linear Correspondences Between Mixed-Effects Regression (MER) and Ratio-of-Differences (ROD) Weight of Advice (WOA) for Experiments 1 to 5 of Rebholz and Hütter (2022)



Note. Plotting is truncted for WOA $\notin [-0.25, 1.25]$. The solid lines indicate the linear correspondences (incl. R^2) between MER- and ROD-WOA across all trials.

A.6 & A.7).[4] In the first four experiments, we have used real items (Experiments 1 to 3: pictures of products for carbon footprint estimation; Experiment 4: pictures of piles of objects for quantity estimation) to foster the ecological value/validity of the judgment task. In contrast, computer-generated material (i.e., images of sets of randomly colored squares) was used for quantity estimation in Experiment 5. Thus, participants' judgment and advice taking behavior might have been different than in the first four experiments for ecological reasons such as being less experienced or trained to estimate the number of colored squares (Larrick & Feiler, 2015).

Conservative inferential judgments in Bayesian belief updating are conceptually similar to egocentric discounting of advice (see Section 1.1). Specifically, conservatism describes the underweighting of external evidence relative to Bayes' rule (Slovic & Lichtenstein, 1971). Open-mindedness (e.g., assimilative processing as induced by high expectations to receive advice in conditions resembling the traditional JAS paradigm; Rebholz & Hütter, 2022) can help overcome biases such as conservatism (Harvey & Harries, 2004). Reasonably, the expectation to receive advice is higher when, for instance, one can think of more people who might be suitable and/or willing to serve as advisors for a certain task. As argued in Section 1.1, advice expectation is inextricably linked to the sampling ecology of the experimental paradigm. The second empirical reinvestigation as presented in the next chapter thus further elaborates on the normative weighting of multiple pieces of sequentially sampled advice.

[4]For the sake of brevity, the same models providing evidence *for* egocentric discounting in both expectation conditions of all other experiments are not reported.

4 Adaptive Sequential Advice Seeking Revisited

In Experiment 5 of Ache (2017), $N = 128$ participants could sample up to $K = 20$ pieces of advice per trial before providing final estimates of the airline distance between $M = 20$ pairs of European cities. In the second manuscript, data of this experiment was reanalyzed to provide initial insights with respect to adaptive strategy selection in sequential advice seeking (Rebholz, Hütter, & Voss, 2023). The prediction performance and selection frequency of a Bayesian account of belief updating (see Section 2.2) was compared to the choosing strategies from single advice taking (Soll & Larrick, 2009). In summary, we found that Bayesian-type compromising is the most frequently selected strategy, followed by no advice taking, and finally choosing the mean of all advisory judgments. On average, the behavior of most participants can be described as relatively more normative than anything else in terms of a high correspondence between actual and predicted final beliefs (i.e., judgment plus confidence; see Rebholz, Hütter, & Voss, 2023, Figure 6). However, from this finding alone it is not clear whether the close correspondence of judgment at the "End-of-Sequence" is indeed due to those participants' weighting behavior resembling the normative updating strategy "Step-by-Step" (cf. Hogarth & Einhorn, 1992). Therefore, comparing empirical MER-weights of sequentially sampled advice to the normative benchmark from the second manuscript extends our research on sequential advice seeking by investigations of adaptive weighting strategies.

4.1 Method

The Step-by-Step Bayesian WOA from Equation 2.9, $\omega^*_{A,ijk}$, is by definition restricted to the $(0, 1)$ interval for participants with reasonable beliefs about their own and advisors' judgment accuracy (see also Rebholz, Hütter, & Voss, 2023, Table 1). In essence, reasonable beliefs correspond to finite and positive initial confidence, C_{ij0}, and advice precision, $\hat{T}_{ijk}^{-2}$, respectively.[5] Hence, the empirical judgment model was fitted as

$$E_{ijK} = \omega_{A,ij}(k)A_{ijk} + [1 - \omega_{A,ij}(k)]E_{ij0} + \varepsilon_{ij}, \tag{4.1}$$

for log-transformed judgments to account for positive skew as in the original study. Corresponding MER-WOA was defined as

$$\omega_{A,ij}(k) = \beta_A + \alpha_{A,i}^S + \alpha_{A,j}^T + \beta_{A\times k}k, \tag{4.2}$$

which implements sequential advice sampling as fixed linear order effect $\beta_{A\times k}$. Alternatively, random order effects could be implemented by adding $\alpha_{A,k}^U \sim N\left(0, \tau_{A,U}^2\right)$ to the standard MER-WOA as defined in Equation 2.11. As fixed and random order effects yielded largely identical results in the sampling analysis of the third manuscript (Rebholz, Biella, & Hütter, 2023), the latter are not implemented here. Trial-wise sum-to-one constraining of MER-WOA with fixed order effects was achieved post hoc by dividing the estimated weights, $\hat{\omega}_{A,ij}(k)$, by the advice sample size that participants realized on a specific trial, K_{ij}. The resulting coefficients were compared to the weights derived from the sequential Bayesian account from the second manuscript, which entails different benchmarks for normative weighting depending on the temporal perspective of the updating process (Rebholz, Hütter, & Voss, 2023). According to Equation 2.8, the influence of earlier advice on posterior judgment is reduced by sequentially multiplying the Step-by-Step Bayesian WOAs from earlier sampling steps by values from the $(0, 1)$ interval in later sampling steps.

4.2 Results

For the sampling experiments of Hütter and Ache (2016), we did not find unequivocal evidence for order effects on weighting sequentially sampled advice in the third manuscript (Rebholz, Biella, & Hütter, 2023). In contrast, for Experiment 5 of Ache (2017) there is evidence for fixed order effects on advice weighting (see the Linear-Distant model in Table 4.1). As $\hat{\beta}_{A\times k} > 0$ there was evidence for recency effects, that is, significantly higher weighting of advice that was sampled later rather than earlier. Although indicating recency effects on average over all sampling positions, the exact functional form of sequential weighting might also be more complex. For instance, serial

[5] Although theoretically possible, the Bayesian account does not align well with both forms of extreme advice weighting. For sequential Bayesian updating, $\hat{\omega}_{A,ijk}^* \in \{0, 1\}$ would require that either the advisee is infinitely confident or that advisory judgments are infinitely precise, respectively. Whereas the latter contradicts empirical findings of judgment accuracy (e.g., Mayer & Heck, 2022; Moussaïd et al., 2013), the former might be justified for knowing the true answer for sure if it exists (e.g., because of searching for it on the internet during an online study). For hierarchical Bayesian updating, internal samples of infinite size would imply no advice weighting, and the opposite (i.e., no internal sampling) would imply complete adoption of advice, respectively. However, the former would require unavailable resources (e.g., an infinite amount of time), and the latter is rather unreasonable for participants in Ache (2017) having to make independent initial judgments before being given the opportunity to sample advice.

Table 4.1: Fixed Effects of Multilevel Models of Final Judgment According to Equation 4.1 With Fixed Linear Order Effects as Specified in Equation 4.2 for Experiment 5 of Ache (2017)

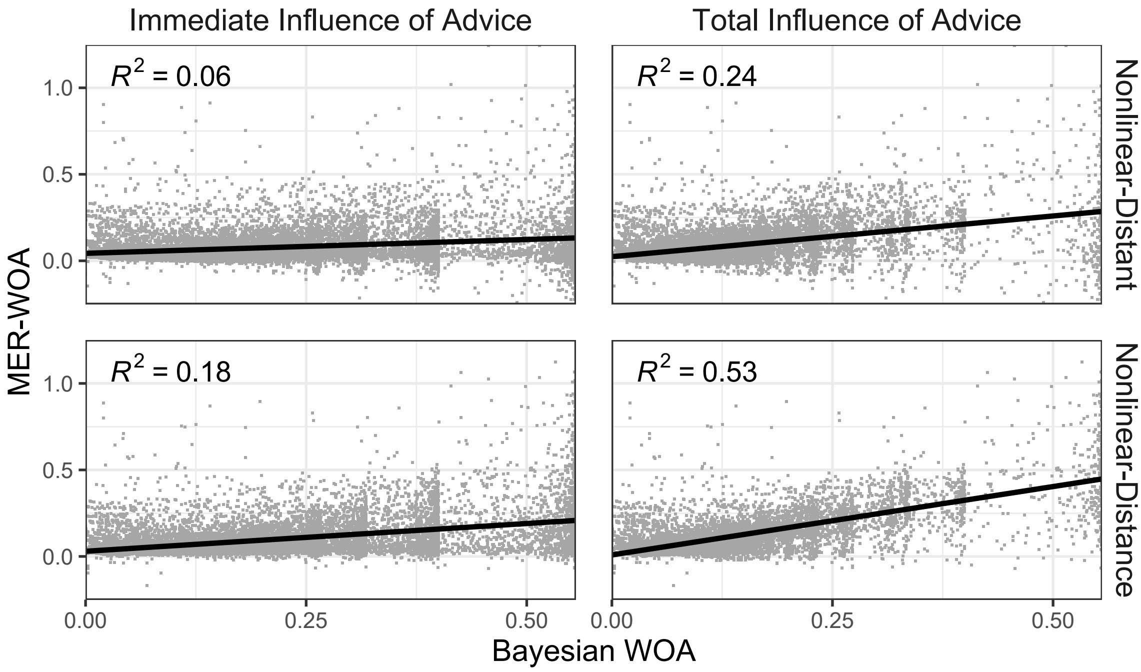
	Linear-Distant		Nonlinear-Distant		Nonlinear-Distance	
	Estimate	*SE*	*Estimate*	*SE*	*Estimate*	*SE*
β_A	0.4539 ***	0.0296	0.4420 ***	0.0298	0.9798 ***	0.1222
$\beta_{A\times C}$	0.2865 ***	0.0116	0.2859 ***	0.0116		
$\beta_{A\times k}$	0.0024 ***	0.0006	0.0081 ***	0.0017	0.0081 ***	0.0018
$\beta_{A\times k^2}$			-0.0003 ***	0.0001	-0.0004 ***	0.0001
$\beta_{A\times D}$					-0.4093 ***	0.1236
$\beta_{A\times log(D)}$					0.4602 **	0.1743

Note. $^*p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$. The full models can be found in Appendix B, Tables B.1 to B.3. In the left and middle panels, fixed treatment effects of contrast-coded advice distance condition (C) are included. In the middle panel, fixed nonlinear order effects are modeled by including a second-order polynomial term of the sampling index k (i.e., k^2). In the right panel, distance condition is replaced by linear and logarithmic effects of the relative absolute distance of advice defined as $D = \frac{|A_k - E_0|}{E_0} + 1$.

positioning curves in free recall typically exhibit a curvilinear shape (Glanzer & Cunitz, 1966). Indeed, polynomial mixed-effects coefficient regressions revealed more systematic patterns of serial positioning for weighting sequentially sampled advice. There is significant evidence for inverse-U-shaped weighting of sequentially sampled advice as $\hat{\beta}_{A\times k^2} < 0$ (see the Nonlinear-Distant model in Table 4.1). The same reasoning about nonlinearity applies to the distance of advice from one's own initial judgments. Much like for other stimulus intensities, peoples' sensitivity diminishes with increasing advice distance (Schultze et al., 2015; see also Stevens, 1957). Accordingly, the effect of distance on single advice weighting was found to be modeled better by a combination of linear and logarithmic terms. A corresponding pattern could also be reproduced for individual, serially nonlinear weights of sequentially sampled advice (see the Nonlinear-Distance model in Table 4.1).

The Nonlinear-Distant model provided the best fit to the data across all models in Table 4.1 according to both Akaike and Bayesian information criteria. However, the correlations with the Step-by-Step Bayesian WOA were stronger for MER-WOA from the Nonlinear-Distance model (see Figure 4.1, top vs. bottom panel). From a sequential or k-wise perspective, Step-by-Step compromising implies primacy effects for the immediate influence of advice, that is, higher immediate weighting of advice that was sampled earlier rather than later (Rebholz, Hütter, & Voss, 2023). The empirical evidence for recency effects as reported above thus suggests a low correspondence between both types of weighting measures for this temporal perspective. Indeed, there

Figure 4.1: Linear Correspondences Between Mixed-Effects Regression (MER) and Step-by-Step Bayesian Weight of Advice (WOA)

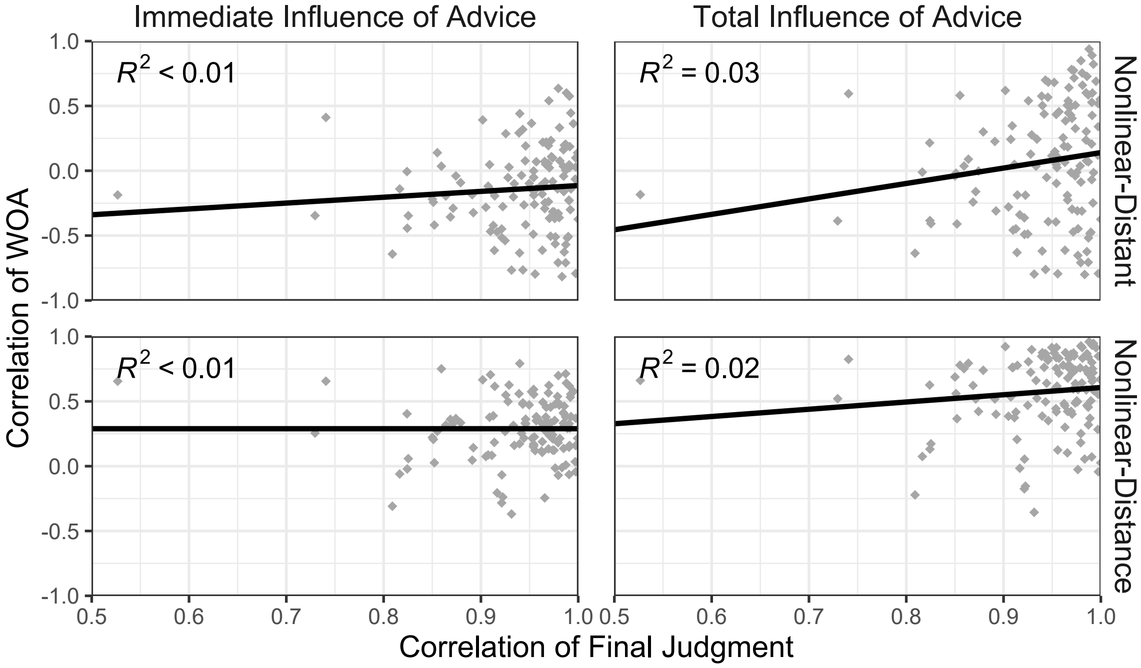


Note. The left (right) panel contains Bayesian WOA measuring the immediate (total) influence of advice. The top (bottom) panel contains MER-WOA based on the Nonlinear-Distant (Nonlinear-Distance) model from Table 4.1. The Step-by-Step Bayesian WOA is restricted to the $(0.00, 0.56)$ interval by assuming internal samples of size larger than one. Plotting is truncated for MER-WOA $\notin [-0.25, 1.25]$. The solid lines indicate the linear correspondences (incl. R^2) between the respective measures of WOA across all sampling trials.

were only small positive correlations between Step-by-Step Bayesian WOAs and empirical MER-weights of sequentially sampled advice from both models (Figure 4.1, left panel). At the End-of-Sequence (i.e., after having finished sampling), however, there is a high likelihood for recency effects in Step-by-Step Bayesian updating for the total influence of advice. The reason is that the impact of earlier pieces of advice is reduced by step-wise multiplications with values strictly smaller than one. Therefore, the correlations were substantially stronger between MER-WOAs and accordingly updated Bayesian WOAs measuring the overall or total influence of a certain piece of advice on final judgment, respectively (Figure 4.1, right panel).

According to the prediction error of actual final beliefs in Experiment 5 of Ache (2017), Bayesian updating is a good description of participants' behavior on many trials (see Rebholz, Hütter, & Voss, 2023, Figures 2 & 5). Actually, some participants' final judgments over $M = 20$ trials were almost perfectly correlated with the Bayesian prediction (see Figure 4.2). However, from the original evidence as reported in the second manuscript alone, it is not clear whether close correspondence is due to those

Figure 4.2: Linear Correspondences Between Participant-Wise Correlations of Mixed-Effects Regression (MER) and Step-by-Step Bayesian Weight of Advice (WOA) as Functions of the Participant-Wise Correlations of Actual and Predicted Final Judgment



Note. The left (right) panel contains Bayesian WOA measuring the immediate (total) influence of advice. The top (bottom) panel contains MER-WOA based on the Nonlinear-Distant (Nonlinear-Distance) model from Table 4.1. The solid lines indicate the linear correspondences (incl. R^2) between the respective participants-wise correlations.

participants' behavior actually resembling the normative weighting strategy. Indeed, Figure 4.2 did not provide evidence for a strong relation between normative weighting and judgment predictability in terms of neither the immediate nor total influence of advice. In other words, there were additional reasons for high correlations between actual and predicted final judgment other than normative weighting of sequentially sampled advice, some of which will be discussed below.

4.3 Discussion

By reanalyzing data from Experiment 5 of Ache (2017), initial evidence was provided for nonlinear serial weighting of sequentially sampled advice. Order effects may reflect perceptions of importance (e.g., Hogarth & Einhorn, 1992), limited central capacities (e.g., Glanzer & Cunitz, 1966), or systematic biases (e.g., Asch, 1946; J. M. Miller & Krosnick, 1998). In any case, the equal weighting approximation as applied in the original study is deemed invalid post hoc. Following from the evidence for recency effects in the weighting of sequentially sampled advice, the correlation is higher between

MER- and Bayesian WOA measuring the total influence as compared to the immediate influence of advice (see Figure 4.1). As specified in Rebholz, Hütter, and Voss (2023), the $(0.00, 0.56)$ interval corresponds to the full range in which the Step-by-Step Bayesian WOA can theoretically lie. This interval contains 94.95% of the empirical MER-weights of sequentially sampled advice from the Nonlinear-Distant model, and even 98.19% of MER-WOAs from the Nonlinear-Distance model. However, the trend lines in the right panel of Figure 4.1 are below the diagonal lines of the plots. Hence, on average, egocentric discounting is even more severe than implemented in the sequential Bayesian account by internal samples of size $L_0 = 4$.

At the End-of-Sequence, participants may retrieve early advice from long-term memory and late advice from working memory (Glanzer & Cunitz, 1966). Such a retrieval process would produce a bimodal serial positioning curve with recency and primacy effects. However, there is significant evidence for unimodal, inverse-U-shaped serial weighting of sequentially sampled advice (see the Nonlinear-Distance model in Table 4.1). For realized advice sample sizes on average slightly larger than working memory capacity ($M = 5.51$, $SD = 5.23$; Cowan, 2010; see also G. A. Miller, 1956), overlapping retrieval processes cannot account for this finding. Instead, Step-by-Step updating as implemented in the sequential Bayesian account can explain inverse-U-shaped serial weighting. As Equation 2.8 relies only on parameter estimates from the current and preceding trials, corresponding judgment formation allows participants to deliberately weight advice in the middle more than at the beginning or end of a sampling chain. Immediate integration is independent of remembering specific pieces of advice when being requested to make a final judgment at the end of a certain trial (Behrens et al., 2007). Consequently, participants in Experiment 5 of Ache (2017) seem to rather have updated their judgments after each single piece of advice instead of relying on their limited (working) memory capacities at the End-of-Sequence.

There is no evidence for strong relations between normative weighting and judgment predictability (see Figure 4.2). Only because predicted and actual final judgments overlap does not guarantee that the updating processes were the same, too. Mathematically, participants' final judgments can be close to the Bayesian prediction at the End-of-Sequence for reasons other than applying normative weighting strategies. For instance, in a sampling sequence of length $K_{ij} = 2$, the first piece of advice A_1 may be a little less distant to the initial judgment E_0 than the second piece of advice A_2 to its corresponding prior $\hat{E}_1$. With non-zero weighting of all available judgments, it is possible that the Bayesian posterior exactly corresponds to the first piece of advice, that is, $\hat{E}_2 = A_1$. Participants' actual final judgment also being equal to the first piece of advice (i.e., $E_2 = A_1$), however, may also be due to choosing advice but knowing for sure that the true value is larger than A_2. Consequently, the second piece of advice is excluded from the set of plausible values (cf. Kahneman, 1992; Wegener et al., 2001, for similar perspectives on extreme anchors). Certainty about lower bounds of plausibility does not per se exclude uncertainty about corresponding upper bounds, and vice versa, which might also explain asymmetric confidence intervals as observed in the data (O'Connor et al., 2001; see also Soll et al., 2022). Moreover, it is easier to predict a single number (e.g., final judgment) than to correctly identify belief distributions due to the "curse of dimensionality" (e.g., Friedman, 1997; Guyon & Elisseeff, 2003). Hence, this simplistic example also nicely demonstrates why we additionally conducted holistic distributional testing in the second manuscript. Our intention was to render the original analyses more robust against coincidental matching and additional sequential advice seeking strategies that were not contained in the model comparison.

5 General Discussion

This dissertation presents multiple approaches based on different statistical philosophies to model ecological constraints such as effects of high versus low expectations and sequential sampling opportunities on advice taking. To draw inferential conclusions, the traditional two-stage approach as applied in the first manuscript utilizes Harvey and Fischer's (1997) weighting index as criterion for statistical optimization at stage two (Rebholz & Hütter, 2022). In contrast, by explicitly telling Thurstonian and Brunswikian sources of information apart in the second manuscript, Bayesian updating implicitly takes the endogeneity (i.e., dependency on external sources of information) of final judgments into consideration (Rebholz, Hütter, & Voss, 2023). However, the Bayesian account in isolation only enables empirical conclusions with respect to judgment strategy selection. Crucially, individual empirical weights of sequentially sampled advice can be derived from the multilevel modeling framework as proposed in the third manuscript (Rebholz, Biella, & Hütter, 2023), which allow insights with respect to weighting strategy selection. By also building on the endogeneity of final judgments, statistical optimization is an integral part of the derivation of corresponding MER-weights. Consequently, the multilevel modeling framework is both more flexible and less sensitive to weighting specifications than the traditional approach. This makes it a valuable tool for countering the reproducibility crisis by reducing researcher degrees of freedom. Moreover, it enables innovative and integrative research such as the novel empirical conclusions presented with respect to expectations and sampling as ecological constraints of (social) information acquisition.

In the first empirical application as presented in Chapter 3, the positive effect of advice expectation on weighting as indicated by traditional modeling is verified by process-consistent multilevel modeling. For a trial-wise contrast of expectation in Rebholz and Hütter (2022, Experiments 3 & 4), expected advice is weighted significantly more strongly than unexpected advice. This constitutes additional support for ongoing mental tasks featuring assimilative processing of advice (cf. Alexopoulos et al., 2012). The second empirical application as presented in Chapter 4 reveals moderate to strong correlations between empirical MER-weights of sequentially sampled advice and Step-by-Step Bayesian weights measuring the total influence of advice. Nevertheless, the new evidence suggests that the predictability of participants' final judgments as reported in Rebholz, Hütter, and Voss (2023) cannot unequivocally be attributed to step-wise belief updating in line with the normative rules as prescribed by the sequential Bayesian account. Instead, participants apply additional belief updating strategies to integrate sequentially sampled advice such as choosing the self or (all) advisors, but also strategies that were not contained in the original model comparison reported in the second manuscript. Additional limitations of the proposed modeling frameworks to account for consequences of expecting external influences, active sampling, and related ecological constraints will be discussed in the context of future research in the following section.

5.1 Merits, Limitations, and Future Research

The ROD formula is intuitive and simple, which makes it easy to calculate, communicate, and comprehend. In contrast, the foundations and interpretations of the proposed Bayesian and regression-based weighting measures involve more complicated statistical and mathematical concepts. On the one hand, sum-to-one constrained mixed-effects regression of final judgment on two different sources of information as specified in Equation 2.10 shares the intuitive interpretability of the traditional approach (Rebholz, Biella, & Hütter, 2023). The corresponding MER-WOA quantifies advice taking relative to the weighting of own initial judgments. On the other hand, partial mixed-effects from unconstrained regression of final judgment on any number and type of exogenous sources of information are less intuitive. The relative interpretation of partial mixed-effects can be recovered post hoc by normalization, that is, dividing the estimated coefficients by the sum of all unrestricted weights. For any model specification, however, random deviations from average weighting as specified in MER-type shrinkage estimates of WOA involve fundamental statistical concepts (i.e., randomness, distributions, and location measures; Snijders & Bosker, 2012). Bayes' theorem additionally involves conditional and marginal probabilities (more generally, densities of continuous random variables for quantitative judgment) including corresponding calculus (e.g., the law of total probability; Gelman et al., 2013). Accordingly, the lack of correspondence between the normativity of advice weighting and the predictability of final judgments as reported in Chapter 4 might also be attributed to a lacking deliberateness of Bayesian-type compromising (Soll & Larrick, 2009).

Step-by-Step Bayesian WOAs derived from normal distributional assumptions match findings from traditional advice taking research. Indeed, relative uncertainty-dependent weights exhibit a rather intuitive and natural interpretation in terms of internal versus external "inconsistency discounting" (Anderson, 1971; Anderson & Jacobson, 1965; Yaniv, 2004a; see also Minson et al., 2011; Rebholz, Hütter, & Voss, 2023, Table 1). In line with Information Integration Theory, distant advice is perceived as external inconsistency and consequently weighted relatively less strongly than close advice in the sequential Bayesian account. Conversely, the self is weighted relatively less strongly for lower as compared to higher levels of confidence as the former reflects greater internal inconsistency in a Thurstonian sense (Sniezek & Buckley, 1995). This interrelationship is in line with the operationalization of advice taking as a combination of judgment and confidence updating (Soll et al., 2022; see also Moussaïd et al., 2013; Schultze et al., 2015). Our Bayesian account from the second manuscript generalizes the corresponding "influence of advice" measure's notion of belief revisions as distributional shifts to sequential advice seeking.

In discrete choice, the mere act of making final decisions (Paese & Sniezek, 1991; Sniezek et al., 1990) as well as the anticipation of having to justify one's choices (Arkes et al., 1987), were found to reduce confidence. Indirectly, both conditions also apply to the JAS paradigm. Specifically, advice that is supposedly centered at the true value of an item—due to participants appreciating the wisdom of crowds (Larrick & Soll, 2006; Mannes, 2009)—may be construed and processed as performance feedback on independent initial judgments (but see Blunden et al., 2019; Brooks et al., 2015, for conceptual differences between advice and feedback). Nevertheless, in the first JAS study, advisor recommendations different from own initial choices did not reduce confidence (Sniezek & Buckley, 1995). One explanation for the previously mentioned negative effect of anticipated justification on confidence is that it triggers more extensive information search (Sniezek & Buckley, 1993). In contrast, receiving multiple diverging pieces of advice made participants in the "Cued" × "Conflict" condition of Sniezek and Buckley (1995) stop their own information search. Therefore, they argued against a "one-to-one correspondence" between diverging opinions and confidence updating. Consequently, strictly growing confidence as well as the assumption that participants have completed internal sampling before starting to consider external information as implemented in the Bayesian account may not reflect participants' actual behavior. Implementing more dynamic weighting not only for judgment but also for confidence updating in Equations 2.6 and 2.7 would resolve both limitations at once. In general, a more ecological Bayesian account would consider the possibility of simultaneous internal and external sampling.

A simple and intuitive concept of weighting is also essential to investigate and identify congruency of actual and desired weighting. In Ache et al. (2020), the original paradigm was extended by yoking real participants for repeated judge-advisor interactions. Participants were randomly assigned one of the two JAS roles. In addition to making judgments (i.e., best guesses) in both roles, participants assigned the advisor role indicated how much they wanted participants assigned the advisee role to weight their advice. Advisors had to specify where between the initial judgment and their advice they wanted the advisee's final judgment to be located. As response format in Experiments 1 and 2, Ache et al. (2020) implemented a visual analogue scale that ranged from 0% or no weighting (i.e., "keep" own judgment) to 100% or full weighting (i.e., "adopt" advisory judgment). Consequently, incongruency could be measured as the discrepancy between actual and desired weighting. In Experiment 1, the visual response format yielded "largely identical" results as asking advisor-participants for their expectations about judgment congruency. That is, how close they expected their advisee's final judgment to be to their advice. As discussed above, the more advanced modeling approaches rely on fundamentally different conceptual notions of how WOA is determined. Nevertheless, they share the construal of judgment formation as an endogenous process (see Chapter 2). Accordingly, for single advice taking, no arithmetic beyond deterministic weighting in line with the definition of ROD-WOA is required for participants to understand, calculate, and formulate (actual and desired) weighting. For sequential advice seeking, however, this holds true only for total weighting of all sampled pieces of advice (e.g., calculated as the ROD-weight of mean advice; Hütter & Ache, 2016), or for implementing and communicating intermediate judgment updating (cf. Hogarth, 1978), respectively.

Ideally, congruency between desired and actual weighting of advice would avoid negative interpersonal consequences of incongruency such as a reduced willingness to provide advice in future interactions (Ache et al., 2020; see also Blunden et al., 2019; Palmeira & Romero Lopez, 2023). In Experiments 3 and 4 of Ache et al. (2020), advisees' initial judgment and confidence was communicated to advisors with the goal to resolve the informational asymmetry between both JAS parties and thus to enable (more) informed expressions of desired weighting. However, there was still informational asymmetry on another consequential dimension as follows from the new and validated evidence for the positive effect of expectation on weighting (see Chapter 3). At least in direct contrast to a relatively low expectation of external influences in the future, a high expectation to receive advice indeed seems to foster assimilative processing of additional information provided by others (Rebholz & Hütter, 2022; see also below). In their considerations about desired weighting, advisors should thus also take into account their advisees' expectations about the sampling ecology, that is, the chances to receive advice from them or any other potential advisor. Hence, not only advisees should take into account their advisors' expectations of how much they want their advice to be used in order to avoid negative interpersonal consequences (Ache et al., 2020). More generally, the presented results hence suggest that perspective taking constitutes another important, genuinely social ecological constraint of information acquisition that has great potential to improve advice interactions in real-world JASs (but see Epley et al., 2006, for increased egocentrism as a result of perspective taking in competitive interactions).

Except for the experiments as reported in the first manuscript, the traditional JAS paradigm usually implements high or even full expectations to receive advice (Rebholz & Hütter, 2022). In contrast to being allowed to revise initial judgments in the light of advice, irreversible decisions trigger coping mechanisms that feature less assimilative processing of external information (Knox & Inkster, 1968; see also Bullens et al., 2011; Liberman & Förster, 2006). Nevertheless, process-consistent multilevel modeling does not change the evidence for egocentric discounting in most high-expectation conditions implemented in the first manuscript, except one. In Experiment 5, the single piece of advice was weighted significantly more strongly than the equal weighting threshold of 0.50 in both expectation conditions. The restricted ecological validity of the judgment task might have increased average advice weighting in this experiment (see Section

3.3). In general, however, the traditional focus on underweighting of advice implies a narrow conception of normative advice taking. For instance, research also found that invalid or misleading advice is weighted too much due to participants' "metacognitive myopia" with respect to advice validity (Fiedler et al., 2019). Essentially, if there are no additional cues available to assess judgment quality, complete adoption of advice (i.e., WOA = 1) is as inappropriate as not taking the advice at all (i.e., WOA = 0). In contrast, the total normative weight of many independent advisors converges to one (Yaniv & Kleinberger, 2000; Yaniv & Milyavsky, 2007). In other words, people should exclusively rely on a large, unbiased crowd's wisdom instead of their own, limited judgment abilities. In the real world, however, advice solicitation depends on many external factors. For instance, advisors take umbrage at judges who consult multiple advisors (Blunden et al., 2019). To that effect, they are less willing to give (good; Bonaccio & Dalal, 2006) advice again in the future. Thus, the sampling ecology constitutes an ecological constraint that is naturally related to the normativity of high versus low expectations to receive advice in a certain judgment environment.

In the second empirical application, process-consistent multilevel modeling was applied to confirm serial positioning implications of the sequential Bayesian account. Initial evidence was provided for nonlinear, inverse-U-shaped serial positioning effects in the weighting of sequentially sampled advice (see Table 4.1). In other words, ordering indeed matters for the processing of multiple pieces of sequentially sampled advice of varying distance. The modeling of serial positioning as fixed versus random order effect is not only a matter of technical concern, but also substantively requires the existence of a corresponding sampling ecology, for instance, by experimental implementation (Rebholz, Biella, & Hütter, 2023). In the original sequential sampling extension of the JAS, advice giving was implemented as the provision of random numbers centered at a certain distance from participants' initial judgments (Hütter & Ache, 2016). However, advice is rarely sampled at random in the real world where people often turn to family members or friends for advice, such as in the introductory example about going to work by bus or bicycle. Close social others often share similar beliefs, for instance, due to encountering the same information (e.g., Soll & Larrick, 2009; Yaniv, 2004b; see also Gino et al., 2009; Schulz & Roessler, 2012). Essentially, positively correlated judgment errors are detrimental for the wisdom of crowds and thus for the accuracy benefits of advice taking (e.g., Broomell & Budescu, 2009; Davis-Stober et al., 2014; Hogarth, 1978; see also Schultze et al., 2019).

Technically, correlated errors of multiple judgments (e.g., due to encountering similar sources of information; Soll & Larrick, 2009; Yaniv, 2004b) are unproblematic for collinearity-stable statistical optimization such as the multilevel modeling framework as proposed in the third manuscript (Rebholz, Hütter, & Voss, 2023). Substantively, however, static JASs are becoming increasingly unlikely in digitalized societies as modern information technology facilitates the total amount and frequency of (social) interactions in a network (e.g., Wang & Wellman, 2010; Wellman, 2012; but see Dunbar, 2012, 2016, for cognitive constraints). As a result, those seeking advice today may themselves be asked for advice on the same or a closely related topic in future interactions—also from their former advisors. Recent research about the consequences of advice solicitation on competency perceptions included self-rated intentions about additional interactions with reversed roles in the future (Brooks et al., 2015). Crucially, role reversals were found to cause the JAS agents to reciprocally increase the weighting of each others' judgment (Mahmoodi et al., 2018). In contrast, people tend to converge back to their initial opinions after leaving small group discussions (see Kerr & Tindale, 2004; Tindale & Kameda, 2000, for reviews). Therefore, future research should also allow for actual repeated interactions in more dynamic JASs with role reversals. In general, the agents' experiences from previous advice interactions (e.g., in terms of incongruency or reciprocity) likely affect their current behavior (Harvey & Fischer, 1997; Kämmer et al., 2023; Mahmoodi et al., 2022). Accordingly, repeated interactions constitute another ecological constraint that is relevant in particular, but not exclusively, for information acquisition in social contexts.

Sequentially alternating between serving as advisor or advisee is also interesting with respect to consequences of (indirect) self-advising (Ariely et al., 2000; Herzog & Hertwig, 2009, 2014). A related phenomenon exists in digitalized judgment environments. Artificial intelligence-based decision support systems often base their recommendations on observations and aggregations of peoples' behavior in the past. This induces a "feedback loop" as the data for (regular) retraining of the algorithm generally also includes its current users' behavior (Chaney et al., 2018). In a corresponding extension of the traditional JAS paradigm, where the sources of information were accordingly manipulated, advisees do not discriminate between artificial and human sources of advice (Hütter & Fiedler, 2019). Hence, human-algorithm interactions can involve indirect self-advising that is akin to repeated human-human interactions with role reversal as described above. Taken together, more advanced cognitive modeling should take peoples' mental models about other human (see Schurz et al., 2021, for a review) or artificial (Logg, 2022) agents' processing of their own, potentially outdated previous beliefs into consideration (cf. "shared metacognition" in group decision-making; Tindale & Kameda, 2000; see also Bonaccio & Dalal, 2006; Mathieu et al., 2000). Particularly, advice that is knowingly based on another agent's processing of one's own previous beliefs might impose corresponding conceptual and computational issues.

5.2 Conclusion

Methodologically, this dissertation presents different modeling frameworks to conduct research on ecological constraints of (social) information acquisition. In general, I propose to rely on process-consistent statistical modeling, that is, to explicitly model the dependency of endogenous judgments (i.e., potentially updated beliefs) on exogenous sources of information (e.g., advice). Essentially, process-consistent modeling renders more differentiated research in terms of modeling flexibility possible. For instance, the assumption of normally distributed beliefs in the second manuscript might be abandoned in favor of a sequential Bayesian account that is also capable of serial positioning other than primacy (recency) effects for the immediate (total) influence of advice. Moreover, individual weights of sequentially sampled advice as derived from the multilevel modeling framework in the third manuscript can easily be extended to discrete choice and multidimensional belief updating (i.e., judgment plus confidence), too.

Substantively, process-consistent modeling is applied to demonstrate how various ecological constraints influence advice taking and phenomena with related cognitive structure. For instance, new evidence is provided with respect to the normativity of weighting external sources of information depending on the sampling ecology. Future research regarding expectations of advice, active information seeking, and related ecological constraints has great potential to extend our understanding of dyadic judgment and decision-making. To mention just one particularly interesting domain, interpersonal consequences of (in)congruency between wanted and actual weighting likely depend on mental models of other agents' information processing and their expectations about advice interactions. In conclusion, the generation of novel and relevant insights about ecological constraints in information sampling and utilization requires integrative research approaches and methods such as those presented in this dissertation.

6 Bibliography

Ache, F. (2017). *Returning advice taking to the wild: Empirical, theoretical, and normative implications of an ecological perspective* [Dissertation]. Eberhard Karls University of Tübingen. Tübingen, Germany. https://doi.org/10.15496/publikation-19538

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

Adjodah, D., Leng, Y., Chong, S. K., Krafft, P. M., Moro, E., & Pentland, A. (2021). Accuracy-risk trade-off due to social learning in crowd-sourced financial predictions. *Entropy*, *23*(7), 801. https://doi.org/10.3390/e23070801

Alexopoulos, T., Fiedler, K., & Freytag, P. (2012). The impact of open and closed mindsets on evaluative priming. *Cognition and Emotion*, *26*(6), 978–994. https://doi.org/10.1080/02699931.2011.630991

Anderson, N. H. (1971). Integration theory and attitude change. *Psychological Review*, *78*(3), 171–206. https://doi.org/10.1037/h0030834

Anderson, N. H. (1981). *Foundations of information integration theory*. Academic Press.

Anderson, N. H., & Jacobson, A. (1965). Effect of stimulus inconsistency and discounting instructions in personality impression formation. *Journal of Personality and Social Psychology*, *2*(4), 531–539. https://doi.org/10.1037/h0022484

Ariely, D., Au, W. T., Bender, R. H., Budescu, D. V., Dietz, C. B., Gu, H., Wallsten, T. S., & Zauberman, G. (2000). The effects of averaging subjective probability estimates between and within judges. *Journal of Experimental Psychology: Applied*, *6*(2), 130–147. https://doi.org/10.1037/1076-898X.6.2.130

Arkes, H. R., & Blumer, C. (1985). The psychology of sunk cost. *Organizational Behavior and Human Decision Processes*, *35*(1), 124–140. https://doi.org/10.1016/0749-5978(85)90049-4

Arkes, H. R., Christensen, C., Lai, C., & Blumer, C. (1987). Two methods of reducing overconfidence. *Organizational Behavior and Human Decision Processes*, *39*(1), 133–144. https://doi.org/10.1016/0749-5978(87)90049-5

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 M. Paquot & S. T. Gries (Eds.), *A Practical Handbook of Corpus Linguistics* (pp. 563–591). Springer International Publishing. https://doi.org/10.1007/978-3-030-46216-1_23

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*, 1–26. 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

Bednarik, P., & Schultze, T. (2015). The effectiveness of imperfect weighting in advice taking. *Judgment and Decision Making*, *10*(3), 265–276. https://doi.org/10.1017/S1930297500004666

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*(9), 1214–1221. https://doi.org/10.1038/nn1954

Biella, M., & Hütter, M. (2023). *Navigating the social environment: Impression formation as function of goal-related information search dynamics*. PsyArXiv. https://doi.org/10.31234/osf.io/m6azf

Blunden, H., Logg, J. M., Brooks, A. W., John, L. K., & Gino, F. (2019). Seeker beware: The interpersonal costs of ignoring advice. *Organizational Behavior and Human Decision Processes*, *150*, 83–100. https://doi.org/10.1016/j.obhdp.2018.12.002

Bolker, B. M. (2015). Linear and generalized linear mixed models. In G. A. Fox, S. Negrete-Yankelevich, & V. J. Sosa (Eds.), *Ecological statistics* (pp. 309–333). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199672547.003.0014

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

Brooks, A. W., Gino, F., & Schweitzer, M. E. (2015). Smart people ask for (my) advice: Seeking advice boosts perceptions of competence. *Management Science*, *61*(6), 1421–1435. https://doi.org/10.1287/mnsc.2014.2054

Broomell, S. B., & Budescu, D. V. (2009). Why are experts correlated? Decomposing correlations between judges. *Psychometrika*, *74*(3), 531–553. https://doi.org/10.1007/S11336-009-9118-Z

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

Brunswik, E. (1952). *The conceptual framework of psychology.* University of Chicago Press.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments* (2nd ed.). University of California Press.

Budescu, D. V., & Yu, H.-T. (2006). To Bayes or not to Bayes? A comparison of two classes of models of information aggregation. *Decision Analysis*, *3*(3), 145–162. https://doi.org/10.1287/deca.1060.0074

Bullens, L., van Harreveld, F., & Förster, J. (2011). Keeping one's options open: The detrimental consequences of decision reversibility. *Journal of Experimental Social Psychology*, *47*(4), 800–805. https://doi.org/10.1016/j.jesp.2011.02.012

Chaney, A. J. B., Stewart, B. M., & Engelhardt, B. E. (2018). How algorithmic confounding in recommendation systems increases homogeneity and decreases utility. In S. Pera, M. Ekstrand, X. Amatriain, & J. O'Donovan (Eds.), *Proceedings of the 12th ACM Conference on Recommender Systems* (pp. 224–232). ACM. https://doi.org/10.1145/3240323.3240370

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

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

Dunbar, R. I. M. (2012). Social cognition on the internet: Testing constraints on social network size. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1599), 2192–2201. https://doi.org/10.1098/rstb.2012.0121

Dunbar, R. I. M. (2016). Do online social media cut through the constraints that limit the size of offline social networks? *Royal Society Open Science*, *3*(1), 150292. https://doi.org/10.1098/rsos.150292

Edwards, J. R. (1994). The study of congruence in organizational behavior research: Critique and a proposed alternative. *Organizational Behavior and Human Decision Processes*, *58*(1), 51–100. https://doi.org/10.1006/obhd.1994.1029

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

Edwards, W. (1962). Dynamic Decision Theory and Probabilistic Information Processings. *Human Factors*, *4*(2), 59–74. https://doi.org/10.1177/001872086200400201

Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review*, *70*(3), 193–242. https://doi.org/10.1037/h0044139

Epley, N., Caruso, E. M., & Bazerman, M. H. (2006). When perspective taking increases taking: Reactive egoism in social interaction. *Journal of Personality and Social Psychology*, *91*(5), 872–889. https://doi.org/10.1037/0022-3514.91.5.872

Feng, B., & MacGeorge, E. L. (2006). Predicting receptiveness to advice: Characteristics of the problem, the advice-giver, and the recipient. *Southern Communication Journal*, *71*(1), 67–85. https://doi.org/10.1080/10417940500503548

Feng, B., & Magen, E. (2016). Relationship closeness predicts unsolicited advice giving in supportive interactions. *Journal of Social and Personal Relationships*, *33*(6), 751–767. https://doi.org/10.1177/0265407515592262

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., & Juslin, P. (2006). Taking the interface between mind and environment seriously. In K. Fiedler & P. Juslin (Eds.), *Information sampling and adaptive cognition* (pp. 3–29). Cambridge University Press. https://doi.org/10.1017/CBO9780511614576.001

Fiedler, K., & Kutzner, F. (2015). Information sampling and reasoning biases. In G. Keren & G. Wu (Eds.), *The Wiley Blackwell handbook of judgment and decision making* (pp. 380–403). John Wiley & Sons. https://doi.org/10.1002/9781118468333.ch13

Friedman, J. H. (1997). On bias, variance, 0/1—loss, and the curse-of-dimensionality. *Data Mining and Knowledge Discovery*, *1*(1), 55–77. https://doi.org/10.1023/A:1009778005914

Galton, F. (1907). Vox Populi. *Nature*, *75*(1949), 450–451. https://doi.org/10.1038/075450a0

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). CRC Press Taylor and Francis Group.

