This article was downloaded by: [155.246.103.35] On: 05 April 2017, At: 20:38
Publisher: Institute for Operations Research and the Management Sciences (INFORMS)
INFORMS is located in Maryland, USA

Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Hit or Miss: What Leads Experts to Take Advice for Long-Term Judgments?

Philipp Ecken, Richard Pibernik

To cite this article:
Philipp Ecken, Richard Pibernik (2016) Hit or Miss: What Leads Experts to Take Advice for Long-Term Judgments?. Management Science 62(7):2002-2021. http://dx.doi.org/10.1287/mnsc.2015.2219

Full terms and conditions of use: http://pubsonline.informs.org/page/terms-and-conditions

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2016, INFORMS

Please scroll down for article—it is on subsequent pages

INFORMS is the largest professional society in the world for professionals in the fields of operations research, management science, and analytics.
For more information on INFORMS, its publications, membership, or meetings visit http://www.informs.org

MANAGEMENT SCIENCE
Vol. 62, No. 7, July 2016, pp. 2002–2021
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)

http://dx.doi.org/10.1287/mnsc.2015.2219
© 2016 INFORMS

Hit or Miss: What Leads Experts to Take Advice for Long-Term Judgments?

Philipp Ecken
Julius-Maximilians-Universität Würzburg, 97070 Würzburg, Germany; and EBS Business School, 65189 Wiesbaden, Germany, philipp.ecken@ebs.edu

Richard Pibernik
Julius-Maximilians-Universität Würzburg, 97070 Würzburg, Germany, richard.pibernik@uni-wuerzburg.de

Managers and policy makers frequently face crucial strategic decisions that inevitably rely on judgments about relevant future events. These judgments are often characterized by very high uncertainty and the absence of experience from previous good or bad judgments. Judgments of other experts are oftentimes an important—sometimes the only—source of additional information to reduce uncertainty and improve judgment accuracy. However, in many practical situations, decision makers have very limited means to evaluate the quality of such "advice" from other experts and could tend to ignore this valid source of information. In this paper, we study what leads decision makers to take advice from an expert panel when judging the probability of far-future events with high economic impact. Our analysis is based on a unique data set that comprises more than 15,000 advice-taking decisions made by almost 1,000 experts from different industries. We find that decision makers have a strong tendency to ignore advice, which pronounces even further when conflicts in terms of beliefs, past experiences, or desires arise.

Keywords: advice taking; belief updating; conflict; long-term judgments; expert judgment
History: Received April 17, 2013; accepted March 9, 2015, by Yuval Rottenstreich, judgment and decision making. Published online in *Articles in Advance* August 19, 2015.

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 20:38 . For personal use only, all rights reserved.

1. Introduction

Managers and policy makers frequently face crucial strategic decisions that inevitably rely on judgments about relevant future events. These judgments are often characterized by very high uncertainty and the absence of experience from previous good or bad judgments. Judgments of other experts are oftentimes an important—sometimes the only—source of additional information to reduce uncertainty and improve judgment accuracy. However, in many practical situations decision makers have very limited means to evaluate the quality of such "advice" from other experts and could tend to ignore this valid source of information.

The following example of the automotive industry aptly depicts the situation. Most car manufacturers have acknowledged that electric vehicles and plug-in hybrids will be vital in the industry's future; their judgment of the market share of electric cars over the next one to two decades will influence how they evaluate a number of strategic options (e.g., investments in technology, acquisitions, or joint ventures with battery suppliers). Apparently, there is some dissent among experts regarding the market share of electric vehicles: "In January, Mr. Ghosn [chief executive officer of Renault and Nissan] stood by his forecast that plug-in cars would account for 10 per cent of the overall car market by 2020—a bullish prognosis not shared by most industry analysts" (Reed 2012). Based on this "bullish prognosis," Renault and Nissan are investing 4 billion Euros in electric cars and batteries. Hence, it is important to understand how the chief executive officer of Renault and Nissan arrived at this (bullish) prognosis and, more specifically, how he revised his prior judgment in light of conflicting "advice" from industry analysts and peers.

So far, there is limited insight concerning how the task characteristics of such "long-term judgments" determine how advice is processed.[1] Previous advice-taking studies were primarily based on intellective judgment tasks, where subjects answered trivia questions for which a demonstrably correct solution existed, e.g., "estimate the average high temperature in January of 20 U.S. cities" (Mannes 2009, p. 1270). A common phenomenon occurs in this domain called "advice discounting" (Bonaccio and Dalal 2006, Harvey and Fischer 1997, Yaniv 2004): people hamper the accuracy of their final judgment accuracy by not giving enough weight to advice

[1] Tetlock (2005) provides a detailed analysis of experts' performance in long-term political forecasting, and in particular studies how they process information about their forecast accuracy *after* uncertainty has been resolved. In our work, we only refer to judgments about events in the future that have not yet materialized and study how advice is processed *before* uncertainty has been resolved.

