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).

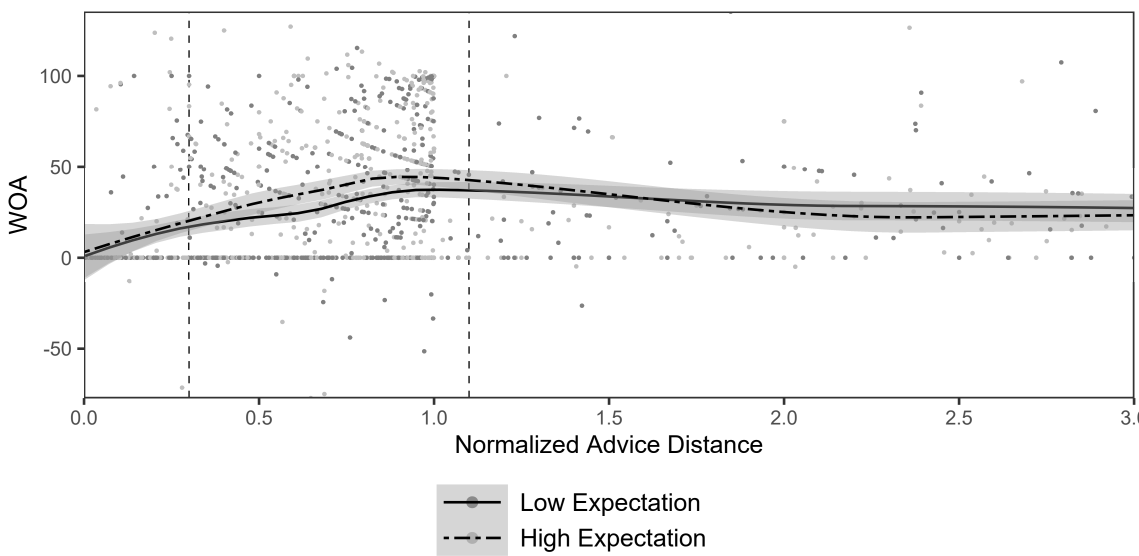

Figure 1: Scatter plots and local polynomial regression fits (incl. 95% confidence bands) for WOA by advice distance as a function of advice expectation in the condition without the dissonance thermometer (N = 98) of Experiment 1. The area enclosed by the thin dashed vertical lines indicates advice of 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

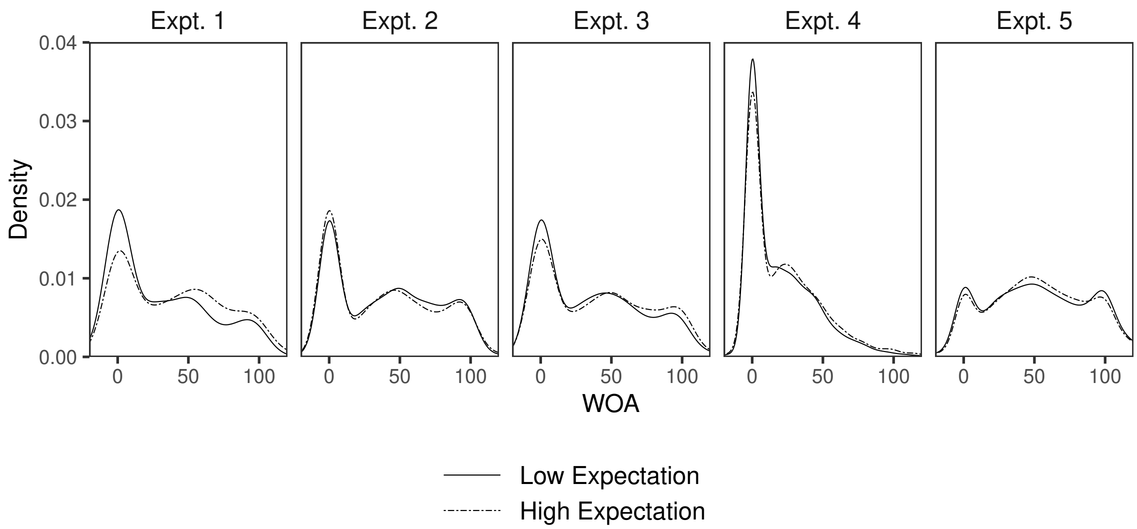

Figure 2: Gaussian kernel density plots of WOA (outliers excluded) as functions of advice expectation in all five experiments. The bandwidth is chosen according to Silverman's (1986) rule of thumb. For Experiment 1, the conditions which yielded positive results post-hoc (i.e., without the dissonance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