Gibbons, M. A., Sniezek, J. A., & Dalal, R. S. (2003–November 10). Antecedents and consequences of unsolicited versus explicitly solicited advice. In D. V. Budescu (Chair) *Symposium in honor of Janet Sniezek* [Symposium]. https://slideplayer.com/slide/12631840/

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., 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

Glanzer, M., & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior*, *5*(4), 351–360. https://doi.org/10.1016/S0022-5371(66)80044-0

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, *3*, 1157–1182.

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

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

Herzog, S. M., & Hertwig, R. (2009). The wisdom of many in one mind: Improving individual judgments with dialectical bootstrapping. *Psychological Science*, *20*(2), 231–237. https://doi.org/10.1111/j.1467-9280.2009.02271.x

Herzog, S. M., & Hertwig, R. (2014). Harnessing the wisdom of the inner crowd. *Trends in Cognitive Sciences*, *18*(10), 504–506. https://doi.org/10.1016/j.tics.2014.06.009

Hoffman, P. J. (1960). The paramorphic representation of clinical judgment. *Psychological Bulletin*, *57*(2), 116–131. https://doi.org/10.1037/h0047807

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

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

Hütter, M., & Ache, F. (2016). Seeking advice: A sampling approach to advice taking. *Judgment and Decision Making*, *11*(4), 401–415. https://doi.org/10.1017/S193029750000382X

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

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

Kahneman, D. (1992). Reference points, anchors, norms, and mixed feelings. *Organizational Behavior and Human Decision Processes*, *51*(2), 296–312. https://doi.org/10.1016/0749-5978(92)90015-Y

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, *3*(3), 430–454. https://doi.org/10.1016/0010-0285(72)90016-3

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263. https://doi.org/10.2307/1914185

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*. https://doi.org/10.1037/dec0000199

Kerr, N. L., & Tindale, R. S. (2004). Group performance and decision making. *Annual Review of Psychology*, *55*, 623–655. https://doi.org/10.1146/annurev.psych.55.090902.142009

Knox, R. E., & Inkster, J. A. (1968). Postdecision dissonance at post time. *Journal of Personality and Social Psychology*, *8*(4), 319–323. https://doi.org/10.1037/h0025528

Koriat, A. (2012a). The self-consistency model of subjective confidence. *Psychological Review*, *119*(1), 80–113. https://doi.org/10.1037/a0025648

Koriat, A. (2012b). When are two heads better than one and why? *Science*, *336*(6079), 360–362. https://doi.org/10.1126/science.1216549

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 107–118. https://doi.org/10.1037/0278-7393.6.2.107

Larrick, R. P., & Feiler, D. C. (2015). Expertise in decision making. In G. Keren & G. Wu (Eds.), *The Wiley Blackwell handbook of judgment and decision making* (pp. 696–721). John Wiley & Sons. https://doi.org/10.1002/9781118468333.ch24

Larrick, R. P., & Soll, J. B. (2006). Intuitions about combining opinions: Misappreciation of the averaging principle. *Management Science*, *52*(1), 111–127. https://doi.org/10.1287/mnsc.1050.0459

Liberman, N., & Förster, J. (2006). Inferences from decision difficulty. *Journal of Experimental Social Psychology*, *42*(3), 290–301. https://doi.org/10.1016/j.jesp.2005.04.007

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. (2022). The psychology of Big Data: Developing a "theory of machine" to examine perceptions of algorithms. In S. C. Matz (Ed.), *The psychology of technology: Social science research in the age of Big Data* (pp. 349–378). American Psychological Association. https://doi.org/10.1037/0000290-011

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

Mahmoodi, A., Bahrami, B., & Mehring, C. (2018). Reciprocity of social influence. *Nature Communications*, *9*(1), 2474. https://doi.org/10.1038/s41467-018-04925-y

Mahmoodi, A., Nili, H., Bang, D., Mehring, C., & Bahrami, B. (2022). Distinct neurocomputational mechanisms support informational and socially normative conformity. *PLoS Biology*, *20*(3), e3001565. https://doi.org/10.1371/journal.pbio.3001565

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

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology*, *85*(2), 273–283. https://doi.org/10.1037/0021-9010.85.2.273

Mayer, M., & Heck, D. W. (2022). Sequential collaboration: The accuracy of dependent, incremental judgments. *Decision*. https://doi.org/10.1037/dec0000193

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. https://doi.org/10.1037/h0043158

Miller, J. M., & Krosnick, J. A. (1998). The impact of candidate name order on election outcomes. *Public Opinion Quarterly*, *62*(3), 291. https://doi.org/10.1086/297848

Minson, J. A., Liberman, V., & Ross, L. (2011). Two to tango: Effects of collaboration and disagreement on dyadic judgment. *Personality & Social Psychology Bulletin*, *37*(10), 1325–1338. https://doi.org/10.1177/0146167211410436

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., 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

Morris, P. A. (1974). Decision analysis expert use. *Management Science*, *20*(9), 1233–1241. https://doi.org/10.1287/mnsc.20.9.1233

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

Oberpriller, J., de Souza Leite, M., & Pichler, M. (2022). Fixed or random? On the reliability of mixed-effects models for a small number of levels in grouping variables. *Ecology and Evolution*, *12*(7), e9062. https://doi.org/10.1002/ece3.9062

O'Connor, M., Remus, W., & Griggs, K. (2001). The asymmetry of judgemental confidence intervals in time series forecasting. *International Journal of Forecasting*, *17*(4), 623–633. https://doi.org/10.1016/S0169-2070(01)00103-0

Paese, P. W., & Sniezek, J. A. (1991). Influences on the appropriateness of confidence in judgment: Practice, effort, information, and decision-making. *Organizational*

Behavior and Human Decision Processes, *48*(1), 100–130. https://doi.org/10.1016/0749-5978(91)90008-H

Palmeira, M., & Romero Lopez, M. (2023). The opposing impacts of advice use on perceptions of competence. *Journal of Behavioral Decision Making*, e2318. https://doi.org/10.1002/bdm.2318

Pronin, E., Gilovich, T., & Ross, L. (2004). Objectivity in the eye of the beholder: Divergent perceptions of bias in self versus others. *Psychological Review*, *111*(3), 781–799. https://doi.org/10.1037/0033-295X.111.3.781

Rader, C. A., Larrick, R. P., & Soll, J. B. (2017). Advice as a form of social influence: Informational motives and the consequences for accuracy. *Social and Personality Psychology Compass*, *11*(8), e12329. https://doi.org/10.1111/spc3.12329

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.). Sage Publications.

Rauhut, H., & Lorenz, J. (2011). The wisdom of crowds in one mind: How individuals can simulate the knowledge of diverse societies to reach better decisions. *Journal of Mathematical Psychology*, *55*(2), 191–197. https://doi.org/10.1016/j.jmp.2010.10.002

Rebholz, T. R., Biella, M., & Hütter, M. (2023). *Mixed-effects regression weights (of advice): Process-consistent modeling of information sampling and utilization.* PsyArXiv. https://doi.org/10.31234/osf.io/x36az

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 (SJDM) Annual Meeting, San Diego, CA. https://sjdm.org/presentations/2022-Poster-Reif-Jessica-advice-anchoring-influence~.pdf

Schrah, G. E., Dalal, R. S., & Sniezek, J. A. (2006). No decision-maker is an island: Integrating expert advice with information acquisition. *Journal of Behavioral Decision Making*, *19*(1), 43–60. https://doi.org/10.1002/bdm.514

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

Schulz, A., & Roessler, P. (2012). The spiral of silence and the internet: Selection of online content and the perception of the public opinion climate in computer-mediated communication environments. *International Journal of Public Opinion Research*, *24*(3), 346–367. https://doi.org/10.1093/ijpor/eds022

Schurz, M., Radua, J., Tholen, M. G., Maliske, L., Margulies, D. S., Mars, R. B., Sallet, J., & Kanske, P. (2021). Toward a hierarchical model of social cognition: A neuroimaging meta-analysis and integrative review of empathy and theory of mind. *Psychological Bulletin*, *147*(3), 293–327. https://doi.org/10.1037/bul0000303

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

Sniezek, J. A., & Buckley, T. (1993). Becoming more or less uncertain. In N. J. Castellan JR. (Ed.), *Individual and group decision making: Current issues* (pp. 87–108). Lawrence Erlbaum Associates.

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

Sniezek, J. A., Paese, P. W., & Switzer, F. S. (1990). The effect of choosing on confidence in choice. *Organizational Behavior and Human Decision Processes*, *46*(2), 264–282. https://doi.org/10.1016/0749-5978(90)90032-5

Sniezek, J. A., Schrah, G. E., & Dalal, R. S. (2004). Improving judgement with prepaid expert advice. *Journal of Behavioral Decision Making*, *17*(3), 173–190. https://doi.org/10.1002/bdm.468

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., & Klayman, J. (2004). Overconfidence in interval estimates. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 299–314. https://doi.org/10.1037/0278-7393.30.2.299

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

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review*, *64*(3), 153–181. https://doi.org/10.1037/h0046162

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

Tindale, R. S., & Kameda, T. (2000). 'Social sharedness' as a unifying theme for information processing in groups. *Group Processes & Intergroup Relations*, *3*(2), 123–140. https://doi.org/10.1177/1368430200003002002

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review*, *117*(2), 440–463. https://doi.org/10.1037/a0018963

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

Van Swol, L. M. (2011). Forecasting another's enjoyment versus giving the right answer: Trust, shared values, task effects, and confidence in improving the acceptance of advice. *International Journal of Forecasting*, *27*(1), 103–120. https://doi.org/10.1016/j.ijforecast.2010.03.002

Wang, H., & Wellman, B. (2010). Social connectivity in America: Changes in adult friendship network size from 2002 to 2007. *American Behavioral Scientist*, *53*(8), 1148–1169. https://doi.org/10.1177/0002764209356247

Wegener, D. T., Petty, R. E., Detweiler-Bedell, B. T., & Jarvis, W. G. (2001). Implications of attitude change theories for numerical anchoring: Anchor plausibility and the limits of anchor effectiveness. *Journal of Experimental Social Psychology*, *37*(1), 62–69. https://doi.org/10.1006/jesp.2000.1431

Wellman, B. (2012). Is Dunbar's number up? *British Journal of Psychology*, *103*(2), 174–176. https://doi.org/10.1111/j.2044-8295.2011.02075.x

Yaniv, I. (2004a). 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. (2004b). The benefit of additional opinions. *Current Directions in Psychological Science*, *13*(2), 75–78. https://doi.org/10.1111/j.0963-7214.2004.00278.x

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: Egocentric discounting and reputation formation. *Organizational Behavior and Human Decision Processes*, *83*(2), 260–281. https://doi.org/10.1006/obhd.2000.2909

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

A Full Multilevel Models of Expectation Effects on Advice Taking

Table A.1: Full Multilevel Model of Final Judgment According to Equation 2.10 (for $k = 1$) With Fixed Treatment Effects of Contrast-Coded Advice Expectation Condition (C) as Specified in Equation 3.1 for Experiment 1 of Rebholz and Hütter (2022)

	Estimate	*95% CI*	*SE*	*t*	*df*	*p*
β_A	0.3550	0.3195 – 0.3905	0.0179	19.7849	126.4680	**<0.001**
$\beta_{A\times C}$	-0.0471	-0.1124 – 0.0182	0.0331	-1.4221	206.5732	0.157
σ	0.6762	0.6590 – 0.6932				
$\tau_{A,S}$	0.2138	0.1902 – 0.2421				
$\tau_{A,T}$	0.0273	0.0018 – 0.0428				
ICC	0.3139					
N	200					
M	16					
Obs.	3200					
R^2_m / R^2_c	0.46 / 0.63					

Note. Bold values indicate $p < 0.05$. Wald 95% CI for fixed and bootstrap 95% CI (with 1,000 iterations) for random effects are shown.

Table A.2: Full Multilevel Model of Final Judgment According to Equation 2.10 (for $k = 1$) With Fixed Treatment Effects of Contrast-Coded Advice Expectation Condition (C) as Specified in Equation 3.1 for Experiment 2 of Rebholz and Hütter (2022)

	Estimate	*95% CI*	*SE*	t	*df*	p
β_A	0.3338	0.2867 – 0.3809	0.0230	14.5335	27.1690	**<0.001**
$\beta_{A\times C}$	0.0173	-0.0332 – 0.0678	0.0257	0.6741	294.5911	0.501
σ	0.6494	0.6352 – 0.6627				
$\tau_{A,S}$	0.1902	0.1701 – 0.2125				
$\tau_{A,T}$	0.0760	0.0466 – 0.1051				
ICC	0.2771					
N	292					
M	16					
Obs.	4672					
R^2_m / R^2_c	0.34 / 0.52					

Note. Bold values indicate $p < 0.05$. Wald 95% CI for fixed and bootstrap 95% CI (with 1,000 iterations) for random effects are shown.

Table A.3: Full Multilevel Model of Final Judgment According to Equation 2.10 (for $k = 1$) With Fixed Treatment Effects of Contrast-Coded Advice Expectation Condition (C) as Specified in Equation 3.1 for Experiment 3 of Rebholz and Hütter (2022)

	Estimate	*95% CI*	*SE*	t	*df*	p
β_A	0.3334	0.2786 – 0.3883	0.0275	12.1407	66.9021	**<0.001**
$\beta_{A\times C}$	0.0374	0.0065 – 0.0683	0.0158	2.3718	1865.4814	**0.018**
σ	0.5314	0.5133 – 0.5495				
$\tau_{A,S}$	0.2117	0.1806 – 0.2426				
$\tau_{A,T}$	0.0764	0.0449 – 0.1063				
ICC	0.3509					
N	119					
M	20					
Obs.	1904					
R^2_m / R^2_c	0.38 / 0.60					

Note. Bold values indicate $p < 0.05$. Wald 95% CI for fixed and bootstrap 95% CI (with 1,000 iterations) for random effects are shown.

Table A.4: Full Multilevel Model of Final Judgment According to Equation 2.10 (for $k = 1$) With Fixed Treatment Effects of Contrast-Coded Advice Expectation Condition (C) as Specified in Equation 3.1 for Experiment 4 of Rebholz and Hütter (2022)

	Estimate	*95% CI*	*SE*	*t*	*df*	*p*
β_A	0.2346	0.2066 – 0.2626	0.0138	16.9614	38.5401	**<0.001**
$\beta_{A \times C}$	0.0273	0.0109 – 0.0437	0.0084	3.2691	4467.5799	**0.001**
σ	0.1566	0.1534 – 0.1598				
$\tau_{A,S}$	0.1325	0.1192 – 0.1461				
$\tau_{A,T}$	0.0479	0.0310 – 0.0660				
ICC	0.1958					
N	297					
M	20					
Obs.	4752					
R^2_m / R^2_c	0.27 / 0.41					

Note. Bold values indicate $p < 0.05$. Wald 95% CI for fixed and bootstrap 95% CI (with 1,000 iterations) for random effects are shown.

Table A.5: Full Multilevel Model of Final Judgment According to Equation 2.10 (for $k = 1$) With Fixed Treatment Effects of Contrast-Coded Advice Expectation Condition (C) as Specified in Equation 3.1 for Experiment 5 of Rebholz and Hütter (2022)

	Estimate	*95% CI*	*SE*	*t*	*df*	*p*
β_A	0.5908	0.5249 – 0.6566	0.0289	20.4361	8.5866	**<0.001**
$\beta_{A \times C}$	-0.0193	-0.0606 – 0.0219	0.0210	-0.9193	1104.7700	0.358
σ	0.2800	0.2757 – 0.2845				
$\tau_{A,S}$	0.2991	0.2813 – 0.3167				
$\tau_{A,T}$	0.0762	0.0336 – 0.1149				
ICC	0.2687					
N	1111					
M	8					
Obs.	8888					
R^2_m / R^2_c	0.42 / 0.58					

Note. Bold values indicate $p < 0.05$. Wald 95% CI for fixed and bootstrap 95% CI (with 1,000 iterations) for random effects are shown.

Table A.6: Full Multilevel Model of Final Judgment According to Equation 2.10 (for $k = 1$) With Fixed Treatment Effects of Dummy-Coded Advice Expectation Condition (C) as Specified in Equation 3.1 for Experiment 5 of Rebholz and Hütter (2022)

	Estimate	*95% CI*	*SE*	*t*	*df*	*p*
β_A	0.6004	0.5328 – 0.6680	0.0307	19.5853	10.8574	**<0.001**
$\beta_{A\times C}$	-0.0193	-0.0606 – 0.0219	0.0210	-0.9193	1104.7700	0.358
σ	0.2800	0.2757 – 0.2845				
$\tau_{A,S}$	0.2991	0.2813 – 0.3167				
$\tau_{A,T}$	0.0762	0.0336 – 0.1149				
ICC	0.2687					
N	1111					
M	8					
Obs.	8888					
R^2_m / R^2_c	0.42 / 0.58					

Note. Bold values indicate $p < 0.05$. C is dummy-coded such that β_A and $\beta_{A\times C}$ measure the effect of low expectations to receive advice and the difference to high expectations, respectively. Wald 95% CI for fixed and bootstrap 95% CI (with 1,000 iterations) for random effects are shown.

Table A.7: Full Multilevel Model of Final Judgment According to Equation 2.10 (for $k = 1$) With Fixed Treatment Effects of Reverse Dummy-Coded Advice Expectation Condition (C) as Specified in Equation 3.1 for Experiment 5 of Rebholz and Hütter (2022)

	Estimate	*95% CI*	*SE*	*t*	*df*	*p*
β_A	0.5811	0.5133 – 0.6489	0.0309	18.8271	11.1541	**<0.001**
$\beta_{A\times C}$	0.0193	-0.0219 – 0.0606	0.0210	0.9193	1104.7700	0.358
σ	0.2800	0.2757 – 0.2845				
$\tau_{A,S}$	0.2991	0.2813 – 0.3167				
$\tau_{A,T}$	0.0762	0.0336 – 0.1149				
ICC	0.2687					
N	1111					
M	8					
Obs.	8888					
R^2_m / R^2_c	0.42 / 0.58					

Note. Bold values indicate $p < 0.05$. C is reverse dummy-coded such that β_A and $\beta_{A\times C}$ measure the effect of high expectations to receive advice and the difference to low expectations, respectively. Wald 95% CI for fixed and bootstrap 95% CI (with 1,000 iterations) for random effects are shown.

B Full Multilevel Models of Adaptive Sequential Advice Seeking

Table B.1: Full Multilevel Model of Final Judgment According to Equation 4.1 With Fixed Linear Order Effects as Specified in Equation 4.2 and Fixed Treatment Effects of Contrast-Coded Advice Distance Condition (C) for Experiment 5 of Ache (2017)

	Estimate	*95% CI*	*SE*	*t*	*df*	*p*
β_A	0.4539	0.3953 – 0.5126	0.0296	15.3164	123.1511	**<0.001**
$\beta_{A\times C}$	0.2865	0.2637 – 0.3093	0.0116	24.6495	14485.3134	**<0.001**
$\beta_{A\times k}$	0.0024	0.0013 – 0.0036	0.0006	4.2948	14507.8709	**<0.001**
σ	0.1452	0.1436 – 0.1470				
$\tau_{A,S}$	0.2721	0.2394 – 0.3071				
$\tau_{A,T}$	0.0706	0.0497 – 0.0937				
ICC	0.5451					
N	127					
M	20					
Obs.	14563					
R^2_m / R^2_c	0.71 / 0.87					

Note. Bold values indicate $p < 0.05$. Wald 95% CI for fixed and bootstrap 95% CI (with 1,000 iterations) for random effects are shown.

Table B.2: Full Multilevel Model of Final Judgment According to Equation 4.1 With Fixed Nonlinear Order Effects as Specified in Equation 4.2 (Incl. Second-Order Polynomial k^2) and Fixed Treatment Effects of Contrast-Coded Advice Distance Condition (C) for Experiment 5 of Ache (2017)

	Estimate	*95% CI*	*SE*	*t*	*df*	*p*
β_A	0.4420	0.3831 – 0.5009	0.0298	14.8422	126.0176	**<0.001**
$\beta_{A \times C}$	0.2859	0.2632 – 0.3087	0.0116	24.6072	14484.4634	**<0.001**
$\beta_{A \times k}$	0.0081	0.0047 – 0.0115	0.0017	4.6791	14495.2890	**<0.001**
$\beta_{A \times k^2}$	-0.0003	-0.0005 – -0.0001	0.0001	-3.4587	14465.6738	**0.001**
σ	0.1452	0.1435 – 0.1470				
$\tau_{A,S}$	0.2712	0.2387 – 0.3060				
$\tau_{A,T}$	0.0707	0.0497 – 0.0938				
ICC	0.5437					
N	127					
M	20					
Obs.	14563					
R^2_m / R^2_c	0.72 / 0.87					

Note. Bold values indicate $p < 0.05$. Wald 95% CI for fixed and bootstrap 95% CI (with 1,000 iterations) for random effects are shown.

Table B.3: Full Multilevel Model of Final Judgment According to Equation 4.1 With Fixed Nonlinear Order Effects as Specified in Equation 4.2 (Incl. Second-Order Polynomial k^2) and Linear and Logarithmic Fixed Effects of Relative Absolute Advice Distance (D) for Experiment 5 of Ache (2017).

	Estimate	*95% CI*	*SE*	*t*	*df*	*p*
β_A	0.9798	0.7402 – 1.2194	0.1222	8.0166	10062.8655	**<0.001**
$\beta_{A \times k}$	0.0081	0.0046 – 0.0116	0.0018	4.5901	14498.3975	**<0.001**
$\beta_{A \times k^2}$	-0.0004	-0.0005 – -0.0002	0.0001	-3.6156	14467.8781	**<0.001**
$\beta_{A \times D}$	-0.4093	-0.6516 – -0.1671	0.1236	-3.3125	14499.4553	**0.001**
$\beta_{A \times log(D)}$	0.4602	0.1185 – 0.8020	0.1743	2.6396	14488.2857	**0.008**
σ	0.1481	0.1464 – 0.1499				
$\tau_{A,S}$	0.2691	0.2367 – 0.3035				
$\tau_{A,T}$	0.0695	0.0486 – 0.0922				
ICC	0.5297					
N	127					
M	20					
Obs.	14563					
R^2_m / R^2_c	0.71 / 0.86					

Note. Bold values indicate $p < 0.05$. Wald 95% CI for fixed and bootstrap 95% CI (with 1,000 iterations) for random effects are shown.

C Acknowledgments

This dissertation was only possible with the valuable support of many people who have accompanied me over the past three years.

First and foremost, I am grateful to Mandy for being a motivating and inspiring mentor who challenged me to grow beyond myself. From her, I learned the importance of concise theorizing for conducting substantive research and how to design and run informative experiments. She helped me find my place in science.

I would also like to thank Andreas and Arndt for being my second and third supervisors and for giving me room to present and critically discuss my work. Moreover, I would like to thank Andreas and Marco for being fantastic collaborators and coauthors of the second and third manuscripts, respectively.

Fortunately, I had the opportunity to be part of the SMiP research training group during my dissertation. Many instructive workshops, enlightening retreats, and informal meetings helped me gain a deeper understanding of the connection between methodology and substantive research, and gave me the chance to establish prolific collaborations. It was a great pleasure to be part of this group together with Marcel, Julian, Maren, Annika, Parker, Robert, Lukas, Julius, Viola, Marcel, Tong, David, Simon, Martin, Fabiola, Constantin, and all other SMiPsters.

My time in Tübingen would not have been the same without my colleagues Marco, Max, Zach, Thomas, Kathrin, Johanna, Bruno, Birka, Niels, and the rest of the Social Cognition and Decision Sciences group. Thank you for making me feel welcome from day one, for the insightful conversations at lunch and during coffee breaks, for scientific collaborations and encouragement, and for the recreational activities on other joyful occasions. Special thanks goes to Zach for proofreading this thesis.

Outside of the office, I found devoted affection, unconditional support, and the necessary balance to work and science in my family and friends. I cannot emphasize enough how grateful I am to my parents for making all this possible in the first place.

Last but not least, I am eternally indebted to Lea for her enduring backing on this path to the doctorate as well as in all other past and future challenges in our life. She is an irreplaceable companion and my most precious source of inspiration.

D Copies of Manuscripts

D.1 Manuscript I

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

Judgment and Decision Making, Vol. 17, No. 4, July 2022, pp. 816–848

The advice less taken: The consequences of receiving unexpected advice

Tobias R. Rebholz* Mandy Hütter†

Abstract

Although new information technologies and social networks make a wide variety of opinions and advice easily accessible, one can never be sure to get support on a focal judgment task. Nevertheless, participants in traditional advice taking studies are by default informed in advance about the opportunity to revise their judgment in the light of advice. The expectation of advice, however, may affect the weight assigned to it. The present research therefore investigates whether the advice taking process depends on the expectation of advice in the judge-advisor system (JAS). Five preregistered experiments (total $N = 2019$) compared low and high levels of advice expectation. While there was no evidence for expectation effects in three experiments with block-wise structure, we obtained support for a positive influence of advice expectation on advice weighting in two experiments implementing sequential advice taking. The paradigmatic disclosure of the full procedure to participants thus constitutes an important boundary condition for the ecological study of advice taking behavior. The results suggest that the conventional JAS procedure fails to capture a class of judgment processes where advice is unexpected and therefore weighted less.

Keywords: advice taking, expectation, judge-advisor system, wisdom-of-the-crowd

1 Introduction

Sometimes it is easy to get support from other people; at other times it can be difficult to find someone who is competent or willing to give advice. Decision problems generally

*Psychology Department, Eberhard Karls University of Tübingen, Schleichstr. 4, 72076 Tübingen, Germany, Email: tobias.rebholz@uni-tuebingen.de, https://orcid.org/0000-0001-5436-0253.

†Psychology Department, Eberhard Karls University of Tübingen, https://orcid.org/0000-0002-0952-3831.

This research was funded by the Deutsche Forschungsgemeinschaft (DFG), grant 2277, Research Training Group "Statistical Modeling in Psychology" (SMiP), and a Heisenberg grant (HU 1978/7-1) awarded to Mandy Hütter.

All data and supplementary materials are available at https://osf.io/bez79.

do not come with all the necessary details to be solved outright. Instead, decision-makers usually engage in building the relevant information bases themselves. As social beings, we often turn to others for their help when we feel uncertain about something. Although new information technologies and social networks make a wide variety of opinions and advice easily accessible, advice taking is still fraught with a high degree of uncertainty. Uncertainty about the information sampling process, whether it concerns the competency of potential advisors or the likelihood of getting any support at all, adds to the uncertainty of the decision problem.

Advice taking is typically studied in the dyadic judge-advisor system (JAS; Bonaccio & Dalal, 2006). As introduced by Sniezek and Buckley (1995), the judge (or advisee) is first asked to give an initial estimate about the unknown true value of a stimulus item. Thereafter, he or she is to render a final estimate in the light of passively presented or actively sampled pieces of information from one or multiple advisors (e.g., Fiedler et al., 2019; Hütter & Ache, 2016; Soll & Larrick, 2009). That is, there is little uncertainty with regard to the information sampling process: Participants are fully aware that they will get the opportunity to revise their initial estimate in the light of external support later in the experiment. This paradigmatic feature is generally neglected in JAS-type studies (for reviews see Bonaccio & Dalal, 2006; Rader et al., 2017).

Research on the effects of unsolicited advice has approached the paradigmatic sampling uncertainty from a decision autonomy perspective. Goldsmith and Fitch (1997) found that autonomy concerns are driven by the degree of (explicit) solicitation of advice. In turn, advice taking intentions (Van Swol et al., 2017; Van Swol et al., 2019) and behaviors are affected (Brehm, 1966; Fitzsimons & Lehmann, 2004; Gibbons et al., 2003; Goldsmith & Fitch, 1997). However, differences in expectation of advice do not necessarily impose differences with respect to decision autonomy: Advice can be equally (un)solicited with or without expecting to receive it. In unsolicited advice taking research, by contrast, being aware of either the opportunity to explicitly solicit advice or the possibility of receiving unsolicited advice, the judge generally expects advice (Gibbons et al., 2003). We thus deem our study of the role of advice expectation complementary to this line of research.

The research at hand posits that an ecological approach to advice taking should take the uncertainty about the external information sampling process into account. To this end, we systematically investigate the effects of advice expectation on quantitative judgment. We thereby assume that the expectation of advice is inextricably linked to the expectation of an opportunity to revise initial judgments. In the following, we elaborate on our perspective of advice taking making a distinction between expected and unexpected advice on the one hand, and between revisable and non-revisable judgments on the other hand (see Bullens & van Harreveld, 2016, for a review on reversible vs. irreversible decisions).

1.1 The Role of Expectation in Advice Taking

Participants who are aware of taking part in an advice taking experiment are naturally aware of the fact that their initial estimates will not be taken as their final say about a particular estimation problem. Essentially, it is very likely that this knowledge about the experimental procedure influences the cognitive processes involved in forming initial and final judgments and thereby the generated estimates and behaviors in those experiments. Put differently, the advance procedural information induces a certain mindset that may influence the impact of advice based on two complementary mechanisms. The first mechanism concerns the generation of initial judgments. Under the expectation that additional evidence can be acquired and incorporated, initial estimates may be made in a provisional manner (see Önkal et al., 2009, for similar influences of the expected source of advice). That is, participants may not apply the same scrutiny to both estimation stages. If one expects to receive additional information, one may not invest as much time and effort to come up with a precise, high-quality estimate, but rather make a rough guess. We assume that someone who invested lots of effort into coming up with an estimate will adopt advice less readily than someone who gave a rough, provisional estimate.

Additionally, the weighting of advice may depend on an assimilative versus contrastive mindset. Ongoing mental tasks should trigger relatively more assimilative processing as compared to tasks that were already completed. That is, we deem the expectation of advice to increase the likelihood that it is accepted as a relevant piece of information that can inform the final judgment. Initial support for our assumptions stems from previous research that has documented stronger assimilative effects of the prime on the target in an evaluative priming paradigm when the processing of the prime was not completed (e.g., by categorizing it as positive or negative before the target is presented; Alexopoulos et al., 2012). Thus, keeping the mental task incomplete leads to stronger priming effects. We believe that similar effects can be expected for the processing of advice in the JAS. As long as participants have not finalized their judgment, they are relatively more open to integrating additional information than when they provided an estimate that they consider final. Expecting advice thus increases the likelihood that advice is included in the universe of pieces of information relevant to form a final judgment.

By contrast, once the mental task was completed and people came up with their final estimate, being presented with a piece of advice may evoke a tendency to defend their own position rather than to adapt it towards the advice. The JAS paradigm thereby relates to research into decision revisability, and thus, cognitive dissonance (Festinger, 1957). If a revision opportunity was not expected, cognitive dissonance may arise (Knox & Inkster, 1968). In research on non-revisable discrete choice, for instance, it was found that the positive aspects of the chosen option remain particularly accessible (e.g., Knox & Inkster, 1968; Liberman & Förster, 2006), in line with the notion that one's views are restructured to be consistent with a decision's outcome (Bullens et al., 2011). If the same effect occurs in the advice taking paradigm, participants with lower expectation of a revision opportunity should reduce post-decisional dissonance by assigning a higher likelihood to their initial estimate being correct. Indeed, previous research shows that greater weight is assigned to judgments of higher quality (Yaniv & Kleinberger, 2000) and to more competent judges (Harvey & Fischer, 1997), especially if it is the self who is perceived higher in expertise (Harvey & Harries, 2004). Reduced weighting of unexpected advice would accordingly be the result of an efficient means to cope with potentially dissonant feelings about the initial, supposedly non-revisable judgment.

1.2 Expectation Effects on Weighting, Accuracy, and Internal Sampling

In line with this reasoning, we expect advice taking to differ between expectation conditions as follows: Weighting of advice is lower in the condition with relatively lower expectation of advice than in the conventional JAS-type condition with relatively higher expectation of advice (Hypothesis 1) due to corresponding instructions. Because advice weights are generally rather small (Harvey & Fischer, 1997; Yaniv & Kleinberger, 2000), this difference may be sufficiently large (and detectable) only in advice distance regions for which comparably high advice weighting and higher variance can be expected (Moussaïd et al., 2013).

As summarized in the notion of the wisdom-of-the-crowd, higher weighting implies an accuracy advantage for the final estimate, not only with advice of high quality but also with advice from non-expert peers due to the sheer increase of the information base (Davis-Stober et al., 2014; Soll & Larrick, 2009). Therefore, the decline in judgment error from initial to final estimation is expected to be attenuated by lower expectation of advice (Hypothesis 2). This assumption is contingent on the reduced weighting of unexpected advice (see Hypothesis 1).

Our reasoning also inspires a minor prediction regarding the quality of the initial estimate, beyond its provisional nature (as argued above). Initial estimates are generated by aggregating various internal viewpoints from internal (Thurstonian) sampling (Juslin & Olsson, 1997; Sniezek & Buckley, 1995; Thurstone, 1927). Internal samples may contain, for instance, sequentially recalled memories or self-constructed feedback (Fiedler & Kutzner, 2015; Henriksson et al., 2010; Stewart et al., 2006). One may thus argue that internal sampling is in the same manner affected by advice expectation as external sampling. In particular, internal samples drawn under the expectation of advice may integrate broader perspectives than internal samples generated without expecting to receive advice from another person (see also Trope & Liberman, 2010). Importantly, the Thurstonian notion relies on random or quasi-random internal sampling (Fiedler & Juslin, 2006). In line with the law of large numbers, we thus expect both less extreme (Hypothesis 3a) and less noisy (Hypothesis 3b) initial estimates when advice is expected.[1]

[1]We preregistered undirected versions of these hypotheses for Experiments 1 and 3. The here discussed

In sum, our aim is to test whether initial estimation and advice weighting, and thereby also the judgment accuracy, depend on the expectation of advice. If support for these hypotheses was found, previous results obtained from JAS-type experiments would have to be reassessed, because indices of advice taking would be inherently biased in conventional JAS paradigms. For instance, "egocentric discounting" (i.e., the propensity to weight one's own judgment more strongly than advice) is observed in large parts of the advice taking literature (Harvey & Fischer, 1997; Yaniv & Kleinberger, 2000). If the expectation of advice indeed artifactually inflates advice weighting in JAS-type experiments, the egocentrism issue would be even more severe than assumed.

2 General Method

We report how we determined our sample sizes, all data exclusions (if any), all manipulations, and all measures[2] for all experiments. All experiments were preregistered. Unless stated otherwise, sample size, manipulations, measures, data exclusions, and analyses adhere to the preregistration. Preregistration documents, materials, surveys, data, analysis scripts, and the online supplement are publicly available at the Open Science Framework (OSF; https://osf.io/bez79).

Across five experiments, we implemented slight variations of the general JAS procedure. For each item, participants provided initial point estimates, received a single piece of advice (presented alongside their own initial estimate) and were given the opportunity to provide a final, possibly revised estimate. The operationalization of advice expectation varied across experiments but was either low or high by means of instruction. There are three dependent variables of interest, the weight of advice (WOA), judgment error (JE), and the (extremity and variance of the) initial estimates. For all experiments, we conducted multilevel modeling for all dependent variables (Baayen et al., 2008; Bates et al., 2015). All models comprise random intercepts for participants and items which were fully crossed by design. Expectation condition was included as a contrast-coded fixed effect, with lower expectation coded as –0.5 and higher expectation as 0.5, for the random intercepts to capture effects in both conditions (Judd et al., 2017). The fixed effect of condition thus indicates the consequences of conventionally high expectation of advice in the JAS. Significance was assessed at the 5% level via one-sided (where justified) p-values computed based on Satterthwaite's (1941) approximation for degrees of freedom in linear models (Luke, 2017); and based on Wald Z-testing in nonlinear models (Bolker et al., 2009). Additionally, Bayes factors comparing the expectation model against the null using the default settings of

directed version was preregistered only for Experiment 4.

[2]We measured level of construal (Trope & Liberman, 2010) in the first two experiments by means of a short version of the Behavioral Identification Form (Vallacher & Wegner, 1989) in Experiment 1 and the Navon (1977) task in Experiment 2. As we neither found treatment effects on "general mindset abstractness" (Krüger et al., 2014) nor on perceptual level of construal (as operationalized by "global dominance;" Liberman & Förster, 2009), the analyses will not be discussed for the sake of brevity.

Makowski et al.'s (2019) bayestestR package are reported to resolve the inconclusiveness of potential null effects.

The measure of our major concern, the amount of advice weighting, is typically calculated as the ratio of judgment shift and advice distance (Bonaccio & Dalal, 2006). We used the most common formalization of Harvey and Fischer (1997) such that advice weighting was measured in percentage points:[3]

$$WOA_{ij} = \frac{FE_{ij} - IE_{ij}}{A_{ij} - IE_{ij}} * 100, \tag{1}$$

where FE_{ij}, IE_{ij}, and A_{ij} indicate the final and initial estimates and advice, respectively, on a given item j in a given participant i. As the WOA is highly sensitive to outliers, we relied on Tukey's (1977) fences to identify and remove outliers on a trial-by-trial basis. For testing of Hypothesis 1, we fitted the following linear multilevel model:

$$WOA_{ij} = \beta_0 + \alpha_i^P + \alpha_j^S + \beta_1 Expectation_{ij} + \varepsilon_{ij}, \tag{2}$$

where subindex i and superscript P refer to participants, subindex j and superscript S to stimulus items, β to fixed effects, α to random effects such that $\mathrm{Var}\left(\alpha_i^P\right) = \sigma_P^2$ and $\mathrm{Var}\left(\alpha_i^S\right) = \sigma_S^2$, and ε to the overall error term. The same formal notation applies to all models throughout.

The WOA distributions were thereupon descriptively explored beyond their measures of central tendency. This is expedient with reference to the findings of Soll and Larrick (2009) who disclosed important systematics in WOA dependent on the level of data aggregation. Specifically, an actual W-shaped distribution of WOA — advice taking consisting of a mixture of strategies including (equal weights) averaging and choosing (the advisor or the self) — is often analytically concealed by focusing on mean differences.

We applied the same statistical criteria as for WOA to identify and remove judgment outliers on a trial-by-trial basis for the calculation of judgment error.[4] For both initial and final estimates, judgment error in percentage points was formally defined as:[5]

$$JE_{ijt} = \frac{\left|T_j - Estimate_{ijt}\right|}{T_j} * 100, \tag{3}$$

where T_j corresponds to the j-th item's true value, $|.|$ denotes the absolute value function, and

$$Estimate_{ijt} = \begin{cases} IE_{ij}, & t = -0.5 \\ FE_{ij}, & t = 0.5 \end{cases} \tag{4}$$

[3]Multiplying the ratio-of-differences formula with a fixed constant of 100 for reporting reasons was only preregistered for Experiment 5 but does not affect the statistical significance of the results.

[4]The statistical significance of the main results was neither affected by removing outliers on the level of judgment nor on the level of judgment error.

[5]Deviating from the preregistrations and multiplying the judgment errors with a fixed constant of 100 for reporting reasons does not affect the statistical significance of the results.

depending on at which point in time t the error is evaluated. Hypothesis 2 can accordingly be tested by adding a time-series interaction to the respective expectation model:

$$JE_{ijt} = \beta_0 + \alpha_i^P + \alpha_j^S + \beta_1 Expectation_{ij} + \beta_2 t + \beta_3 Expectation_{ij} * t + \varepsilon_{ijt}. \quad (5)$$

For the α coefficients to capture random effects of both points in time — in the same vein as for both levels of advice expectation as discussed above — t was contrast-coded with the initial judgment phase as –0.5 and the final one as 0.5 (Judd et al., 2017).

To account for extensive differences in truth (e.g., the highest true value in the stimulus sets for Experiments 1 to 3 as introduced below was 34,000, whereas the lowest one was 0.48), initial estimates were normalized as follows prior to analyzing them:

$$NIE_{ij} = \frac{IE_{ij}}{T_j} \quad (6)$$

(Moussaïd et al., 2013). In line with the extremity/noise foundation of Hypotheses 3a and 3b, no exclusion criteria were applied to the normalized initial estimates. Following the recommendations of Lo and Andrews (2015), instead of transforming the response itself, a generalized multilevel model with log-link on the Gamma distribution was implemented to account for positively skewed estimates. Accordingly, the model for testing of Hypothesis 3a can be written as:

$$NIE_{ij} = \exp\left(\beta_0 + \alpha_i^P + \alpha_j^S + \beta_1 Expectation_{ij} + \varepsilon_{ij}\right), \quad (7)$$

where exp(.) denotes the exponential function. Explicit testing of the variance part (Hypothesis 3b) was preregistered for the fourth experiment to be based on a Fligner-Killeen (median) test of variance homogeneity on the log-transformed values. This test does not allow for one-sided hypothesis testing but is comparably robust against departures from normality (Conover et al., 1981). Both extremity and noise results were corroborated by two-sample Kolmogorov-Smirnov testing (not preregistered). That is, compound testing of both hypotheses took place by usage of the complete distributional information to check whether the normalized initial estimates follow the same sampling distributions in both expectation conditions.

3 Experiment 1

Experiment 1 was designed to delineate the juxtaposition of the condition with the conventional full expectation of advice and a less informed group of participants that does not expect to receive advice. There were no further restrictions on the advice stimuli. This allowed us to explore typically reported patterns of advice taking over the entire distance scale, that is, the inverse U-shaped relation between WOA and advice distance that peaks in the region of intermediately distant advice values with its corresponding effects on judgment accuracy (Schultze et al., 2015).