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 20:38 . For personal use only, all rights reserved.

or, more precisely, because they ignore advice too often (Soll and Larrick 2009). To understand the advice discounting phenomenon, previous literature has explored the relationship between advice taking and several characteristics of the person judging the advice (the "judge"), the advisor, and the task itself (for a review, see Bonaccio and Dalal 2006; for recent contributions, see, e.g., Soll and Larrick 2009, Tost et al. 2012). These studies indicate that objective measures of high quality of advice (e.g., good past performance of the advisor) can counteract the judge's tendency to discount or ignore advice (Sniezek and Van Swol 2001, Tost et al. 2012, Yaniv and Kleinberger 2000). Unfortunately, in many important decision-making contexts in the real world, there is no demonstrably correct solution in a short period of time, and judges have very limited access to objective measures of advice quality (Bonaccio and Dalal 2010). This is particularly true in strategic and long-term decision making with high complexity, little regularity, and no demonstrably correct solution in the near future (Kahneman and Klein 2009, Tetlock 2005, Van Swol 2011). In such situations, it is extremely difficult or even impossible for the judge to evaluate advisors' past performance or to obtain other objective cues to the quality of advice.

This paper analyzes how industry experts take advice when assessing important events in the far-future—e.g., the market share of plug-in cars, as in our example—and do not have objective cues to advice quality. In our study, industry experts made initial judgments about important events in the far future, received collective advice from a group of other experts, and had the opportunity to update their prior belief judgment in light of this additional information. We argue that in the absence of objective quality measures, the judge will resort to some surrogate "soft" measures to develop a perception of the quality of advice. We identify such cues to advice quality from research in social psychology and from prior advice-taking literature, which focuses on other task domains.[2] Based on naïve realism and social comparison theory, we suggest that these cues indicate some form of conflict and that a perceived conflict lowers the judge's "willingness to take advice" (measured as probability to take advice). In our study, these conflicts can occur in two dimensions: (1) *between* the judge and the advisor group, e.g., in terms of the deviation of the collective group judgment from the judge's initial judgment, and (2) *within* the group of advisors, e.g., the dissent of the advisors' judgments.

We conduct an empirical analysis to explore the effects of soft cues to advice quality on judges' advice-taking behavior. Our analysis is based on a unique data set comprising more than 15,000 advice-taking decisions made by almost 1,000 expert judges with diverse backgrounds in terms of industry, nationality, and occupation. Each expert evaluated multiple events in the far future in their respective field of expertise that would have a high impact on a specific industry; they received aggregated advice from an expert panel and had the opportunity to revise their initial judgment, i.e., take advice. Based on this advice, 491 of these expert judges decided to revise at least one of their prior judgments. We use hierarchical generalized linear models (HGLMs) to analyze whether (and to what extent) these judges take advice. The results of our analyses suggest that soft cues to advice quality explain a substantial part of the variance in judges' willingness to take advice. We find that conflicts *between* the judge and the advisor group and *within* the group of advisors increase judges' tendency to ignore advice. These insights provide a better understanding of what leads experts to take more or less advice in the professionally relevant domain of long-term judgments. This is important because a tendency to discount or ignore advice for long-term judgments will most likely have a detrimental impact on the accuracy of long-term forecasts and, ultimately, on the quality of strategic decisions (Tetlock 2005, McDonald and Westphal 2003, McDonald et al. 2008).

2. Research Hypotheses

2.1. Distance to Advice

The first potential cue to advice quality is the distance between the judge's prior belief and the advice the judge receives—we will simply term this difference as "distance to advice" or "distance." It is intuitive to assume that any judge will observe and evaluate how much the advice she receives deviates from her prior judgment. As such, distance to advice is ubiquitous and arguably utmost salient to any advice-taking paradigm. Although previous research provides strong evidence for a significant relationship between distance to advice and advice taking, valid arguments exist for both a positive and a negative effect on willingness to take advice.

In support of a positive relationship, it can be argued that the judge's benefit of taking advice into account increases in distance. Liberman et al. (2012,

[2] We focus on advice/advisor characteristics for two reasons: First, past advice-taking research has investigated numerous judge characteristics that influence advice taking. We have no means to believe that they should function very different in our domain (e.g., judges that feel more powerful take less advice; See et al. 2011). Second, for long-term judgments in particular, Tetlock (2005) provides extensive analyses of how judge characteristics impact experts' willingness to consider other's point of view. In our study, we do not explicitly operationalize subject characteristics, but capture their influence as unobserved subject level effects.