3.1 Method

3.1.1 Design and Participants

A 2 (advice expectation: yes vs. no) × 2 (judgment phase: initial vs. final) × 2 (dissonance measure: administered vs. not administered) mixed design with repeated measures on the second factor was implemented. The experiment was conducted online with the link distributed via the general mailing list of the University of Tübingen. In compensation for a median duration of 20.08 minutes (IQR = 7.50), participants could take part in a raffle for five €20 vouchers of a German grocery chain. More accurate estimates (±25% around the true value) were rewarded with additional raffle tickets. Participants were informed that their participation is voluntary, and that any personal data will be stored separate from their experimental data. At the end of the experiment, they were debriefed and thanked.

We conducted a-priori power analyses to determine the required sample sizes in all five experiments. Power analyses focused on Hypothesis 1, that is, on detecting treatment effects on WOA. For the first experiment, we based our calculation on repeated measures ANOVA designs (Faul et al., 2007). Detecting a small effect (d = 0.30) with sufficient power ($1–\beta$ = 0.80) required collecting data of at least 188 participants. Based on our expectations about the exclusion rate, we preregistered collecting data of 209 participants. At that point, the exclusion rate turned out to be more than twice as high than expected (21% vs. 10%). We therefore did not stop recruiting participants until we had reached a sample of size N = 250 to make up for the additional exclusions. After applying the preregistered exclusion criteria, we ended up with a final sample of N = 200 (123 female, 76 male, 1 diverse). Their median age was 25 years (IQR = 7.25).

3.1.2 Materials and Procedure

Participants were asked to estimate the Product Carbon Footprint (PCF) in kilograms of carbon dioxide equivalents (kg-CO2e), a classic measure to quantify the ecological life cycle efficacy of products.[6] For that purpose, they were presented with pictures of products most of which were taken from the database of Meinrenken et al. (2020; https://carboncatalogue.coclear.co). In order to introduce a higher variability in product categories, additional stimulus items from other sources were included as well. To calibrate participants, they were provided with background knowledge and went through a practice phase of three trials with feedback about the true values at the beginning of the experiment. Moreover, only those 16 of 50 products for which the participants of a pretest[7] performed best on median

[6]Given that our estimation task was bounded by zero from below, it involved a proportionally higher chance for over- as opposed to underestimation (positive skew) — not necessarily of the true value but relative to the total sample mean. Extremity (Hypothesis 3a) is accordingly characterized by higher estimate values due to the interplay of the law of large numbers and the positive skew in the underlying judgment domain.

[7]The pretest was conducted online via the general mailing list of the University of Tübingen. We collected PCF estimates and confidence intervals of N = 107 participants (73 female, 33 male, 1 diverse) on 50 items (plus three items on practice trials). Participants' median age was 24 years (IQR = 4.00). The pretest data and

were used to ensure the existence of a wise crowd.

Between-participants manipulations of advice expectation required a blocked design. Participants were asked to provide all initial point estimates as well as lower and upper bounds building an 80% confidence interval for the full set of items in the first block. In the second block, they received a single piece of advice (i.e., "the judgment of a randomly selected previous participant") presented alongside their own initial estimate and were asked to give a final, possibly revised estimate and confidence. Stimulus items were presented in the same order across blocks, which was randomized across participants. Participants in the conventional JAS-type condition were informed about the revision phase in which they will be provided with advice prior to the initial estimation block. By contrast, participants who did not expect to receive advice were informed that they will estimate the PCF of products in the first block of the experiment without further notice of the second block. Only upon completing the initial judgment phase, the opportunity to adjust their initial estimate given a single piece of advice in a second judgment phase was revealed to them. In order to provide ecological advice, the median estimates and interquartile ranges from the pretest determined the location and spread of the truncated (at the admissible response range from 0.001 to 999999.999) normal distributions from which the artificial advisory estimates were drawn.

After the practice phase, we administered an instructional manipulation check. Participants were asked to indicate how often they will make an estimate for a certain product. Those who responded incorrectly (18.40%) were excluded from the analysis as preregistered. Many participants misunderstood the question and responded with the total number of items to be judged (i.e., 16), or completely unrelated values (e.g., 10). We nevertheless carried out exclusions as planned in Experiment 1 (and results do not change if we deviate from the preregistration), but we clarified instructions and did not preregister to carry out exclusions based on instructional manipulation checks in later experiments.

3.2 Results

A summary of the fixed effects of the multilevel models for Experiment 1 is given in Table 1. Means and standard deviations by expectation condition are presented in Table 2. The full models and model comparison statistics can be found in Table S1 of the online supplement.

3.2.1 WOA

We excluded trials with a WOA < -77.17 and WOA > 137.25 (Tukey, 1977). In total, we excluded 142 of 3200 trials (4.44%). For testing of Hypothesis 1 that advice weighting is lower for participants who did not expect to receive advice, we fitted the multilevel model of WOA on contrast-coded advice expectation as defined in Equation 2. The fixed effect of

materials can be found on the OSF repository.

TABLE 1: Fixed effects (and standard errors) of multilevel models of weight of advice (WOA), judgment error (JE), and normalized initial estimates (NIE) on contrast-coded advice expectation for all five experiments. The full models and model comparison statistics can be found in the online supplement.

		Predictor	Expt. 1		Expt. 2		Expt. 3		Expt. 4		Expt. 5	
WOA	β_0	Intercept	31.33	***	39.92	***	39.09	***	20.69	***	54.76	***
			(1.45)		(1.41)		(1.82)		(1.28)		(2.59)	
	β_1	Expectation	–0.99		–1.63		4.62	**	2.29	***	0.23	
			(2.70)		(2.28)		(1.60)		(0.66)		(1.65)	
JE	β_0	Intercept	92.74	***	121.50	***	73.58	***	43.40	***	55.89	***
			(4.41)		(6.00)		(2.15)		(2.88)		(1.93)	
	β_1	Expectation	4.31		–7.36		1.65		–0.24		–0.73	
			(4.70)		(6.30)		(1.90)		(0.40)		(1.28)	
	β_2	t	–26.77	***	–47.61	***	–10.71	***	–4.78	***	–8.67	***
			(2.43)		(3.12)		(1.88)		(0.40)		(0.28)	
	β_3	Expectation	–2.98		5.46		–0.44		–0.38		–0.32	
		* t	(4.86)		(6.24)		(3.77)		(0.80)		(0.56)	
NIE	β_0	Intercept	2.75	***	4.02	***	2.80	***	0.65	***	0.49	***
			(0.46)		(0.67)		(0.64)		(0.05)		(0.04)	
	β_1	Expectation	1.09		0.68	*	1.15	***	1.00		1.01	
			(0.22)		(0.12)		(<0.01)		(0.01)		(0.07)	

Note. Two-sided p-values with * $p < .05$, ** $p < .01$, *** $p < .001$.

expectation thus indicated the consequences of receiving expected advice. Advice expectation had no significant effect on participants' WOA (β_1 = –0.99, 95% CI [–6.28, 4.30], SE = 2.70, d = –0.03, t(198.31) = –0.37, p = .643, BF_{10} = 0.131).

3.2.2 Accuracy

We merged the two block-separated hypotheses about judgment error from the preregistration into one joint accuracy shift hypothesis (Hypothesis 2). We excluded 15.94% of trials based on either normalized initial or final estimates being outliers (Tukey, 1977) and fitted the multilevel model as defined in Equation 5. The significant reduction in judgment error from initial to final estimation (β_2 = –26.77, 95% CI [–31.54, –22.01], SE = 2.43, d = –0.28, t(5151.88) = –11.01, p < .001) indicated collectively beneficial advice weighting as expected. The negative trend did however not significantly interact with expectation (β_3 = –2.98, 95% CI [–12.51, 6.54], SE = 4.86, d = –0.03, t(5151.88) = –0.61, p = .270,

Table 2: Means (and standard deviations) of weight of advice (WOA), judgment error (JE), and normalized initial estimates (NIE) by expectation condition in all five experiments.

	Phase	Expectation	Expt. 1	Expt. 2	Expt. 3	Expt. 4	Expt. 5
WOA		low	31.84	40.77	36.62	19.43	54.33
			(34.16)	(37.87)	(37.23)	(25.49)	(39.62)
		high	30.68	39.14	41.41	21.93	54.83
			(34.04)	(38.46)	(39.27)	(26.92)	(38.31)
JE	initial	low	100.07	142.99	77.11	46.11	61.46
			(105.34)	(185.98)	(60.13)	(24.90)	(27.00)
		high	106.37	134.65	78.97	46.70	60.98
			(117.52)	(178.42)	(65.47)	(25.04)	(26.85)
	final	low	74.79	92.65	66.62	41.51	52.95
			(71.25)	(95.08)	(42.81)	(25.22)	(28.02)
		high	78.11	89.77	68.05	41.73	52.16
			(74.42)	(90.10)	(43.19)	(25.37)	(28.59)
NIE		low	14.24	31.91	10.64	0.76	1.01
			(90.42)	(266.89)	(56.46)	(0.82)	(2.41)
		high	10.36	11.26	26.72	0.76	1.13
			(74.28)	(74.18)	(462.50)	(0.86)	(3.16)

BF_{10} = 0.049). Hence descriptively, the decline in judgment error from initial to final estimation was stronger with expectation of advice as expected, but this difference fell short of statistical significance.

3.2.3 Initial Belief Formation

Normalized initial estimates were modeled by multilevel gamma models with log-link as defined in Equation 7 (Lo & Andrews, 2015). We did not exclude any outliers to capture the hypothesized extremity/noise patterns in initial estimation. The fixed effect of contrast-coded advice expectation failed to reach statistical significance (β_1 = 1.09, 95% CI [0.74, 1.61], d = 0.03, SE = 0.22, t = 0.46, p = .677, BF_{10} = 0.020; Hypothesis 3a).[8] Neither did Fligner-Killeen testing of variance homogeneity, $\widehat{\sigma}^2_{low}$ = 4.37, $\widehat{\sigma}^2_{high}$ = 3.65, χ^2_{FK} (1) =

[8]In line with the reporting conventions for nonlinear models with log-link, we report exponentiated coefficients here and in the following. The coefficient of contrast-coded expectation thus being significantly different from 1 corresponds to a rejection of the null hypothesis of equally extreme initial estimates across expectation conditions. Coefficients on the original scale of the model in Equation 7 can be retrieved by taking the log of the reported values.

2.46, p = .117 (Hypothesis 3b), nor two-sample Kolmogorov-Smirnov testing, D = 0.03, p = .350, support differences in initial belief formation.

3.2.4 Post-hoc Analyses

Dissonance Thermometer. About half of the participants received parts of the so-called dissonance thermometer, a self-report measure of affect that asks participants to reflect on their current feelings on 7-point scales (Devine et al., 1999; Elliot & Devine, 1994), between initial and final judgment. The dissonance thermometer (contrast-coded with presence as 0.5 and absence as –0.5) significantly interacted with our expectation manipulation (β_3 = –11.84, 95% CI [–22.18, –1.50], SE = 5.28, d = –0.35, t(196.16) = –2.24, p = .026, BF_{10} = 10.909; online supplement, Table S2, left panel). As such, reflecting on their feelings made participants in the low-expectation condition take significantly more advice (β_1 = 12.81, 95% CI [5.59, 20.04], SE = 3.69, d = 0.38, t(195.98) = 3.48, p = .001; Table S2, middle panel). Essentially, the dissonance thermometer is criticized for not only measuring, but most likely also reducing dissonance (Martinie et al., 2013). Hence, if cognitive dissonance was indeed induced by the alleged non-revisability of initial judgments in the low-expectation group, it might have been reduced by filling out the dissonance thermometer, in turn, increasing advice weighting. In contrast, there was little evidence for an effect of the dissonance thermometer in the high-expectation condition (β_1 = 0.98, 95% CI [–6.42, 8.37], SE = 3.77, d = 0.03, t(196.34) = 0.26, p = .796; Table S2, right panel). We thus accounted for this influence by considering the dissonance thermometer as an additional factor in the following analyses.

Advice Distance. The effect of expectation on WOA was descriptively opposite to our hypothesis (β_2 = –1.28, 95% CI [–6.45, 3.89], SE = 2.64, d = –0.04, t(196.16) = –0.49, p = .687; Table S2, left panel). However, advice taking is typically found to vary with the distance of advice from a participant's initial beliefs (e.g., Hütter & Ache, 2016; Schultze et al., 2015). In particular, weighting is most pronounced for advice of "intermediate distance" as categorized by Moussaïd et al. (2013) and flattening out for both closer and more distant values (Figure 1). The same advice distance region was characterized by most pronounced differences in WOA across expectation conditions. This visual impression was confirmed by building multilevel models that took the advice distance categorization from the literature into account (online supplement, Table S3): For participants who did not fill out the dissonance thermometer, weighting of unexpected advice of intermediate distance was significantly lower (β_6 = 7.06, 95% CI [0.68, 13.43], SE = 3.25, d = 0.21, t(3009.72) = 2.17, p = .015, BF_{10} = 0.113). Given that participants received ecological advice from the pretest, significantly reduced advice weighting should have impaired judgment accuracy (Davis-Stober et al., 2014; Soll & Larrick, 2009). Nevertheless, we found no significant attenuation of the reduction in judgment error (β_{14} = –19.57, 95% CI [–45.95, 6.80], SE = 13.46, d = –0.22, t(5137.13) = –1.46, p = .073, BF_{10} = 0.004).

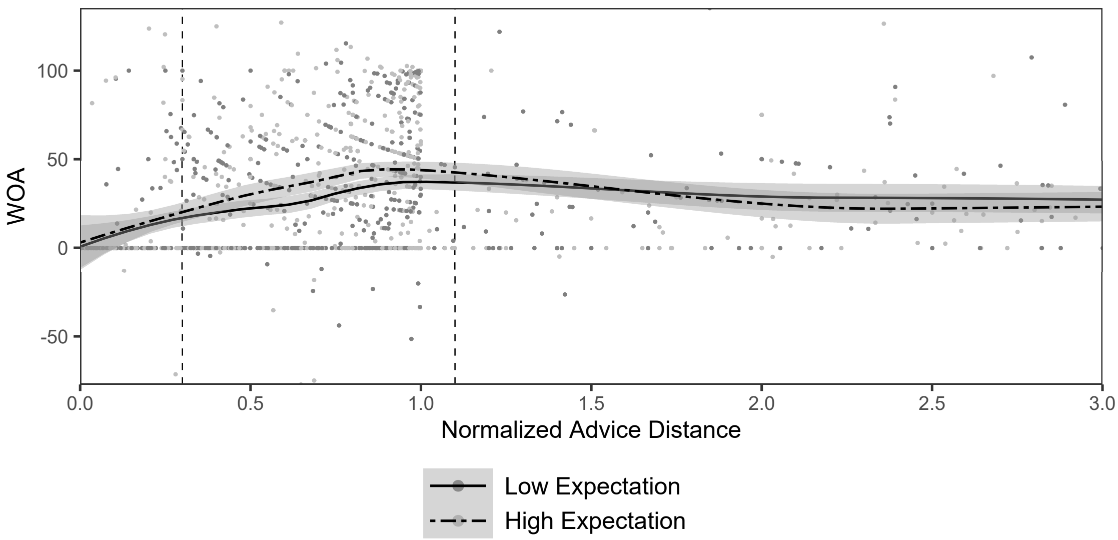


Figure 1: Scatter plots and local polynomial regression fits (incl. 95% confidence bands) for WOA by advice distance as a function of advice expectation in the condition without the dissonance thermometer (N = 98) of Experiment 1. The area enclosed by the thin dashed vertical lines indicates advice of intermediate normalized distance (Moussaïd et al., 2013). Plotting is truncated for outliers of WOA (Tukey, 1977) and normalized advice distance larger than 3.

3.3 Discussion

In the data set that considers all levels of advice distance, we did not obtain evidence for an influence of advice expectation. This was shown to be partly due to the influence of the dissonance thermometer. Moreover, the exploratory post-hoc analysis for advice of intermediate distance provides good reasons for a distance-qualified investigation of the proposed expectation effects on advice weighting. For advice of intermediate distance, there was a significant reduction in advice weighting of around seven percentage points in the low-expectation condition. Unfortunately, as assessed by means of simulation (Green & MacLeod, 2016; see also below), this post-hoc test lacked power ($1–\beta = 0.59$, 90% CI [0.56, 0.62]). These limitations will be addressed in Experiment 2.

4 Experiment 2

In Experiment 2, we focused on the intermediate distance region which typically exhibits highest WOA. That is, advice was neither too close nor too distant from a participant's initial estimate. Experiment 2 was thus designed to enable a confirmatory, sufficiently powered version of the post-hoc analysis of Experiment 1.

4.1 Method

4.1.1 Design and Participants

A 2 (advice expectation: yes vs. no) × 2 (judgment phase: initial vs. final) mixed design with repeated measures on the second factor was implemented. The experiment was again conducted online, and the link was distributed via the general mailing list of the University of Tübingen. In compensation for a median duration of 22.71 (*IQR* = 9.79) minutes, participants could take part in a raffle for five €10 vouchers of a German bookstore chain and receive course credit. More accurate estimates (±25% around the true value) were rewarded with additional raffle tickets. Moreover, one tree per complete participation was donated to the Trillion Tree Campaign (https://trilliontreecampaign.org). Participants were informed that their participation is voluntary, and that any personal data will be stored separate from their experimental data. At the end of the experiment, they were debriefed and thanked.

We assumed a smaller effect size of $d = 0.25$ in Experiment 2 due to regression to the mean for replications on the one hand (Fiedler & Prager, 2018), and the reduced variation in advice distance and hence supposedly less diagnostic external information on the other hand. Moreover, we utilized data from the preceding experiment to conduct a-priori power analysis for multilevel modeling by means of simulation (Green & MacLeod, 2016). Based on 1000 iterations, sufficient power (95% confidence that $1–\beta \geq 0.80$) required at least N = 284 participants. The experiment was preregistered to automatically stop recruitment when the last required participant with valid data reached the final page. As further participants could have entered and start working on the experiment at that point, a sample of size N = 292 (209 female, 81 male, 2 diverse) was eventually recruited. Those participants' median age was 23 years (*IQR* = 8.00).

4.1.2 Materials and Procedure

The procedure of Experiment 2 resembled Experiment 1 with three exceptions. First, the critical instructions in the low-expectation condition mentioned the existence of a second part of the experiment without providing specific information on the task. Second, we omitted the dissonance thermometer. Third, the experiment focused on the region of advice distance where the exploratory post-hoc analyses of Experiment 1 revealed significant treatment effects on WOA. The critical region of intermediate distance corresponds to the region for which advice weighting is typically reported to peak if examined dependent on advice distance (Hütter & Ache, 2016; Moussaïd et al., 2013; Schultze et al., 2015). The mechanism attempted to generate an intermediately distant value from a truncated normal distribution as specified by the pretest parameters for a maximum of 1000 times. This corresponds to an alleged drawing of advisors from a hypothetical pretest sample of corresponding size. If no congenial advisor could be drawn, that is, no intermediately distant advice value could be generated upon reaching this threshold, a fallback mechanism randomly generated an intermediately distant value without drawing from the distributions as defined by the pretest parameters. Participants who received fallback advice at least once (6.07%) were preregistered to be not counted towards the final sample size and to be excluded from the analysis.

4.2 Results

A summary of the fixed effects of the multilevel models for Experiment 2 is given in Table 1 with the corresponding means and standard deviations by expectation condition as presented in Table 2. The full models and model comparison statistics can be found in Table S4 of the online supplement.

4.2.1 WOA

We excluded trials with a WOA < –100.81 and WOA > 172.27 (Tukey, 1977). In total, we excluded 82 of 4672 trials (1.76%). The fixed effect of expectation indicated the consequences of receiving unexpected advice. The effect was descriptively opposite to our prediction, that is, expected advice was slightly less taken, but this effect failed to reach statistical significance (β_1 = –1.63, 95% CI [–6.11, 2.84], SE = 2.28, d = –0.04, t(290.29) = –0.71, p = .763, BF_{10} = 0.109).

4.2.2 Accuracy

The lack of an effect of advice expectation on advice weighting once more anticipates the results of the judgment accuracy analysis. We excluded 17.79% of trials based on either normalized initial or final estimates being outliers (Tukey, 1977). The significant reduction in judgment error from initial to final estimation (β_2 = –47.61, 95% CI [–53.73, –41.50], SE = 3.12, d = –0.33, t(7336.50) = –15.26, p < .001) did not depend on advice expectation (β_3 = 5.46, 95% CI [–6.77, 17.69], SE = 6.24, d = 0.04, t(7336.50) = 0.88, p = .809, BF_{10} = 0.093). Although the sign of the interaction is consistent with the WOA effects of opposite direction than expected, the results do not support Hypothesis 2.

4.2.3 Initial Belief Formation

Normalized initial estimates were modeled by multilevel gamma models with log-link (Lo & Andrews, 2015). We did not exclude trials of normalized initial estimates to capture the hypothesized extremity/noise patterns. The significant fixed effect of contrast-coded advice expectation (β_1 = 0.68, 95% CI [0.49, 0.96], SE = 0.12, d = –0.13, t = –2.20, p = .014, BF_{10} = 0.162) was evident in favor of treatment effects on (mean) initial belief formation. Initial estimation was more extreme with lower expectation of advice as expected (Hypothesis 3a). Moreover, the Fligner-Killeen test indicated significantly higher variance (“noise”) in the initial estimates of the low-expectation group ($\widehat{\sigma}^2_{low}$ = 4.66, $\widehat{\sigma}^2_{high}$ = 3.75, χ^2_{FK} (1) = 15.68, p <

.001; Hypothesis 3b). The results were corroborated by two-sample Kolmogorov-Smirnov testing which suggested significant differences in the sampling distributions of the groups' initial estimates ($D = 0.06$, $p < .001$).

4.2.4 Post-hoc Analysis Beyond the Means

Once more, there was no evidence for an WOA effect of practical importance. If so, it would have even been in the opposite direction. This second descriptive reversal led us to explore this null effect more deeply. It is possible that factually distinctive advice taking behavior was concealed by focusing on mean differences. For instance, egocentric discounting is a consequence of taking the means across a mixture of averaging and choosing strategies (Soll & Larrick, 2009). While many people actually follow the normative rule of equal weights *averaging* (Mannes, 2009), a non-negligible amount of people prefers to *choose* one of both sources of information (WOA = 0 or WOA = 100).

The reverse pattern from the aggregate analysis of Experiment 2 also materialized on the disaggregate level (Figure 2). Unexpectedly, there was a slightly higher share of trials where advice was not used at all in the high-expectation condition and a relatively left-skewed averaging distribution centered at equal weighting in favor of the low-expectation group. Overall, however, the characteristic W-shaped WOA distributions were fairly congruent across advice expectation conditions. Under the conditions of the post-hoc analysis of Experiment 1, by contrast, the expectation effect is accrued by a reduction in the propensity to stick to one's initial judgment (WOA = 0) in favor of a relatively more left-skewed averaging distribution in the high-expectation group. Across all experiments, there was least evidence for the hypothesized effect of advice expectation on WOA in Experiment 2.

4.3 Discussion

Evidence from the post-hoc analysis of Experiment 1, which indicated significant treatment effects on the weighting of advice of intermediate distance, could not be corroborated. There is no additional support for an effect on WOA of practical importance given presence versus absence of advice expectation. The null effects on WOA may be due to our modification of the traditional paradigm which implemented initial and final estimates in two blocks in order to enable the between-participants manipulation of advice expectation. Although Experiment 2 lent support to advice expectation effects on internal sampling (Hypotheses 3a and 3b: more extreme and more noisy initial estimates in the low-expectation group), this effect failed to extend to external sampling in the second estimation phase. This limitation will be addressed in Experiment 3 that dissolves the blocked design.

The circumstances of Experiment 2 allow some speculation as to why the effect on WOA was descriptively opposite to our prediction both on the aggregate as well as on the disaggregate level. This outcome may be attributed to the incentives announced for participation, namely, the donation of trees. Thereby, we may have inadvertently recruited

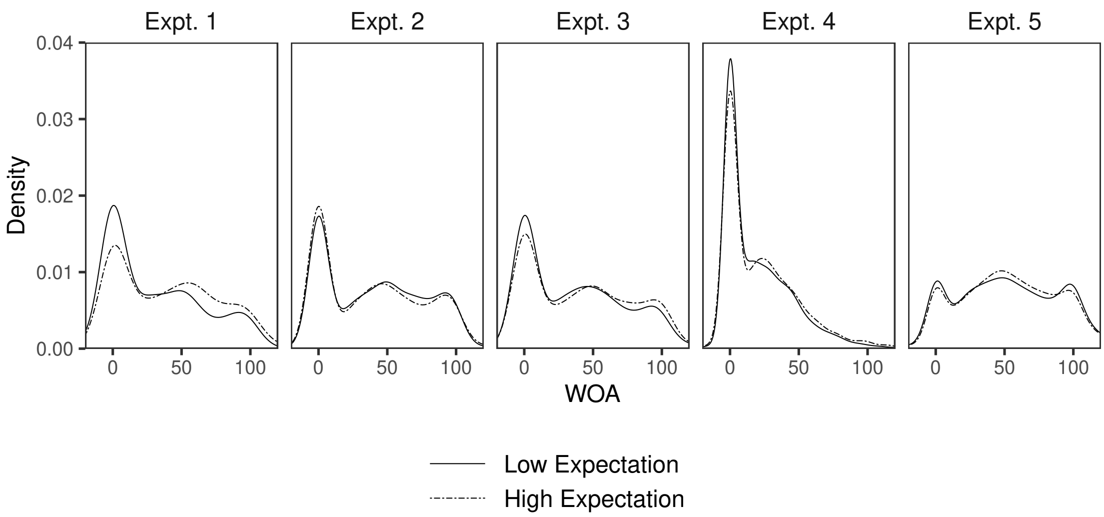


FIGURE 2: Gaussian kernel density plots of WOA (outliers excluded) as functions of advice expectation in all five experiments. The bandwidth is chosen according to Silverman's (1986) rule of thumb. For Experiment 1, the conditions which yielded positive results post-hoc (i.e., without the dissonance thermometer, N = 98, and for advice of intermediate distance) are shown.

a sample which held relatively high believes about their own competencies for the life cycle assessment of consumer products compared to the average recipient of our invitation. Such an eco-conscious sample is supposedly less reluctant to advice on the given judgment domain (i.e., PCF) such that participants' advice taking behavior may be less sensitive to our manipulation. Therefore, we switched to monetary compensation in Experiment 3.

5 Experiment 3

The blocked design which was necessary to implement between-participants manipulations in the previous two experiments is a nontrivial component of the original paradigm. For instance, participants may have had doubts as to whether the values marked as initial estimates in the final judgment phase were actually their own, thereby affecting their advice taking behavior (Soll & Mannes, 2011). This could explain why we only obtained a significant treatment effect on initial estimates in Experiment 2. Moreover, the blocked design may be incompatible with the notion of ongoing mental tasks. Specifically, the succession of initial estimates might force participants to mentally close the preceding task in order to focus on the current one, thus not affecting the assimilative processing between expectation conditions. These issues were addressed by switching to a within-participants manipulation and thereby to a sequential version of the paradigm in Experiment 3.

5.1 Method

5.1.1 Design and Participants

This experiment implemented a 2 (advice expectation: high vs. low) × 2 (judgment phase: initial vs. final) within-participants design with repeated measures on both factors. This time, participants were recruited via Prolific (https://prolific.co). Median monetary compensation amounted to £5.75 per hour for an experiment with median duration of 17.43 minutes (*IQR* = 4.41). Additionally, participants could take part in a raffle for three £10 Amazon vouchers. More accurate estimates (±25% around the true value) were rewarded with additional raffle tickets. Participants were informed that their participation is voluntary, and that any personal data will be stored separate from their experimental data. At the end of the experiment, they were debriefed, thanked, and redirected to Prolific for compensation.

Although the within-participants operationalization of advice expectation may make the manipulation particularly salient (i.e., changing from trial to trial), somewhat smaller effects ($d = 0.125$) on the dependent measure are expected from a less extreme difference on the probabilistic dimension (see below). A-priori power simulation was based on the data from Experiment 2. Due to the more powerful within-participants design, at least N = 109 participants were required to reach sufficient power (95% confidence that $1–\beta \geq 0.80$). Whereas advice was provided on a fixed set of 16 items per participant in Experiments 1 and 2, it was provided on random 16 of 20 items per participant in Experiment 3. Therefore, we preregistered to aim for 20% more data than needed ($N = 131$) to guarantee sufficient power. Based on our expectations about the exclusion rate, we preregistered collecting data of 150 participants. Prolific eventually recruited 151 participants. After applying the preregistered exclusion criteria, we ended up with a final sample of $N = 119$ participants (57 female, 58 male, 4 diverse). Their median age was 24 years (*IQR* = 7.50).

5.1.2 Materials and Procedure

In Experiment 3, advice expectation was manipulated within-participants. For that purpose, the block-structure present in Experiments 1 and 2 was resolved. For each item, participants first gave their initial estimate, then received advice and gave their final estimate before they continued on to the next item. Moreover, the operationalization of low and high levels of expectation was less extreme than in the previous experiments. Participants were informed that the probability of receiving advice on the next product will be “80% (i.e., very likely)” on nine high- versus “20% (i.e., very unlikely)” on eleven low-expectation trials. This information was provided at the beginning of each trial. In fact, they received advice of intermediate distance on eight of nine and eleven “advice trials,” respectively. On the remaining four “no-advice trials,” neither did they receive advice, nor did they get the opportunity to provide a final estimate. Instead, they directly continued with initial estimation of the next product. Confidence ratings were elicited on extremum-labeled (*very unconfident* to *very confident*) sliders to avoid confounding of the expectation manipulation with secondary probabilistic instructions. Advice was again of intermediate distance (Moussaïd et al., 2013).

The procedure was intended to reflect real-world uncertainty in judgment and decision making in an ecological setup. To that effect, the global imbalance of advice and no-advice trials was implemented to increase participants' trust in the diagnosticity of the information about advice probability, while ensuring a balanced set of advice trials per participant and condition for the main analysis.[9] The assignment of products to advice versus no-advice and high versus low-expectation trials was fully random. We added the next four best items from the pretest to our selection of estimation tasks to include four no-advice trials on top of the 16 advice trials per participant.

5.2 Results

A summary of the fixed effects of the multilevel models for Experiment 3 is given in Table 1 with the corresponding means and standard deviations by expectation condition as presented in Table 2. The full models and model comparison statistics can be found in Table S5 of the online supplement. Moreover, see Figure 2 for the WOA distributions.

5.2.1 WOA

We excluded no-advice trials and trials with a WOA < –100.00 and WOA > 171.43 (Tukey, 1977). In total, we excluded 26 of 1904 advice trials (1.37%). We fitted the multilevel model of WOA on contrast-coded advice expectation with –0.5 for low and 0.5 for high expectation of advice. The fixed effect of expectation thus indicated the consequences of higher advice expectation. WOA was now significantly reduced on low-expectation trials (β_1 = 4.62, 95% CI [1.48, 7.76], SE = 1.60, d = 0.12, t(1755.44) = 2.88, p = .002, BF_{10} = 5.899). Moreover, the Bayes factor indicates moderate evidence for Hypothesis 1.

5.2.2 Accuracy

We excluded no-advice trials and 17.96% of advice trials based on either normalized initial or final estimates being classified as outliers according to Tukey's (1977) fences. The significant reduction in judgment error from initial to final estimation (β_2 = –10.71, 95% CI [–14.40, –7.02], SE = 1.88, d = –0.20, t(2971.45) = –5.69, p < .001) did not significantly interact with expectation (β_3 = –0.44, 95% CI [–7.82, 6.94], SE = 3.77, d = –0.01, t(2971.45) = –0.12, p = .454, BF_{10} = 0.021).

[9]Extending Equation 2 by a fixed effect of trial number $s = 1, \ldots, 20$ and its interaction with contrast-coded expectation (see also Equation 5), there is no significant effect of time on WOA in Experiment 3 (β_s = –0.36, 95% CI [–0.75, 0.03], SE = 0.20, d = –0.01, t(1787.02) = –1.80, p = .072, BF_{10} = 0.002) and Experiment 4 (β_s = –0.08, 95% CI [–0.24, 0.09], SE = 0.08, d = 0.00, t(4477.05) = –0.90, p = .367, BF_{10} < 0.001). That is, there is no evidence for advice taking changing over the course of the two experiments. This suggests that participants either did not notice the implemented mismatch in expectation versus outcome or construed it as a merely less representative personal outcome of the communicated probabilities.

5.2.3 Initial Belief Formation

Normalized initial estimates were modeled by multilevel gamma models with log-link (Lo & Andrews, 2015). We neither excluded no-advice trials nor any outliers to capture the hypothesized extremity/noise patterns. Opposite to our prediction, the fixed effect of contrast-coded advice expectation indicated more extreme initial estimates on trials in which advice was expected ($\beta_1 = 1.15$, 95% CI [1.14, 1.15], $SE < 0.01$, $d = 0.04$, $t = 81.66$, $p > .999$, $BF_{10} = 0.233$). However, those judgment extremity results (Hypothesis 3a) of unexpected direction could not be corroborated by the judgment noise results (Hypothesis 3b) as the Fligner-Killeen test did not support differences in initial estimation variance ($\widehat{\sigma}^2_{low} = 4.26$, $\widehat{\sigma}^2_{high} = 4.46$, $\chi^2_{FK}(1) = 0.47$, $p = .494$). Moreover, a two-sample Kolmogorov-Smirnov test did not allow rejecting the null of indifferent sampling distributions ($D = 0.03$, $p = .603$).

5.3 Discussion

The results of Experiment 3 indicate less (rather than more) extreme initial estimates on low-expectation trials (Hypothesis 3a). However, due to conflicting evidence from variance (Hypothesis 3b) and distributional shape testing, we deem this analysis inconclusive. Importantly, consistent with our expectation, on trials characterized by low expectation the amount of advice weighting was significantly reduced in Experiment 3 (Hypothesis 1). One reason why this effect may have failed to affect estimation accuracy (Hypothesis 2) is that there was least relative improvement — and hence room for expectation effects — in judgment error from initial to final estimation in Experiment 3: As derived from the coefficients of the JE-models in Table 1, relative accuracy improvement across both advice expectation conditions was only 13.57% in Experiment 3 whereas it amounted to 25.22% and 32.77% in Experiments 1 and 2, respectively.

Another explanation lies in the stimulus material used in the first three experiments. Assessment of judgment accuracy largely depends on the true values of the items (see Equation 3) and so do the results of the respective analyses. Admittedly, differences in laws and (international) standards make the objective quantification of PCFs as selected for the estimation tasks quite complex. In the database from which most products were taken, 70% of the footprints were determined by using three different PCF standards (Meinrenken et al., 2020). For another 21%, the standard used was not specified. Accordingly, stimulus quality can be improved by switching to an easier, more tangible judgment domain with less problematic objective ground truth. For instance, Galton (1907), who first documented the wisdom-of-the-crowd phenomenon, analyzed the estimates of an ox's weight made by visitors of a country fair. We thus switched to a simpler, more accessible estimation task in Experiment 4.

6 Experiment 4

Experiment 4 constituted a higher-powered conceptual replication of Experiment 3 with different stimulus material for which the ground truth was less problematic than for PCFs. Thereby, we aim to provide generalized evidence for the existence of the hypothesized expectation effect and extend it to practical relevance in terms of judgment accuracy.

6.1 Method

6.1.1 Design and Participants

The experiment realized a 2 (advice expectation: high vs. low) × 2 (judgment phase: initial vs. final) within-participants design with repeated measures on both factors. The experiment was again conducted online. Participants were recruited via the general mailing list of the University of Tübingen. In compensation for a median duration of 20.03 (IQR = 7.07) minutes, participants could take part in a raffle for five €20 vouchers of a German bookstore chain and receive course credit. More accurate estimates (±25% around the true value) were rewarded with additional raffle tickets. Participants were informed that their participation is voluntary, and that any personal data will be stored separate from their experimental data. At the end of the experiment, they were debriefed and thanked.

We increased the threshold for sufficient power (95% confidence that $1-\beta \geq 0.95$), and — anticipating regression to the mean for the replicated effect (Fiedler & Prager, 2018) — based our simulations on a smaller effect size of d = 0.10. Power analysis resulted in N = 243. However, we observed that the power simulation results became more unstable for higher thresholds on a-priori power. Therefore, we preregistered to aim for 10% more data than needed (N = 270) to guarantee sufficient power for Experiment 4. The experiment was designed to automatically stop recruitment when the last required participant with valid data reached the final page. As further participants could have entered and start working on the experiment at that point, a sample of size N = 297 (180 female, 113 male, 4 diverse) was eventually recruited.[10] Participants' median age was 23 years (IQR = 8.00).

[10]To encourage English-speaking students to take part in the experiment, an English version was also administered. The data of 25 additional participants can be found on the OSF repository but were not included in the main analysis. Qualitatively, the results do not change if the analyses were carried out on the combined sample. Moreover, one participant from the German version was excluded due to a technical error.

6.1.2 Materials and Procedure

The procedure was identical to Experiment 3 except for the material used.[11] Participants were asked to provide estimates about the number of items in a pile of objects photographed against a white background (the stimuli can be found on the OSF). Twenty objects were chosen from several distinct categories: foods, toys, sanitary and household articles, and natural products. For instance, participants were asked to judge the number of breakfast cereals or thistles in a picture. The (integer) true values for those stimulus items ranged from 2,533 in the former example to 59 in the latter one. The exact number could not have been determined by counting for any of the 20 items. As the new material was not pretested, we could not use other persons' estimates to generate ecological advisory estimates. Instead, the advice values were randomly generated in accordance with the fallback mechanism of the previous experiments.[12] That is, intermediately distant values pointing in the direction of the true value were randomly drawn from uniform distributions.

6.2 Results

A summary of the fixed effects of the multilevel models for Experiment 4 is given in Table 1 with the corresponding means and standard deviations by expectation condition as presented in Table 2. The full models and model comparison statistics can be found in Table S6 of the online supplement. Moreover, see Figure 2 for the WOA distributions.

6.2.1 WOA

We excluded no-advice trials and trials with a WOA < –100.00 and WOA > 171.43.[13] In total, we excluded 25 of 4752 advice trials (0.53%). We fitted the multilevel model of WOA on contrast-coded advice expectation. The significant fixed effect of expectation once more indicated that WOA is significantly reduced on low-expectation trials (β_1 = 2.29, 95% CI

[11]In addition to the confidence question, participants were asked to indicate satisfaction with each of their estimates on extremum-labeled (*very unsatisfied* to *very satisfied*) sliders. This served a test of the cognitive dissonance account. For instance, assigning a higher likelihood to the correctness of a judgment just because it is perceived rather unlikely to change should be reflected in higher self-reported satisfaction. In contrast with this notion, there were no treatment effects on satisfaction. However, many participants reported that they were confused about the difference between confidence and satisfaction, so that the present results should not be interpreted.

[12]While the multipliers for initial estimates were randomly drawn from the intervals [0.0001, 0.7000] and [1.3000, 2.1000] to generate advice of intermediate distance in previous experiments, the lower bound of the former interval was changed to 0.4333 in Experiments 4 and 5 to yield more ecological advisory judgments of quantities.

[13]With the preregistered outlier detection according to Tukey's (1977) fences, the exclusion criteria would have been WOA < –51.46 and WOA > 88.24. With reference to Figure 2 and the results of Soll and Larrick (2009) who found that advice taking consists of a mixture of averaging and choosing strategies, it would be nontrivial to exclude trials which fall into one of those strategy categories. Instead, we applied the exclusion criteria from Experiment 3. Analysis based on the exclusion criteria as preregistered does not change the results qualitatively.

[0.99, 3.59], SE = 0.66, d = 0.09, t(4415.14) = 3.45, p < .001, BF_{10} = 9.369). The Bayes factor is on the verge of indicating strong evidence for Hypothesis 1.

6.2.2 Accuracy

As preregistered, we excluded no-advice trials and 6.36% of advice trials based on either normalized initial or final estimates being classified as outliers according to Tukey's (1977) fences.[14] The significant reduction in judgment error from initial to final estimation (β_2 = –4.78, 95% CI [–5.56, –4.00], SE = 0.40, d = –0.19, t(8580.22) = –12.02, p < .001) did not significantly interact with expectation (β_3 = –0.38, 95% CI [–1.94, 1.18], SE = 0.80, d = –0.01, t(8580.22) = –0.47, p = .318, BF_{10} < 0.001). Hence descriptively, the decline in judgment error from initial to final estimation was attenuated by the absence of advice expectation, but there still was decisive evidence against Hypothesis 2.

6.2.3 Initial Belief Formation

Normalized initial estimates were modeled by multilevel gamma models with log-link (Lo & Andrews, 2015). We neither excluded no-advice trials nor any outliers to capture the hypothesized extremity/noise patterns. The fixed effect of contrast-coded advice expectation was not significant (β_1 = 1.00, 95% CI [0.97, 1.03], SE = 0.01, d = 0.00, t = 0.09, p = .537, BF_{10} = 0.013). The Fligner-Killeen test also did not indicate differences in initial estimation variance ($\widehat{\sigma}^2_{low} = 0.56$, $\widehat{\sigma}^2_{high} = 0.59$, $\chi^2_{FK}(1) = 0.16$, $p = .686$), and the two-sample Kolmogorov-Smirnov test for differences in sampling distributions was insignificant as well (D = 0.02, p = .735). Hence, there was no evidence for treatment effects on initial belief formation (see also Footnote 14).

6.3 Discussion

Experiment 4 replicated the main finding of Experiment 3 with respect to advice weighting in a larger sample and with more power. Overall, we observed a strong reduction in weighting of advice, which was only about half as high than in the (intermediate conditions of) previous experiments. This outcome suggests that the amount of knowledge required by the task exerts an influence on advice weighting. There is much less — or even no — previous knowledge required for successfully completing a quantity estimation task than a PCF estimation task.[15] As a consequence, two mechanisms could explain generally lower advice weighting. First, the task may have been perceived as less difficult than in the previous experiments (Gino & Moore, 2007; Schrah et al., 2006). Second, it is very unlikely that

[14]This is less than half of the exclusion rates for all other experiments and most likely due to the participants having more practice in number of items estimation tasks than PCF estimation tasks which negatively affects the extremity/noise patterns of their estimates.

[15]Empirical backing for this assumption can be derived from Table 2. On average, NIE in Experiment 4 is much closer to 1 (indicating perfectly accurate initial judgment) than in the previous three experiments.

participants assumed a previous participant (the advisor) could have been better equipped to estimate the number of items (Harvey & Fischer, 1997; Sniezek & Buckley, 1995).

More important, the negative effect of unexpected advice on WOA (Hypothesis 1) is also significant with the new material. Nevertheless, the effect was small ($d = 0.09$), so that effects of advice expectation on accuracy (Hypothesis 2) were again not obtained. Therefore, this experiment too fails to corroborate the practical relevance of advice expectation — at least in terms of judgment accuracy from a wisdom-of-the-crowd perspective.

7 Experiment 5

The major difference between Experiments 1 and 2 on the one hand, and Experiments 3 and 4 on the other hand, concerns the block- versus trial-wise implementation of the estimation tasks. However, we introduced an additional change that complicates the interpretation of the differences between the two types of designs: a deterministic versus probabilistic manipulation of advice expectation. Therefore, this last experiment implements probabilistic expectation in a blocked design to differentiate between the influences of sequencing and extremity of expectation on our findings.

7.1 Method

7.1.1 Design and Participants

A 2 (advice expectation: high vs. low) × 2 (judgment phase: initial vs. final) mixed design with repeated measures on the second factor was realized. Participants were recruited via MTurk (https://www.mturk.com) with median monetary compensation (incl. up to \$0.40 bonus) of \$8.36 per hour for an experiment with median duration of 4.57 minutes (IQR = 2.37). More accurate estimates (±10% around the true value) were rewarded with \$0.05 bonus payment each. Participants gave their informed consent and were debriefed and thanked at the end of the experiment.

A-priori power simulation was conducted to detect treatment effects on WOA (Hypothesis 1) of the size from Experiments 3 and 4 combined ($d = 0.10$). Based on 1000 iterations, we preregistered collecting data of at least N = 1080 participants to reach sufficient power (95% confidence that $1–\beta \geq 0.80$). The experiment was again designed to automatically stop recruitment when the last required participant with valid data reached the final page. A sample of size N = 1111 (486 female, 618 male, 7 diverse) with median age of 38 years (IQR = 18.00) was eventually recruited.

7.1.2 Materials and Procedure

The procedure was similar to Experiment 2 with two major differences. First, we chose a new judgment domain from which estimation tasks were drawn. Second, advice expectation was manipulated in a probabilistic manner like in the within-participants designs in Experiments

3 and 4. Participants were told that the study comprised two groups. The "advice group" would be given advice and the opportunity to revise their initial judgment. The "solo group" would not receive advice and only form one judgment that is final. Participants in the high-expectation condition were told that the likelihood of being in the advice group is "80% (i.e., very likely)." In the low-expectation condition, this probability was stated as "20% (i.e., very unlikely)." In fact, all participants were assigned to the advice group to obtain the measures required for hypothesis testing. To make sure that participants read the relevant instructions, we preregistered spending less than five seconds on the respective page as an exclusion criterion.

Participants' task was to estimate the number of uniformly distributed, randomly colored squares in eight pictures (the stimuli and stimulus generation script can be found on the OSF; true values ranged from 527 to 11062 squares). We did not measure confidence in this experiment. As this task was again not very demanding in terms of knowledge (see also Footnote 15), we presented the stimuli for a maximum of ten seconds in both blocks to prevent the strong reduction in advice weighting as observed for quantity estimation in Experiment 4. The same random uniform, intermediately distant values pointing in the direction of the true value were provided as advice.

7.2 Results

A summary of the fixed effects of the multilevel models for Experiment 5 is given in Table 1 with the corresponding means and standard deviations by expectation condition as presented in Table 2. For the sake of brevity, we will only discuss the results for WOA (Hypothesis 1) here. However, the full models and model comparison statistics for all three dependent variables can be found in Table S7 of the online supplement. Moreover, see Figure 2 for the WOA distributions.

As preregistered, we excluded trials with a WOA < –67.11 and WOA > 179.11. In total, we excluded 274 of 8888 trials (3.08%). The multilevel regression of WOA on contrast-coded advice expectation did not yield evidence for reduced weighting in the low-expectation group (β_1 = 0.23, 95% CI [–3.01, 3.47], SE = 1.65, d = 0.01, $t(1091.14)$ = 0.14, p = .890, BF_{10} = 0.045). The size of the descriptively positive effect is negligible, and the Bayes factor even indicates strong evidence in favor of no differences in advice weighting across expectation conditions.

7.3 Discussion

There is no difference in weighting of unexpected and expected advice despite the probabilistic (i.e., less extreme) implementation of differences in expectation. Thus, the sequencing of judgments into blocks (Experiments 1, 2, and 5) versus trial-by-trial advice taking (Experiments 3 and 4) seems to be responsible for inconsistencies across between- and within-participants designs with respect to the weighting of unexpected advice in the results as reported thus far. Positive effects of expectation on weighting (Hypothesis 1) apparently are restricted to more ecological sequential judgment and expectation.

8 General Discussion

We set out to answer the question of whether peoples' judgment processes — specifically their advice weighting — depend on the expectation of advice prior to initial belief formation. In fact, in conventional JAS-type experiments participants can generally be sure to receive advice before providing a final, possibly revised judgment (for reviews see Bonaccio & Dalal, 2006; Rader et al., 2017). On a methodological dimension, the present project relates to the question of whether commonly reported levels of advice taking are bound to paradigmatic features of the JAS. In other words, we were interested in whether advice taking is robust towards variations in advice expectation.

We obtained support for the hypothesis that unexpected advice is less taken than expected advice (Hypothesis 1). For unexpected advice of intermediate distance as defined by Moussaïd et al. (2013), this effect was significant in the two sequential designs that manipulated advice expectation within-participants ($d = 0.12$ in Experiment 3 and $d = 0.09$ in Experiment 4) as well as in the post-hoc analysis of Experiment 1 ($d = 0.21$). However, two insignificant replications ($d = –0.04$ in Experiment 2 and $d = 0.01$ in Experiment 5) and the corresponding Bayes factors ($BF_{10} = 0.113$ in the post-hoc analysis of Experiment 1, $BF_{10} = 0.109$ in Experiment 2, and $BF_{10} = 0.045$ in Experiment 5) constitute rather strong evidence for a "reliable null effect" (Lewandowsky & Oberauer, 2020) of advice expectation on weighting in blocked designs that implement expectation manipulations between-participants.

Experiment 5 substantiated the null results' independence of the extremeness of expectations. Instead, segmenting the estimation process into separate blocks apparently suppresses expectation effects. At this point, we can only offer some speculations as to why this is the case. First, the blocked design might counteract the notion of ongoing mental tasks. Specifically, the succession of initial estimates might force participants to mentally close the preceding task in order to focus on the current one, thus not affecting the assimilative processing between expectation conditions. Second, blocked designs increase the temporal distance between the final and initial judgments. Relative to the initial judgment, advice is presented closer to the final judgment, potentially increasing the weight it receives in final judgments (Hütter & Fiedler, 2019). Thus, advice weighting in this version of the paradigm may profit less from the expectation of advice.

All experiments failed to give a clear indication of treatment effects on judgment accuracy (Hypothesis 2). There was no evidence that the overall significant decline in judgment error from initial to final estimation depends on advice expectation in any experiment. That is, participants expecting advice do not benefit from their significantly increased weighting of advice in Experiments 3 and 4. One reason may lie in the inherently problematic objective ground truth of product carbon footprints (Meinrenken et al., 2020) on which the judgment accuracy analysis in Experiment 3 relies. For both experiments, however, the generally small effects observed on the WOA counteract strong benefits in terms of wisdom-of-the-crowd.

Overall, we did not obtain support for Hypotheses 3a and 3b (no effects in Experiments 1, 4, and 5; positive effects in Experiment 2; mixed effects in Experiment 3). Consequently, there is currently no unequivocal evidence for effects of expecting advice on internal sampling, that is, on the way in which initial estimates are generated by aggregating various internal viewpoints (Juslin & Olsson, 1997; Sniezek & Buckley, 1995; Thurstone, 1927).

8.1 Limitations and Future Research

Our manipulation of advice expectation is naturally confounded with the expectation of an opportunity to revise one's estimate. Without an opportunity to revise their estimate, the judgment presented to participants could hardly serve as advice. Likewise, revising one's judgment is most useful if new information (e.g., in the form of advice) is considered. The present research thus cannot discern the effects of advice expectation proper and the mere revisability of one's estimate. Investigating this question requires an additional condition in which participants merely expect to revise their judgments at a second stage and are then surprised with advice. In such a condition, a post-decisional dissonance-based influence on advice weighting should be eliminated (Knox & Inkster, 1968). If advice expectation proper is responsible for the present effects, a difference should be observed between our high-expectation condition and the mere-expectation-of-revision condition with lower weighting of advice in the latter condition.

A fourth condition with high expectation of advice but low expectation of the opportunity to revise one's estimate would complete this more advanced design. Given that we found no unambiguous evidence for expectation effects on the initial estimates, however, such an additional condition likely provides insights only with respect to expectation effects on the weighting of (supposedly) mere "post-decisional feedback" (Zeelenberg, 1999). Advice taking in such a scenario thus closely relates to the literature on (performance) feedback acceptance. This opens up new research opportunities such as investigating the moderating role of self-efficacy in advice taking (Nease et al., 1999). In return, the WOA in this condition could enrich the literature with a well-studied *behavioral* measure of feedback acceptance (Bell & Arthur, 2008; Ilgen et al., 1979).

In our derivation of the hypotheses tested in the present research, we discussed possible mechanisms mediating the effect of our manipulation of advice expectation on advice weighting. The present research, however, does not provide evidence in support of these mechanisms. That is, the dissonance thermometer intended as a measure of cognitive dissonance in Experiment 1 instead seemed to affect advice weighting in the low-expectation condition.[16] Therefore, we focused the present research on our ontological claim regarding

[16]In line with the criticized potential of the dissonance thermometer to function as a coping mechanism

the existence of the effect rather than its mediation by certain cognitive processes. Future research should investigate the underlying mechanisms more carefully. Thereby, we would also gain a better understanding of the factors that influence the size of the expectation effect.

Theoretically, none of the delineated explanations requires dichotomous manipulations of expectation as implemented in the present research. For instance, cognitive dissonance is typically regarded as a continuum (Elliot & Devine, 1994). One might thus conceive of expectation as a continuous, probabilistic dimension of psychological experience. Accordingly, experimenting with randomly generated probabilities of advice receipt would be more informative (Cumming, 2014) and has the potential to enhance the ecology of the experimental setting that covers a broader spectrum of advice expectation. However, such an operationalization requires participants to memorize and use this information on a trial-by-trial basis, increasing the attentional demands of the experiment. Moreover, the analysis of the relationship between stated advice probability and advice weighting would have to account for the fact that humans generally do not construe probability as a linear dimension (Kahneman & Tversky, 1979).

Our experiments yielded first evidence for effects of advice expectation on a single advice taking measure. However, advice taking may serve additional functions and thereby extend to other measures. For instance, close advice may increase confidence and does not necessarily result in an adaptation of one's estimate, although it was assimilated to one's information base (Schultze et al., 2015). Thus, instead of restricting advice to be of intermediate distance, future research should investigate whether the paradigmatic expectation affects other dimensions of advice taking such as shifts in confidence or the sampling of external information. It would be worthwhile investigating whether participants would still actively sample unexpected advice (Hütter & Ache, 2016). According to our reasoning, we expect the effects documented for the WOA to extend to this measure, resulting in smaller sample sizes when participants did not expect to be able to sequentially sample advice.

8.2 Conclusion

Advice weighting can be reduced in situations in which it is unlikely that advice will be available compared to situations in which people expect to receive advice. Theories of advice taking should thus consider the role of advice expectation. Advice taking is likely less effective — in terms of the normative rule of equal weights averaging (Mannes, 2009) — if people are not prepared for it. As there is uncertainty about getting support in many real-life judgment situations, we recommend interpreting observed levels of advice weighting in light of the advice expectation conveyed by the experimental set-up.

(Martinie et al., 2013), this unidirectional confounding (see also online supplement, Table S2, middle vs. right panel) actually constitutes evidence for post-decisional cognitive dissonance (Knox & Inkster, 1968) indeed playing a role only in the low-expectation condition.

References

Alexopoulos, T., Fiedler, K., & Freytag, P. (2012). The impact of open and closed mindsets on evaluative priming. *Cognition and Emotion*, *26*(6), 978–994. https://doi.org/10.1080/02699931.2011.630991

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

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bell, S. T., & Arthur, W. (2008). Feedback acceptance in developmental assessment centers: The role of feedback message, participant personality, and affective response to the feedback session. *Journal of Organizational Behavior*, *29*(5), 681–703. https://doi.org/10.1002/job.525

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J.-S. S. (2009). Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology and Evolution*, *24*(3), 127–135. https://doi.org/10.1016/j.tree.2008.10.008

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

Brehm, J. W. (1966). *A theory of psychological reactance*. Academic Press.

Bullens, L., & van Harreveld, F. (2016). Second thoughts about decision reversibility: An empirical overview. *Social and Personality Psychology Compass*, *10*(10), 550–560. https://doi.org/10.1111/spc3.12268

Bullens, L., van Harreveld, F., & Förster, J. (2011). Keeping one's options open: The detrimental consequences of decision reversibility. *Journal of Experimental Social Psychology*, *47*(4), 800–805. https://doi.org/10.1016/j.jesp.2011.02.012

Conover, W. J., Johnson, M. E., & Johnson, M. M. (1981). A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data. *Technometrics*, *23*(4), 351–361. https://doi.org/10.1080/00401706.1981.10487680

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*(1), 7–29. https://doi.org/10.1177/0956797613504966

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

Devine, P. G., Tauer, J. M., Barron, K. E., Elliot, A. J., & Vance, K. M. (1999). Moving beyond attitude change in the study of dissonance-related processes. In E. Harmon-Jones & J. Mills (Eds.), *Cognitive dissonance: Progress on a pivotal theory in social psychology* (pp. 297–323). American Psychological Association. https://doi.org/10.1037/10318-012

Elliot, A. J., & Devine, P. G. (1994). On the motivational nature of cognitive dissonance: Dissonance as psychological discomfort. *Journal of Personality and Social Psychology*, *67*(3), 382–394. https://doi.org/10.1037/0022-3514.67.3.382

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.

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., & Juslin, P. (2006). Taking the interface between mind and environment seriously. In K. Fiedler & P. Juslin (Eds.), *Information sampling and adaptive cognition* (pp. 3–29). Cambridge University Press. https://doi.org/10.1017/CBO9780511614576.001

Fiedler, K., & Kutzner, F. (2015). Information sampling and reasoning biases. In G. Keren & G. Wu (Eds.), *The Wiley Blackwell handbook of judgment and decision making* (pp. 380–403). Wiley Blackwell. https://doi.org/10.1002/9781118468333.ch13

Fiedler, K., & Prager, J. (2018). The regression trap and other pitfalls of replication science — Illustrated by the report of the open science collaboration. *Basic and Applied Social Psychology*, *40*(3), 115–124. https://doi.org/10.1080/01973533.2017.1421953

Fitzsimons, G. J., & Lehmann, D. R. (2004). Reactance to Recommendations: When Unsolicited Advice Yields Contrary Responses. *Marketing Science*, *23*(1), 82–94. https://doi.org/10.1287/mksc.1030.0033

Galton, F. (1907). Vox Populi. *Nature*, *75*(1949), 450–451. https://doi.org/10.1038/075450a0

Gibbons, M. A., Sniezek, J. A., & Dalal, R. S. (2003, November 8–10). Antecedents and consequences of unsolicited versus explicitly solicited advice. In D. V. Budescu (Chair), *Symposium in honor of Janet Sniezek,* Society for Judgment and Decision Making Annual Meeting, Vancouver, BC.

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

Goldsmith, D. J., & Fitch, K. (1997). The normative context of advice as social support. *Human Communication Research*, *23*(4), 454–476. https://doi.org/10.1111/j.1468-2958.1997.tb00406.x

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

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

Hütter, M., & Ache, F. (2016). Seeking advice: A sampling approach to advice taking. *Judgment and Decision Making*, *11*(4), 401–415. https://doi.org/10.15496/publikation-13917

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

Ilgen, D. R., Fisher, C. D., & Taylor, M. S. (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology*, *64*(4), 349–371. https://doi.org/10.1037/0021-9010.64.4.349

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

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263. https://doi.org/10.2307/1914185

Knox, R. E., & Inkster, J. A. (1968). Postdecision dissonance at post time. *Journal of Personality and Social Psychology*, *8*(4), 319–323. https://doi.org/10.1037/h0025528

Krüger, T., Fiedler, K., Koch, A. S., & Alves, H. (2014). Response category width as a psychophysical manifestation of construal level and distance. *Personality and Social Psychology Bulletin*, *40*(4), 501–512. https://doi.org/10.1177/0146167213517009

Lewandowsky, S., & Oberauer, K. (2020). Low replicability can support robust and efficient science. *Nature Communications*, *11*(1), 358. https://doi.org/10.1038/s41467-019-14203-0

Liberman, N., & Förster, J. (2006). Inferences from decision difficulty. *Journal of Experimental Social Psychology*, *42*(3), 290–301. https://doi.org/10.1016/j.jesp.2005.04.007

Liberman, N., & Förster, J. (2009). The effect of psychological distance on perceptual level of construal. *Cognitive Science*, *33*(7), 1330–1341. https://doi.org/10.1111/j.1551-6709.2009.01061.x

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, *6*, 1171. https://doi.org/10.3389/fpsyg.2015.01171

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, *49*(4), 1494–1502. https://doi.org/10.3758/s13428-016-0809-y

Makowski, D., Ben-Shachar, M., & Lüdecke, D. (2019). bayestestR: Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework. *Journal of Open Source Software*, *4*(40), 1541. https://doi.org/10.21105/joss.01541

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

Martinie, M.-A., Milland, L., & Olive, T. (2013). Some theoretical considerations on attitude, arousal and affect during cognitive dissonance. *Social and Personality Psychology Compass*, *7*(9), 680–688. https://doi.org/10.1111/spc3.12051

Meinrenken, C. J., Chen, D., Esparza, R. A., Iyer, V., Paridis, S. P., Prasad, A., & Whillas, E. (2020). Carbon emissions embodied in product value chains and the role of Life Cycle Assessment in curbing them. *Scientific Reports*, *10*(1), 6184. https://doi.org/10.1038/s41598-020-62030-x

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

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, *9*(3), 353–383. https://doi.org/10.1016/0010-0285(77)90012-3

Nease, A. A., Mudgett, B. O., & Quiñones, M. A. (1999). Relationships among feedback sign, self-efficacy, and acceptance of performance feedback. *Journal of Applied Psychology*, *84*(5), 806–814. https://doi.org/10.1037/0021-9010.84.5.806

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

Rader, C. A., Larrick, R. P., & Soll, J. B. (2017). Advice as a form of social influence: Informational motives and the consequences for accuracy. *Social and Personality Psychology Compass*, *11*(8), e12329. https://doi.org/10.1111/spc3.12329

Satterthwaite, F. E. (1941). Synthesis of variance. *Psychometrika*, *6*(5), 309–316. https://doi.org/10.1007/BF02288586

Schrah, G. E., Dalal, R. S., & Sniezek, J. A. (2006). No decision-maker is an island: Integrating expert advice with information acquisition. *Journal of Behavioral Decision Making*, *19*(1), 43–60. https://doi.org/10.1002/bdm.514

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.

Sniezek, J. A., & Buckley, T. (1995). Cueing and cognitive conflict in judge-advisor decision making. *Organizational Behavior and Human Decision Processes*, *62*(2), 159–174. https://doi.org/10.1006/obhd.1995.1040

Soll, J. B., & Larrick, R. P. (2009). Strategies for revising judgment: How (and how well) people use others' opinions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(3), 780–805. https://doi.org/10.1037/a0015145

Soll, J. B., & Mannes, A. E. (2011). Judgmental aggregation strategies depend on whether the self is involved. *International Journal of Forecasting*, *27*(1), 81–102. https://doi.org/10.1016/j.ijforecast.2010.05.003

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

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, *34*(4), 273–286. https://doi.org/10.1037/h0070288

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review*, *117*(2), 440–463. https://doi.org/10.1037/a0018963

Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley.

Vallacher, R. R., & Wegner, D. M. (1989). Levels of personal agency: Individual variation in action identification. *Journal of Personality and Social Psychology*, *57*(4), 660–671. https://doi.org/10.1037/0022-3514.57.4.660

Van Swol, L. M., MacGeorge, E. L., & Prahl, A. (2017). Advise with Permission? The Effects of Advice Solicitation on Advice Outcomes. *Communication Studies*, *68*(4), 476–492. https://doi.org/10.1080/10510974.2017.1363795

Van Swol, L. M., Prahl, A., MacGeorge, E., & Branch, S. (2019). Imposing advice on powerful people. *Communication Reports*, *32*(3), 173–187. https://doi.org/10.1080/08934215.2019.1655082

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

Zeelenberg, M. (1999). Anticipated regret, expected feedback and behavioral decision making. *Journal of Behavioral Decision Making*, *12*(2), 93–106. https://doi.org/10.1002/(SICI)1099-0771(199906)12:2<93::AID-BDM311>3.0.CO;2-S

D.2 Manuscript II

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

Bayesian Advice Taking: Adaptive Strategy Selection in Sequential Advice Seeking

Tobias R. Rebholz[1], Mandy Hütter[1], and Andreas Voss[2]

[1]Psychology Department, Eberhard Karls University of Tübingen

[2]Institute of Psychology, Heidelberg University

Author Note

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

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

Andreas Voss https://orcid.org/0000-0002-4499-3660

Data and reproducible analysis scripts are publicly available at the Open Science Framework repository (https://osf.io/s9j8q). We have no known conflicts of interest to disclose. This research was funded by the Deutsche Forschungsgemeinschaft (DFG), grant 2277, Research Training Group "Statistical Modeling in Psychology" (SMiP).

Correspondence concerning this article should be addressed to Tobias R. Rebholz, Psychology Department, Eberhard Karls University of Tübingen, Schleichstr. 4, 72076 Tübingen, Germany. Email: tobias.rebholz@uni-tuebingen.de

Abstract

In sampling approaches to advice taking, participants can sequentially sample multiple pieces of advice before making a final judgment. To contribute to the understanding of active advice seeking, we develop and compare different strategies for information integration from external sources, including Bayesian belief updating. In a reanalysis of empirical data, we find that participants most frequently compromise between their initial beliefs and the distributions of multiple pieces of advice sampled from others. Moreover, across all participants, compromising predicts their final beliefs better than choosing one of the two sources of information. However, participants' willingness to integrate external opinions is relatively higher for multiple pieces of reasonably distant as compared to close advice. Nevertheless, egocentrism is as pronounced as in the traditional paradigm where only a single piece of external evidence is provided. Crucially, there are large inter- and intra-individual differences in strategy selection for sequential advice taking. On the one hand, some participants choose their own or others' judgments more often, and other participants are better described as compromisers between internal and external sources of information. On the other hand, virtually all participants apply different advice taking strategies for different items and trials. Our findings constitute initial evidence of the adaptive utilization of multiple, sequentially sampled external opinions.

Keywords: advice seeking, sequential sampling, Bayesian belief updating, information integration, judge-advisor system

Bayesian Advice Taking: Adaptive Strategy Selection in Sequential Advice Seeking

People often turn to others for advice. Sometimes they sequentially ask multiple persons for advice on the same subject. In fact, information sampling evolves in a sequential manner in many other situations, too. In social contexts, for instance, people make serial adjustments for affect inferences about other people (Yik et al., 2019), or in impression formation during experience sampling (Denrell, 2005). However, serial patterns can also be observed in visual perception (Fischer & Whitney, 2014), or causal inferences in sports, health, marketing, and other domains (Hogarth & Einhorn, 1992). The first piece of external information may not be enough to feel sufficiently confident for making a final judgment or decision. In that case, sampling continues until a more comprehensive informational basis is reached. The decision about the sufficiency of information likely depends on the current beliefs. Those, in turn, may have been sequentially updated based on the previously encountered pieces of information (Denrell, 2005; Hogarth & Einhorn, 1992).

In the dominant experimental paradigm of advice taking research, the judge-advisor system (JAS) of Sniezek and Buckley (1995), participants first judge a set of items on a certain dimension without external help. For instance, they are asked to estimate the caloric content of food (Hütter & Ache, 2016; Schultze et al., 2015; Yaniv et al., 2009), airline distances between cities (Ache et al., 2020; Schultze et al., 2012, 2017), dates of historic events (Gino, 2008; Hütter & Fiedler, 2019; Yaniv, 2004a), or carbon footprints of products (Rebholz & Hütter, 2022). In most of these examples, participants can revise their initial judgments after having been given access to a single piece of advice from another person. In sampling extensions of the original paradigm, however, judges do not passively receive a certain amount of advice from others. Instead, before providing a possibly revised final judgment they have the opportunity to sequentially sample as many additional opinions as they like (e.g., up to 20 in Ache, 2017, Experiment 5, and Hütter & Ache, 2016, Experiments 2 & 3).

In previous research, the sequential nature of the sampling process was often not taken explicitly into consideration when investigating peoples' advice taking behavior in the sampling-JAS. Exceptions are the procedures implemented by Scheunemann et al. (2020, 2021), where intermediate judgments after each sampling step were requested. However, receiving a certain amount of advice per item was mandatory (i.e., the procedure involved no active sampling) in the later study. More importantly, asking participants to explicitly express potentially updated intermediate judgments bears a higher risk of demand characteristics. That is, requesting repeated judgments for the same item is a meta-communicative act that increases the pressure on participants to make serial adjustments. Without data on intermediate judgments, multiple advice taking can be approximated as simple (i.e., unweighted) averaging across all pieces of advice sampled during a trial (Ache, 2017; Hütter & Ache, 2016). However, for trials in which more pieces of advice are sampled than can be kept in working memory (Cowan, 2010; see also G. A. Miller, 1956), this cumulative averaging assumption would require central capacities that most individuals cannot mobilize.

Weighted averaging is more in line with Information Integration Theory according to which the "diagnosticity" of a certain piece of advice is proportional to its serial positioning in the sample (Anderson, 1971; Shanteau, 1970, 1972). The multilevel regression-based model of Rebholz et al. (2023) enables the estimation of individual weights per individually sampled advice despite the lack of intermediate judgments. Instead of implicitly imposing equal-weighting constraints, they propose a multilevel modeling framework for "estimating" advice- or advisor-specific weights as sampling-related deviations (i.e., random effects) from the overall weighting tendency (i.e., fixed effect; Bates et al., 2015; Brown et al., 2018; Raudenbush & Bryk, 2002). Alternatively, sequential weights might take the judges' respective uncertainties into account as suggested by Soll et al. (2022) for an extension of their “influence of advice” measure. In the underlying modeling approach, influence of advice is operationalized as a mixture of judgment shift and confidence change in the belief formation process. We will extend this concept to sequential sampling by applying Bayes' theorem to approximate latent intermediate judgments.

The main aim of the present research is to generate insights with respect to adaptive strategy selection in sequential advice seeking. Accordingly, we are predominantly interested in inter- and intra-individual differences, which reflect participants' idiosyncratic selections of strategies for updating their beliefs about certain items (Payne et al., 1993; Schrah et al., 2006). For that purpose, we first develop a sequential Bayesian model of multiple advice taking. Moreover, we show that a hierarchical Bayesian model implies judgment updating that is equivalent to simple cumulative compromising. In a reanalysis of empirical data from Ache (2017), the two compromising models derived from the Bayesian account are compared with the two choosing strategies that were established in traditional JAS research, that is, choosing the self or choosing others (Soll & Larrick, 2009). The results of this model comparison add to the understanding of adaptive advice seeking by enabling insights with respect to individual differences in strategy selection for sequential advice taking. Finally, in the Discussion we will highlight important conceptual differences between different kinds of compromising, which is the most frequently applied multiple advice taking strategy in our empirical example.

A Bayesian Account of Sequential Advice Seeking

Sampling processes may rely on internal (“Thurstonian”) or external (“Brunswikian”) sources of information (Juslin & Olsson, 1997). According to the notion of Thurstonian sampling, participants' initial judgments about an unknown true value are the result of an internal sampling process (e.g., self-constructed feedback, sequentially recalled memories; Fiedler & Kutzner, 2015; Henriksson et al., 2010; Stewart et al., 2006). In contrast, their final judgments represent the combination of both Thurstonian and Brunswikian sampling (i.e., pieces of advice; Sniezek & Buckley, 1995; Yaniv &

Kleinberger, 2000; see also Hertwig et al., 2019). The Thurstonian notion relies on random or quasi-random sampling (Fiedler & Juslin, 2006; Juslin & Olsson, 1997; but see Herzog & Hertwig, 2014; Rauhut & Lorenz, 2011; Soll & Klayman, 2004). Therefore, we find a residual degree of uncertainty about the unknown true state of the world at the end of an internal sampling chain (as induced by, for instance, participants' impatience or the implemented experimental procedure; Juslin & Olsson, 1997). Uncertainty is formally often expressed as

$$\theta \sim N(E_0, C_0), \tag{1}$$

where the lower-case Greek character denotes a quantity unknown to the judge, and the upper-case Latin characters denote known quantities (the same notation applies throughout). Specifically, a judge's initial uncertainty about the unknown true value θ of an item is parameterized as normally distributed prior belief centered at a "best guess" E_0 with subjective confidence C_0 (Hemmer et al., 2015; B. Miller & Steyvers, 2011; Soll et al., 2022).[1]

Uncertainty motivates sampling of Brunswikian sources of information (Ache, 2017, Experiments 4 & 5; see also Gino & Moore, 2007; Yaniv, 2004b). An external piece of advice A_k sampled at time $k = 1, ..., K$ is assumed to provide additional information about the true value θ with likelihood

$$A_k \sim N(\theta, \pi), \tag{2}$$

[1] The normal approximation often works reasonably well (Adjodah et al., 2021; Moussaïd et al., 2013; Soll & Klayman, 2004). Its parametrization with precision C, which is inversely related to the variance as follows $S^2 = C^{-1}$, is convenient in the sense that larger values represent higher confidence in a certain judgment. From a Thurstonian point of view, the amount and validity of focal information in participants' working memory determines subjective confidence C (Hütter & Fiedler, 2019; Koriat, 2012a, 2012b; Koriat et al., 1980). In general, however, more flexible distributional assumptions could easily be implemented at the expense of modeling parsimony and analytical solutions (cf. Molleman et al., 2020).

where π denotes the population precision of advice (i.e., higher values indicating more trustworthy or knowledgeable sources of information), or $\tau^2 = \pi^{-1}$ the variance of advice around the true value of an item, respectively. Essentially, judges are supposed to believe that the advice they receive is unbiased on average but imprecise (to a certain degree π) and accordingly error-prone if considering a single value A_k (randomly) drawn from the population of advisors. From a wisdom of crowds perspective, such beliefs are justified if certain conditions (e.g., independence of judgments to reduce “shared error” or “redundancy”) are met in the decision-making environment (Hogarth, 1978; Soll & Larrick, 2009; see also Hertwig, 2012; Koriat, 2012b; but see Davis-Stober et al., 2014). More importantly, people tend to—albeit imperfectly—appreciate the wisdom of crowds (Larrick & Soll, 2006; Mannes, 2009). That is, individuals treat advice in the form of judgments of others as information that is “more or less trustworthy” (p. 635) and thus relevant for assessing the true state of the world (Deutsch & Gerard, 1955).

Sequential Processing of Advice

Because the normal distribution is the conjugate prior of itself, Bayes’ rule yields closed form solutions for judgment and confidence updated in response to having sampled a certain piece of advice A_k:

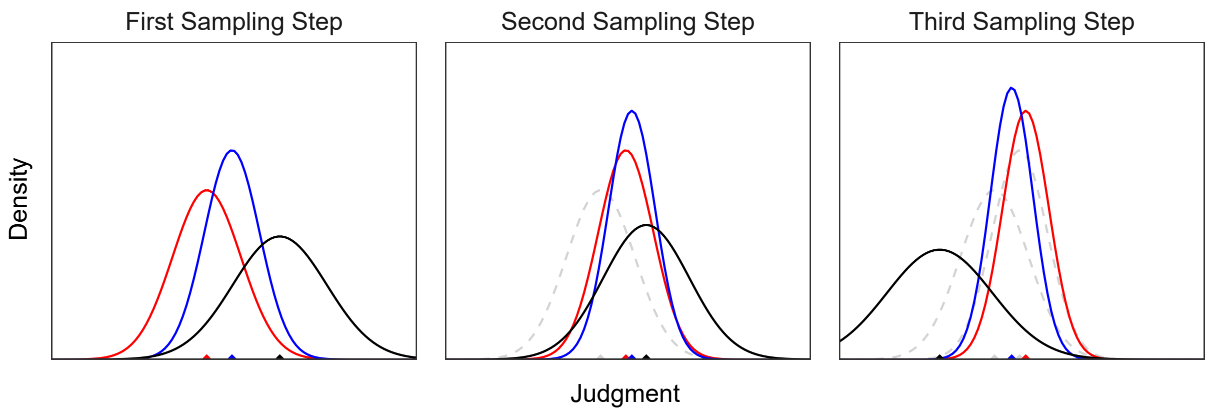
$$\hat{E}_k = \frac{\hat{C}_{k-1}}{\hat{C}_k}\hat{E}_{k-1} + \frac{\pi}{\hat{C}_k}A_k, \tag{3}$$

$$\hat{C}_k = \hat{C}_{k-1} + \pi \tag{4}$$

(Gelman et al., 2013), where $\hat{E}_k$ and $\hat{C}_k$ are the estimated best guess and confidence after having received advice A_k (see also Bednarik & Schultze, 2015, for incorporating a single piece of advice). Moreover, $\hat{E}_0 = E_0$ and $\hat{C}_0 = C_0$ are observed quantities. Accordingly, for instance, after encountering the first piece of external information A_1, Bayesian advice takers update their internal judgment E_0 to the mean $\hat{E}_1$ of the normal posterior belief distribution (see Figure 1, for an example with artificial data, which additionally implements sequential advice variance updating as specified below). Simultaneously, advice

Figure 1

Artificial Example of Step-by-Step Compromising with Three Sampling Steps



Note. Three samples of advice and their corresponding likelihoods (incl. advice variance updating according to Equations 8 to 10) are shown in black. The first two pieces of advice lie above and the last one below the center of the current (sequentially updated) belief distribution, which is plotted in red. The predicted posterior belief (blue) becomes the new prior of the next sampling step. Belief distributions from preceding sampling steps are grayed out.

seekers become more confident (as $\pi > 0$ by definition of normal precision) in their updated judgment by the amount they consider the external evidence to be informative about the truth (Gigerenzer et al., 1991; Koriat, 2012a). When encountering further pieces of external information A_k, any previous judgment $\hat{E}_{k-1}$ and confidence $\hat{C}_{k-1}$ are updated to the parameters of the k-th normal posterior belief distribution. Updating is repeated along the entire advice sampling chain (of trial-specific length K) until reaching a final judgment $\hat{E}_K$ (confidence $\hat{C}_K$). Borrowing terminology of Hogarth and Einhorn (1992), to compromise sequentially between internal and external sources of information as defined by Equations 3 and 4 will be referred to as "Step-by-Step" compromising strategy throughout.

Step-by-Step Bayesian judgment updating corresponds to a linear combination of the prior judgment and new data, where the respective weights are inversely related to the relative uncertainties associated with both judgments. The traditional advice taking index of Harvey and Fischer (1997) determines the relative influence of a single piece (or the average of multiple pieces; Ache, 2017; Hütter & Ache, 2016) of advice on participants' updated judgment as

$$W_{A_k} = \frac{E_k - E_{k-1}}{A_k - E_{k-1}}, \tag{5}$$

where the judgments $E_k \forall k \in [1, K-1]$ are unobserved. Rearranging this formula to account for endogenous judgment formation (i.e., the dependence of posterior beliefs on prior beliefs and external information; Rebholz et al., 2023) suggests that Step-by-Step Bayesian advice taking implies a weighting strategy that is close to the shared definition of this index:

$$E_k = (1 - W_{A_k})E_{k-1} + W_{A_k}A_k \tag{6}$$

where the normative weight of advice (WOA) corresponds to

$$\hat{W}^*_{A_k} = \frac{\pi}{\hat{C}_{k-1} + \pi} \tag{7}$$

(see Equation 3). The distributional shift from prior to posterior beliefs as induced by sampling a certain piece of advice may hence serve as a benchmark for optimal advice taking. Moreover, because uncertainty changes with the exploration of the environment (Hertwig et al., 2019; Lejarraga & Hertwig, 2021), we consider relative uncertainty-dependent (in our terminology: "Bayesian") weighting an adaptive advice taking strategy (see also Bednarik & Schultze, 2015).

In reality, optimal weighting is infeasible as the population variance of advice τ^2 is unknown to the judges. Instead, judges may update their beliefs about the validity of advice together with their judgment and confidence during sampling (cf. Landrum et al., 2015). The scaled inverse-χ^2 distribution is a "convenient" conjugate prior for beliefs about the variability of data:

$$\tau^2 \sim \text{Inv-}\chi^2\Big(L_0, T_0^2\Big) \tag{8}$$

(Gelman et al., 2013), where L_0 denotes the internal sample size, and T_0^2 participants' initial expectations about the variability of external samples of information. For S_0^2 denoting the inverse of participants' initial confidence, setting $T_0^2 = S_0^2$ implements initial expectations about the similarity of external and internal sampling experiences (cf. Koriat, 2012b). Bayes' rule again yields a closed form solution for updating the parameters of the conjugate prior of normal advice variance:

$$L_k = L_{k-1} + 1, \tag{9}$$

$$\hat{T}_k^2 = \frac{L_{k-1}}{L_k}\hat{T}_{k-1}^2 + \frac{1}{L_k}\Big(A_k - \hat{E}_{k-1}\Big)^2, \tag{10}$$

where $\hat{T}_0^2 = T_0^2$ is observed by implementing corresponding assumptions about participants' sampling expectations (i.e., $T_0^2 = S_0^2$; see above). Accordingly, beliefs about the variability of advice (which is inversely related to its perceived precision $\hat{P}_k = \hat{T}_k^{-2}$) are updated by its squared distance to the judges' current beliefs. Sequentially seeking advice to extend the internal sample of size L_0 hence entails comparing any additionally sampled data point A_k to all other, previously encountered judgments E_0, A_1, ..., and A_{k-1} represented as $\hat{E}_{k-1}$ (cf. Molleman et al., 2020, for simultaneous presentation of multiple pieces of advice). The diagnosticity of advisors' deviations from the rest of the sample is thereby assessed relative to their position in the growing sampling chain. It can be shown that advice distance is a valid cue to judgment accuracy (Soll et al., 2022), which thus justifies both the advice variance updating as specified above as well as confidence updating by adding the advice precision.

Subsequently, judgment and confidence are updated as defined in Equations 3 and 4, respectively. Replacing the unknown population precision π by judges' beliefs about it denoted as $\hat{P}_k$ yields

$$\hat{W}_{A_k} = \frac{\hat{P}_k}{\hat{C}_{k-1} + \hat{P}_k}, \tag{11}$$

which represents an admissible Bayesian advice weighting strategy. In four experiments without advice sampling, Soll and Larrick (2009) found a characteristic W-shaped pattern of weighting with three modes, two at the extremes of choosing the self ($W_{A_1} = 0$) and choosing the advisor ($W_{A_1} = 1$), and one at the midpoint representing simple averaging ($W_{A_1} = 0.5$). Depending on the relative uncertainties associated with internal and external sources of information, the Step-by-Step Bayesian WOA asymptotically captures all three documented strategies (see also Table 1). When confidence is relatively high compared to the perceived advice precision, $\hat{W}_{A_k}$ is close to the lower bound of 0. In contrast, when confidence is relatively low compared to the perceived advice precision, $\hat{W}_{A_k}$ is close to 1. For similar confidence and perceived advice precision, we have $\hat{W}_{A_k} \approx 0.5$. Accordingly, Step-by-Step Bayesian updating *dynamically* represents all three judgment strategies found in traditional JAS research.

Hierarchical Processing of Advice

Technically, two-stage processing implies that updating sequentially alternates between (a) an assessment of the perceived validity of social others in stage one and (b) updating one's beliefs about the truth in stage two. Essentially, the outcomes of parameter updating within the respectively other process are known at a certain stage. Hence, for the likelihood functions we have $A_k \sim N\left(\hat{E}_{k-1}, \pi\right)$ in stage one and $A_k \sim N\left(\theta, \hat{P}_k\right)$ in stage two, with the lower-case Greek characters specifying unknown quantities and the upper-case Latin characters specifying quantities known to the judge. Alternatively, hierarchical processing of advice variability and judgment plus confidence updating can be implemented by means of specifying a corresponding Bayesian account.

Assuming exchangeability of judgment and inverse confidence, the joint prior of a corresponding hierarchical model has a normal-inverse-χ^2 distribution denoted as

$$\left(\theta, \tau^2\right) \sim N\text{-Inv-}\chi^2\left(E_0, L_0; \lambda_0, T_0^2\right) \tag{12}$$

(Gelman et al., 2013). For the same reason as setting $T_0 = S_0$ above (i.e., participants expecting similar conditions for external and internal sampling; see above), we implemented equality of initial degrees of freedom and internal sample size, that is,

Table 1

Step-by-Step Bayesian Weight of Advice (WOA) as a Function of the Current Level of Confidence and Advice Distance

Confidence	Distance		Marginal Effect
	Close	Distant	
Low	$\hat{C}_{k-1} \ll \hat{P}_k \Rightarrow \hat{W}_{A_k} \to 1$	$\hat{C}_{k-1} \approx \hat{P}_k \Rightarrow \hat{W}_{A_k} \approx 0.5$	$\hat{W}_{A_k} \gtrsim 0.5$
High	$\hat{C}_{k-1} \approx \hat{P}_k \Rightarrow \hat{W}_{A_k} \approx 0.5$	$\hat{C}_{k-1} \gg \hat{P}_k \Rightarrow \hat{W}_{A_k} \to 0$	$\hat{W}_{A_k} \lesssim 0.5$
Marginal Effect	$\hat{W}_{A_k} \gtrsim 0.5$	$\hat{W}_{A_k} \lesssim 0.5$	$\hat{W}_{A_k} \in (0, 1)$

Note. For confidence much smaller than perceived advice precision, the Step-by-Step Bayesian WOA denoted as $\hat{W}_{A_k}$ converges to 1 in the top left cell. In contrast, for confidence much larger than perceived advice precision, $\hat{W}_{A_k}$ converges to 0 in the bottom right cell. That is, a judge with relatively low confidence tends to fully adopt closer advice whereas a judge with relatively high confidence tends to fully neglect more distant advice. For confidence $\hat{C}_{k-1}$ and advice precision $\hat{P}_k$ of comparable size, $\hat{W}_{A_k}$ converges to 0.5 in the top right and bottom left cells. That is, distant (close) advice is weighted about equally than the prior by judges with low (high) confidence. In summary, the marginal effects reflect "inconsistency discounting" (Anderson, 1971; Anderson & Jacobson, 1965; Yaniv, 2004a; see also Minson et al., 2011). Specifically, internal (external) inconsistency reflected in low confidence (high advice distance) implies weighting the advice (self) relatively more strongly.