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 20:38 . For personal use only, all rights reserved.

p. 508) suggest that when large distances between the assessments of judge and advisor "reflect differences in the sources of information prompting those assessments, the benefits of taking those assessments [advice] into account are likely to be greatest." Indeed, from a purely statistical perspective, aggregating the advice with the prior judgment should be increasingly beneficial because the correct answer is more likely to lie between the two judgments the more they differ (Larrick and Soll 2006, Minson et al. 2011). Furthermore, presuming that advice is based on valid information, the distance can be viewed as a *learning signal* received by the judge, and the potential benefit of learning from the advisor's valid belief increases the more distant the advice is (cf. Sharot et al. 2011). Simply speaking, the judge's prior belief is challenged more and more as the distance to the (potentially valid) advice increases. As Bonaccio and Dalal (2006) conclude from their extensive literature review, advice-taking can be understood as a function of the judge's estimates of the potential costs and benefits of taking advice. From this perspective, willingness to take advice should increase in distance because the benefit increases while the fundamental costs of taking advice remain constant, e.g., perceived devaluation of the self by giving up one's own belief and cognitive effort to process and react to the advisor's recommendation (Soll and Mannes 2011, Bonaccio and Dalal 2006). Summarizing these arguments, willingness to take advice should increase in distance.

However, as distance increases, advice becomes more conflicting with the judge's prior belief, which could also decrease her willingness to take advice. This notion reflects the tenet of naïve realism (for a review, see Pronin 2004): people generally assume that they see the world objectively and take for granted that others share their view; faced with another's conflicting opinion, "people's faith in their own objectivity often prompts them to view those others as biased" (Pronin 2008, p. 1178). Correspondingly, previous advice-taking studies have suggested that distant advice signals negative advice quality to the judge and reduces advice taking; i.e., distance can act as a negative *quality signal* to the judge. Ravazzolo and Røisland (2011) propose a Bayesian updating model that considers distance to advice as a signal that judges use to update their own, as well as advisor's perceived competence. Assuming asymmetric information about competence, i.e., the judge has better knowledge of her own competence than of her advisor's competence, their model predicts that larger distances decrease advice taking. In an empirical contribution by Yaniv (2004), undergraduate students answered trivia questions (e.g., "In what year was the Suez Canal first opened for use?"; Yaniv 2004, Table 1), received advice from a single source (e.g., another participant; for multiple sources, see Yaniv and Milyavsky 2007), and made their final judgments. Yaniv (2004) observed that participants placed less weight on advice as distance to advice increased and conjectured that large distances lower the perceived quality of advice. Finally, a recent study by Minson et al. (2011) provides empirical support for both a negative effect of distance on advice taking and judges' perception of large distances as negative signals to advice quality.

We argue that these apparently contradicting findings can be explained by taking the value of distance into account. In other words, we suggest that the sign of the relationship between distance to advice and willingness to take advice may change as distance increases. At first, we expect that the judge's willingness to take advice increases in distance. Assuming that the advisor's judgment is based (to a certain extent) on valid differential information, the benefit of learning from advice increases the more distant it is (Liberman et al. 2012, Sharot et al. 2011). However, this argument critically relies on the assumption that judges perceive the advice as valid. If distance increases to an extent that judges perceive advice as being in conflict (rather than in general agreement) with their prior belief, their perceived quality of advice may decrease (Minson et al. 2011, Pronin 2004). This could attenuate and eventually reverse the positive effect of distance on willingness to take advice. Therefore, the positive relationship between willingness to take advice and distance should be curvilinear and reach a context-specific inflection point upon which it turns negative. In short, we expect willingness to take advice to have an inverted U-shaped relationship with distance to advice. Empirical support for this notion is provided in classic studies of attitude change, which have shown that a persuasive message (e.g., an essay arguing for a reduction in the number of sleeping hours; Bochner and Insko 1966, p. 616) becomes more influential (i.e., leads to more opinion change) the more it differs from people's initial view (i.e., the larger distance); however, unless the source is perfectly credible, this positive relationship is curvilinear and turns negative when distance grows too large, resulting in an inverted U-shaped relationship (Aronson et al. 1963, Bochner and Insko 1966, Insko et al. 1966).

Given the particular task characteristics of long-term judgments, an inverted U-shaped relationship should be observable: With low levels of experience and very high uncertainty, the benefit of learning from the (valid) group advice should increase the more distant it is. However, the lack of objective measures to advice quality should lead the judge to attribute a (too) large distance to shortcomings of the advice (e.g., lack of advisor's expertise), rather than

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 20:38 . For personal use only, all rights reserved.

to acknowledge this distance as valuable differential information.

Hypothesis 1. *Ceteris paribus, the judge's willingness to take advice has an inverted U-shaped relationship with distance to advice.*

2.2. Discrepancy with the Advisor

In the context of long-term judgments, judges often receive advice from the same (group of) advisors on multiple occasions (e.g., the judge receives advice from an industry report that assesses multiple future events). From multiple observations of distance to advice, the judge may infer a cue to the quality of the *advisor* that goes beyond the *advice* quality signal of distance in individual instances. In particular, a judge may attribute one or multiple instances of distant advice not just to low individual advice quality (Hypothesis 1), but to the advisor's general lack of competence or trustworthiness. The following illustration motivates our argument (note that we term the average distance between judge and advisor as "discrepancy"): Judge A faces on average larger distances to advice than judge B across several judgments; i.e., judge A's discrepancy with her advisor is higher. This suggests that judge A has, on average, received more conflicting advice than judge B, leading to a different—presumably lower—perception of the quality of the advisor's judgments (Kennedy and Pronin 2008, Yaniv and Kleinberger 2000). Hence, judges with a higher discrepancy with their advisor may perceive the quality of their advisor's judgments as generally lower.