$\lambda_0 = L_0$. Consequently, judgment and inverse confidence are updated as follows:

$$\hat{E}_K = \frac{L_0}{L_K} E_0 + \frac{K}{L_K} \bar{A}, \tag{13}$$

$$L_K = L_0 + K, \tag{14}$$

$$\hat{S}_K^2 = \frac{L_0}{L_K} S_0^2 + \frac{L_0 K}{L_K^2} \Big(\bar{A} - E_0\Big)^2 + \frac{K-1}{L_K} S_A^2, \tag{15}$$

where $\bar{A} = \frac{1}{K} \sum_{k=1}^{K} A_k$ and $S_A^2 = \frac{1}{K-1} \sum_{k=1}^{K} \Big(A_k - \bar{A}\Big)^2$. As in the sequential Bayesian account, the two updating processes take the respective uncertainties of both sources of information into account. Specifically, final judgment reflects a weighted linear combination of initial judgment and the simple (i.e., *unweighted*) average of advice. Moreover, posterior variance also combines prior variance and uncertainty conveyed by advice distance by adding the first two terms (cf. Equation 10). Additionally, the sample variance is taken into consideration by adding the last term.

For internal samples of size $L_0 = 1$, judgment updating as specified in Equation 15 is mathematically equivalent to simple cumulative averaging. Therefore, it is referred to as "End-of-Sequence" compromising strategy throughout (Hogarth & Einhorn, 1992). More generally, that is, for $L_0 > 1$, the hierarchical Bayesian account implements egocentric discounting. In other words, corresponding Bayesian updating provides an adapted version of the default compromising strategy that can be applied to multiple advice taking. Essentially, simple cumulative averaging is sampling invariant. That is, it yields the same predictions irrespective of the sequence in which samples (e.g., advice) are drawn from a population. For independent data probabilities, Bayes' rule also does not depend on the sequence in which evidence is sampled (Kruschke, 2015). Hence, the hierarchical Bayesian updating model is sampling invariant, too. Hogarth and Einhorn (1992) criticize this property of Bayesian cognitive modeling, which is inconsistent with their findings of serial weighting patterns. By contrast, the (temporal) dissociation of updating beliefs about advice variability and about the true value of an item makes the sequential Bayesian account non-invariant with respect to the sequence in which advice is sampled.

Accordingly, an empirical comparison of both the Step-by-Step and End-of-Sequence compromising strategies is apposite to generate insights with respect to participants' adaptive strategy selection in sequential advice taking scenarios.

Empirical Application

We applied the Bayesian models to reanalyze data of Experiment 5 presented by Ache (2017), because the data structure and experimental procedure are appropriate for our modeling approach. Specifically, in contrast to other experiments employing a sampling approach to advice taking (e.g., Hütter & Ache, 2016; Scheunemann et al., 2020), this experiment involved distributional measures of confidence regarding participants' initial and final judgments. Previous research suggests that the impact of weighting fallacies (i.e., base rate neglect, conservatism) varies not only between individuals but also between situations (Howe et al., 2022). Therefore, our aim is to generate insights mainly with respect to individual differences in participants' information aggregation strategies in judgment tasks that allow to sequentially sample external evidence. For that purpose, the Step-by-Step and End-of-Sequence compromising models were compared to adapted versions of the choosing strategies that were established in traditional JAS research without active sampling (Soll & Larrick, 2009; see also Himmelstein, 2022). Specifically, choosing the self corresponded to repeating one's initial judgment and confidence, and choosing others was implemented as stating the simple average and sample variance of all pieces of advice sampled as final beliefs. A significance level of 5% was used for statistical testing throughout. Data and reproducible analysis scripts are publicly available online (https://osf.io/s9j8q/).

Method

Experiment 5 of Ache (2017) implemented a 2 (judgment phase: initial vs. final judgment) × 2 (advice distance: close vs. intermediate) × 2 (sampling: free vs. costly)[2]

[2] In the free sampling condition, advice was provided immediately after clicking the sampling button. In addition to the natural costs consisting of the time spent in the laboratory, sampling costs were increased

mixed design with repeated measures on the first two factors. Participants' task was to estimate the airline distance between 20 pairs of European cities for which they received a performance-contingent bonus based on the accuracy of their final judgments. On the first screen of a new trial, participants were asked to provide an independent initial judgment as well as lower and upper bounds to create an 80% confidence interval. Confidence intervals were transformed to unidimensional measures of spread as follows:

$$S_k = \frac{1}{2} \sum_{p \in \{0.1, 0.9\}} \frac{C_k^p - E_k}{z_p}, \tag{16}$$

where C_k^p denotes the lower ($p = 0.10$) and upper ($p = 0.90$) confidence bounds, respectively, and z_p the corresponding quantiles of the standard normal distribution (Soll & Klayman, 2004). The mean was taken to account for (potential) asymmetries in belief distributions.

During the subsequent sampling phase, participants could sequentially sample up to 20 pieces of advice allegedly stemming from participants of a previous study. In fact, advisory judgments were randomly drawn from normal distributions with means manipulated to be relatively more distant from participants' initial judgments in the intermediate as compared to the close distance condition. The centers of the advice distributions pointed in the direction of the true value of an item with constant standard deviation across distance conditions. By clicking the final judgment button at any stage of the sampling process (or having sampled the maximum amount of advice for a certain item), participants reached the final screen where they were asked for their final judgments and corresponding lower and upper bounds of 80% confidence (again transformed to S_K by applying Equation 16).

---

by six seconds waiting time after clicking the sampling button for participants assigned to the costly condition. We did not find notable differences except for significantly more sampling in the free condition ($M = 9.2392$, 95% CI $[8.9018, 9.5766]$, $SD = 5.9676$) as compered to the costly one ($M = 2.6783$, 95% CI $[2.5827, 2.7740]$, $SD = 1.7475$), $t(1395.7880) = 36.7013$, $p < .001$ (see also Figure 8 in the Discussion). Therefore, we will not make this distinction here.

Ache (2017, Experiment 5) collected data of $N = 128$ participants. We excluded 48 trials (1.88%; incl. one participant completely) on which participants' confidence intervals did not contain their best guesses. The reason is that the median or 50-th quantile, which is equal to the mean of a normal distribution, constitutes the optimal judgment strategy in an environment with symmetric loss/reward structure (Raiffa & Schlaifer, 1970; Soll et al., 2022). Therefore, best guesses should lie in between confidence bounds that represent the 10-th and 90-th quantiles of a belief distribution. Because of unreasonably small lower confidence bounds, we also excluded 18 trials (0.70%) on which participants assigned 10% probability to city distances of 1 km and below (cf. Soll & Klayman, 2004). Moreover, 6 trials (0.23%) were excluded on which participants stated deterministic beliefs, that is, infinite initial or final confidence.

Step-by-Step and End-of-Sequence Bayesian updating were compared to choosing strategies for final belief formation. Picking up on the finding of a characteristic W-shaped weighting pattern (Soll & Larrick, 2009), multiple advice taking was operationalized as simple averaging (Ache, 2017; see also Hütter & Ache, 2016). We extend this original approach to both parameters of the belief distribution (cf. Soll et al., 2022). Hence, choosing others was implemented as unweighted averaging over *all* external sources of information and stating the sample precision as final confidence. In contrast, choosing the self was implemented as repeating one's initial judgment and confidence. Simple cumulative averaging represents a mixture of these two strategies that is inherently part of the hierarchical Bayesian account. However, the two compromising strategies derived from the Bayesian account build on internal samples of size $L_0 = 4$.[3] In general, $L_0 > 1$ follows from a "privileged access" to one's own reasoning, whereas only one judgment per advisor

[3] The improvement by repeatedly generating own judgments was shown to quickly level off at around three to five iterations (see Rauhut & Lorenz, 2011, Figures 1 & 2). At this point, nearly the same performance was reached as both a hypothetical, infinitely large inner crowd as well as one to three individuals from an external crowd, respectively. Interestingly, this elbow in performance improvement overlaps with the working memory capacity of young adults (Cowan, 2010). Therefore, internal samples of size $L_0 = 4$ seem

is observed (Yaniv & Kleinberger, 2000). Consequently, the Bayesian compromising strategies entail egocentrically biased judgment (and variance) updating. In End-of-Sequence Bayesian judgment formation (Equation 13), the relative weight of the initial judgment E_0 is directly proportional to the relative amount of internal (L_0) versus external ($L_K - L_0 = K$) sampling. For Step-by-Step Bayesian updating, the upper left cell of the judgment formation matrix in Table 1 (see also Figure 10 in the Discussion) is more sparsely populated for relatively higher weighting of internal than external uncertainty (i.e., larger L_0) via Equation 10. Together with the conceptual difficulties of classifying confidence as high versus low (e.g., participant-wise comparisons of initial confidence to the mean across all trials), this sparsity is the reason for not explicitly investigating marginal effects of confidence in the following.

Prediction performance was assessed as absolute prediction error (APE) of actual final judgment (log-transformed to account for positive skew; Moussaïd et al., 2013) and standard deviation. By contrast to alternative performance metrics (e.g., squared error) or analyzing confidence (instead of standard deviation), this procedure ensured error measurement on the original scale, which facilitates interpretation. Accordingly, larger values of final standard deviation indicate less confident final beliefs and vice versa (see also Footnote 1), and the terms confidence and standard deviation will be used interchangeably in the following. Additionally, parameter-wise performance evaluation is complemented by holistic distributional testing.[4] The Kullback-Leibler divergence (KLD) is defined as

$$KLD = \int_{-\infty}^{\infty} \phi(\theta) \log\left(\frac{\phi(\theta)}{\hat{\phi}(\theta)}\right) d\theta, \tag{17}$$

to be a reasonable choice without information on internal sampling. Moreover, the results are qualitatively similar with setting $L_0 = 3$ or $L_0 = 5$.

[4] Although only the Bayesian account explicitly relies on (normal) distributional assumptions about participants' belief updating, the variance calculation in Equation 16 implies normality of beliefs for all other strategies, too.

where ϕ and $\hat{\phi}$ denote the normal density functions of the actual and predicted posterior beliefs, $\theta \sim N(E_K, C_K)$ and $\theta \sim N\left(\hat{E}_K, \hat{C}_K\right)$, respectively. Accordingly, KLD served as a unidimensional measure of posterior belief prediction accuracy.

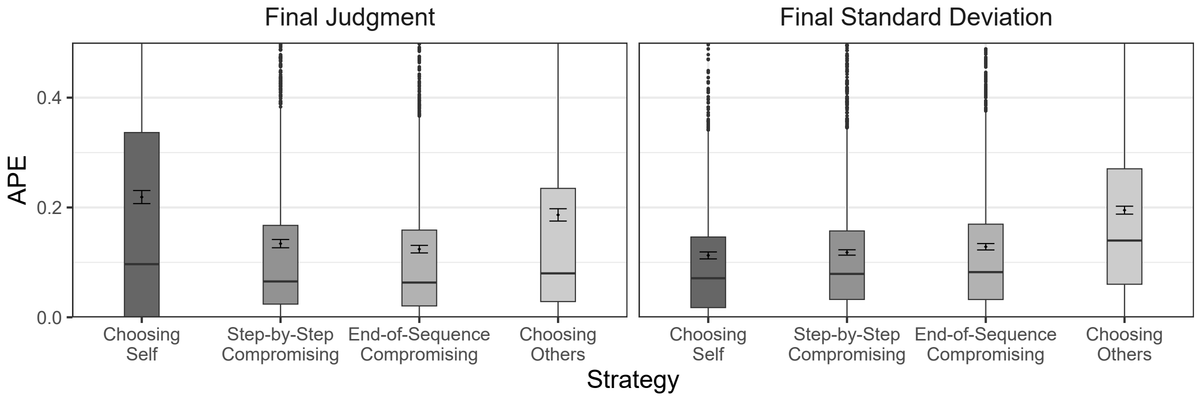
Results

On mean (median), prediction error of judgment as measured by APE was below 0.25 (0.10) for all choosing and compromising strategies (Figure 2, left panel). As indicated by non-overlapping 95% CIs, the two compromising strategies performed significantly better than predicting participants' final judgments as either sticking to their own initial estimate, or combining only the estimates of all others (Appendix A, Table A1). Moreover, APE was slightly and significantly lower for End-of-Sequence than Step-by-Step compromising. Prediction performance for standard deviation or confidence, respectively, was in the same range for all strategies but choosing the self (Figure 2, right panel). Specifically, a decrease in mean APE by more than half as well as a strong reduction in performance variability was observed for repeating one's initial confidence as compared to choosing one's initial judgment. Moreover, the ranking of the two compromising strategies reverses with APE for final confidence slightly lower for Step-by-Step than End-of-Sequence compromising.

Empirically, participants' strategies diverge for taking close versus distant advice. Close advice boosts confidence rather than triggering judgment shifts, whereas opposite patterns are generally found for more distant advice up to an intermediate degree (Hütter & Ache, 2016; Schultze et al., 2015; see also Soll et al., 2022). Whereas the prediction performances for final confidence are approximately the same across distance conditions, all strategies perform much better in predicting final judgments in response to closer advice (Figure 3). Impaired performance—particularly in terms of increased variability—could be attributed to significantly larger absolute shifts of judgment in the intermediate condition ($M = 0.3824$, 95% CI $[0.3636, 0.4012]$, $SD = 0.3390$) than in the close condition ($M = 0.0550$, 95% CI $[0.0482, 0.0617]$, $SD = 0.1214$), $t(1560.0570) = 32.0925$, $p < .001$.

Figure 2

Absolute Prediction Error (APE) per Belief Distribution Parameter and Strategy

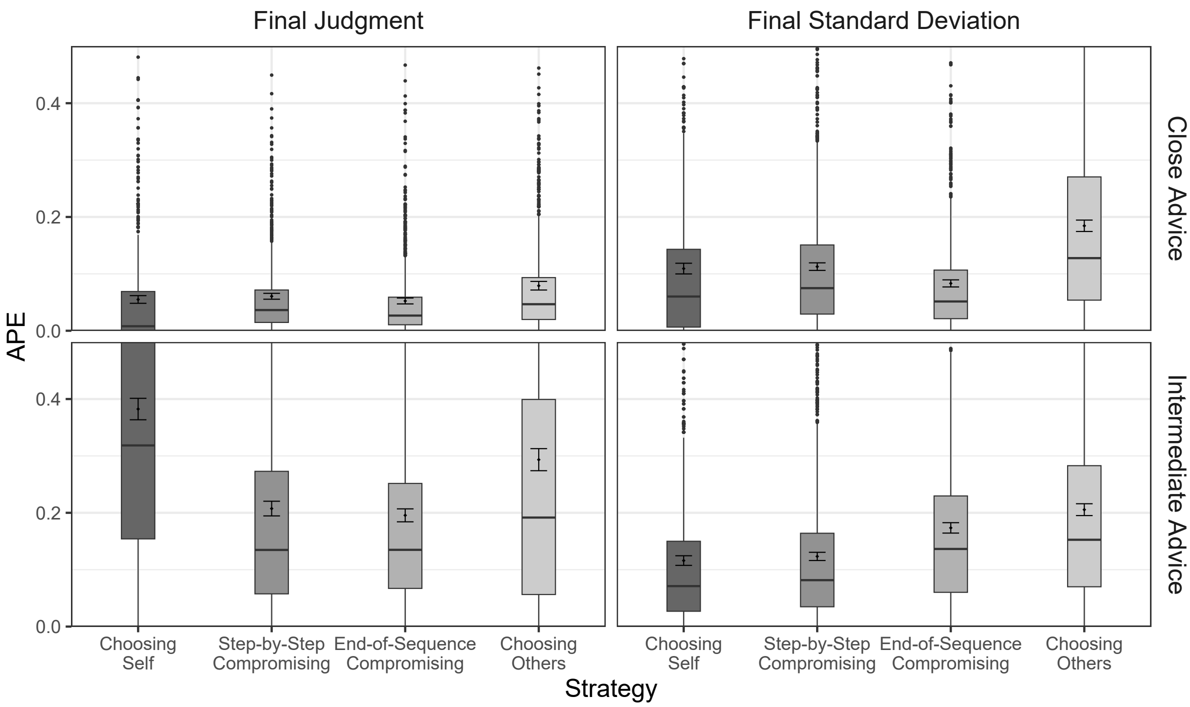


Note. Plotting is truncated for $APE > 0.5$ for both belief distribution parameters. Box plots are accompanied by means and 95% CIs. Summary statistics can be found in Appendix A, Table A1.

Shifting decreases the odds for accurate point predictions as compared to situations without any change because it can go in either direction by any amount. Indeed, the ranking of the two choosing strategies reverses across advice distance conditions with the choosing-others strategy being a better (worse) prediction than the choosing-self strategy in the intermediate (close) advice distance condition. Also in line with previous empirical evidence, the standard deviations of participants' belief distributions decreased significantly more in the close condition ($M = -0.0912$, 95% CI $[-0.1011, -0.0812]$, $SD = 0.1792$) as compared to the intermediate one ($M = -0.0592$, 95% CI $[-0.0694, -0.0491]$, $SD = 0.1823$), $t(2485.5060) = -4.4056$, $p < .001$. That is, opposite treatment effects of distance on judgment and confidence updating can be observed and point in the expected directions. In contrast to judgment, however, a ranking reversal of the compromising strategies but not of the choosing strategies was observed across the distance conditions for confidence prediction.

Figure 3

Absolute Prediction Error (APE) per Belief Distribution Parameter, Strategy, and Advice Distance Condition

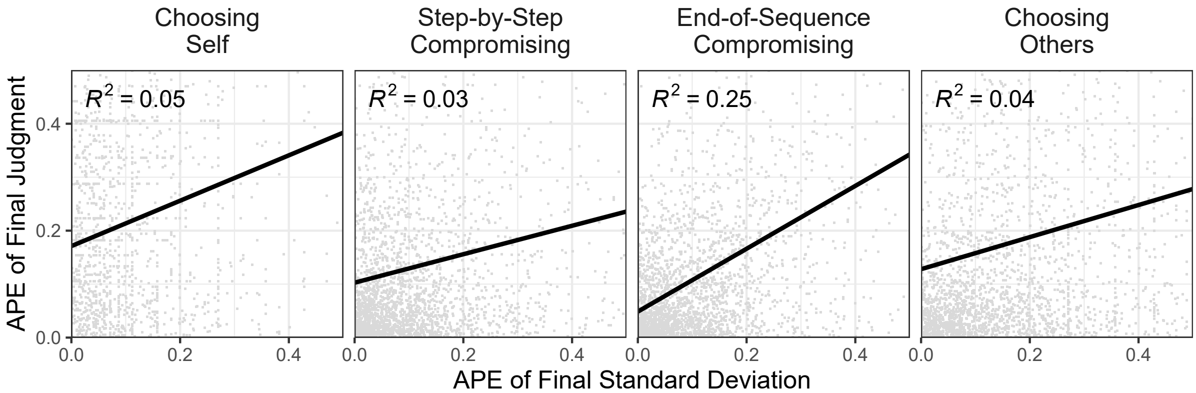


Note. Plotting is truncated for $APE > 0.5$ for both belief distribution parameters. Box plots are accompanied by means and 95% CIs. Summary statistics can be found in Appendix A, Tables A2 and A3 for close and intermediate advice distance, respectively.

Good fit of one parameter of the belief distribution does not necessarily imply good fit of the other. Indeed, there are only weak relations between predictability of judgment and predictability of confidence for all strategies except End-of-Sequence compromising (Figure 4). On median, End-of-Sequence and Step-by-Step compromising achieve the best belief distribution predictions from a holistic perspective, followed by choosing the self and choosing others in this order (Figure 5; see Appendix B for corresponding results across advice distance conditions). Note that the large performance variability of the choosing-self strategy for holistic distributional predictions is inherited from the judgment parameter. In

Figure 4

Performance Relations Across Belief Distribution Parameters as Measured by Absolute Prediction Errors (APE)



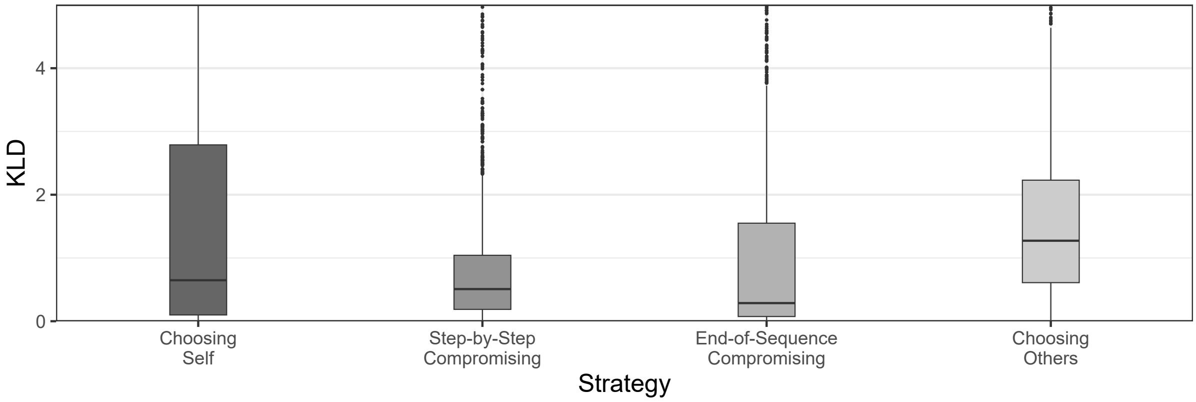
Note. Plotting is truncated for $APE > 0.5$ for both belief distribution parameters. The solid lines indicate the linear correspondences (incl. R^2) between APE of final judgment and APE of final standard deviation.

summary, the overall pattern from Figure 2 was replicated for measuring belief distribution divergence by KLD. Therefore, we will exclusively rely on holistic distributional testing in the following (see Appendix C for corresponding parameter-wise results).

Previous research suggests that the existence of certain weighting fallacies (i.e., base rate neglect, conservatism) varies between individuals and situations (Howe et al., 2022). Although participants' behavior over all trials of an experiment may be best described by a certain strategy, they may still apply different strategies for updating their beliefs about different items. The relative shares of best strategy fits per participant are plotted in Figure 6. On average and median, End-of-Sequence compromising is the best description of most participants' strategy selections, followed by Step-by-Step compromising, choosing the self, and finally choosing others. Moreover, the ranking reversal of the two compromising strategies across distance conditions was also observed in terms of shares of best fits (Figure 7). Indeed, external evidence of intermediate distance is integrated relatively more

Figure 5

Performance for Holistic Belief Distribution Predictions per Strategy as Measured by Kullback-Leibler Divergence (KLD)



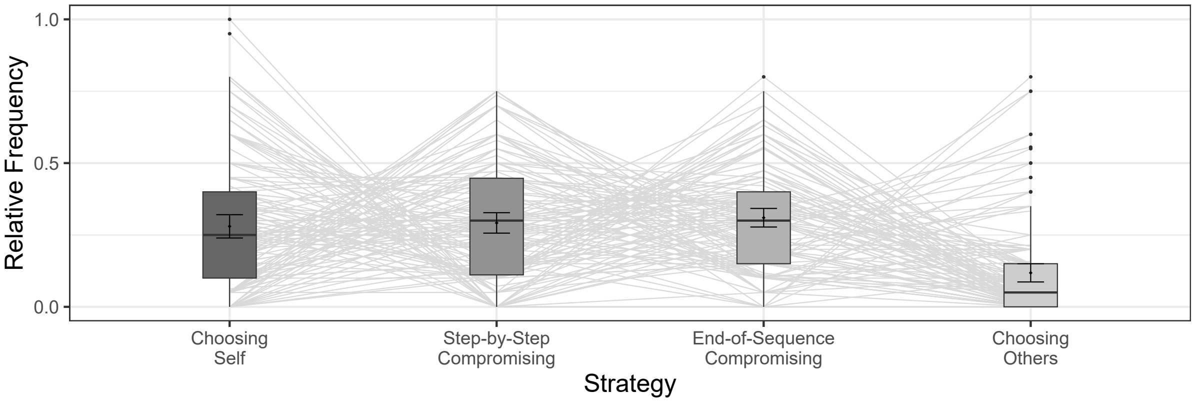
Note. Plotting is truncated for $KLD > 5$. As the advice variance is zero for $K = 1$, infinite values of KLD are excluded for choosing others. Summary statistics can be found in Appendix A, Table A4.

strongly and chosen relatively more often than close advice. More importantly, however, is that for most participants the best-fitting strategy varies between items as indicated by the thin gray lines representing participants' individual strategy selection dynamics (i.e., intra-individual differences) in Figures 6 and 7. Consequently, the relative shares of best strategy fits are strictly smaller than 100% for most participants and strategies.

In summary, participant-wise comparisons of strategy selection on an aggregated level match with the performance relations according to trial-wise prediction error measurement as described above. Nevertheless, there are also large inter-individual differences in strategy selection. For instance, one participant was perfectly described as a self-chooser (i.e., advice non-taker) across all items and irrespective of distance condition (see Figure 6). Notably, this participant did not sample more than the by default provided first piece of advice on all but the first trial and never changed his initial judgment. In other words, this special case did not really engage in sequential advice seeking. In

Figure 6

Relative Shares of Best Strategy Fits per Participant as Measured by Holistic Distributional Testing



Note. The thin gray lines correspond to individual participants and thus capture inter- and intra-individual idiosyncrasies. Box plots are accompanied by means and 95% CIs. Summary statistics can be found in Appendix A, Table A5.

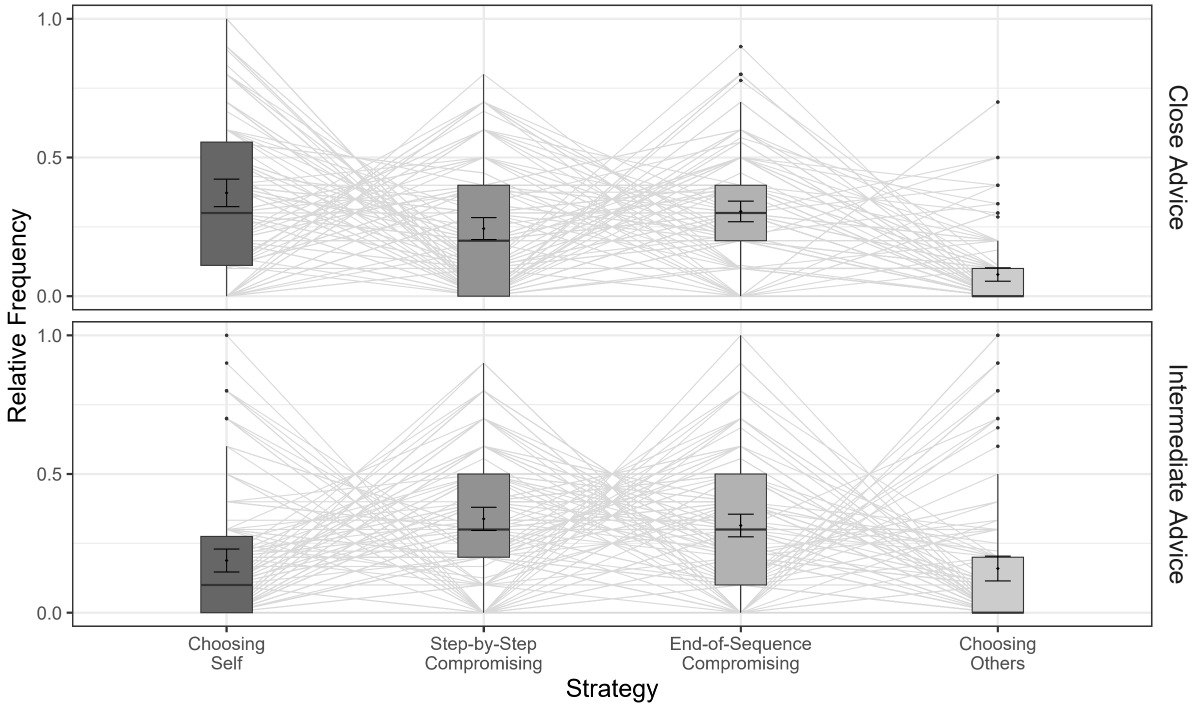
contrast, one participant was perfectly described as End-of-Sequence compromiser and two as other-choosers across all trials of the intermediate advice distance condition (see Figure 7). They gave up their initial beliefs on (almost) all trials. More generally, whereas some participants are better described as Step-by-Step or End-of-Sequence compromisers, others more often prefer internal over external sources of information or vice versa as indicated by the participant-wise ranking reversals of strategies in Figures 6 and 7. In conclusion, there is no strategy that fits all participants' data equally well and thus not a single strategy that dominates most participants' adaptive belief formation behavior.

Discussion

Different belief updating strategies for situations in which advice could be sampled sequentially were compared to generate insights with respect to adaptive advice seeking. The model comparison included the two established choosing strategies, that is, sticking to one's own judgment or choosing other peoples' judgments as well as two different

Figure 7

Relative Shares of Best Strategy Fits per Participant and Advice Distance Condition as Measured by Holistic Distributional Testing



Note. The thin gray lines correspond to individual participants and thus capture inter- and intra-individual idiosyncrasies. Box plots are accompanied by means and 95% CIs. Summary statistics can be found in Appendix A, Tables A6 and A7 for close and intermediate advice distance, respectively.

compromising strategies. Those compromising strategies were derived from sequential and hierarchical Bayesian updating, respectively.

Overall, End-of-Sequence Bayesian compromising provides the best description of participants' belief updating behavior. This holds true in terms of average judgment prediction error, holistic distributional testing, and the share of participants' strategy selections (see Figures 2, 5, & 6, respectively). However, the variance of participants' final belief distributions is better predicted by Step-by-Step Bayesian compromising and their initial confidence (see Figure 2, right panel). As participants' final judgments are congruent with their initial judgments on 28.78% of trials, egocentrism is as pronounced in multiple as in single advice taking situations (Soll & Larrick, 2009; see also below). Moreover, advice distance influences participants' strategies (see Figures 3 & 7). For instance, confidence updating is not as much affected as participants' integration of close versus more distant advice into their own judgments (cf. Moussaïd et al., 2013; Schultze et al., 2015). In summary, participants apply a mix of strategies in sequential advice taking comparable to single advice taking scenarios (Soll & Larrick, 2009).

Step-by-Step Bayesian compromising almost reaches the prediction performance of its End-of-Sequence counterpart and the average shares of participants' selections of those two strategies are about the same (see also Appendix A, Table A5). On average, a non-negligible share of 29.19% of participants' strategies is best described by the sequential Bayesian account. Consequently, participants often behave like sequential advice takers applying a dynamic, confidence-dependent weighting scheme. In total, more than half of participants' strategy selections (on average 60.19%) are best described by some sort of compromising between the self and others. Although the two advice weighting measures are strongly and significantly correlated, Pearson's $r(14561) = 0.7632$, $p < .001$, we observed many and partly striking ranking reversals between the two compromising strategies for some participants (see the thin gray lines in Figure 6). Therefore, the remainder of this discussion will mainly focus on the properties and differences between Step-by-Step and End-of-Sequence compromising as the two most popular strategies for multiple advice taking.

Inconsistency Discounting

The implementation of End-of-Sequence compromising such that all judges apply the same weighting strategies across items j and time k precludes adaptive judgment aggregation. Whereas this strategy entails a minor advantage over Step-by-Step compromising in terms of prediction performance, it is rather unrealistic in light of findings from the traditional JAS (e.g., differential weighting with respect to advice distance; Moussaïd et al., 2013; Schultze et al., 2015, or relative expertise; Harvey & Harries, 2004; Sniezek et al., 2004). In contrast, Step-by-Step Bayesian judgment updating takes the relative uncertainties of internal and external sources of information into account. In the terminology of Information Integration Theory, uncertainty-dependent weights imply inconsistency discounting (Anderson, 1971; Anderson & Jacobson, 1965; Yaniv, 2004a; see also Minson et al., 2011). Specifically, discounting of internal inconsistency (i.e., low confidence) corresponds to weighting the advice more strongly than one's own judgment (see Table 1). In contrast, discounting of external inconsistency (i.e., high advice distance) corresponds to weighting the self more strongly than others. The consequences for adaptive advice seeking will be discussed in separate subsections in the following.

External Inconsistency Discounting and Stopping. Recent evidence suggests that advice of intermediate distance as implemented in Experiment 5 of Ache (2017) is most influential (Moussaïd et al., 2013; Schultze et al., 2015). More specifically, these authors showed a nonlinear, inverse-U-shaped relation between advice weighting and distance. According to the Bayesian account, however, higher distance between advice and prior judgment implies, ceteris paribus, more variance. More imprecise advice is less influential (see Equations 3 & 10) in Step-by-Step compromising but does not affect judgmental integration in End-of-Sequence compromising (see Equation 13). For tasks that require the combination of internal and external sources of information as defined in Equations 1 and 2, respectively, this makes sense intuitively. More distant advice provides less reliable information about the truth, which is supposed to be temporarily centered at one's current beliefs. This is in line with findings of taking multiple pieces of advice that were presented simultaneously (Yaniv & Milyavsky, 2007). And indeed, the relatively strong increase in the mean share of best strategy fits of Step-by-Step as compared to End-of-Sequence compromising for more distant advice (see Figure 7) can be taken as evidence in support of external inconsistency discounting.

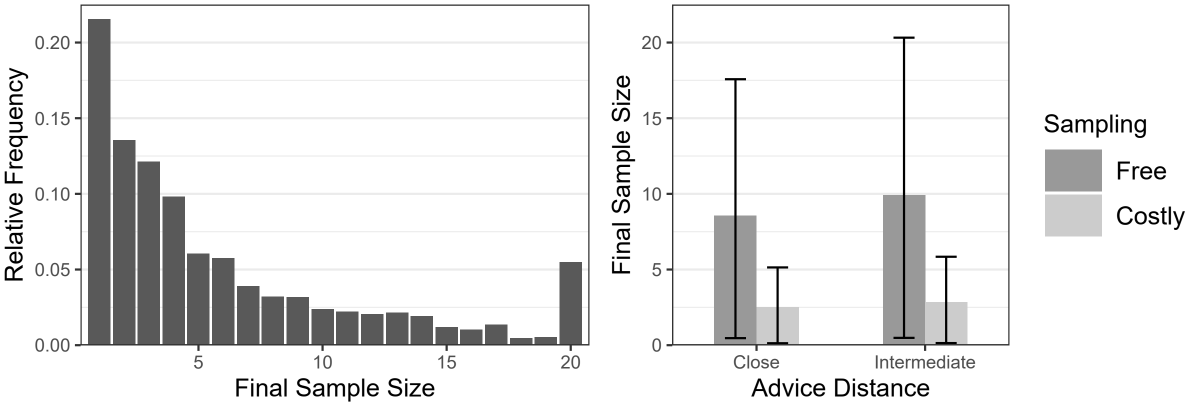
In cumulative averaging, the influence of advice decays in harmonic progression as it is weighted inversely proportional to the number of cumulation steps. In other words, the effects of individual pieces of advice on final judgment are, ceteris paribus, smaller in longer as compared to shorter sampling chains. Therefore, differences in sampling behavior may be responsible for the partly pronounced ranking reversals between the two compromising strategies within participants as plotted in Figure 6. For instance, harmonic progression predicts comparably low weights for each individual piece of advice in large samples of close advice whereas external inconsistency discounting predicts the opposite (i.e., comparably high individual weights of close advice). Similarly, larger inconsistencies in small samples of the more distant condition imply relatively low Step-by-Step weights but higher End-of-Sequence weights.

From a Thurstonian perspective, additional sampling suggests that advice seekers are not sufficiently confident in their judgment (Hütter & Fiedler, 2019; Koriat, 2012a, 2012b; Koriat et al., 1980; see also Footnote 1). On most trials of Experiment 5 of Ache (2017), participants did not sample more than the first piece of advice provided by default (Figure 8, left panel). Nevertheless, some of them exhaustively sampled all 20 pieces of advice available as well as everything in between no and full sampling on other trials. As a quasi-experimental factor, advice sample size depended on the within-subjects distance manipulation and the between-subjects cost factor (Figure 8, right panel). Specifically, more distant and lower-cost advice was sampled more than closer and higher-cost advice. Essentially, close advice typically induces confidence boosts rather than judgment shifts (Hütter & Ache, 2016; Schultze et al., 2015; Soll et al., 2022; see also Equations 4, 10, & 15). Therefore, posterior confidence thresholds are natural candidates for defining informative stopping criteria to predict participants' sampling decisions in future research (cf. Hausmann & Läge, 2008).

The average advice sample size of those trials best described by the sequential Bayesian account is significantly smaller than the sample size of End-of-Sequence

Figure 8

Sampling Descriptives for Experiment 5 of Ache (2017)

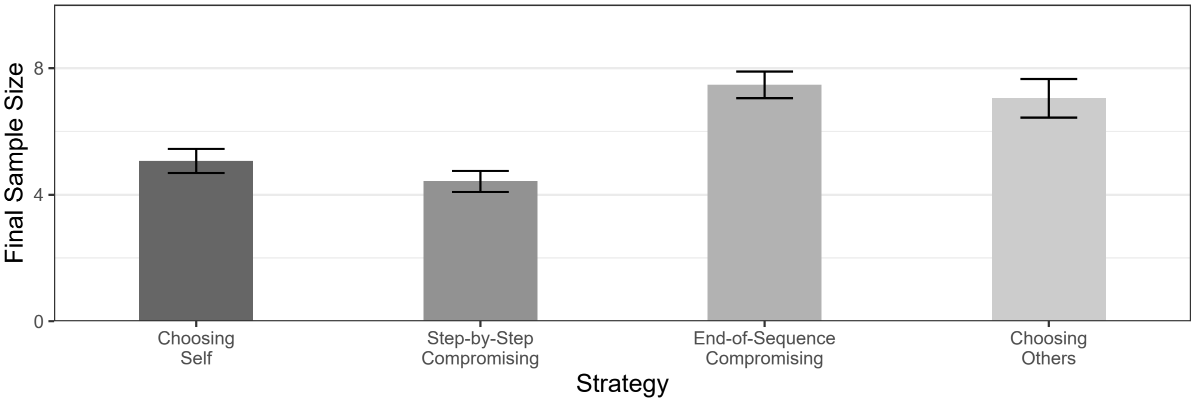


Note. Error bars show the 95% CIs.

compromising and choosing others (see Figure 9 and Appendix A, Table A8). Put differently, uncertainty-dependent Step-by-Step weighting is a better predictor for taking less advice. Conversely, cumulative averaging describes the treatment of larger samples of advice best. Both observations are counterintuitive from a working memory perspective, because Step-by-Step Bayesian updating should be less demanding than simple cumulative averaging in terms of working memory capacities (Behrens et al., 2007), making it easier to aggregate information accurately (Luan et al., 2020). Alternatively, this finding may be interpreted as a reduced need for additional pieces of external evidence when compromising (i.e., reaching “consensus”) is conducted Step-by-Step. Choosing the self, however, is obviously the least resource-demanding strategy. Moreover, the realized advice sample size was shown to have a significant positive effect on the total weighting of advice (Hütter & Ache, 2016). Therefore, the average sample size in those trials which are best described by the choosing-self strategy is surprisingly high relative to more resource-demanding strategies like compromising. This inefficiency (e.g., in terms of opportunity costs; Ostwald et al., 2015) suggests that participants also engage in confirmatory sampling, that is, sampling external evidence (merely) for the sake of confirming their initial opinions

Figure 9

Final Sample Size per Best-Fitting Strategy as Measured by Kullback-Leibler Divergence (KLD)



Note. Error bars show the 95% CIs. Summary statistics can be found in Appendix A, Table A8.

(Fiedler, 2000; see also Rader et al., 2015). The alternative explanation that those samples of advice were assessed as too inconsistent for containing any signal is less plausible as the distribution of advice was manipulated to rather consistently point in the same direction.[5]

Internal Inconsistency Discounting and Confidence Measurement. Theoretically, relative uncertainty-dependent weights of advice are also higher if confidence is relatively low (e.g., as compared to other trials within participants; see Table 1). That is, the discounting of external inconsistency goes hand in hand with the discounting of internal inconsistency. More generally, the Step-by-Step Bayesian compromising strategy dynamically incarnates all three judgment strategies depending on the relative uncertainties associated with internal and external sources of information (cf. Soll & Larrick, 2009). For $L_0 > 1$ (see also Footnote 3) and factorial (i.e., time-invariant) advice distance manipulations in Experiment 5 of Ache (2017), the Step-by-Step Bayesian WOA

[5] Constant spread of the advice distributions as implemented in Experiment 5 of Ache (2017) implies reduced diagnosticity of samples on trials with close as compared to more distant advice. However, the sampling pattern is qualitatively similar across advice distance conditions (see Appendix D, Figure D1).

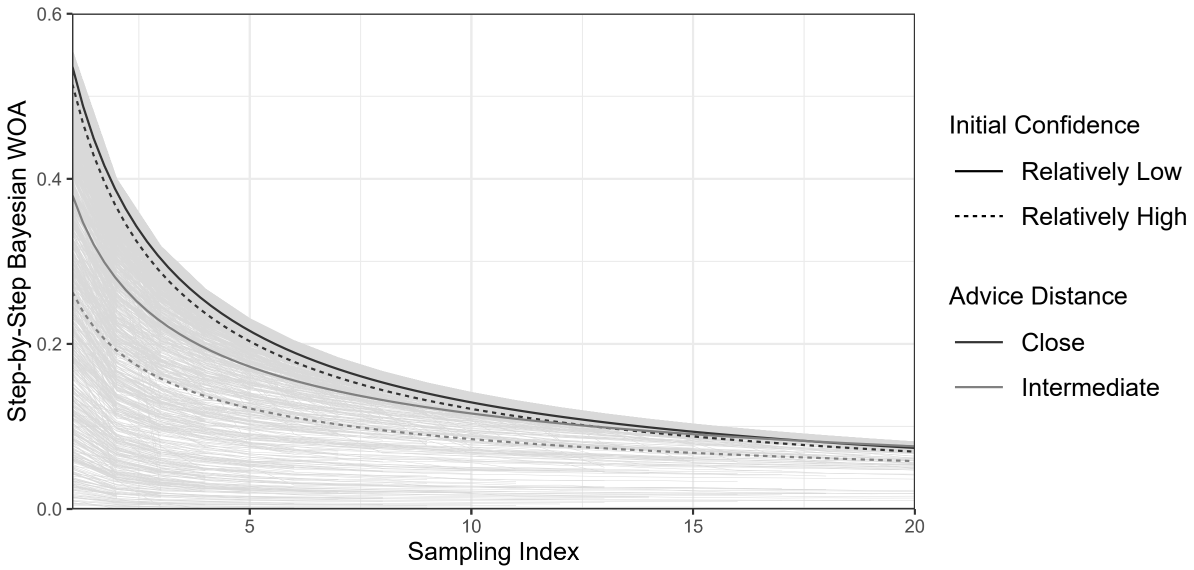
implements egocentric discounting (see also Figure 10). Specifically, we never observed $\hat{C}_{k-1} < \hat{P}_k$ in the data such that the top left cell of the judgment formation matrix in Table 1 is unpopulated for $k > 1$. Moreover, we find that $\hat{W}_{A_k} < \frac{1}{k+1}$ on 31.63% of sampling trials, that is, a relative share of egocentrism in multiple advice taking that is comparable to observations for single advice taking (e.g., Harvey & Fischer, 1997; Lim & O'Connor, 1995; Yaniv & Kleinberger, 2000; but see Soll & Larrick, 2009). Conversely, the fitted decays in Figure 10 level off at around 5% weighting of each (additional) piece of advice in rather long sampling chains. To that effect, Step-by-Step Bayesian updating predicts nearly complete adoption of large samples of advice. This is in line with findings of Hütter and Ache (2016) that more pieces of advice are weighted more strongly in sum.