Although it is intuitive to assume that a higher discrepancy should consequently lower the judge's willingness to take advice, this could function as both a direct and an indirect effect. To the best of our knowledge, prior empirical evidence is limited to a recent study by Minson et al. (2011), who investigated the effect of distance to advice and controlled for the average distance between the judge and the advisor, i.e., discrepancy with the advisor, across multiple judgments. Minson et al. (2011) do not find a significant relationship between discrepancy and advice taking (measured as the probability that individuals move more than halfway toward advice), but do not report whether there was an interaction effect between distance and discrepancy. We argue that although discrepancy may not have a direct relationship with advice taking, it could influence how judges process the learning signal they receive from distance to advice. Referring back to our illustration, this entails that judge A processes the learning signal of distance to advice more conservatively since her general perception of advice quality is at a lower level (higher discrepancy with the advisor). Vice versa, judge B should be more inclined to learn from distant advice since her perception of the quality of advisor's judgment is higher (lower discrepancy with the advisor). Therefore we suggest the following:

Hypothesis 2. *Ceteris paribus, the relationship between distance to advice and willingness to take advice is moderated by discrepancy, so that higher discrepancy with the advisor reduces the positive effect of distance to advice on willingness to take advice.*

In conclusion, willingness to take advice should increase with distance to advice, but this effect should be curbed and eventually reversed for larger distance (Hypothesis 1) and should be smaller when judge and advisor conflict with each other more on average (Hypothesis 2).

2.3. Dissent of the Advisor Group

When advice is obtained from a group, the judge usually receives additional information beyond advice (i.e., beyond some aggregate of the group judgment). Most frequently, this additional information will include some measure of dissent that captures the degree to which individual group members hold disagreeing beliefs. Obvious candidates are the within-group standard deviation of the judgments or the interquartile range (IQR). Research on the "wisdom of crowds" effect suggests that dissent is beneficial when it reflects independent judgments from a group of individuals. In this case, dissent indicates diversity of information and methods used by group members for their estimates. Statistically, the collective judgment of a group improves with such diversity up to an extent that it can be more important than the expertise of its individuals (Lamberson and Page 2012). In accordance, much empirical work highlights that dissent can improve judgment accuracy and, thus, advice quality (Lorenz et al. 2011, Yaniv 2011, Yaniv et al. 2009). Partly, this is due to the informational value of independent opinions, which tend to disagree under uncertainty. Consensus, on the other hand, could be the result of interdependent and collectively biased group opinions.

Although the aforementioned research views dissent as a positive signal, people frequently underestimate the value of independent opinions and adhere to consensus instead. Recent studies have found that judges have less confidence in more dissenting group advice, and this has a negative impact on their willingness to take advice (Lorenz et al. 2011, Yaniv et al. 2009; see also Budescu and Rantilla 2000, Yaniv and Milyavsky 2007). Indeed, for advice that is presented as the collective judgment of a group, disagreeing beliefs among group members could be interpreted as uncertainty inherent to the advice. Judges could interpret such crowd uncertainty as reason to question the reliability of the aggregated group judgment. In particular, they may (falsely) believe the group aggregate

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 20:38 . For personal use only, all rights reserved.

to be just as "bad" as the judgments of its "uncertain" individuals (cf. Larrick and Soll 2006) and infer from the dissent of the group a negative cue to the quality of the collective advice. With uncertainty being substantial for long-term judgments, disagreeing expert beliefs are common, and arguably dissent could be a pronounced cue. Hence, we propose the following:

Hypothesis 3. *Ceteris paribus, the judge's willingness to take advice decreases as dissent of the advisor group increases.*

2.4. Differential Desirability

As suggested by social comparison theory, people have a basic tendency to compare their opinions to those they deem similar to themselves (Festinger 1954, Suls et al. 2002). This notion transfers to the advice-taking paradigm insofar that advisor dissimilarity has been found to decrease the influence of advice (Gino et al. 2009, Van Swol 2011, Yaniv et al. 2011). A judge may question the quality of advice from an advisor with dissimilar characteristics if she believes that these characteristics have an influence on advice. When choosing a movie, for example, a judge may perceive advice from someone with dissimilar values to be less informative since the judge may feel that different values indicate a different taste in movies (Van Swol 2011). A comparable effect may be observable in the domain of long-term judgments.