Despite ignoring asymmetry for the transformation of confidence intervals to standard deviations in Equation 16, the predictions of all strategies were slightly better for this second parameter that measures the spread of normal final belief distributions. This is true both on average and in terms of performance variance and thus provides additional support for the merit of confidence thresholds as informative stopping criteria. However, although confidence is a crucial part of the traditional JAS paradigm (Sniezek & Buckley, 1995), it is not collected per default in advice taking experiments. If initial confidence was measured on scales (e.g., Scheunemann et al., 2020) or not collected at all (e.g., Hütter & Ache, 2016), the variance of (standardized) judgments over multiple trials may serve as a good proxy for confidence in a given judgment domain. In Experiment 5 of Ache (2017), by contrast, participants entered their best guess together with the lower and upper bounds of confidence on the same screen. This implementation supports the notion of *simultaneous* judgment and confidence updating as specified in the Bayesian account (cf. Gigerenzer et al., 1991).

According to research on overconfidence, the concept and measurement of judgment certainty is a matter of ongoing debate (Soll & Klayman, 2004; Soll et al., 2022; see also Lisi et al., 2021). As confidence is a seminal component of the sequential Bayesian account

Figure 10

Decays of Step-by-Step Bayesian Weight of Advice (WOA) as Functions of Sampling Index by Advice Distance Condition and Relative Level of Initial Confidence



Note. The light-gray lines show the decay of Bayesian WOA per trial. The four solid and dashed lines display the downward log-logistic trajectories estimated as $\hat{W}_{A_k} = \frac{1}{1+(k/\alpha)^{\beta}}$, where α and β denote the scale and shape parameters, respectively. Initially being relatively less (more) confident is determined within participants for confidence below (above) the median over all 20 trials.

only, this debate implies two important limitations for the corresponding Step-by-Step belief updating strategy. First, empirical findings of overconfidence restrict the applicability of a normative model that implements step-wise judgment updating exclusively based on relative confidence. In other words, the typically overestimated level of confidence relative to advice precision threatens the applicability or validity of Equation 3. This might also explain why the prediction performance of initial confidence for final confidence is so much better than the prediction performance of initial judgment for final judgment (Figure 2). Second, the mathematical operations necessary to transform the confidence intervals from Experiment 5 of Ache (2017) to unidimensional measures of spread (see Equation 16) introduce error. For instance, standard normal quantiles are not in line with assuming small internal samples of size $L_0 = 4$ in the Bayesian account. Empirically testing such assumptions (e.g., by surveying internal sampling experiences)[6] adds scrutiny and thus has the potential to provide valuable insights with respect to both internal and external sampling.

Serial Positioning and Order Effects

Beyond its relative precision, the serial positioning of advice often plays a crucial role for its weighting. According to an extensive literature review that compares descriptive and experiential research paradigms, the Bayesian lens can be adequate for judgments in learning tasks such as typically used in advice taking research (Lejarraga & Hertwig, 2021). It is often applied as cognitive modeling framework for information integration in social contexts such as wisdom of crowds. Typical tasks with multiple sources of information involve quantitative judgments and predictions (e.g., Adjodah et al., 2021; Molleman et al., 2020), ranking (e.g., B. Miller & Steyvers, 2011), or probabilistic inference (e.g., Budescu & Yu, 2006). Wisdom of crowds research, however, rarely deals with sequential information sampling but is more focused on group decision-making. Indeed, simple (cumulative) averaging is better justified for all these examples as multiple pieces of advice were provided simultaneously. Consequently, averaging represents an efficient information reduction strategy without restrictions, for instance, by limited working memory capacities (Cowan,

[6] Directly asking participants about their perceived internal sample sizes is likely biased and unreliable. Moreover, it would require instructions (and maybe even training) on the concept of "internal sampling"—similar to the alternative of directly asking for variance estimates. Nevertheless, in future research adaptive (i.e., capable of individual differences) End-of-Sequence compromising could be implemented by specifying sampling priors according to participants' internal sample sizes L_0. Importantly, however, the Bayesian account does not need to subscribe to a Thurstonian perspective. Distributional assumptions about beliefs can also be justified by participants' uncertainty about their judgments for other than internal sampling-related reasons (e.g., diffidence, deliberateness, or simply inattentiveness or ignorance of the judgment domain).

2010; see also Behrens et al., 2007). In contrast, systematic patterns of serial positioning such as order effects are often observed for sequential sampling (e.g., Hogarth & Einhorn, 1992; Shanteau, 1972). Actually, serial positioning effects in belief updating are supposed to capture the "natural presumption that order reflects the importance of the information provided" (Hogarth & Einhorn, 1992, p. 7). Although order effects may also be viewed as biases (e.g., Asch, 1946; J. M. Miller & Krosnick, 1998), an informative model of belief updating should be capable of accounting for corresponding empirical observations.

From a k-wise perspective, Step-by-Step Bayesian evidence accumulation implies primacy effects due to strictly growing posterior confidence. That is, for a belief distribution that becomes monotonically narrower, Step-by-Step Bayesian advice takers' (relative) sensitivity to advice decreases with increasing k. For instance, with same distance of advice A_2 to the updated judgment $\hat{E}_1$ as of advice A_1 to the initial judgment E_0, the effect of A_2 on $\hat{E}_1$ is smaller than the effect of A_1 on E_0. This is reasonable as, in contrast to the situation when initially encountering advice A_1, one has additional knowledge about less extreme (in terms of distance to the updated judgment $\hat{E}_1$) external opinions (i.e., A_1) when encountering more extreme advice A_2. In general, earlier pieces of advice have, ceteris paribus, more *immediate* influence on (intermediate) posteriors than later ones. Accordingly, recency (e.g., due to limited working memory capacities; Cowan, 2010; Glanzer & Cunitz, 1966) should not matter for the k-wise integration of advice in Step-by-Step compromising.

At the End-of-Sequence (i.e., for $k = K$), however, chances are high to observe recency effects in Step-by-Step Bayesian updating. Plugging all previous judgments into Equation 3 for $k = K$ yields

$$\hat{W}_{A_k}^{K} = \prod_{l=k+1}^{K} \left(1 - \hat{W}_{A_l}\right) \hat{W}_{A_k}, \tag{18}$$

for the updated Step-by-Step WOAs which quantify the *total* normative influence of advice A_k on final judgment $\hat{E}_K$. By definition of Equations 11 and 13, the Bayesian WOAs are restricted to the $[0,1]$ interval.[7] Hence, at stage K, $\hat{W}_{A_1}$ was eventually multiplied by $K-1$ values $\left(1-\hat{W}_{A_l}\right) \in [0,1]$, $l = 2,\ldots,K$, $\hat{W}_{A_2}$ by $K-2$ values $\left(1-\hat{W}_{A_l}\right) \in [0,1]$, $l = 3,\ldots,K$, and so on. Consequently, the Step-by-Step weights of earlier pieces of advice are updated in each stage of a sampling sequence by a factor strictly smaller than one. Therefore, the longer the chain, the smaller is the total or End-of-Sequence weight of advice that was sampled earlier as compared to advice that was sampled later. Nevertheless, we observe $\hat{W}^K_{A_{k+1}} < \hat{W}^K_{A_k}$ for 16.24% of sampling trials in our empirical application. In summary, the sequential Bayesian account entails opposing kinds of serial weighting patterns depending on the time perspective. Future research could draw on empirically observed order effects (also from domains other than advice taking) to render the Step-by-Step compromising model more flexible with respect to serial positioning.

On the Completeness of the Model Universe

Whenever models are compared to gain insights about human cognition and behavior, one obvious question concerns the selection and specification of the models. While the present selection was based on previous knowledge about (single) advice taking behavior, we also obtained initial evidence suggesting that the four models may not capture all factors influencing confidence and weighting in multiple advice taking. The mean share of best strategy fits of choosing one's initial judgment (0.3844, 95% CI $[0.3459, 0.4229]$; Appendix C, Table C1) is significantly larger than the actual share of $E_K = E_0$ on 28.78% of all trials as indicated by the corresponding 95% CI of the former. Specifically, also relatively small shifts from E_0 are still better described by choosing-self

[7] Sum-to-one constraining implies that the Bayesian account does not allow for over- or underweighting of advice. Therefore, Bayesian updating has similar conceptual problems as the ratio-of-differences formula (see Equation 5) with absolute distances or post-hoc truncation to the $[0,1]$ interval (Rebholz et al., 2023). For instance, if it encourages additional search there can be good psychological reasons for "pushing away" from a certain piece of advice (Rader et al., 2015). Moreover, note that according to Equation 3 the initial judgment E_0 is weighted by $\prod_{k=1}^{K}(1-\hat{W}_{A_k})$ in $\hat{E}_K$.

than any other strategy. Unless (mean) advice is equal to initial judgment, which was rather unlikely with advice distance manipulations as in Experiment 5 of Ache (2017), all other strategies generally suggest shifting. Put differently, this divergence of predicted and actual behavior suggests that participants may apply additional strategies that were not contained in our model comparison. Empirically, equal weights averaging and the two choosing strategies are most often observed for single advice taking (Soll & Larrick, 2009). In multiple advice taking situations, however, there exist additional meaningful strategies. For instance, the choice of a specific subset (e.g., only one) of multiple advisors may be a consequence of memory constraints, completely at random, or caused by any other (unobserved) preferences. For instance, "consensus-based trimming" corresponds to ignoring the most extreme advisory judgments, where extremity may be defined with an egocentric bias as largest distance to one's own judgment (Yaniv & Milyavsky, 2007). An extended model comparison in future research should also include further reasonable strategies for sequential advice seeking.

Conclusion

In the present research, we theoretically and empirically compared different strategies of information integration to develop a better understanding of advice seeking. By reanalyzing data of an advice sampling experiment reported by Ache (2017), we found that End-of-Sequence compromising (i.e., simple cumulative averaging) predicts participants' actual final judgments best, followed by Step-by-Step compromising, choosing other peoples' judgments, and choosing one's own judgment. Interestingly, holistic distributional testing revealed pronounced inter- and intra-individual differences in strategy selection. Whereas some participants chose more often between their own and others' judgments, other participants were better described as compromisers between internal and external sources of information. Moreover, virtually all participants applied different advice taking strategies on different items and trials. The current approach provides insights into why this might be the case. For instance, the taking of multiple pieces of relatively more distant advice is better described by compromising than the taking of multiple pieces of close advice. Indeed, compromising often improves judgment—especially with a larger variance increasing the chances of two or more judgments bracketing the truth—due to the wisdom of crowds (Mannes, 2009; Soll & Larrick, 2009). In a nutshell, our data suggest that people make adaptive use of multiple, sequentially sampled external opinions.

References

Ache, F. (2017). *Returning advice taking to the wild: Empirical, theoretical, and normative implications of an ecological perspective* [Dissertation]. Eberhard Karls University of Tübingen. Tübingen, Germany. https://doi.org/10.15496/publikation-19538

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

Adjodah, D., Leng, Y., Chong, S. K., Krafft, P. M., Moro, E., & Pentland, A. (2021). Accuracy-risk trade-off due to social learning in crowd-sourced financial predictions. *Entropy*, *23*(7), 801. https://doi.org/10.3390/e23070801

Anderson, N. H. (1971). Integration theory and attitude change. *Psychological Review*, *78*(3), 171–206. https://doi.org/10.1037/h0030834

Anderson, N. H., & Jacobson, A. (1965). Effect of stimulus inconsistency and discounting instructions in personality impression formation. *Journal of Personality and Social Psychology*, *2*(4), 531–539. https://doi.org/10.1037/h0022484

Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal Psychology*, *41*, 258–290. https://doi.org/10.1037/h0055756

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

Bednarik, P., & Schultze, T. (2015). The effectiveness of imperfect weighting in advice taking. *Judgment and Decision Making*, *10*(3), 265–276. https://doi.org/10.1017/S1930297500004666

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*(9), 1214–1221. https://doi.org/10.1038/nn1954

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

Budescu, D. V., & Yu, H.-T. (2006). To Bayes or not to Bayes? A comparison of two classes of models of information aggregation. *Decision Analysis*, *3*(3), 145–162. https://doi.org/10.1287/deca.1060.0074

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

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

Denrell, J. (2005). Why most people disapprove of me: Experience sampling in impression formation. *Psychological Review*, *112*(4), 951–978. https://doi.org/10.1037/0033-295X.112.4.951

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

Fiedler, K. (2000). Beware of samples! A cognitive-ecological sampling approach to judgment biases. *Psychological Review*, *107*(4), 659–676. https://doi.org/10.1037//0033-295X.107.4.659

Fiedler, K., & Juslin, P. (2006). Taking the interface between mind and environment seriously. In K. Fiedler & P. Juslin (Eds.), *Information sampling and adaptive cognition* (pp. 3–29). Cambridge University Press. https://doi.org/10.1017/CBO9780511614576.001

Fiedler, K., & Kutzner, F. (2015). Information sampling and reasoning biases. In G. Keren & G. Wu (Eds.), *The Wiley Blackwell handbook of judgment and decision making* (pp. 380–403). John Wiley & Sons. https://doi.org/10.1002/9781118468333.ch13

Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nature Neuroscience*, *17*(5), 738–743. https://doi.org/10.1038/nn.3689

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). CRC Press Taylor and Francis Group.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, *98*(4), 506–528. https://doi.org/10.1037/0033-295X.98.4.506

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

Glanzer, M., & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior*, *5*(4), 351–360. https://doi.org/10.1016/S0022-5371(66)80044-0

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

Hausmann, D., & Läge, D. (2008). Sequential evidence accumulation in decision making: The individual desired level of confidence can explain the extent of information acquisition. *Judgment and Decision Making*, *3*(3), 229–243. https://doi.org/10.1017/S1930297500002436

Hemmer, P., Tauber, S., & Steyvers, M. (2015). Moving beyond qualitative evaluations of Bayesian models of cognition. *Psychonomic Bulletin & Review*, *22*(3), 614–628. https://doi.org/10.3758/s13423-014-0725-z

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

Hertwig, R. (2012). Tapping into the wisdom of the crowd—with confidence. *Science*, *336*(6079), 303–304. https://doi.org/10.1126/science.1221403

Hertwig, R., Pleskac, T. J., & Pachur, T. (2019). Reckoning with uncertainty: Our program of research. In R. Hertwig, T. J. Pleskac, & T. Pachur (Eds.), *Taming Uncertainty* (pp. 3–25). The MIT Press. https://doi.org/10.7551/mitpress/11114.003.0004

Herzog, S. M., & Hertwig, R. (2014). Harnessing the wisdom of the inner crowd. *Trends in Cognitive Sciences*, *18*(10), 504–506. https://doi.org/10.1016/j.tics.2014.06.009

Himmelstein, M. (2022). Decline, adopt or compromise? A dual hurdle model for advice utilization. *Journal of Mathematical Psychology*, *110*, 102695. https://doi.org/10.1016/j.jmp.2022.102695

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

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

Howe, P. D. L., Perfors, A., Walker, B., Kashima, Y., & Fay, N. (2022). Base rate neglect and conservatism in probabilistic reasoning: Insights from eliciting full distributions. *Judgment and Decision Making*, *17*(5), 962–987. https://doi.org/10.1017/S1930297500009281

Hütter, M., & Ache, F. (2016). Seeking advice: A sampling approach to advice taking. *Judgment and Decision Making*, *11*(4), 401–415. https://doi.org/10.1017/S193029750000382X

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

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

Koriat, A. (2012a). The self-consistency model of subjective confidence. *Psychological Review*, *119*(1), 80–113. https://doi.org/10.1037/a0025648

Koriat, A. (2012b). When are two heads better than one and why? *Science*, *336*(6079), 360–362. https://doi.org/10.1126/science.1216549

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 107–118. https://doi.org/10.1037/0278-7393.6.2.107

Kruschke, J. K. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2nd ed.). Academic Press.

Landrum, A. R., Eaves, B. S., & Shafto, P. (2015). Learning to trust and trusting to learn: A theoretical framework. *Trends in Cognitive Sciences*, *19*(3), 109–111. https://doi.org/10.1016/j.tics.2014.12.007

Larrick, R. P., & Soll, J. B. (2006). Intuitions about combining opinions: Misappreciation of the averaging principle. *Management Science*, *52*(1), 111–127. https://doi.org/10.1287/mnsc.1050.0459

Lejarraga, T., & Hertwig, R. (2021). How experimental methods shaped views on human competence and rationality. *Psychological Bulletin*, *147*(6), 535–564. https://doi.org/10.1037/bul0000324

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

Lisi, M., Mongillo, G., Milne, G., Dekker, T., & Gorea, A. (2021). Discrete confidence levels revealed by sequential decisions. *Nature Human Behaviour*, *5*(2), 273–280. https://doi.org/10.1038/s41562-020-00953-1

Luan, S., Schooler, L. J., & Tan, J. H. (2020). Improving judgment accuracy by sequential adjustment. *Psychonomic Bulletin & Review*, *27*(1), 170–177. https://doi.org/10.3758/s13423-019-01696-5

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

Miller, B., & Steyvers, M. (2011). The wisdom of crowds with communication. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *33*, 1292–1297. https://escholarship.org/uc/item/4jt6q62c

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. https://doi.org/10.1037/h0043158

Miller, J. M., & Krosnick, J. A. (1998). The impact of candidate name order on election outcomes. *Public Opinion Quarterly*, *62*(3), 291. https://doi.org/10.1086/297848

Minson, J. A., Liberman, V., & Ross, L. (2011). Two to tango: Effects of collaboration and disagreement on dyadic judgment. *Personality & Social Psychology Bulletin*, *37*(10), 1325–1338. https://doi.org/10.1177/0146167211410436

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

Ostwald, D., Starke, L., & Hertwig, R. (2015). A normative inference approach for optimal sample sizes in decisions from experience. *Frontiers in Psychology*, *6*, 1342. https://doi.org/10.3389/fpsyg.2015.01342

Payne, J. W., Johnson, E. J., & Bettman, J. R. (1993). *The adaptive decision maker.* Cambridge University Press.

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

Raiffa, H., & Schlaifer, R. (1970). *Applied statistical decision theory* (5th ed.). Harvard University.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Sage Publications.

Rauhut, H., & Lorenz, J. (2011). The wisdom of crowds in one mind: How individuals can simulate the knowledge of diverse societies to reach better decisions. *Journal of Mathematical Psychology*, *55*(2), 191–197. https://doi.org/10.1016/j.jmp.2010.10.002

Rebholz, T. R., Biella, M., & Hütter, M. (2023). *Mixed-effects regression weights (of advice): Process-consistent modeling of information sampling and utilization.* PsyArXiv. https://doi.org/10.31234/osf.io/x36az

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

Scheunemann, J., Fischer, R., & Moritz, S. (2021). Probing the hypersalience hypothesis—An adapted judge-advisor system tested in individuals with psychotic-like experiences. *Frontiers in Psychiatry*, *12*, 612810. https://doi.org/10.3389/fpsyt.2021.612810

Scheunemann, J., Gawęda, Ł., Reininger, K.-M., Jelinek, L., Hildebrandt, H., & Moritz, S. (2020). Advice weighting as a novel measure for belief flexibility in people with psychotic-like experiences. *Schizophrenia Research*, *216*, 129–137. https://doi.org/10.1016/j.schres.2019.12.016

Schrah, G. E., Dalal, R. S., & Sniezek, J. A. (2006). No decision-maker is an island: Integrating expert advice with information acquisition. *Journal of Behavioral Decision Making*, *19*(1), 43–60. https://doi.org/10.1002/bdm.514

Schultze, T., Mojzisch, A., & Schulz-Hardt, S. (2012). Why groups perform better than individuals at quantitative judgment tasks: Group-to-individual transfer as an alternative to differential weighting. *Organizational Behavior and Human Decision Processes*, *118*(1), 24–36. https://doi.org/10.1016/j.obhdp.2011.12.006

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., 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

Shanteau, J. C. (1970). An additive model for sequential decision making. *Journal of Experimental Psychology*, *85*(2), 181–191. https://doi.org/10.1037/h0029552

Shanteau, J. C. (1972). Descriptive versus normative models of sequential inference judgment. *Journal of Experimental Psychology*, *93*(1), 63–68. https://doi.org/10.1037/h0032509

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

Sniezek, J. A., Schrah, G. E., & Dalal, R. S. (2004). Improving judgement with prepaid expert advice. *Journal of Behavioral Decision Making*, *17*(3), 173–190. https://doi.org/10.1002/bdm.468

Soll, J. B., & Klayman, J. (2004). Overconfidence in interval estimates. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 299–314. https://doi.org/10.1037/0278-7393.30.2.299

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

Yaniv, I. (2004a). 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. (2004b). The benefit of additional opinions. *Current Directions in Psychological Science*, *13*(2), 75–78. https://doi.org/10.1111/j.0963-7214.2004.00278.x

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: Egocentric discounting and reputation formation. *Organizational Behavior and Human Decision Processes*, *83*(2), 260–281. https://doi.org/10.1006/obhd.2000.2909

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

Yik, M., Wong, K. F. E., & Zeng, K. J. (2019). Anchoring-and-adjustment during affect inferences. *Frontiers in Psychology*, *9*, 2567. https://doi.org/10.3389/fpsyg.2018.02567

Appendix A

Summary Statistics of Main Figures

Table A1

Summary Statistics of Absolute Prediction Error (APE) per Belief Distribution Parameter in Figure 2

Parameter	*Strategy*	*Obs.*	*M*	*95% CI*	*SD*	*Mdn*	*IQR*
E_K	CS	2488	0.2189	[0.2070, 0.2309]	0.3028	0.0967	0.3365
	SbS	2488	0.1341	[0.1266, 0.1416]	0.1915	0.0653	0.1434
	EoS	2488	0.1241	[0.1172, 0.1309]	0.1746	0.0632	0.1381
	CO	2488	0.1864	[0.1752, 0.1977]	0.2857	0.0800	0.2062
S_K	CS	2488	0.1127	[0.1063, 0.1190]	0.1609	0.0711	0.1286
	SbS	2488	0.1180	[0.1131, 0.1229]	0.1247	0.0790	0.1248
	EoS	2488	0.1284	[0.1226, 0.1342]	0.1478	0.0822	0.1372
	CO	2488	0.1950	[0.1878, 0.2023]	0.1838	0.1398	0.2103

Note. E_K = Final Judgment, S_K = Final Standard Deviation, CS = Choosing Self, SbS = Step-by-Step Compromising, EoS = End-of-Sequence Compromising, CO = Choosing Others.

Table A2

Summary Statistics of Absolute Prediction Error (APE) per Belief Distribution Parameter for Close Advice in the Top Panel of Figure 3

Parameter	*Strategy*	*Obs.*	*M*	*95% CI*	*SD*	*Mdn*	*IQR*
E_K	CS	1242	0.0550	[0.0482, 0.0617]	0.1214	0.0080	0.0690
	SbS	1242	0.0606	[0.0554, 0.0657]	0.0930	0.0364	0.0571
	EoS	1242	0.0524	[0.0472, 0.0575]	0.0932	0.0267	0.0485
	CO	1242	0.0791	[0.0716, 0.0867]	0.1356	0.0467	0.0737
S_K	CS	1242	0.1093	[0.0999, 0.1187]	0.1688	0.0601	0.1367
	SbS	1242	0.1127	[0.1061, 0.1194]	0.1195	0.0749	0.1215
	EoS	1242	0.0832	[0.0770, 0.0895]	0.1120	0.0515	0.0854
	CO	1242	0.1845	[0.1745, 0.1945]	0.1801	0.1277	0.2166

Note. E_K = Final Judgment, S_K = Final Standard Deviation, CS = Choosing Self, SbS = Step-by-Step Compromising, EoS = End-of-Sequence Compromising, CO = Choosing Others.

Table A3

Summary Statistics of Absolute Prediction Error (APE) per Belief Distribution Parameter for Intermediate Advice in the Bottom Panel of Figure 3

Parameter	*Strategy*	*Obs.*	*M*	*95% CI*	*SD*	*Mdn*	*IQR*
E_K	CS	1246	0.3824	[0.3636, 0.4012]	0.3390	0.3185	0.3567
	SbS	1246	0.2074	[0.1945, 0.2203]	0.2321	0.1347	0.2154
	EoS	1246	0.1955	[0.1841, 0.2069]	0.2049	0.1349	0.1845
	CO	1246	0.2934	[0.2740, 0.3128]	0.3490	0.1916	0.3429
S_K	CS	1246	0.1160	[0.1075, 0.1245]	0.1526	0.0711	0.1230
	SbS	1246	0.1233	[0.1161, 0.1305]	0.1295	0.0816	0.1292
	EoS	1246	0.1735	[0.1643, 0.1826]	0.1646	0.1364	0.1694
	CO	1246	0.2055	[0.1951, 0.2159]	0.1870	0.1526	0.2129

Note. E_K = Final Judgment, S_K = Final Standard Deviation, CS = Choosing Self, SbS = Step-by-Step Compromising, EoS = End-of-Sequence Compromising, CO = Choosing Others.

Table A4

Summary Statistics of Kullback-Leibler Divergence (KLD) for Holistic Distributional Testing in Figure 5

Strategy	*Obs.*	*M*	*95% CI*	*SD*	*Mdn*	*IQR*
CS	2488	199.3867	[-115.2138, 513.9872]	8002.4899	0.6118	2.5823
SbS	2488	68.4433	[-19.6484, 156.5350]	2240.7873	0.5081	0.8552
EoS	2488	230.4140	[-10.7980, 471.6259]	6135.7059	0.2864	1.4756
CO	1947	6.0356	[1.1574, 10.9139]	109.7566	1.2725	1.6200

Note. CS = Choosing Self, SbS = Step-by-Step Compromising, EoS = End-of-Sequence Compromising, CO = Choosing Others. As the advice variance is zero for $K = 1$, infinite values of KLD are excluded for CO.

Table A5

Summary Statistics of the Shares of Best Strategy Fits in Figure 6

Strategy	*Obs.*	*M*	*95% CI*	*SD*	*Mdn*	*IQR*
CS	127	0.2800	[0.2394, 0.3205]	0.2307	0.2500	0.3000
SbS	127	0.2919	[0.2564, 0.3275]	0.2024	0.3000	0.3361
EoS	127	0.3100	[0.2776, 0.3423]	0.1842	0.3000	0.2500
CO	127	0.1181	[0.0866, 0.1497]	0.1798	0.0500	0.1500

Note. CS = Choosing Self, SbS = Step-by-Step Compromising, EoS = End-of-Sequence Compromising, CO = Choosing Others.

Table A6

Summary Statistics of the Shares of Best Strategy Fits for Close Advice in the Top Panel of Figure 7

Strategy	*Obs.*	*M*	*95% CI*	*SD*	*Mdn*	*IQR*
CS	127	0.3725	[0.3229, 0.4220]	0.2822	0.3000	0.4444
SbS	127	0.2438	[0.2042, 0.2833]	0.2254	0.2000	0.4000
EoS	127	0.3056	[0.2685, 0.3427]	0.2113	0.3000	0.2000
CO	127	0.0782	[0.0540, 0.1024]	0.1379	0.0000	0.1000

Note. CS = Choosing Self, SbS = Step-by-Step Compromising, EoS = End-of-Sequence Compromising, CO = Choosing Others.

Table A7

Summary Statistics of the Shares of Best Strategy Fits for Intermediate Advice in the Bottom Panel of Figure 7

Strategy	*Obs.*	*M*	*95% CI*	*SD*	*Mdn*	*IQR*
CS	127	0.1881	[0.1468, 0.2294]	0.2354	0.1000	0.2750
SbS	127	0.3383	[0.2965, 0.3801]	0.2381	0.3000	0.3000
EoS	127	0.3144	[0.2736, 0.3552]	0.2323	0.3000	0.4000
CO	127	0.1593	[0.1146, 0.2040]	0.2545	0.0000	0.2000

Note. CS = Choosing Self, SbS = Step-by-Step Compromising, EoS = End-of-Sequence Compromising, CO = Choosing Others.

Table A8

Summary Statistics of Final Sample Size per Best-Fitting Strategy in Figure 9

Strategy	*Obs.*	*M*	*95% CI*	*SD*	*Mdn*	*IQR*
CS	698	5.0659	[4.6826, 5.4492]	5.1582	3.0000	6.0000
SbS	730	4.4219	[4.0907, 4.7531]	4.5584	3.0000	4.0000
EoS	771	7.4734	[7.0509, 7.8959]	5.9759	6.0000	8.0000
CO	289	7.0484	[6.4402, 7.6567]	5.2537	5.0000	6.0000

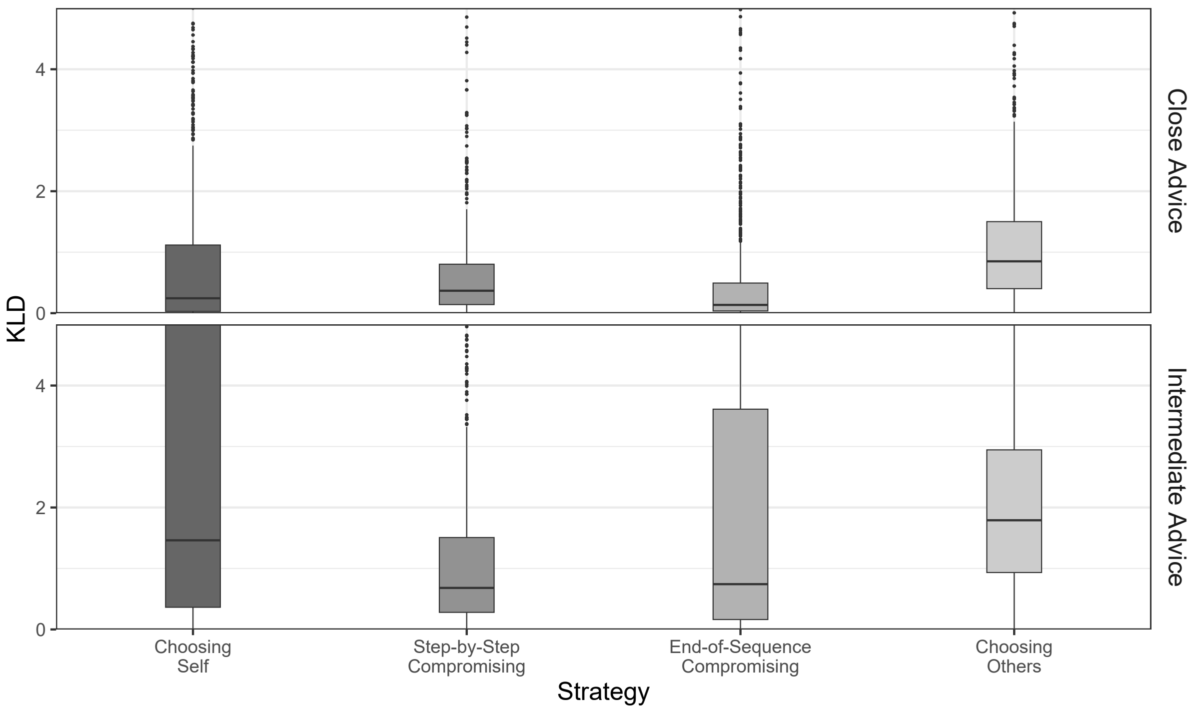
Note. CS = Choosing Self, SbS = Step-by-Step Compromising, EoS = End-of-Sequence Compromising, CO = Choosing Others.

Appendix B

Additional Results of Holistic Distributional Testing

Figure B1

Performance for Holistic Belief Distribution Predictions per Strategy and Advice Distance Condition as Measured by Kullback-Leibler Divergence (KLD)



Note. Plotting is truncated for $KLD > 5$. As the advice variance is zero for $K = 1$, infinite values of KLD are excluded for choosing others. Summary statistics can be found in Tables B1 and B2.

Table B1

Summary Statistics of Kullback-Leibler Divergence (KLD) for Holistic Distributional Testing of Close Advice in the Top Panel of Figure B1

Strategy	*Obs.*	*M*	*95% CI*	*SD*	*Mdn*	*IQR*
CS	1242	4.4042	[2.0479, 6.7606]	42.3282	0.2249	1.0714
SbS	1242	0.9242	[0.6730, 1.1753]	4.5113	0.3679	0.6618
EoS	1242	250.6925	[-112.3893, 613.7743]	6522.1891	0.1352	0.4576
CO	920	1.3977	[1.1209, 1.6745]	4.2783	0.8501	1.0989

Note. CS = Choosing Self, SbS = Step-by-Step Compromising, EoS = End-of-Sequence Compromising, CO = Choosing Others. As the advice variance is zero for $K = 1$, infinite values of KLD are excluded for CO.

Table B2

Summary Statistics of Kullback-Leibler Divergence (KLD) for Holistic Distributional Testing of Intermediate Advice in the Bottom Panel of Figure B1

Strategy	*Obs.*	*M*	*95% CI*	*SD*	*Mdn*	*IQR*
CS	1246	393.7433	[-234.6886, 1022.1752]	11306.9827	1.4497	5.8265
SbS	1246	135.7456	[-40.1958, 311.6871]	3165.6050	0.6811	1.2266
EoS	1246	210.2005	[-108.1079, 528.5090]	5727.1250	0.7429	3.4496
CO	1027	10.1904	[0.9455, 19.4352]	150.9818	1.7905	2.0115

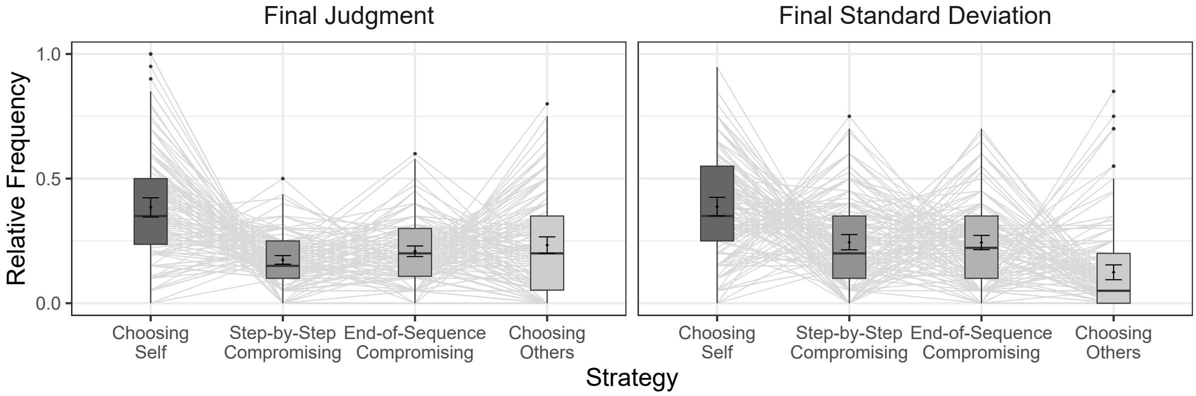
Note. CS = Choosing Self, SbS = Step-by-Step Compromising, EoS = End-of-Sequence Compromising, CO = Choosing Others. As the advice variance is zero for $K = 1$, infinite values of KLD are excluded for CO.

Appendix C

Additional Results of Parameter-Wise Model Comparisons

Figure C1

Relative Shares of Best Strategy Fits per Belief Distribution Parameter and Participant



Note. The thin gray lines correspond to individual participants and thus capture inter- and intra-individual idiosyncrasies. Box plots are accompanied by means and 95% CIs. Summary statistics can be found in Table C1.

Table C1

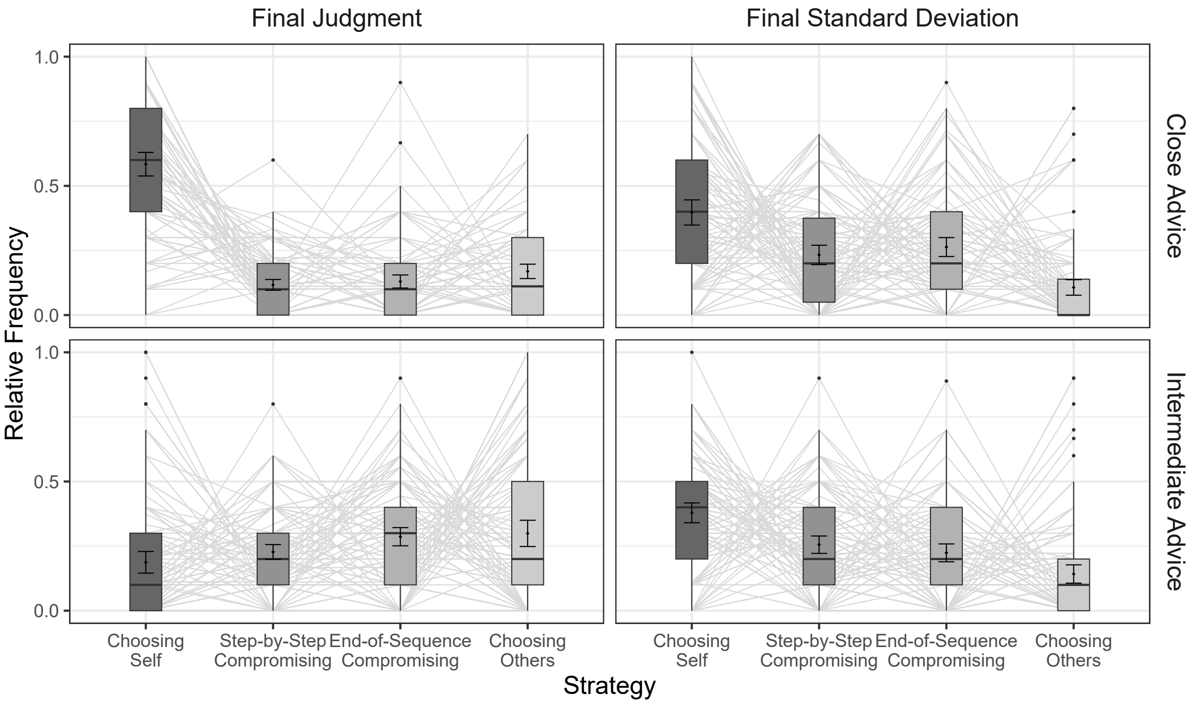
Summary Statistics of the Shares of Best Strategy Fits in Figure C1

Parameter	*Strategy*	*Obs.*	*M*	*95% CI*	*SD*	*Mdn*	*IQR*
E_K	CS	127	0.3844	[0.3459, 0.4229]	0.2193	0.3500	0.2639
	SbS	127	0.1739	[0.1568, 0.1910]	0.0975	0.1500	0.1500
	EoS	127	0.2085	[0.1872, 0.2298]	0.1212	0.2000	0.1918
	CO	127	0.2332	[0.2001, 0.2663]	0.1884	0.2000	0.2974
S_K	CS	127	0.3877	[0.3504, 0.4250]	0.2124	0.3500	0.3000
	SbS	127	0.2449	[0.2143, 0.2754]	0.1739	0.2000	0.2500
	EoS	127	0.2434	[0.2147, 0.2721]	0.1636	0.2222	0.2500
	CO	127	0.1241	[0.0943, 0.1539]	0.1697	0.0500	0.2000

Note. E_K = Final Judgment, S_K = Final Standard Deviation, CS = Choosing Self, SbS = Step-by-Step Compromising, EoS = End-of-Sequence Compromising, CO = Choosing Others.

Figure C2

Relative Shares of Best Strategy Fits per Belief Distribution Parameter, Participant, and Advice Distance Condition



Note. The thin gray lines correspond to individual participants and thus capture inter- and intra-individual idiosyncrasies. Box plots are accompanied by means and 95% CIs. Summary statistics can be found in Tables C2 and C3 for close and intermediate advice distance, respectively.

Table C2

Summary Statistics of the Shares of Best Strategy Fits for Close Advice in the Top Panel of Figure C2

Parameter	*Strategy*	*Obs.*	*M*	*95% CI*	*SD*	*Mdn*	*IQR*
E_K	CS	127	0.5839	[0.5383, 0.6296]	0.2599	0.6000	0.4000
	SbS	127	0.1169	[0.0961, 0.1378]	0.1186	0.1000	0.2000
	EoS	127	0.1300	[0.1046, 0.1553]	0.1442	0.1000	0.2000
	CO	127	0.1692	[0.1413, 0.1970]	0.1584	0.1111	0.3000
S_K	CS	127	0.3970	[0.3484, 0.4457]	0.2772	0.4000	0.4000
	SbS	127	0.2328	[0.1953, 0.2702]	0.2134	0.2000	0.3250
	EoS	127	0.2633	[0.2266, 0.3000]	0.2088	0.2000	0.3000
	CO	127	0.1069	[0.0769, 0.1369]	0.1708	0.0000	0.1389

Note. E_K = Final Judgment, S_K = Final Standard Deviation, CS = Choosing Self, SbS = Step-by-Step Compromising, EoS = End-of-Sequence Compromising, CO = Choosing Others.

Table C3

Summary Statistics of the Shares of Best Strategy Fits for Intermediate Advice in the Bottom Panel of Figure C2

Parameter	*Strategy*	*Obs.*	*M*	*95% CI*	*SD*	*Mdn*	*IQR*
E_K	CS	127	0.1872	[0.1453, 0.2291]	0.2386	0.1000	0.3000
	SbS	127	0.2273	[0.1989, 0.2558]	0.1621	0.2000	0.2000
	EoS	127	0.2863	[0.2511, 0.3215]	0.2004	0.3000	0.3000
	CO	127	0.2992	[0.2484, 0.3500]	0.2892	0.2000	0.4000
S_K	CS	127	0.3787	[0.3402, 0.4171]	0.2189	0.4000	0.3000
	SbS	127	0.2554	[0.2217, 0.2891]	0.1920	0.2000	0.3000
	EoS	127	0.2239	[0.1893, 0.2585]	0.1969	0.2000	0.3000
	CO	127	0.1420	[0.1066, 0.1774]	0.2014	0.1000	0.2000

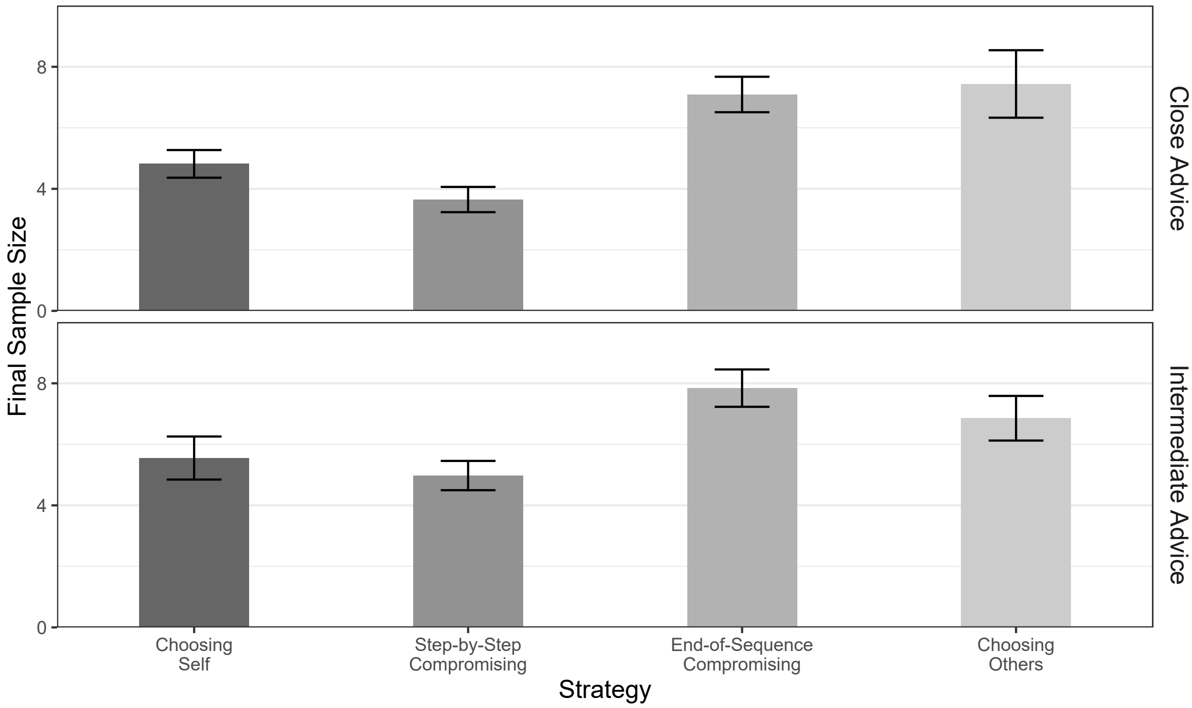
Note. E_K = Final Judgment, S_K = Final Standard Deviation, CS = Choosing Self, SbS = Step-by-Step Compromising, EoS = End-of-Sequence Compromising, CO = Choosing Others.

Appendix D

Additional Sampling Results

Figure D1

Final Sample Size per Advice Distance Condition and Best-Fitting Strategy as Measured by Kullback-Leibler Divergence (KLD)



Note. Error bars show the 95% CIs. Summary statistics can be found in Tables D1 and D2 for close and intermediate advice distance, respectively.

Table D1

Summary Statistics of Final Sample Size per Best-Fitting Strategy for Close Advice in the Top Panel of Figure D1

Strategy	*Obs.*	*M*	*95% CI*	*SD*	*Mdn*	*IQR*
CS	462	4.8182	[4.3646, 5.2718]	4.9612	3.0000	6.0000
SbS	305	3.6492	[3.2354, 4.0629]	3.6722	2.0000	3.0000
EoS	379	7.0923	[6.5113, 7.6734]	5.7533	5.0000	8.0000
CO	96	7.4375	[6.3317, 8.5433]	5.4576	5.0000	8.0000

Note. CS = Choosing Self, SbS = Step-by-Step Compromising, EoS = End-of-Sequence Compromising, CO = Choosing Others.

Table D2

Summary Statistics of Final Sample Size per Best-Fitting Strategy for Intermediate Advice in the Bottom Panel of Figure D1

Strategy	*Obs.*	*M*	*95% CI*	*SD*	*Mdn*	*IQR*
CS	236	5.5508	[4.8452, 6.2565]	5.5022	3.5000	7.0000
SbS	425	4.9765	[4.4967, 5.4562]	5.0319	3.0000	4.0000
EoS	392	7.8418	[7.2293, 8.4544]	6.1686	6.0000	9.2500
CO	193	6.8549	[6.1233, 7.5865]	5.1528	5.0000	6.0000

Note. CS = Choosing Self, SbS = Step-by-Step Compromising, EoS = End-of-Sequence Compromising, CO = Choosing Others.

D.3 Manuscript III

Rebholz, T. R., Biella, M., & Hütter, M. (2023). *Mixed-effects regression weights (of advice): Process-consistent modeling of information sampling and utilization.* PsyArXiv. https://doi.org/10.31234/osf.io/x36az

Mixed-Effects Regression Weights (of Advice): Process-Consistent Modeling of Information Sampling and Utilization

Tobias R. Rebholz, Marco Biella, and Mandy Hütter

Psychology Department, Eberhard Karls University of Tübingen

Author Note

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

Marco Biella is now at the Faculty of Business and Economics, University of Basel, Switzerland.

Reproducible analysis scripts are publicly available at the Open Science Framework repository (https://osf.io/6gmhs). We have no known conflicts of interest to disclose. This research was funded by the Deutsche Forschungsgemeinschaft (DFG), grant 2277, Research Training Group "Statistical Modeling in Psychology" (SMiP).

Correspondence concerning this article should be addressed to Tobias R. Rebholz, Psychology Department, Eberhard Karls University of Tübingen, Schleichstr. 4, 72076 Tübingen, Germany. Email: tobias.rebholz@uni-tuebingen.de

Abstract

Advice taking and related research are dominated by deterministic weighting indices such as ratio-of-differences-based formulas for investigating informational influence. They are intuitively simple but entail various measurement problems and restrict research to a certain paradigmatic approach. As a solution, we propose process-consistent mixed-effects regression modeling for specifying how strongly peoples' judgment is influenced by external information. Our formal derivation of the proposed weighting measures is accompanied by a detailed elaboration on their most important technical and statistical subtleties. Essentially, the approach differentiates between components of endogenous (i.e., final judgment) and exogenous (e.g., initial judgment and advice) nature by relying on accordingly specified multilevel models. Corresponding mixed-effects regression coefficients of various exogenous sources of information hence also reflect individual weighting but are based on a conceptually consistent representation of the endogenous judgment process. We use this modeling approach to revisit empirical findings from sequential collaboration and advice taking paradigms. Specifically, whereas we do not obtain evidence for systematic order effects in sequential collaboration, we document recency effects in the weighting of sequentially sampled advice. We argue that process-consistent modeling of information sampling and utilization has the potential to increase the replicability of our science and opens up new avenues for innovative research. Moreover, the proposed method is relevant beyond sequential collaboration and advice taking. Mixed-effects regression weights can also inform research on related cognitive phenomena such as multidimensional belief updating, anchoring effects, hindsight biases, or attitude change.

Keywords: weight of advice, advice taking, belief updating, information sampling, judge-advisor system, multilevel modeling

Mixed-Effects Regression Weights (of Advice): Process-Consistent Modeling of Information Sampling and Utilization

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.

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 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 own initial beliefs and advice is supposedly taken as evidence for a certain level of initial bias which may be compensated by assimilating external evidence into one's own 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 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; but 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, 2004a; Yaniv & Milyavsky, 2007) due to conceptual and paradigmatic similarities.

Harvey and Fischer's (1997) advice taking index dominated the recent literature (e.g., Hütter & Ache, 2016; Schultze et al., 2015; Soll & Larrick, 2009; see Bailey et al., 2022, for a review). Their index, as formally introduced below, reflects how strongly people adapted their judgment (i.e., the endogenous component) in units of the distance between the initial judgment and the advice (i.e., exogenous components). We suspect that it is the simplicity of this ratio-of-differences (ROD) formula paired with its capability to capture (inter- and intra-)individual differences[1] that is responsible for its popularity. The same reasoning applies to other popular 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). However, the specific arithmetics of building intermixed (i.e., final judgment vs. advice and initial judgment) ratios of intermixed difference scores imply a number of conceptual and measurement problems (Cronbach, 1943; Cronbach & Furby, 1970; Edwards, 1995; Firebaugh & Gibbs, 1985). For instance, implicit equal weighting conceals the relative variance contributions of the difference score components which implies conceptual ambiguity (Edwards, 1994, 1995). Moreover, outcome ambiguity occurs when

[1] With "individual differences" capability we refer to enabling the calculation of individual values that describe behavior on the level of trials (Baayen et al., 2008; Bauer, 2011). Consequently, corresponding measures can capture idiosyncrasies of persons (and items etc.) in specific situations (see also Kämmer et al., 2023).

separate effects of independent variables on the difference score components are reduced to a single coefficient.

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 & 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 own initial judgments in the forecasting literature (e.g., Lim & O'Connor, 1995). The major limitation of regression-based approaches as available back then, however, was their aggregate data analysis scheme (Bonaccio & Dalal, 2006). Mixed-effects regression[2] instead allows to simultaneously model participant and stimulus item variation (Baayen et al., 2008; Raudenbush & Bryk, 2002). Hence 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. Additionally, regression can easily handle unbalanced or missing data, accommodate arbitrary types of predictor and response variables,[3] and has many more desirable properties.

[2] The terms multilevel, mixed-effects, and hierarchical modeling or regression all refer to the same statistical procedure in which coefficients comprise fixed and random (i.e., mixed) components.

[3] The ROD formula is extremely limited in explanatory power for choices among a set of discrete, qualitatively different alternatives (Bailey et al., 2022). For instance, it merely describes "matching" in terms of acceptance versus disregard of advice in binary choice (Sniezek & Buckley, 1995). Instead, implementing appropriate link functions (e.g., logit) for modeling qualitative decisions in a generalized multilevel regression framework enables more informative weighting parameters, for instance, in terms of choice probabilities or odds ratios. Therefore, we will 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 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 approach.

A Mixed-Effects Regression Model of 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:

$$\omega_{A,ij} = \frac{F_{ij} - I_{ij}}{A_{ij} - I_{ij}} \tag{1}$$

where I_{ij} and F_{ij} indicate the initial and final judgments of a participant $i = 1, \ldots, N$ about a given stimulus item $j = 1, \ldots, M$, 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. Hence, $\omega_{A,ij} = 1$ indicates complete adoption, $\omega_{A,ij} = 0$ entire disregard, and everything in between corresponding weighting of advice with $\omega_{A,ij} \in (0, 1)$.[4] Exhibiting conceptual

[4] Theoretically, $\omega_{A,ij} \notin [0, 1]$ 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).

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 response linear processing scheme as becomes apparent by rearranging its terms to account for endogenous formation of final judgments:

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

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

(Hoffman, 1960). Accordingly, 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}). Hence 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 response linear processing that relies on a weighting policy that (adaptively) compromises 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.

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

$$F_{ij} = \omega_{A,ij}A_{ij} + (1 - \omega_{A,ij})I_{ij} + \varepsilon_{ij}, \tag{4}$$

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 overall error ε

such that individual regression coefficients of the form

$$\omega_{A,ij} = \beta_A + \alpha^S_{A,i} + \alpha^T_{A,j}, \tag{5}$$

become admissible (Baayen et al., 2008; Bates et al., 2015; Brown et al., 2018; Raudenbush & Bryk, 2002).[5] On the weighting level, β_A denotes the fixed effect of advice on final judgment and $\alpha^q_A \sim N\left(0, \tau^2_{A,q}\right)$ the random effects of participants S and stimulus items T, respectively, with $\tau^2_{A,S}$ and $\tau^2_{A,T}$ mutually independent.

The multilevel regression model as specified in Equations 4 and 5 formalizes the assumption 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, people differ in their selection of advice taking strategies from a disjunct set including complete adoption, disregard, and equal weighting (Soll & Larrick, 2009). However, idiosyncratic characteristics of stimulus items may also influence strategy selection or weighting. Specifically, participants may be more knowledgeable in judging a certain item due to experiences with the underlying judgment domain and thus need and/or accept less help. For instance, in Experiment 2 of Ache et al. (2020) it is easier for participants to judge 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.

Thus far we have mostly argued *how* it is feasible to account for endogeneity and individual differences in advice taking by means of multilevel regression. Most important, however, there are multiple methodological and theoretical considerations that give rise to

[5] Formally, the multilevel regression-based estimator of WOA is defined as $\hat{\omega}_{A,ij} = \hat{\beta}_A + \hat{\alpha}^S_{A,i} + \hat{\alpha}^T_{A,j}$, which denotes (empirical Bayes) estimated mixed-effects regression 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 $\hat{\alpha}^S_{A,i}$ and $\hat{\alpha}^T_{A,j}$. As such, the individual weight estimates $\hat{\omega}_{A,ij}$ are treated as observed draws of random variables on the respective grouping level.

the practical relevance of the proposed multilevel modeling approach. For instance, sum-to-one constrained weighted averaging on the judgment level (Equation 4) and additivity on the weighting level (Equation 5) eventually appear overly restrictive. Accordingly, we will discuss two of the most valuable extensions of the multilevel modeling framework to address such restrictions in the following. Thereby relying on the reanalysis of empirical examples will showcase the practical applicability and substantiate the merit of process-consistent regression modeling of advice taking behavior. A significance level of 5% will be used for statistical testing throughout. Reproducible analysis scripts are publicly available online (https://osf.io/6gmhs).

Beyond Relativity to the Self

Harvey and Fischer (1997) called it "perverse" advice taking when a final judgment does not lie strictly in between initial judgment and advice. For Equation 1, negative weights result from eventually shifting in the opposite direction than advised. Similarly, WOA larger than one indicates overshooting the advice. Researchers often either conceal perverse advice taking behavior by taking a ratio-of-absolute-differences approach with weighting specified as

$$\tilde{\omega}_{A,ij} = \frac{|F_{ij} - I_{ij}|}{|A_{ij} - I_{ij}|} \tag{6}$$

(e.g., Gino, 2008; Yaniv, 2004b) or discard it actively by winsorizing ("truncating") negative values to zero and values larger than one to one (e.g., Gino & Schweitzer, 2008; Gino et al., 2009; Schultze et al., 2015; Soll & Larrick, 2009). Both approaches suffer from potentially yielding undefined or ambiguous values. Essentially, the same observation can be interpreted in strongly contrasting ways. For instance, shifting from 100 to 90 in spite of advice of 110 is considered non-weighting in the truncation approach whereas it is considered full weighting in the ratio-of-absolute-differences approach. Therefore, Bonaccio and Dalal (2006) recommend analyzing the data twice, with and without "problematic"

WOA values, to guarantee 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 conclusion, we argue against pre-processing or excluding "odd" observations in terms of weighting.

Unconstrained Regression

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 individual weights for each individual source of information. The system of Equations 4 and 5 is rewritten as:

$$F_{ij} = \omega_{A,ij} A_{ij} + \omega_{I,ij} I_{ij} + \varepsilon_{ij} \tag{7}$$

and

$$\omega_{p,ij} = \beta_p + \alpha^S_{p,i} + \alpha^T_{p,j}, \tag{8}$$

where $p \in \{A, I\}$ indicates the weights of advice and the self with fixed effects β_p and random effects $\alpha^q_p \sim N_{\|p\|}(\mathbf{0}_{2\times1}, \mathbf{\Sigma}_q)$, $q \in \{S, T\}$, where $\mathbf{0}_{2\times1}$ is the zero vector. By assumption, the covariance matrices of participants $q = S$ and stimulus items $q = T$

$$\boldsymbol{\Sigma}_q = \begin{bmatrix} \tau^2_{A,q} & \tau_{IA,q} \\ \tau_{AI,q} & \tau^2_{I,q} \end{bmatrix} \tag{9}$$

are mutually independent.

The coefficient regressions of Equation 8 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. Hence, the approach still allows comparing the weight of advice to the weight of the initial judgment. Moreover, sum-to-one constraining can be restored as follows:

$$\tilde{\omega}_{A,ij} = \frac{\hat{\omega}_{A,ij}}{\hat{\omega}_{A,ij} + \hat{\omega}_{I,ij}} \tag{10}$$

Applying the divide-by-total principle post hoc (cf. Harvey et al., 2000) restores the original scaling and hence 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 oneself by estimating WOAs from unconstrained regressions, it is possible to consider alternative formulations of the judgment formation model in Equation 7.

Partial effects $\omega_{p,ij}$ are particularly relevant under the following three circumstances. First, more than the classic 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; Prahl & 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., Rebholz & Hütter, 2022; Schultze et al., 2015). 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 those situations, ROD-calculus cannot be applied, which is the main focus of the following empirical application.

Empirical Application

A blind spot of advice taking research is informational influence without prior formulation of independent judgments. Although it is not explicitly framed like this, the "sequential collaboration" experiments of Mayer and Heck (2022) are procedurally equivalent to the traditional JAS. In their paradigm, participants had to answer general knowledge questions (Experiments 1 & 2; e.g., "How tall is the Eiffel Tower?") or locate cities on maps (Experiment 3) receiving the estimate, that is, essentially advice of a previous participant in the sequential collaboration condition. One major difference to classic advice taking studies is that no independent initial judgments were required. Consequently, informational influence could not be assessed by indices of relative positioning such as ROD.[6] Instead, by applying a performance perspective, the original study provided evidence for increasing judgment accuracy along sequential collaboration chains. By reanalyzing data of their first two experiments with the proposed regression approach, we will contribute to resolving an important blind spot of advice taking research.

Method

Essentially, individual weights for individual sources of information can be calculated by means of mixed-effects regression also when advice is the only *observed*

[6] For the original accuracy or an anchoring perspective, an index that is informative in terms of relative positioning could be calculated with respect to the true value of an item (e.g., "0-1-scores;" Röseler et al., 2022).

source of information. In a bivariate regression model specified as

$$F_{ij} = \omega_{A,ij} A_{ij} + \varepsilon_{ij}, \tag{11}$$

unobserved initial judgments are part of the residual terms ε_{ij}. Most important,

$$\omega_{A,ij} = \beta_A + \alpha^S_{A,i} + \alpha^T_{A,j} \tag{12}$$

captures the partial effect of advice on final judgment in this model. Put differently, weighting was operationalized as the change in F in response to a unit change in A.[7] 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 complete adoption of advice or, in other words, congruence of both estimates. 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 over-responsiveness to) advice.

Changes in informational influence are the most interesting comparison for advice taking in sequential chains. Whereas Mayer and Heck (2022) found that the change probability decreases and that, relative to the true value, estimates are changed less in absolute terms for later collaboration, partial effects may tell a different story. Therefore, we also built a model that included an interaction term of advice and chain position $c_i = 2, \ldots, C$ (with $C = 4$ in Experiment 1 and $C = 6$ in Experiment 2) such that

$$\omega_{A,ij}(c_i) = \beta_A + \alpha^S_{A,i} + \alpha^T_{A,j} + \beta_{A\times c}(c_i - 1). \tag{13}$$

Subtracting one chain position implies that β_A measures weighting during the first inter-individual interactions where estimates of participants at position 1 were provided as advice for participants at position 2 of a chain. Observations with $c_i = 1$ were excluded as

[7] To promote model convergence, we applied item-wise z-standardization to the raw judgments instead of their distances from the true values as applied in the original study. Consequently, unit changes were measured in standard deviations and the intercept could be set to zero in Equation 11.

no advice was provided. Technically, only Experiment 2 of Mayer and Heck (2022) satisfies the practical recommendations about a minimum number of five factor levels for precise estimation of random effects variances (Bolker, 2015; see also Oberpriller et al., 2022). Therefore, the results for alternatively estimating random positioning effects via

$$\omega_{A,ijc} = \beta_A + \alpha^S_{A,i} + \alpha^T_{A,j} + \alpha^U_{A,c}, \tag{14}$$

where $\alpha^U_A \sim N\left(0, \tau^2_{A,U}\right)$, can be found in Appendix A.

Results

To extend the original research beyond the investigation of accuracy gains for general knowledge questions via sequential collaboration (Mayer & Heck, 2022), we applied the model in Equations 11 and 12 to investigate informational influences. In Experiment 2, the mean weighting was slightly larger than in Experiment 1 as indicated by $\hat{\beta}_A$ (Table 1). 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. In classic advice taking research, the distribution of weighting is of 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 are left-skewed with modes at full responsiveness (i.e., $\omega_{A,ij} = 1$) to judgments of previous participants in both sequential collaboration experiments (Figure 1). Nevertheless, mean responsiveness was significantly smaller than one as indicated by the corresponding 95% *CI* (Appendix A, Tables A1 & A2). Although participants are quite responsive to advice in most trials and almost never completely unresponsive, the distribution is a little wider in Experiment 2.

Replacing Equation 12 by Equation 13 in the multilevel regression models as specified in Equation 11, there was no evidence for a fixed positioning effect on MER-WOA. In both experiments, the effect of chain position on weighting, $\hat{\beta}_{A\times c}$, was not significantly different from zero (Table 2). Descriptively, the positioning effects pointed in

Table 1

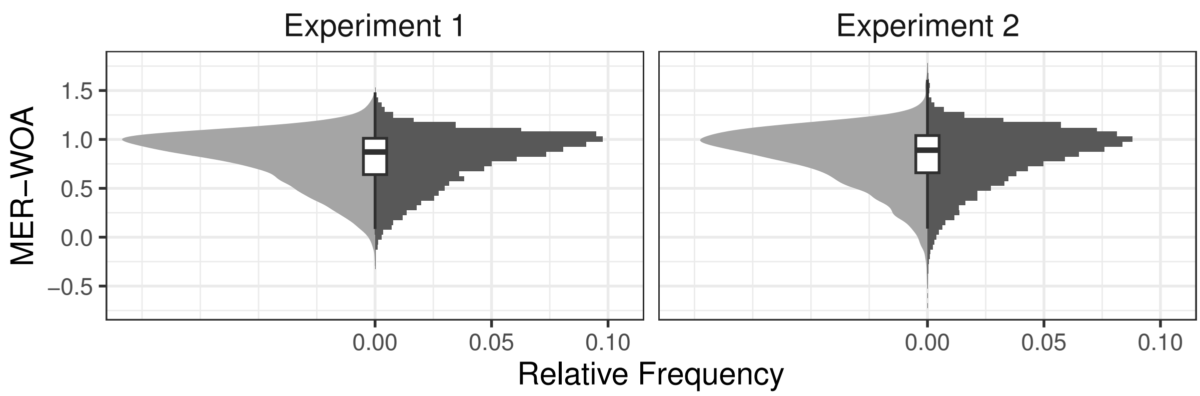
Fixed Effects of Multilevel Models of Final Judgment According to Equations 11 and 12 of Experiments 1 and 2 of Mayer and Heck (2022)

	Experiment 1		Experiment 2	
	Estimate	*SE*	*Estimate*	*SE*
β_A	0.8099 ***	0.0304	0.8241 ***	0.0248

Note. $^*p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$. The full models can be found in Appendix A, Tables A1 and A2.

Figure 1

Distributions of Mixed-Effects Regression Weights of Advice (MER-WOA) in all Trials of Experiments 1 and 2 of Mayer and Heck (2022)



Note. Outliers are not displayed in the box plots.

the same directions 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 exclusively indicated for initial or earlier interactions, respectively. In later interactions, the variances of the distributions stabilize at a lower level than for initial interactions.

Table 2

Fixed Effects of Multilevel Models of Final Judgment According to Equation 11 With Fixed Positioning Effects as Specified in Equation 13 for Experiments 1 and 2 of Mayer and Heck (2022)

	Experiment 1		Experiment 2	
	Estimate	*SE*	*Estimate*	*SE*
β_A	0.7064 ***	0.0704	0.7756 ***	0.0462
$\beta_{A\times c}$	0.0520	0.0320	0.0165	0.0133

Note. *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$. The full models can be found in Appendix A, Tables A3 and A4.

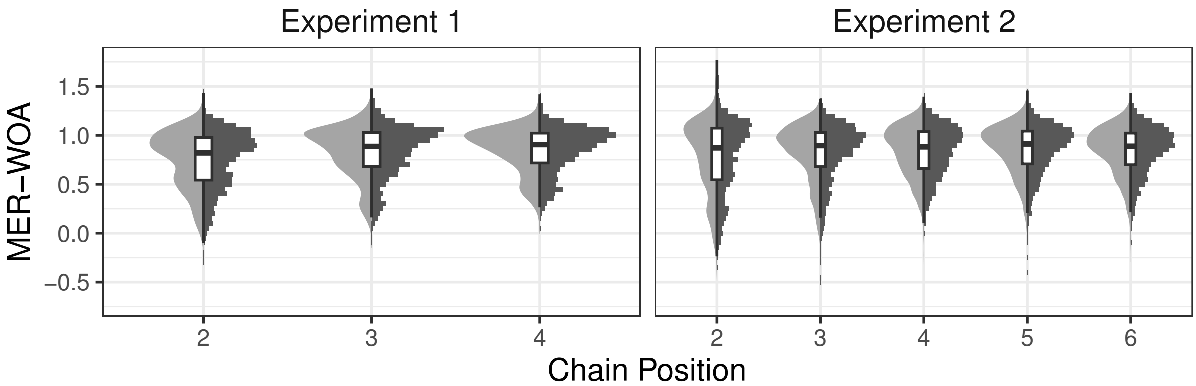
Discussion

Independent initial judgments are irrelevant from a (group) 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. Hence while the original study found that accuracy increased with the number of collaborations, the proposed regression approach shows that this was not due to changes in informational influence along the chains. This contradicts findings that advice of (objectively) higher quality (i.e., for later positions in the scenario at hand) is taken relatively more (e.g., Yaniv & Kleinberger, 2000; but see Schultze et al., 2017).[8] Accordingly, the new evidence

[8] The number of previous contributors was unknown to participants. Although judgments were not independent, increasing advice quality is likely a consequence of crowd wisdom increasing with the number

Figure 2

Distributions of Mixed-Effects Regression Weights of Advice (MER-WOA) With Fixed Positioning Effects per Chain Position in all Trials of Experiments 1 and 2 of Mayer and Heck (2022)



suggests that the positive effect of advice quality on weighting may be driven by having implemented a reference point (i.e., own initial judgment) for advice quality assessment in the traditional paradigm (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 efficiency of sequential collaboration.

Being immediately confronted with collaborators' estimates (i.e., anchors) in sequential collaboration may also trigger an anchoring and adjustment process in the classic sense (Tversky & Kahneman, 1974; see also Minson & Mueller, 2012; Schultze et al., 2019). Consequently, participants (insufficiently) adjusted away from earlier anchors of lower quality as much as from later anchors of higher quality (Schultze et al., 2017; but see Hütter & Fiedler, 2019; Röseler et al., 2023; Schweickart et al., 2021). In the traditional paradigm, however, independent initial judgments also constitute anchors from which

of contributors (Hogarth, 1978; but see Davis-Stober et al., 2014). This might explain why advice non-taking (and larger variance in advice taking) was a particularly prominent feature of initial collaborations (Figure 2).

participants adjust away by considering advice (Bonaccio & Dalal, 2006; Harvey & Fischer, 1997; Lim & O'Connor, 1995). Omitting the initial judgment phase accordingly avoids self-anchoring in the JAS. Essentially, informational influences can still be investigated by means of the proposed method. Thus, MER-WOAs can also help avoiding anchoring effects and resolve a paradigmatic peculiarity of the classic JAS.

Although not recorded, this does not imply that participants did not form independent initial judgments. Accordingly, *unobserved* initial judgment being represented in the overall error term ε may eventually imply omitted variable bias for the model as specified in Equation 11 (see also Raudenbush & Bryk, 2002, Chapter 9; Snijders & Bosker, 2012, Chapter 14). Specifically, if independent judgments were formed without explicit request by the experimenter, the reported effect of advice on final judgment would incorporate the effects of such missing variables if they are correlated with advice. Indeed, the influence of advice on judgment was surprisingly strong here as compared to traditional advice taking research (e.g., Harvey & Fischer, 1997; Soll & Larrick, 2009; Yaniv & Kleinberger, 2000). On the one hand, however, simultaneously presenting stimuli for the first time and judgments of collaborators was a strong procedural barrier for forming independent beliefs in the experiments of Mayer and Heck (2022). On the other hand, omitted variable bias can also not be ruled out if initial judgment is recorded and incorporated into the model (i.e., in both Equation 7 and the traditional model implied by ROD calculus in Equation 4). Part of the reason are some issues already mentioned above, for instance, missing additional cues that might have been inferred and/or processed by participants such as expertise or quality. In general, the inevitable temporal distance between initial and final judgments in the original paradigm might itself cause (small) discrepancies in participants' judgments (e.g., because of limited memory capacities or deliberately having changed one's mind for other unobservable reasons).

Individual Weights of Sequentially Sampled Advice

In sampling approaches to advice taking, the traditional paradigm is extended by a (free) sampling phase. For instance, in Experiments 2 and 3 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. By lack of more advanced techniques, however, the factually sequential taking of multiple pieces of advice was modeled as the taking of the mean of all advisory judgments within a trial:

$$\omega_{\bar{A},ij} = \frac{F_{ij} - I_{ij}}{\bar{A}_{ij} - I_{ij}} \tag{15}$$

where $\bar{A}_{ij} = \frac{1}{K_{ij}} \sum_{k=1}^{K_{ij}} A_{ijk}$, for $k = 1, \ldots, K_{ij}$ sampled pieces of advice per trial ij. Plugging $\bar{A}_{ij}$ into Equation 15 and rearranging it in terms of response linearity yields:

$$F_{ij} = \frac{\omega_{\bar{A},ij}}{K_{ij}} A_{ij1} + \cdots + \frac{\omega_{\bar{A},ij}}{K_{ij}} A_{ijK_{ij}} + (1 - \omega_{\bar{A},ij}) I_{ij}. \tag{16}$$

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 $k, l = 1, ..., K_{ij}$.[9] Put differently, $\omega_{\bar{A},ij}$ indicates the *total* weight of advice which is defined as the *sum* of all single, equally weighted pieces of advice that are sampled on a specific trial.

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 differences in (perceived or actual) expertise among a set of distinct advisors (Brehmer & Hagafors, 1986; Harvey & Harries, 2004). Moreover, there could be order effects (e.g., primacy or recency) in the weighting of sequentially sampled advice as were found for the weighting of sequentially sampled anchors (Hogarth & Einhorn, 1992). However, presupposing equal weights of

[9] Conversely, 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, Equation 5). To the best of our knowledge, there are no published findings about the relevance of this implicit assumption.

multiple advisors suppresses differential weighting with respect to, for instance, expertise or sampling position.

Fixed versus Random Order Effects

In mixed-effects regressions, each sampled piece of advice may be construed as a set of additional predictors on the level of trials. Treating the advice coefficients of Equation 16 as free parameters, individual weights of individual pieces of advice can be estimated by fitting the following model:

$$F_{ij} = \sum_{k=1}^{K_{ij}} \omega_{A_{ijk}} A_{ijk} + \left(1 - \sum_{k=1}^{K_{ij}} \omega_{A_{ijk}}\right) I_{ij} + \varepsilon_{ij}, \tag{17}$$

with $\omega_{A_{ijk}}$ denoting the sum-to-one constrained weights of the k-th piece of advice encountered or sampled during the ij-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). Technically, however, this approach has several limitations. First, sequential sampling with self-determined sample sizes can produce sparse predictor sets $A_{ij} = \begin{bmatrix} A_{ij1} & \cdots & A_{ijK_{ij}} \end{bmatrix}^{\mathrm{T}}$. The reason is that adaptively sampling participants may not realize an experiment's full sampling potential (i.e., $K_{ij} < K$ where on most trials $K_{ij} \neq K_{i'j'}$; Hütter & Ache, 2016). To that effect, estimation uncertainty would be higher for weights of advice later in the sampling chain due to an increasing lack of data for increasing k. Second, and more important, the potential for linear dependencies ("redundancy;" Soll & Larrick, 2009; Yaniv et al., 2009) between advice values A_{ijk} and $A_{ijk'}$, $k \neq k'$, or I_{ij} naturally increases in K. Accordingly, the reliability of weights estimated from a model with fixed order effects implemented as separate coefficient regressions likely suffers from multicollinearity for large K (see also Hoffman, 1960).

If the effect of advice on final judgment varies systematically (e.g., linearly) along the sampling chain, the summation from Equation 17 could be replaced by a product term.

For instance, the model

$$F_{ij} = \omega_{A,ij}(k)A_{ijk} + [1 - \omega_{A,ij}(k)]I_{ij} + \varepsilon_{ij} \tag{18}$$

includes only one coefficient regression specified as follows:

$$\omega_{A,ij}(k) = \beta_A + \alpha^S_{A,i} + \alpha^T_{A,j} + \beta_{A\times k}k. \tag{19}$$

In words, adding the interaction of advice and its respective sampling index k as a fixed effect enables us to measure individual weights of sampled advice as a function of k. Moreover, trial-wise sum-to-one constraining can be achieved by normalizing (i.e., dividing) the estimated weights by the realized advice sample sizes K_{ij} of a certain trial.

Instead, advice may be construed as the third crossed clustering instance on the weighting level, which is particularly relevant for real-world applications. Just as participants and stimulus items often are (randomly) drawn instances of larger respective populations, advisors may stem from larger populations of potential candidates (e.g., Soll et al., 2022)—and should hence be modeled as such. In a correspondingly simplified judgment model

$$F_{ij} = \omega_{A,ijk}A_{ijk} + (1 - \omega_{A,ijk})I_{ij} + \varepsilon_{ijk}, \tag{20}$$

the weight of the k-th piece of advice may be calculated as

$$\omega_{A,ijk} = \beta_A + \alpha^S_{A,i} + \alpha^T_{A,j} + \alpha^U_{A,k}, \tag{21}$$

where $\alpha^U_A \sim N\left(0, \tau^2_{A,U}\right)$ denotes the random effects of advisors U (see also Appendix B). A critical, substantive evaluation of the implicit assumption of static advisor chains versus global patterns of serial positioning is provided in the context of an empirical application.

Empirical Application

Originally, advice that was more distant (but not too distant; as defined by Moussaïd et al., 2013) from participants' initial judgments was found to be sampled more frequently than closer advice and, partly for that reason, weighted more strongly *in total*

(Hütter & Ache, 2016). This indicates that people seem to balance the informational asymmetry in favor of own judgments (Yaniv, 2004b) by sampling additional advisory judgments. Reanalyzing the publicly available data (https://journal.sjdm.org/15/151110a/) of Experiments 2 and 3 of Hütter and Ache (2016) by means of the proposed mixed-effects regression approach allows insights with respect to individual differences at the level of advice. In other words, MER-WOAs enable us to clarify whether *individual* pieces of relatively more distant advice were indeed weighted more strongly (e.g., because of its relatively high informational value; Schultze et al., 2015). If not, evidence for increased total weighting may be an artifact of distant advice boosting the sampling of additional evidence.

Method

Technically, there are the same two options as for sampling (or chain position in the first empirical application) to model treatment effects on weighting. The first one is to include advice distance as a fixed effect, more specifically, as an interaction term of advice and distance condition $C_{ij} \in \{\text{close}, \text{distant}\}$ such that $\beta_{A\times C}$ captures the fixed treatment effect on weighting. The second option is to model treatment effects as additional crossed random effect $\alpha_A^U \sim N\left(0, \tau_{A,U}^2\right)$ on the grouping level. The distributional assumptions as well as practical recommendations for a minimum number of five factor levels are in favor of option one for distance as binary treatment factor (Bolker, 2015; Oberpriller et al., 2022). Moreover, significance testing of group differences as a fixed effect more closely resembled the original analyses.

Following the same reasoning, implementing random order effects (i.e., Equation 21) would be apposite as participants realized sufficiently large samples of advice K_{ij} (Experiment 2: $M = 8.05$, $SD = 7.01$; Experiment 3: $M = 10.26$, $SD = 5.98$). In addition, free advice sampling renders uneven sampling across factor levels very likely and thus would be in favor of random effects estimation (Bolker, 2015). However, Oberpriller et al.

(2022) recommended to implement fixed (sampling) effects in response to singular fits with random (sampling) effects, which yield almost identical results (see Appendix B).[10]

Results

On average, advice weighting was higher in Experiment 2 than in Experiment 3 of Hütter and Ache (2016) as indicated by $\hat{\beta}_A$ (Table 3). In both experiments, there was no significant fixed effect of the sampling index k on WOA as indicated by $\hat{\beta}_{A\times k}$. That is, the weighting of sequentially sampled advice did not significantly change along the sampling chain. Descriptively, both fixed effects indicated tiny recency effects, that is, slightly higher weighting of advice that was sampled later in the sequential chain. Moreover, those individual advice weights were on average significantly higher for distant than close advice (implemented as fixed treatment effect) as $\hat{\beta}_{A\times C} > 0$ in both experiments. This effect of advice distance on weighting was qualitatively similar to the descriptive one based on traditional ROD modeling (see Figure 3) and the one with random order effects (see Appendix B).

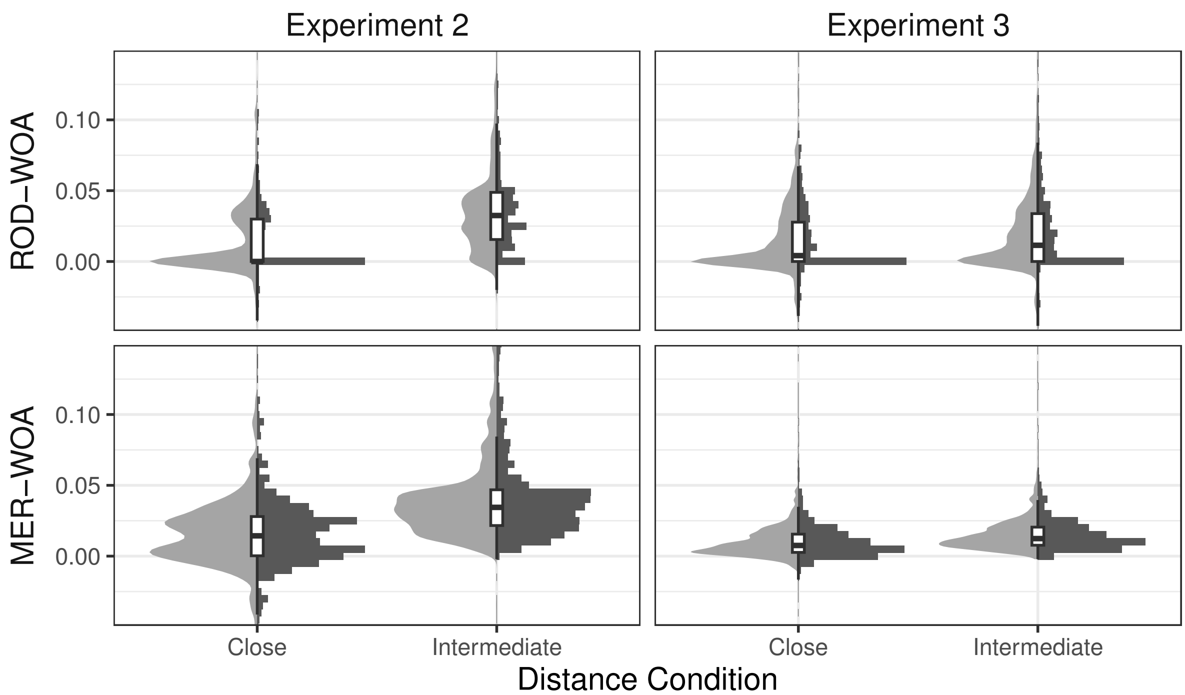
Previous research generally obtained an inverse-U-shaped relation between weighting and the relative absolute distance of advice

$$D = \frac{|A - I|}{I} + 1 \tag{22}$$

[10] In our experience, variables from different levels being measured on different scales are one of the major pitfalls for model convergence. Instead of z-standardizing judgments as in the first empirical application (see Footnote 7), aligning them with the weighting scale prior to fitting the models often guarantees convergent solutions without disturbing the scaling (i.e., center and variance) of the data. Indeed, dividing all judgments by ten to the power of the maximum number of digits observed for a certain experiment yielded stable results.

Figure 3

Distributions of Normalized Weights of Advice (WOA) Calculated by Means of the Ratio-of-Differences (ROD) Formula (Top) Versus Estimated by Means of Mixed-Effects Regression (MER) With Fixed Linear Order Effects and Fixed Treatment Effects of Contrast-Coded Advice Distance Condition (Bottom) in Experiments 2 and 3 of Hütter and Ache (2016)



Note. WOA is normalized by dividing by the realized advice sample sizes K_{ij}. Outliers are not displayed in the box plots.

Table 3

Fixed Effects of Multilevel Models of Final Judgment According to Equation 18 With Fixed Linear Order Effects as Specified in Equation 19 and Fixed Treatment Effects of Contrast-Coded Advice Distance Condition (C) for Experiments 2 and 3 of Hütter and Ache (2016)

	Experiment 2		Experiment 3	
	Estimate	*SE*	*Estimate*	*SE*
β_A	0.2494 ***	0.0420	0.1357 ***	0.0160
$\beta_{A\times k}$	0.0012	0.0007	0.0008	0.0005
$\beta_{A\times C}$	0.3055 ***	0.0263	0.0700 ***	0.0053

Note. *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$. The full models can be found in Appendix B, Tables B1 and B2.

(Moussaïd et al., 2013). Including linear and logarithmic terms of distance as in Schultze et al. (2015),[11] this characteristic pattern could be replicated for normalized (i.e., divided by the respective realized advice sample sizes K_{ij}) ROD-WOA in both experiments (Table 4). However, the curvature is less pronounced in Experiment 3 (see also Figure 4, top panel). More interesting, however, is the question of whether the same curvilinear relationship is also observed for individual weights of sampled advice. For fixed linear order effects, the characteristic inverse-U-shaped relation was indeed also observed in Experiment 2 but not in Experiment 3 (Table 5). Descriptively, even the opposite pattern was observed in the latter (see also Figure 4, bottom panel). By accounting for nonlinear effects of advice distance, there was significant evidence for recency effects in Experiment 2 as $\hat{\beta}_{A\times k} > 0$. That is, advice that was sampled later in the sequential chain was weighted

[11] To avoid undefined values of $\log(D)$ for zero advice distance, which was not implemented in Schultze et al. (2015), we added a unit constant to the relative absolute distance D as defined in Equation 22. Moreover, the results were qualitatively similar for a combination of linear and quadratic trends.

slightly more strongly than advice that was sampled earlier. However, this was not replicated for random order effects (see Appendix B).