Judges may fear advice to be biased and question its quality when the advisor is dissimilar in terms of the *desirability* for future events to occur. Such information is often available (e.g., industry reports state the desirability of certain developments) or could be inferred from an advisor's background and peers. First, judges may interpret a differential desirability as an indicator for a potential self-interest bias and question advisors' motives (Eisenhardt 1989, Van Swol 2011). This would undermine advisors' trustworthiness and could prompt judges to ignore advice (Bonaccio and Dalal 2010, Gino and Schweitzer 2008, Sniezek and Van Swol 2001, Van Swol 2011). Second, judges may interpret a differential desirability as an indicator for advisors' wishful thinking, i.e., an optimistic bias expressed by a positive relationship between outcome desirability and the probability judgment made by the advisor (for a review, see Krizan and Windschitl 2007). People frequently use naïve theories of biases in judging others (Pronin 2007); one is the perception that others are optimistically biased (Armor et al. 2008, Russo and Yong 2011). Accordingly, people have been found to attribute different opinions to other's wishful thinking (Liberman et al. 2012). Clearly, a perceived bias from wishful thinking would lower the perceived quality of advice and, as a consequence, decrease judges' willingness to accept it.

Both a self-interest bias and a wishful thinking bias may coexist, and it is difficult to distinguish between the two. However, any evidence that such biases may have influenced the advisor's judgment will be perceived as an indicator for low quality of advice. This effect should be particularly pronounced for long-term judgments: given the absence of a demonstrably correct solution, accuracy-related cues are less pronounced, whereas soft cues indicating trustworthiness gain importance in determining the quality of advice (Van Swol 2011, Zarnoth and Sniezek 1997). In sum, we can expect a differential desirability between judge and advisor to derogate perceived quality of advice and thus to decrease judge's willingness to take advice:

Hypothesis 4. *Ceteris paribus, the judge's willingness to take advice decreases as the difference between judge's and advisor's outcome desirability increases.*

3. Method and Measures

3.1. Background and Data

For our analysis we use a unique data set compiled from a series of 21 expert surveys that were conducted between 2008 and 2011. Each expert survey was carried out as part of a so-called "future study," which had the objective of assessing long-term developments and deriving future scenarios for different industries. Most of these future studies were commissioned by private sector companies (e.g., leading consulting firms) for commercial or market intelligence reasons. All of the studies were conducted by a group of researchers at a private business school in Germany. Each one of the studies was carried out by a team of three to four members of this research group (one of the authors collaborated as a team member in four of the surveys).

The context and content of these studies can be exemplified with one particular study that addressed the future of the automotive industry (Warth et al. 2011), especially the future development and the market potential of alternative power train technologies (see the example in the introduction). To this end, 138 automotive experts (19% from original equipment manufacturers, 18% from automotive suppliers, 15% from academia, 9% automotive experts from consulting firms, and the rest from various stakeholder groups such as government, nongovernmental organizations, utility companies, etc.) participated in a survey in which they were asked to evaluate the probability of occurrence of 20 high-impact future events, e.g., "2030: Cars powered by internal combustion engines (gasoline and diesel engines, including efficiency technologies and alternative fuels) still dominate the number of new registrations" and

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 20:38 . For personal use only, all rights reserved.

"2030: Energy for new power train technologies is derived mainly from renewable sources" (a full list of future events of this study is included in Table A.1 in the appendix). The evaluations of the experts were used to develop probable scenarios of the future of the automotive industry and to derive recommendations for different stakeholder groups.

In total 21 of such future studies (with corresponding expert surveys) were conducted for different industries, such as aerospace, financial services, and pharmaceuticals, and had different regional foci, such as the global economy or different developed and developing countries. The number of experts varied between 13 and 138 (mean = 47.1; SD = 34.0), with a total of 966 experts taking part in the different surveys. The number of future events that were evaluated varied between 13 and 30 (mean= 19.6; SD= 3.5) with a total of 411 far-future events. All future events described either economic, political, social, or technological events occurring in 10 to 20 years time (mean = 14; SD= 4.5) with a high relevance. An overview of the different studies, including their industries, regional foci, number of experts, and further information is provided in Table A.2 in the appendix.

All of the expert surveys followed a standardized procedure (for developing future events, recruitment of experts, etc.) and, even more importantly, utilized the same tool for data collection: an online version of the Delphi method (Dalkey and Helmer 1963), the so called "real-time Delphi" (popularized by Gordon and Pease 2006; for a review, see von der Gracht et al. 2011). This method reflects essential elements of the "judge–advisor system" paradigm, which is frequently used in experimental advice-taking research (Bonaccio and Dalal 2006, Sniezek and Van Swol 2001). Specifically, experts made probability judgments about the occurrence of various far-future events (e.g., "2030: Cars powered by internal combustion engines . . . ;" see above); after each initial judgment, experts immediately received feedback (advice) from a panel of other experts and had the opportunity to update their initial judgments. Details regarding the survey design and procedure are provided in the following section.

Since all expert surveys in our data set were based on a standardized research design, the comparability of the results of all 21 expert surveys is ensured. The heterogeneity resulting from the diverse foci of the future studies, in terms of industry and geographical region, provides us with confidence that our analysis is based on a representative data set.

3.2. Survey Design and Procedure

All of the 21 expert surveys were designed and carried out according to a standardized procedure that can be described by the following phases.

3.2.1. Definition of Future Events. First, the members of the research team conducted literature reviews and derived a list of relevant future events for the industry in focus. Thereafter, multiple workshops were held with academics and practitioners to review and consolidate these events. Finally, in the last workshop, participants selected a comprehensive set of relevant future events as the basis for the expert surveys. Events were phrased in short and clear terms, and their descriptions followed established quality guidelines (Rowe and Wright 2001). It was ensured that future events were unambiguous, did not include conditional statements, and used common definitions instead of technical terms. Specific attention was paid so that the described future events were higher ranking in their causes and, thus, not under the expert judges' personal control. Table A.1 in the appendix lists the future events evaluated in the survey that focused on the future of the automotive industry.

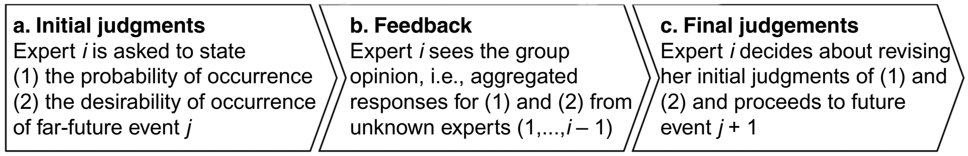
3.2.2. Expert Recruitment. Extensive effort was made to identify and recruit experts with substantial knowledge in the respective industry of study. Potential experts were identified based on prior contributions to the field (e.g., keynote speeches, publications) and their involvement and standing in relevant organizations (e.g., their management level). As highlighted in the previous section, panels included experts from multiple stakeholder groups, e.g., from academia, public institutions, and industry. Identified experts were invited via phone and email to participate in the study. No financial reward was offered for participation. Experts responded to the survey invitation voluntarily, most likely driven by their personal interest or involvement in the industry in focus.

3.2.3. Presurvey Information. Experts were invited to participate in an anonymous expert survey to study future developments in their industry of expertise. The invitation email described some of the key challenges for the industry's future and laid out the main objectives of the study (e.g., developing future scenarios for power train technologies in the automotive industry). Invited experts received a personalized hyperlink directing them to a welcome screen with a description of the survey procedure. Experts were told that they would participate in a "real-time Delphi survey" in which they would be required to evaluate a certain number of future events for their specific industry. It was stated that they would be asked to evaluate each future event's probability and desirability of occurrence; that upon providing their initial assessments, a statistical group opinion of all participants would be calculated momentarily and provided thereafter; and that they would then have the opportunity to reassess and potentially adapt their responses in light of the views of their expert peers.

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 20:38 . For personal use only, all rights reserved.

Figure 1 **Data Collection Process**



Experts were furthermore told that anonymity was assured throughout the survey process. Finally, it was recommended to "consult the tutorial before starting the survey," which would "provide a brief guided tour." The tutorial was a flash animation of the procedure that exemplified the individual steps, i.e., providing an initial assessment, display of the group opinion, revising one's initial judgment, and moving to the next future event. Upon seeing the welcome screen, experts were able to launch the survey.

3.2.4. Data Collection (**Real-Time Delphi**). Each expert $i \in \{1, \ldots, 966\}$ evaluated far-future events j ($j = 1, \ldots, J_i$), with J_i denoting the number of events evaluated by expert i, according to the sequence of activities displayed in Figure 1:

(a) *Initial judgments*. Expert i saw event j at the top of the screen (e.g., "2030: Cars powered by internal combustion engines . . . ;" see above) and was asked to enter "Your answer" for the "Probability of occurrence" in a text field (any integer between 0 and 100) and choose the "desirability of occurrence" on a five-point Likert scale (1 indicating very low and 5 indicating very high). In the following, we denote by $PROBINIT_{ij}$ the initial judgment of expert i for the probability of occurrence of event j, and by $DESINIT_{ij}$ her initial judgment of the desirability of occurrence.

(b) *Feedback*. Upon confirmation of her initial judgment, expert i immediately saw the aggregated group opinion of experts $1, \ldots, i-1$ who already completed the survey. This information included the median of the final judgments of experts $1, \ldots, i-1$ for both probability and desirability of occurrence, displayed as a single number. Box plots additionally visualized the median and the first and third quartiles of the group opinions (i.e., the IQR), as well as the expert i's own prior judgments $PROBINIT_{ij}$ and $DESINIT_{ij}$. An additional visual stimulus to consider the group opinion was provided by the appearance of different color schemes, which framed the group opinion. The color was dependent on how much the expert's initial judgments differed from the group opinion relative to the IQR: (1) green if the absolute distance between the expert's initial judgment and median of the group was smaller than 10% of the IQR of the group's judgments, (2) yellow if the distance was between 10% and 20% of the IQR, (3) orange if the distance was between 20% and 40% of the IQR, or (4) red if the distance was larger than 40% of the IQR. Experts did not receive any further information about the group composition or its size. They had only been informed that they would receive the responses from a group of other high-level industry experts.