Table 4

Fixed Effects of Multilevel Models of Normalized ROD-WOA with Participant- and Item-Wise Random Intercepts and Linear and Logarithmic Fixed Effects of Relative Absolute Advice Distance (D) for Experiments 2 and 3 of Hütter and Ache (2016)

	Experiment 2		Experiment 3	
	Estimate	*SE*	*Estimate*	*SE*
β_0	0.1983 ***	0.0170	0.0271 ***	0.0034
β_D	-0.1659 ***	0.0142	-0.0056 *	0.0024
$\beta_{\log(D)}$	0.2920 ***	0.0203	0.0160 ***	0.0045

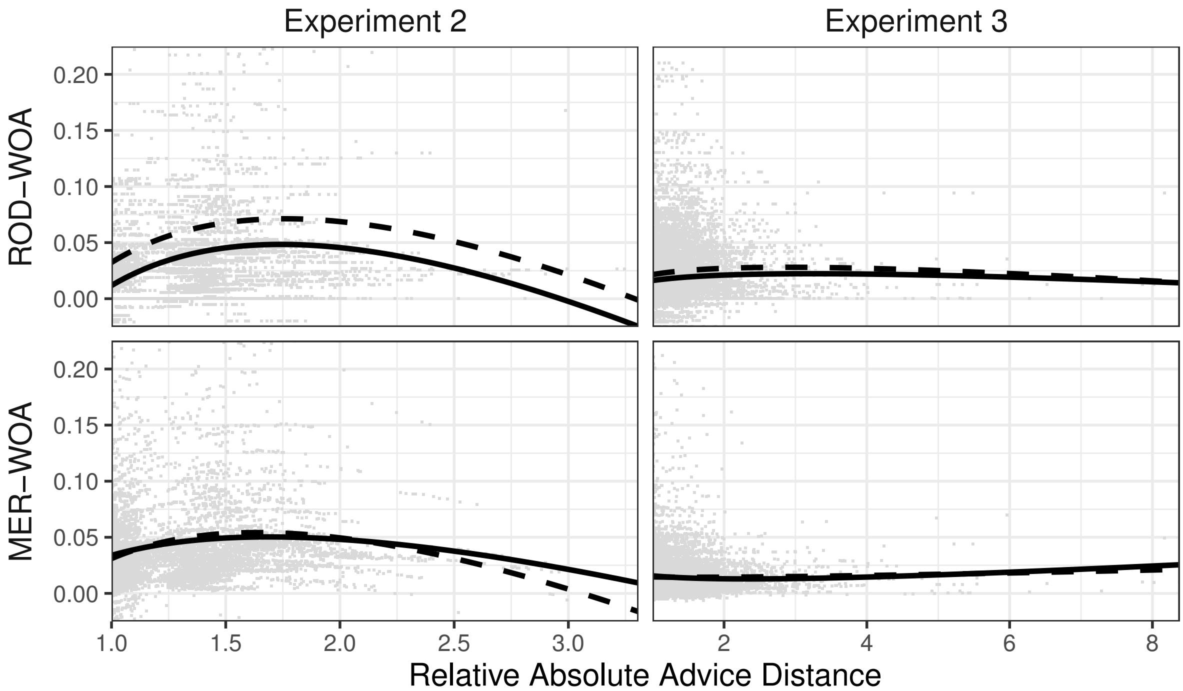
Note. *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$. WOA is normalized by dividing by the realized advice sample sizes K_{ij}. The full models can be found in Appendix B, Tables B3 and B4.

Discussion

Without accounting for nonlinear effects of distance on advice weighting, there was neither evidence for fixed nor random order effects in both experiments of Hütter and Ache (2016) that implemented sequential advice seeking. In other words, weighting of sequentially sampled advice does not systematically vary along the sampling chain. Therefore, the original approximation to assume equal weighting of all sampled pieces of advice (or mean advice taking depending on which interpretation is chosen; see above) is deemed valid for binary advice distance manipulations. In line with the notion of the wisdom of crowds, averaging idiosyncrasies out on the sampling level as done in the original analyses presumably worked well because participants sampled sufficiently many pieces of advice per trial (i.e., $M > 8$ for K_{ij} in both experiments; Hogarth, 1978). Before we conducted the reported analyses, however, there was uncertainty regarding the

Figure 4

Normalized Weights of Advice (WOA) Calculated by Means of the Ratio-of-Differences (ROD) Formula (Top) Versus Estimated by Means of Mixed-Effects Regression (MER) as Specified in Table 5 (Bottom) as Functions of Relative Absolute Advice Distance (D) in Experiments 2 and 3 of Hütter and Ache (2016)



Note. WOA is normalized by dividing by the realized advice sample sizes K_{ij}. Plotting of WOA is truncated outside the interval $[-0.0250, 0.2250]$. The solid lines display the linear plus logarithmic fits of normalized weights. The dashed lines display the functional forms according to the estimated fixed effects as reported in Tables 4 and 5 (for the fixed effects divided by and k set equal to the respective average realized advice sample sizes), respectively.

Table 5

Fixed Effects of Multilevel Models of Final Judgment According to Equation 18 With Fixed Linear Order Effects as Specified in Equation 19 and Linear and Logarithmic Fixed Effects of Relative Absolute Advice Distance (D) for Experiments 2 and 3 of Hütter and Ache (2016)

	Experiment 2		Experiment 3	
	Estimate	*SE*	*Estimate*	*SE*
β_A	1.3344 ***	0.0943	0.1281 ***	0.0193
$\beta_{A\times k}$	0.0014 *	0.0007	0.0005	0.0005
$\beta_{A\times D}$	-1.0958 ***	0.1010	0.0176	0.0152
$\beta_{A\times \log(D)}$	1.7950 ***	0.1725	-0.0289	0.0360

Note. *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$. The full models can be found in Appendix B, Tables B5 and B6.

appropriateness of this practice. Moreover, we also found first evidence for significant, albeit small, recency effects in weighting sequentially sampled advice. Both facts highlight the importance of the mixed-effects regression approach as a tool to estimate separate weights of individually sampled advice.

Advice giving was implemented as drawing random numbers in the experiments under investigation (Hütter & Ache, 2016). However, alternative models impose different assumptions about serial positioning. Random order effects imply that the K advisors have something in common, for instance, are the same and/or sampled in the same sequence across all trials ij. Similarly, fixed order effects implemented via an interaction of advice and sampling position k presuppose the existence of global (i.e., no individual differences), linear patterns of serial positioning (e.g., primacy or recency) in participants' advice weighting strategies. Fixed order effects from separate coefficient regressions for individual covariates as formalized in Equation 17 provide relief at the cost of convergence and multicollinearity issues for large K. Hence the decision for implementing fixed versus random order effects in empirical practice not only depends on mathematical constraints, but also on the sampling propensities of participants and the experimental implementation of advice giving. The fixed effects versions are better suited for environments with low sampling potential K or relatively small realized sample sizes K_{ij} in spite of large K, respectively. In contrast, the random effects version should be preferred for larger K and K_{ij}. This is mainly to account for practical model fitting recommendations such as the above-mentioned minimum number of five factor levels per clustering instance (to precisely estimate the respective variance terms) or including at most three sources of randomness (e.g., participants, items, and advisors; Bolker, 2015; Oberpriller et al., 2022). Although implemented as such, it is rather unlikely that people randomly sample advice in the real world. Eventually, the estimated weights reflect their corresponding substantive interpretations (e.g., in terms of serial positioning) and ecological foundations.

In the sampling experiments of Hütter and Ache (2016), potentially updated judgments were expressed only once after having finished advice sampling. Therefore, each sampled value's distance could merely be defined relative to participants' initial judgments. However, participants may immediately update their judgments in response to additional evidence even if not explicitly requested to do so (Hogarth & Einhorn, 1992).[12] If this was the case, our reanalysis would be constrained by an impoverished definition of advice distance that does not take intermediate updating in the course of sampling into account. Accordingly, a Bayesian modeling framework may be a more natural account of sequentially sampled advice which will enable investigations of participants' stopping decisions in sampling extensions of the JAS (Rebholz et al., 2023). Future research should

[12] A "Step-by-Step procedure" requests intermediate judgments after each piece of advice was sampled (Hogarth & Einhorn, 1992). Instead, an "End-of-Sequence procedure" requires participants to express their final judgments only once at the end of a sampling chain. For implementations of the former, the possibility to calculate the ROD formula step-by-step would indeed enable research on individual differences on the sampling level.

thus develop a model of advice taking that is consistent with both sequential sampling of information and corresponding endogenous formation of judgment.

General Discussion

We proposed mixed-effects regression weights of advice (MER-WOAs) for analyzing 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 capitalizes on the multilevel structure of most data collected in these paradigms. Specifically, the experimental crossing of the grouping factors (i.e., participants, stimulus items, and—potentially—advisors). In contrast to state-of-the-art ratio-of-differences (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. Moreover, it has many technical merits, some of which will be discussed in more detail below. Essentially, the more advanced modeling approach proved to be practically applicable in multiple reanalyses of empirical data. For instance, the blind spot of informational influences without recording initial judgments as in Mayer and Heck (2022) was mitigated. Moreover, we have provided first evidence for systematic order effects (i.e., recency effects) in weighting sequentially sampled advice. Actually, our findings validate the aggregated approach used by Hütter and Ache (2016) for binary advice distance manipulations, but also demonstrate potential pitfalls of such simplistic statistical modeling.

In methodological research, "more advanced" is often connoted with "more complex" modeling. Indeed, instead of calculating individual weights per trial by means of simple arithmetics, a statistical optimization procedure is applied to estimate mean weights and predict individual deviations thereof. However, at least parsimony within the multilevel regression framework can be controlled by means of model specification. In sampling extensions, for instance, modeling sampling as random effects of a third, crossed clustering factor reduces the estimation problem of K individual mixed-effects coefficients to merely one additional variance term on the weighting level (with the corresponding implications for interpretation as elaborated above). Given the presented benefits, sacrificing parsimony in favor of more complex model building and statistical optimization appears worthwhile. The prevailing demand for inferential conclusions anyway challenges the existence of factual differences in model(ing) complexity between ours and the more established approach. In the latter, researchers have to rely on a two-step procedure: Merely descriptive ROD-WOAs from step one are utilized as dependent variables in statistical testing procedures (e.g., multilevel modeling: Ache et al., 2020; Hütter & Fiedler, 2019; Minson & Mueller, 2012; Schultze et al., 2015) on step two. Therefore, those and other important merits and limitations of the two methods will be critically evaluated in the following.

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 mixed-effects regression approach, by contrast, provides well-defined—in the sense of person- and item-specific deviations from mean weighting tendencies—estimates of weights also for those situations. 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 hence 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, which is caused by its specific arithmetics as elaborated above, makes exclusions or truncation often inevitable in established empirical practice (e.g., Hütter & Ache, 2016; Schultze et al., 2015). By contrast, shrunken MER-WOAs do not require to implement often complex and ambiguous (see also the example in Beyond Relativity to the Self) outlier specifications. As data pre-processing implies researcher degrees of freedom, the proposed method can hence also be seen as a tool to counter the reproducibility crisis. This is achieved in addition to enabling extensions of existing empirical findings or developing new substantive research domains.

Admissible Interpretation of (Shrunken) Regression Coefficients as "Weights"

On balance, Bonaccio and Dalal (2006) favor regression modeling, more specifically, dominance weights of advice.[13] 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 individual 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). Hence, a limitation of the new approach is its proposal of potentially poor measures

[13] General 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 hence require comparing dominance weights between different regressions models, Önkal et al. (2009) favored ROD-WOAs instead.

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 mixed-effects regression 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 bivariate model with only advice distance as predictor (see Equation 4). Even with multiple pieces of advice, a collinearity-stable regression model could be derived for random order effects (see Equations 21 & 20).

Multilevel modeling is mathematically equivalent to ridge regression (Baayen & Linke, 2020; Brown et al., 2018). In classic 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 stabilize 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.

Cognitive Psychometrics and Nonlinear Advice Weighting

Additional predictors can be added at any level of a multilevel regression model. Modeling in terms of response linearity hence allows for consideration of additional predictors on one 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 on the respective level, researchers can investigate the weighting behavior more holistically. In general, any number and type of main or (cross-level) interactive effects of additional predictors on judgment or weighting can be assessed. An accordingly extended system of Equations 4 and 5 hence also allows for cognitive psychometrics in the sense of moderation effects on the weight of advice.

Often people combine different sources of information in a nonlinear way (e.g., Ganzach, 1995, for clinical judgment). In empirical research, ROD-WOA usually features an inverse-U-shaped relation with advice distance: Close as well as more distant advice is typically taken less than advice of "intermediate" distance (Hütter & Ache, 2016; Moussaïd et al., 2013; Schultze et al., 2015; see also Figure 4). As the regression framework allows to incorporate any number and type of additional predictors on any level, researchers no longer have to assume response *linear* processing of advice for calculating the weights. To model nonlinear weighting as requested by, for instance, Bailey et al. (2022), the set of additional predictors may contain higher order terms such as polynomials (e.g., Schreiner et al., 2023) or nonlinear transformations of advice (distance; see also the second empirical application above). Base expansions provide an even more flexible, so-called "generalized additive modeling" framework (e.g., Baayen & Linke, 2020). In essence, as already requested by Slovic and Lichtenstein (1971), the proposed multilevel modeling approach renders statistical tests of alternative theoretical accounts about external information utilization feasible.

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. 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 own 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 that could also be replicated for individual MER-weights of sampled advice in Experiment 2 of Hütter and Ache (2016). In a regression framework, variables from different scales can be included side by side in the same statistical model. Hence although both variables are measured on very different scales (i.e., ordinal confidence and numeric judgments), MER-*confidence* weights of advice can be derived for confidence as response variable of interest.

Extension to multivariate multilevel regression—that is, multiple response variables modeled collectively in one larger regression model—are equally straightforward (Snijders & Bosker, 2012, Chapter 16). A multivariate multilevel regression model captures advice taking simultaneously on multiple dimensions such as judgmental shift *plus* (over-)confidence change (see also Soll et al., 2022). In other words, we can capture the duality of advice taking by accounting for dimensional interdependencies of the coefficient regressions. Judgment accuracy may further enrich corresponding 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 multidimensional weights of advice and hence ultimately a regression-based procedure.

For the sake of brevity, we will not discuss further valuable extensions such as generalized MER-WOAs for discrete choice (see also Footnote 3) or multilevel quantile regression to investigate parts of advice weighting distributions other than means.

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 (Brysbaert & Stevens, 2018; Green & MacLeod, 2016). More important, traditional modeling of advice taking is responsible for a multitude of limitations (see Bonaccio & Dalal, 2006, for a review). For instance, there may be more (or less; see also the first empirical application above) sources of information than only advice and initial judgment. Without recording intermediate judgments for sampling information from multiple external sources, calculation of the established weighting index of Harvey and Fischer (1997) is also infeasible. Earlier regression-based approaches already accounted for 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 weighted linear combinations of exogenous sources of information. In contrast to those aggregated analyses schemes, however, the proposed multilevel regression framework still enables researchers to estimate individual weights for each experimental trial.

Conclusion

In summary, the main strength of the proposed mixed-effects regression weights (of advice) is to 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 social networks make it increasingly convenient to access multiple opinions and advisors. Therefore, more flexible tools that enable investigations of informational influences in such situations are precious.

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.

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 M. Paquot & S. T. Gries (Eds.), *A Practical Handbook of Corpus Linguistics* (pp. 563–591). Springer International Publishing. https://doi.org/10.1007/978-3-030-46216-1_23

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*, 1–26. 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

Bolker, B. M. (2015). Linear and generalized linear mixed models. In G. A. Fox, S. Negrete-Yankelevich, & V. J. Sosa (Eds.), *Ecological statistics* (pp. 309–333). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199672547.003.0014

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

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

Brysbaert, 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

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

Chung, Y., Rabe-Hesketh, S., Dorie, V., Gelman, A., & Liu, J. (2013). A nondegenerate penalized likelihood estimator for variance parameters in multilevel models. *Psychometrika*, *78*(4), 685–709. https://doi.org/10.1007/S11336-013-9328-2.

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

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

Edwards, J. R. (1994). The study of congruence in organizational behavior research: Critique and a proposed alternative. *Organizational Behavior and Human Decision Processes*, *58*(1), 51–100. https://doi.org/10.1006/obhd.1994.1029

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

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.socec.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

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

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

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. https://doi.org/10.1017/S193029750000382X

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

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*. https://doi.org/10.1037/dec0000199

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

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*. https://doi.org/10.1037/dec0000193

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

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

Oberpriller, J., de Souza Leite, M., & Pichler, M. (2022). Fixed or random? On the reliability of mixed-effects models for a small number of levels in grouping variables. *Ecology and Evolution*, *12*(7), e9062. https://doi.org/10.1002/ece3.9062

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

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

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.). Sage Publications.

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 (SJDM) 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

Röseler, L., Weber, L., Helgerth, K., Stich, E., Günther, M., Tegethoff, P., Wagner, F., Antunovic, M., Barrera-Lemarchand, F., Halali, E., Ioannidis, K., Genschow, O., Milstein, N., Molden, D. C., Papenmeier, F., Pavlovic, Z., Rinn, R., Schreiter, M. L., Zimdahl, M. F., … Schütz, A. (2022). The open anchoring quest dataset: Anchored estimates from 96 studies on anchoring effects. *Journal of Open Psychology Data*, *10*(1), 1–16. https://doi.org/10.5334/jopd.67

Schreiner, M. R., Quevedo Pütter, J., & Rebholz, T. R. (2023). *Time for an update: Belief updating based on ambiguous scientific evidence* [Manuscript in preparation], Department of Psychology, University of Mannheim.

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

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

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

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.

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

Yaniv, I. (2004a). 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. (2004b). The benefit of additional opinions. *Current Directions in Psychological Science*, *13*(2), 75–78. https://doi.org/10.1111/j.0963-7214.2004.00278.x

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: Egocentric discounting and reputation formation. *Organizational Behavior and Human Decision Processes*, *83*(2), 260–281. https://doi.org/10.1006/obhd.2000.2909

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

Appendix A

Additional Results for the Reanalysis of Mayer and Heck (2022)

Full Multilevel Models Without Positioning Effects

Table A1

Full Multilevel Model of Final Judgment Without Positioning Effects According to Equations 11 and 12 for Experiment 1 of Mayer and Heck (2022)

	Estimate	*95% CI*	*SE*	*t*	*df*	*p*
β_A	0.8099	0.7498 – 0.8700	0.0304	26.6179	151.2516	**<0.001**
σ	0.5650	0.5555 – 0.5748				
$\tau_{A,S}$	0.2629	0.2250 – 0.3009				
$\tau_{A,T}$	0.1231	0.0958 – 0.1495				
ICC	0.2072					
N	111					
M	65					
Obs.	6621					
R^2_m / R^2_c	0.62 / 0.70					

Note. Bold values indicate $p < 0.05$. Wald 95% *CI* for fixed and bootstrap 95% *CI* (with 1,000 iterations) for random effects are shown.

Table A2

Full Multilevel Model of Final Judgment Without Positioning Effects According to Equations 11 and 12 for Experiment 2 of Mayer and Heck (2022)

	Estimate	*95% CI*	*SE*	*t*	*df*	*p*
β_A	0.8241	0.7752 – 0.8729	0.0248	33.2511	207.2645	**<0.001**
σ	0.5703	0.5638 – 0.5768				
$\tau_{A,S}$	0.2828	0.2556 – 0.3109				
$\tau_{A,T}$	0.1305	0.1048 – 0.1564				
ICC	0.2289					
N	254					
M	65					
Obs.	15076					
R^2_m / R^2_c	0.62 / 0.70					

Note. Bold values indicate $p < 0.05$. Wald 95% CI for fixed and bootstrap 95% CI (with 1,000 iterations) for random effects are shown.

Full Multilevel Models for Fixed Positioning Effects

Table A3

Full Multilevel Model of Final Judgment According to Equation 11 With Fixed Positioning Effects as Specified in Equation 13 for Experiment 1 of Mayer and Heck (2022)

	Estimate	*95% CI*	*SE*	*t*	*df*	*p*
β_A	0.7064	0.5670 – 0.8458	0.0704	10.0330	119.7243	**<0.001**
$\beta_{A\times c}$	0.0520	-0.0113 – 0.1154	0.0320	1.6267	111.5095	0.107
σ	0.5650	0.5556 – 0.5748				
$\tau_{A,S}$	0.2607	0.2229 – 0.2989				
$\tau_{A,T}$	0.1225	0.0953 – 0.1488				
ICC	0.2047					
N	111					
M	65					
Obs.	6621					
R^2_m / R^2_c	0.62 / 0.69					

Note. Bold values indicate $p < 0.05$. Wald 95% CI for fixed and bootstrap 95% CI (with 1,000 iterations) for random effects are shown.

Table A4

Full Multilevel Model of Final Judgment According to Equation 11 With Fixed Positioning Effects as Specified in Equation 13 for Experiment 2 of Mayer and Heck (2022)

	Estimate	*95% CI*	*SE*	*t*	*df*	*p*
β_A	0.7756	0.6847 – 0.8665	0.0462	16.7937	301.4366	**<0.001**
$\beta_{A\times c}$	0.0165	-0.0096 – 0.0426	0.0133	1.2420	263.4371	0.215
σ	0.5703	0.5638 – 0.5768				
$\tau_{A,S}$	0.2825	0.2553 – 0.3098				
$\tau_{A,T}$	0.1305	0.1048 – 0.1566				
ICC	0.2286					
N	254					
M	65					
Obs.	15076					
R^2_m / R^2_c	0.61 / 0.70					

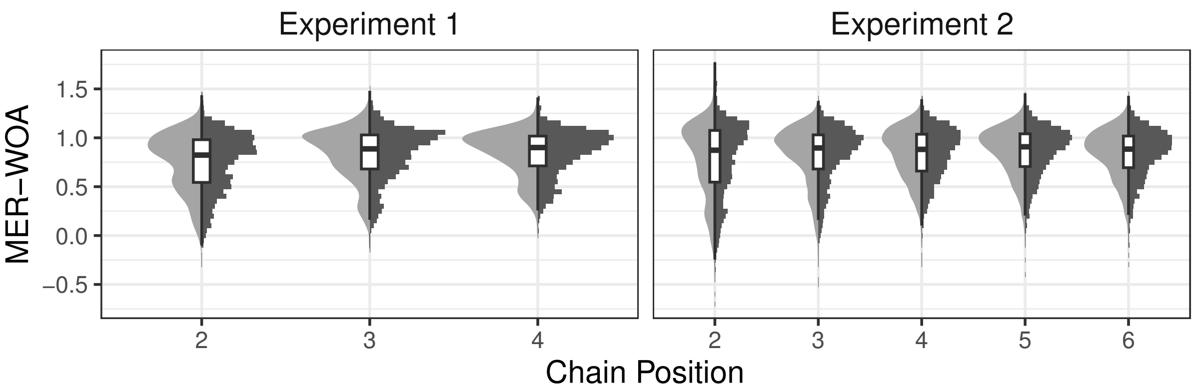
Note. Bold values indicate $p < 0.05$. Wald 95% CI for fixed and bootstrap 95% CI (with 1,000 iterations) for random effects are shown.

Results and Full Multilevel Models for Random Positioning Effects

Technically, only Experiment 2 of Mayer and Heck (2022) just satisfies practical recommendations about minimum number of five factor levels for precise random effects variance estimation (Bolker, 2015; Oberpriller et al., 2022). For completeness, we also report the results for Experiment 1 in the following. Estimating multilevel regression models as specified in Equations 11 and 14, there was no evidence for a random positioning effect (i.e., $\hat{\tau}_{A,U} > 0$) on MER-WOA in Experiment 1, $\chi^2(1) = 0.0197$, $p = .444$ (Table A5), and in Experiment 2, $\chi^2(1) = 0.0000$, $p = .500$ (Table A6)[14]. The descriptive interpretations of random positioning effects qualitatively replicate the ones for fixed positioning effects (e.g., Figure A1 vs. Figure 2).

Figure A1

Distributions of Mixed-Effects Regression Weights of Advice (MER-WOA) With Random Positioning Effects per Chain Position in all Trials of Experiments 1 and 2 of Mayer and Heck (2022)



[14] Likelihood ratio tests may not be appropriate in case of singularity (i.e., estimated variance-covariance matrices with less than full rank) due to random effects variances of zero (Bates et al., 2015). Alternatively, (partially) Bayesian methods could be applied to force the respective variance terms away from zero using regularizing priors (Chung et al., 2013).

Table A5

Full Multilevel Model of Final Judgment According to Equation 11 With Random Positioning Effects as Specified in Equation 14 for Experiment 1 of Mayer and Heck (2022)

	Estimate	*95% CI*	*SE*	*t*	*df*	*p*
β_A	0.8101	0.6955 – 0.9247	0.0360	22.5247	2.9934	**<0.001**
σ	0.5650	0.5561 – 0.5753				
$\tau_{A,S}$	0.2615	0.2203 – 0.2977				
$\tau_{A,T}$	0.1229	0.0957 – 0.1486				
$\tau_{A,U}$	0.0336	0.0000 – 0.1001				
ICC	0.2078					
N	111					
M	65					
C	3					
Obs.	6621					
R^2_m / R^2_c	0.62 / 0.70					

Note. Bold values indicate $p < 0.05$. Wald 95% CI for fixed and bootstrap 95% CI (with 1,000 iterations) for random effects are shown.

Table A6

Full Multilevel Model of Final Judgment According to Equation 11 With Random Positioning Effects as Specified in Equation 14 for Experiment 2 of Mayer and Heck (2022)

	Estimate	*95% CI*	*SE*	*t*	*df*	*p*
β_A	0.8241	0.7752 – 0.8729	0.0248	33.2511	207.2641	**<0.001**
σ	0.5703	0.5642 – 0.5770				
$\tau_{A,S}$	0.2828	0.2564 – 0.3106				
$\tau_{A,T}$	0.1305	0.1054 – 0.1549				
$\tau_{A,U}$	0.0000	0.0000 – 0.0000				
ICC	0.2289					
N	254					
M	65					
C	5					
Obs.	15076					
R_m^2 / R_c^2	0.62 / 0.70					

Note. Bold values indicate $p < 0.05$. Wald 95% CI for fixed and bootstrap 95% CI (with 1,000 iterations) for random effects are shown.

Appendix B

Additional Results for the Reanalysis of Hütter and Ache (2016)

Full Multilevel Models for Fixed Order Effects

Table B1

Full Multilevel Model of Final Judgment According to Equation 18 With Fixed Linear Order Effects as Specified in Equation 19 and Fixed Treatment Effects of Contrast-Coded Advice Distance Condition (C) for Experiment 2 of Hütter and Ache (2016)

	Estimate	*95% CI*	*SE*	*t*	*df*	*p*
β_A	0.2494	0.1655 – 0.3332	0.0420	5.9343	66.5845	**<0.001**
$\beta_{A\times k}$	0.0012	-0.0002 – 0.0026	0.0007	1.7413	7063.7388	0.082
$\beta_{A\times C}$	0.3055	0.2539 – 0.3571	0.0263	11.6051	7035.3042	**<0.001**
σ	0.0053	0.0053 – 0.0054				
$\tau_{A,S}$	0.2438	0.1939 – 0.2987				
$\tau_{A,T}$	0.0648	0.0420 – 0.0860				
ICC	0.4558					
N	44					
M	20					
Obs.	7086					
R^2_m / R^2_c	0.54 / 0.75					

Note. Bold values indicate $p < 0.05$. Wald 95% CI for fixed and bootstrap 95% CI (with 1,000 iterations) for random effects are shown.

Table B2

Full Multilevel Model of Final Judgment According to Equation 18 With Fixed Linear Order Effects as Specified in Equation 19 and Fixed Treatment Effects of Contrast-Coded Advice Distance Condition (C) for Experiment 3 of Hütter and Ache (2016)

	Estimate	*95% CI*	*SE*	*t*	*df*	*p*
β_A	0.1357	0.1039 – 0.1675	0.0160	8.4899	76.1509	**<0.001**
$\beta_{A\times k}$	0.0008	-0.0003 – 0.0018	0.0005	1.4127	11802.1405	0.158
$\beta_{A\times C}$	0.0700	0.0596 – 0.0805	0.0053	13.1359	11811.0680	**<0.001**
σ	0.0063	0.0062 – 0.0064				
$\tau_{A,S}$	0.0945	0.0766 – 0.1124				
$\tau_{A,T}$	0.0392	0.0249 – 0.0529				
ICC	0.1218					
N	58					
M	20					
Obs.	11896					
R^2_m / R^2_c	0.21 / 0.31					

Note. Bold values indicate $p < 0.05$. Wald 95% CI for fixed and bootstrap 95% CI (with 1,000 iterations) for random effects are shown.

Full Multilevel Models of Total ROD-WOA and Individual MER-WOA (With Fixed Order Effects) for Continuous Advice Distance

Table B3

Full Multilevel Model of Normalized ROD-WOA With Participant- and Item-Wise Random Intercepts and Linear and Logarithmic Fixed Effects of Relative Absolute Advice Distance (D) for Experiments 2 of Hütter and Ache (2016)

	Estimate	*95% CI*	*SE*	*t*	*df*	*p*
β_0	0.1983	0.1648 – 0.2318	0.0170	11.6390	315.4017	**<0.001**
β_D	-0.1659	-0.1937 – -0.1382	0.0142	-11.7191	5597.6327	**<0.001**
$\beta_{\log(D)}$	0.2920	0.2521 – 0.3318	0.0203	14.3782	5722.2338	**<0.001**
σ	0.0452	0.0444 – 0.0460				
$\tau_{0,S}$	0.0675	0.0535 – 0.0822				
$\tau_{0,T}$	0.0042	0.0022 – 0.0061				
ICC	0.6915					
N	44					
M	20					
Obs.	6041					
R^2_m / R^2_c	0.03 / 0.70					

Note. Bold values indicate $p < 0.05$. WOA is normalized by dividing by the realized advice sample sizes K_{ij}. Wald 95% CI for fixed and bootstrap 95% CI (with 1,000 iterations) for random effects are shown.

Table B4

Full Multilevel Model of Normalized ROD-WOA With Participant- and Item-Wise Random Intercepts and Linear and Logarithmic Fixed Effects of Relative Absolute Advice Distance (D) for Experiments 3 of Hütter and Ache (2016)

	Estimate	*95% CI*	*SE*	*t*	*df*	*p*
β_0	0.0271	0.0203 – 0.0339	0.0034	7.8830	159.8289	**<0.001**
β_D	-0.0056	-0.0103 – -0.0008	0.0024	-2.2940	10328.6358	**0.022**
$\beta_{\log(D)}$	0.0160	0.0073 – 0.0248	0.0045	3.5821	10329.3467	**<0.001**
σ	0.0330	0.0325 – 0.0334				
$\tau_{0,S}$	0.0196	0.0162 – 0.0234				
$\tau_{0,T}$	0.0028	0.0016 – 0.0039				
ICC	0.2652					
N	58					
M	20					
Obs.	10380					
R_m^2 / R_c^2	0.00 / 0.27					

Note. Bold values indicate $p < 0.05$. WOA is normalized by dividing by the realized advice sample sizes K_{ij}. Wald 95% CI for fixed and bootstrap 95% CI (with 1,000 iterations) for random effects are shown.

Table B5

Full Multilevel Model of Final Judgment According to Equation 18 With Fixed Linear Order Effects as Specified in Equation 19 and Linear and Logarithmic Fixed Effects of Relative Absolute Advice Distance (D) for Experiments 2 of Hütter and Ache (2016)

	Estimate	*95% CI*	*SE*	*t*	*df*	*p*
β_A	1.3344	1.1495 – 1.5193	0.0943	14.1576	1418.4685	**<0.001**
$\beta_{A\times k}$	0.0014	0.0000 – 0.0028	0.0007	2.0001	7063.3715	**0.046**
$\beta_{A\times D}$	-1.0958	-1.2938 – -0.8978	0.1010	-10.8471	7059.6869	**<0.001**
$\beta_{A\times \log(D)}$	1.7950	1.4568 – 2.1332	0.1725	10.4041	7005.4686	**<0.001**
σ	0.0053	0.0053 – 0.0054				
$\tau_{A,S}$	0.2416	0.1922 – 0.2959				
$\tau_{A,T}$	0.0625	0.0404 – 0.0834				
ICC	0.4499					
N	44					
M	20					
Obs.	7086					
R^2_m / R^2_c	0.55 / 0.75					

Note. Bold values indicate $p < 0.05$. Wald 95% CI for fixed and bootstrap 95% CI (with 1,000 iterations) for random effects are shown.

Table B6

Full Multilevel Model of Final Judgment According to Equation 18 With Fixed Linear Order Effects as Specified in Equation 19 and Linear and Logarithmic Fixed Effects of Relative Absolute Advice Distance (D) for Experiments 3 of Hütter and Ache (2016)

	Estimate	*95% CI*	*SE*	*t*	*df*	*p*
β_A	0.1281	0.0900 – 0.1662	0.0193	6.6430	151.8270	**<0.001**
$\beta_{A\times k}$	0.0005	-0.0006 – 0.0015	0.0005	0.8626	11794.1238	0.388
$\beta_{A\times D}$	0.0176	-0.0122 – 0.0474	0.0152	1.1583	11782.8411	0.247
$\beta_{A\times \log(D)}$	-0.0289	-0.0996 – 0.0417	0.0360	-0.8023	11838.8181	0.422
σ	0.0063	0.0063 – 0.0064				
$\tau_{A,S}$	0.0945	0.0770 – 0.1123				
$\tau_{A,T}$	0.0403	0.0259 – 0.0541				
ICC	0.1212					
N	58					
M	20					
Obs.	11896					
R^2_m / R^2_c	0.20 / 0.30					

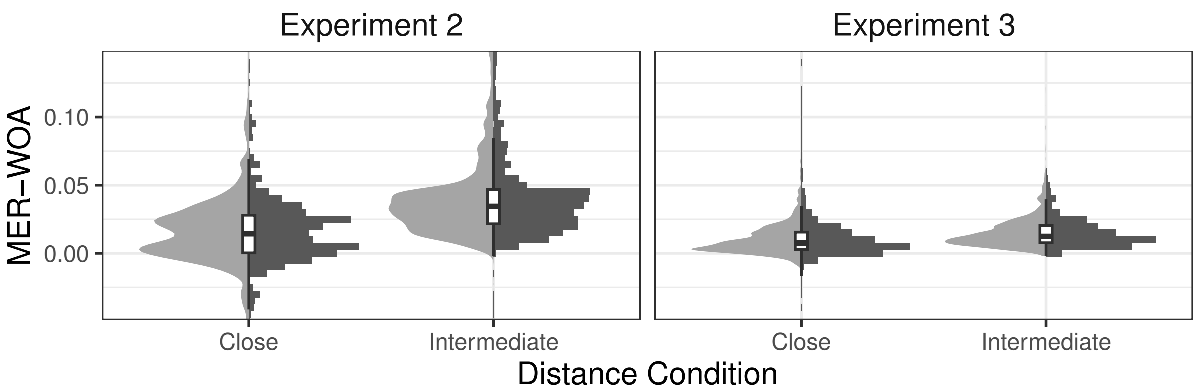
Note. Bold values indicate $p < 0.05$. Wald 95% CI for fixed and bootstrap 95% CI (with 1,000 iterations) for random effects are shown.

Results and Full Multilevel Models for Random Order Effects

There was no significant random effect of the sampling index k, $\hat{\tau}_{A,U}$, on WOA in Experiment 2, $\chi^2(1) = 0.0000$, $p = .500$ (Table B7), and in Experiment 3, $\chi^2(1) = 0.0000$, $p = .500$ (Table B8; but see Footnote 14). More specifically, in both experiments we found that $\sum_{k=1}^{K} \left|\hat{\alpha}_{A,k}^{U}\right| = 0.0000$. Nevertheless, for close advice coded as -0.50 and distant advice coded as 0.50, the individually estimated advice weights were on average significantly higher for distant advice in Experiments 2 and 3 (see also Figure B1). Replicating the results for fixed linear order effects, the characteristic inverse-U-shaped relation between weighting and the relative absolute advice distance as defined in Equation 22 (Moussaïd et al., 2013; Schultze et al., 2015) was also observed in Experiment 2 (Table B9) but not in Experiment 3 (Table B10; see also Figure B2). However, there was no significant random order effect on WOA, $\chi^2(1) = 0.0000$, $p = .500$, in both experiments.

Figure B1

Distributions of Normalized Weights of Advice (WOA) Estimated by Means of Mixed-Effects Regression (MER) With Random Order Effects and Fixed Treatment Effects of Contrast-Coded Advice Distance Condition in Experiments 2 and 3 of Hütter and Ache (2016)



Note. WOA is normalized by dividing by the realized advice sample sizes K_{ij}. Outliers are not displayed in the box plots.

Table B7

Full Multilevel Model of Final Judgment According to Equation 20 With Random Order Effects as Specified in Equation 21 and Fixed Treatment Effects of Contrast-Coded Advice Distance Condition (C) for Experiment 2 of Hütter and Ache (2016)

	Estimate	*95% CI*	*SE*	*t*	*df*	*p*
β_A	0.2555	0.1715 – 0.3396	0.0421	6.0698	65.3792	**<0.001**
$\beta_{A\times C}$	0.3063	0.2547 – 0.3579	0.0263	11.6354	7036.2957	**<0.001**
σ	0.0053	0.0053 – 0.0054				
$\tau_{A,S}$	0.2455	0.1945 – 0.2979				
$\tau_{A,T}$	0.0647	0.0432 – 0.0868				
$\tau_{A,U}$	0.0000	0.0000 – 0.0126				
ICC	-					
N	44					
M	20					
K	20					
Obs.	7086					
R^2_m / R^2_c	0.68 / -					

Note. Bold values indicate $p < 0.05$. Wald 95% CI for fixed and bootstrap 95% CI (with 1,000 iterations) for random effects are shown.

Table B8

Full Multilevel Model of Final Judgment According to Equation 20 With Random Order Effects as Specified in Equation 21 and Fixed Treatment Effects of Contrast-Coded Advice Distance Condition (C) for Experiment 3 of Hütter and Ache (2016)

	Estimate	*95% CI*	*SE*	*t*	*df*	*p*
β_A	0.1406	0.1094 – 0.1718	0.0156	8.9941	68.9344	**<0.001**
$\beta_{A\times C}$	0.0698	0.0593 – 0.0802	0.0053	13.0911	11813.4455	**<0.001**
σ	0.0063	0.0062 – 0.0064				
$\tau_{A,S}$	0.0948	0.0766 – 0.1135				
$\tau_{A,T}$	0.0393	0.0258 – 0.0525				
$\tau_{A,U}$	0.0000	0.0000 – 0.0096				
ICC	-					
N	58					
M	20					
K	20					
Obs.	11896					
R^2_m / R^2_c	0.23 / -					

Note. Bold values indicate $p < 0.05$. Wald 95% CI for fixed and bootstrap 95% CI (with 1,000 iterations) for random effects are shown.

Table B9

Full Multilevel Model of Final Judgment According to Equation 20 With Random Order Effects as Specified in Equation 21 and Linear and Logarithmic Fixed Effects of Relative Absolute Advice Distance (D) for Experiments 2 of Hütter and Ache (2016)

	Estimate	*95% CI*	*SE*	*t*	*df*	*p*
β_A	1.3411	1.1560 – 1.5261	0.0943	14.2167	1388.6422	**<0.001**
$\beta_{A \times D}$	-1.0947	-1.2928 – -0.8967	0.1010	-10.8346	7060.3014	**<0.001**
$\beta_{A \times \log(D)}$	1.7932	1.4550 – 2.1315	0.1726	10.3923	7005.8724	**<0.001**
σ	0.0053	0.0053 – 0.0054				
$\tau_{A,S}$	0.2436	0.1931 – 0.2956				
$\tau_{A,T}$	0.0625	0.0421 – 0.0838				
$\tau_{A,U}$	0.0000	0.0000 – 0.0126				
ICC	-					
N	44					
M	20					
K	20					
Obs.	7086					
R^2_m / R^2_c	0.69 / -					

Note. Bold values indicate $p < 0.05$. Wald 95% CI for fixed and bootstrap 95% CI (with 1,000 iterations) for random effects are shown.

Table B10

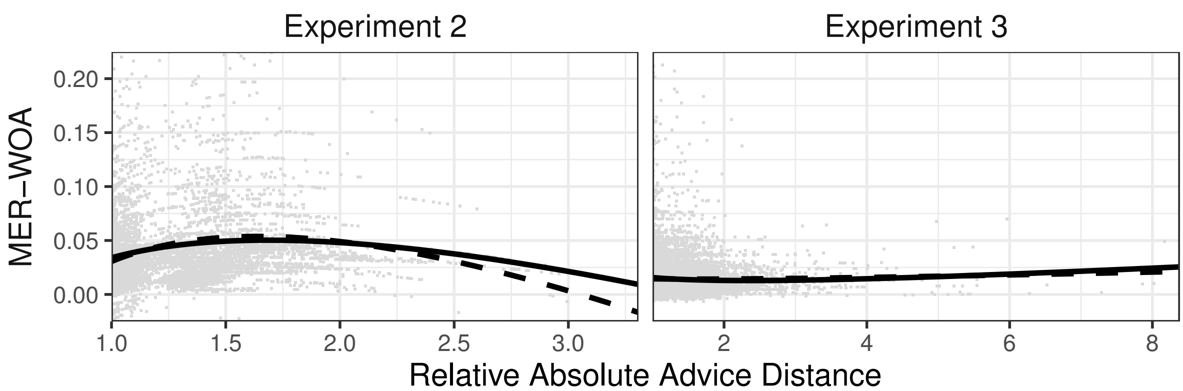
Full Multilevel Model of Final Judgment According to Equation 20 With Random Order Effects as Specified in Equation 21 and Linear and Logarithmic Fixed Effects of Relative Absolute Advice Distance (D) for Experiments 3 of Hütter and Ache (2016)

	Estimate	*95% CI*	*SE*	*t*	*df*	*p*
β_A	0.1309	0.0933 – 0.1685	0.0190	6.8783	142.9552	**<0.001**
$\beta_{A\times D}$	0.0177	-0.0121 – 0.0475	0.0152	1.1629	11785.5846	0.245
$\beta_{A\times \log(D)}$	-0.0287	-0.0994 – 0.0419	0.0360	-0.7967	11839.8614	0.426
σ	0.0063	0.0063 – 0.0064				
$\tau_{A,S}$	0.0947	0.0766 – 0.1135				
$\tau_{A,T}$	0.0403	0.0267 – 0.0540				
$\tau_{A,U}$	0.0000	0.0000 – 0.0096				
ICC	-					
N	58					
M	20					
K	20					
Obs.	11896					
R^2_m / R^2_c	0.22 / -					

Note. Bold values indicate $p < 0.05$. Wald 95% CI for fixed and bootstrap 95% CI (with 1,000 iterations) for random effects are shown.

Figure B2

Normalized Weights of Advice (WOA) Estimated by Means of Mixed-Effects Regression (MER) as Specified in Tables B9 and B10 as Functions of Relative Absolute Advice Distance (D) in Experiments 2 and 3 of Hütter and Ache (2016)



Note. WOA is normalized by dividing by the realized advice sample sizes K_{ij}. Plotting of WOA is truncated outside the interval $[-0.0250, 0.2250]$. The solid lines display the linear plus logarithmic fits of normalized weights. The dashed lines display the functional forms according to the estimated fixed effects as reported in Tables B9 and B10 (for the fixed effects divided by the respective average realized advice sample sizes), respectively.

This dissertation was supported by a grant from the Deutsche Forschungsgemeinschaft (DFG) to the Research Training Group "Statistical Modeling in Psychology" (GRK 2277).