(c) *Final judgments*. Based on the information about the group opinion, expert i was able to revise her initial judgments. The instruction on the screen read, "Compare your estimates with the group opinion and revise your first round answers." Specifically, the expert was able to revise $PROBINIT_{ij}$ by changing the integer value she had previously entered in the text field and $DESINIT_{ij}$ by changing the value previously chosen on the five-point Likert scale. We denote the final judgments by $PROBFIN_{ij}$ and $DESFIN_{ij}$. Thereafter, the expert proceeded to evaluate the next event $j+1$.

In sum, the data resulting from these surveys provide a valid basis for our investigation of experts' advice-taking behavior. In particular, the instructions and the survey procedure clearly framed the goal of the task: revising one's prior judgment in light of the group opinion. Hence, even in the absence of financial incentives for accuracy, participants should have perceived their task as deciding whether and how to revise their prior judgments based on the opinion of their expert peers.[3] Furthermore, experts received feedback immediately after confirming their initial judgments; i.e., they had to decide to revise their initial judgment before they could proceed to the next future event. This controlled for external sources of learning between initial and final judgments—the group opinion was the only source of learning that may (or may not) have led experts to revise their initial judgments.

[3] In our research design—as in the real world—it is not possible to provide financial incentives for accuracy of long-term judgments because there is no objectively correct answer on which to base incentives ex ante. Even ex post, it can be quite difficult to hold decision makers accountable for relatively inaccurate long-term forecasts since they have numerous ways to defend their belief systems. For example, it is easily argued that "although the predicted event did not occur, it eventually will (off on timing) or it nearly did (the close call) and would have but for . . . (the exogenous shock)" (Tetlock 2005, p. 22). In addition to these real-world constraints, several past advice-taking studies omitted performance-based incentives (e.g., Gino et al. 2009, Van Swol 2011) or found them to have no effect (e.g., See et al. 2011, Study 4). Furthermore, performance-based incentives could also introduce their own problems in research settings, including biased information processing and overconfidence (Meloy et al. 2006).

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 20:38 . For personal use only, all rights reserved.

3.3. Measures

3.3.1. Dependent Measure. We say that expert i took advice if she decided to change her initial judgment of the probability of event j in light of the group opinion. Formally, we capture this decision by the variable $ADVTAKE_{ij}$, which equals one if $PROBINIT_{ij} \neq PROBFIN_{ij}$ and zero otherwise. From these advice-taking decisions, we derive our main dependent variable, *willingness to take advice*, which we define as the probability that *ADVTAKE* equals one, i.e., $\Pr(ADVTAKE_{ij} = 1)$.

Occasionally, a revised judgment moves into the opposite direction of advice relative to the prior judgment. This happened for only 0.3% of all *ADVTAKE* observations. As is common practice, we interpret those incidences as no advice taking and changed *ADVTAKE* from one to zero (e.g., Gino and Moore 2007, Soll and Larrick 2009, Soll and Mannes 2011).

3.3.2. Independent Measures. In line with past studies, we measure distance to advice by the absolute distance between the judge's prior judgment, $PROBINIT_{ij}$, and the advice, i.e., the median of the group judgments on probability of occurrence, and denote this by $DISTANCE_{ij}$. In accordance with Minson et al. (2011), we measure discrepancy with the advisor by expert i's mean absolute distance over all evaluated far-future events j and define this as $DISCREPA_i = (1/J_i)\sum_{j=1}^{J_i}(DISTANCE_{ij})$. Note that this is a subject-level (i.e., level 2) variable since it only varies between experts i. We denote by $DISSENT_{ij}$ the IQR of group probability judgments that experts saw in the box plot (in addition to group median and their prior judgment) and use it as a measure for dissent of the advisor group. Finally, we measure differential desirability by the absolute difference between the judge's (initial) desirability and the median desirability of the advisor group, denoted as $DIFFDESI_{ij}$.

3.3.3. Controls. Recall that experts, in addition to taking advice on probability of occurrence, were able to change their desirability in light of the group opinion. Consequently, it is possible that decision makers' update of probability is caused by a change in their desirability for it to happen. We control for this by $CHNGDESI_{ij}$, which equals one if experts changed their desirability and zero if they maintained it. Hypothesis 4 should depend on whether judges are confident of their desires and retain them in the light of social influences. A judge with undiscriminating desires should find it hard to compute differential desires (Yaniv et al. 2011). Furthermore, judges were able to observe in the box plot whether their prior judgment lay inside or outside the (interquartile) range of group opinions. To separate a potential inside/outside group effect from the effect of group dissent, we include the control variable $WITHINGR_{ij}$, which equals one (zero) if an expert's prior judgment lies inside (outside) the IQR of group judgments. Moreover, we control for two ordering effects of the survey design. First, since experts participated one after another, later participants received group advice from a larger group. As groups grow wiser with size, though at a diminishing rate, this could impact advice taking positively (Hogarth 1978, Mannes 2009). We control for this possibility using the logarithm of the size of the group that generated the advice. We denote this control variable by $GROUPSIZ_i$. Second, participants' effort could decline over time as they advance in the survey. We control for this possibility by including their progress in the survey, defined by $PROGRESS_{ij} = j$. Finally, we use the level 2 means of our independent variables as controls, since this is recommended for the particular model we use for our analyses (see §4.2). Thus, in addition to *DISCREPA* (which is the level 2 mean of *DISTANCE*) we control for expert i's average of $DISSENT_{ij}$ and $DIFFDESI_{ij}$, denoted by $DISSENT_i^{\sim}$ and $DIFFDESI_i^{\sim}$, respectively. To keep the model lean and comprehensible, we focus on these key control variables. However, we checked the robustness of our results with respect to additional control variables, e.g., color schemes, differences between studies and future events, and extremity of experts' initial judgments.

3.4. Sample

The data set described above already omitted the very first "expert" in every study because these were the facilitators of the surveys and their advice-taking behavior is not representative. Furthermore, in line with the advice-taking literature, we exclude observations where advice equals the prior judgment because an advice-taking decision is hardly measurable in such cases (Gino and Moore 2007, See et al. 2011, Soll and Larrick 2009); this omits 11.8% of observations. The resulting "full sample" contains 15,831 observations from 966 experts and provides a strong basis for a rigorous test of our hypotheses. Yet, in an effort to enhance rigor as much as possible, we propose an additional subsample of this.

Recall that every expert evaluated approximately 20 far-future events on average and made as many advice-taking decisions. A rigorous test of our hypotheses commonly requires an idiographic statistical approach that considers the *within* variance of an individual's decisions to take advice over several judgments (see Cooksey 1996; e.g., Mannes 2009, Minson et al. 2011). However, the variance at the individual level can be zero if experts decide to ignore advice on all occasions. We call those individuals *holdouts* and deliberately exclude them from our sample (this omits 50.1% of the sample). First, holdouts

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 20:38 . For personal use only, all rights reserved.

Table 1 Descriptive Statistics

Variable name	Mean	(SD)	Min	Max	(1)	(2)	(3)	(4)	(5)	(6)
Judgment level (level 1); Advice taker, $n = 7{,}897$; full sample, $n = 15{,}831$										
1. *ADVTAKE*[a]	**0.27**	(**0.44**)	**0**	**1**						
	0.13	(0.34)	0	1						
2. *DISTANCE*	**20.72**	(**13.52**)	**1**	**85**	**0.33****					
	21.95	(14.50)	1	100	0.17**					
3. *DISSENT*	**29.54**	(**12.21**)	**0**	**89**	**−0.06****	**0.10****				
	29.48	(12.67)	0	94	−0.03**	0.09**				
4. *DIFFDESI*	**0.78**	(**0.78**)	**0**	**4**	**0.04****	**0.16****	**0.03****			
	0.82	(0.80)	0	4	0.01	0.19**	0.03**			
5. *CHNGDESI*[a]	**0.16**	(**0.36**)	**0**	**1**	**0.27****	**0.06****	**−0.02***	**0.40****		
	0.09	(0.28)	0	1	0.33**	0.01	−0.01	0.26**		
6. *PROGRESS*	**9.85**	(**5.65**)	**1**	**30**	**−0.05****	**−0.03****	**−0.03****	**−0.06****	**−0.03***	
	9.97	(5.71)	1	30	−0.04**	−0.04**	−0.05**	−0.05**	−0.03**	
7. *WITHINGR*[a]	**0.29**	(**0.45**)	**0**	**1**	**−0.27****	**−0.48****	**0.33****	**−0.07****	**−0.04****	**0.01**
	0.27	(0.44)	0	1	−0.16**	−0.47**	0.31**	−0.09**	−0.01	0.01
Subject level (level 2); Advice taker, $N = 491$; full sample, $N = 966$										
1. *ADVTAKE*	**0.27**	(**0.21**)	**0.03**	**1**						
	0.14	(0.20)	0	1						
2. *DISCREPA*	**20.52**	(**4.82**)	**9.68**	**40.26**	**−0.05**					
	21.70	(6.01)	5.50	63.65	−0.15**					
3. *DISSENT*$^{\sim}$	**29.16**	(**6.75**)	**0**	**54.66**	**−0.09***	**0.22****				
	29.12	(7.75)	0	68.32	−0.04	0.15**				
4. *DIFFDESI*$^{\sim}$	**0.77**	(**0.26**)	**0.22**	**1.77**	**0.02**	**0.37****	**0.06**			
	0.81	(0.29)	0.07	2.63	−0.09**	0.47**	0.07*			
5. *GROUPSIZ*	**3.16**	(**1.01**)	**0**	**4.93**	**0.22****	**−0.08**	**0.02**	**−0.12***		
	3.12	(1.08)	0	4.93	0.13**	−0.08*	0.18**	−0.11**		

Notes. Numbers in bold show the advice taker sample. Values are shown for uncentered variables.
[a]Dummy.
$^{*}p < 0.05$; $^{**}p < 0.01$.

provide less rigorous tests for our hypotheses since they offer no variance (i.e., no information) at the judgment level. Second, the only nontrivial variance they provide lies between holdouts and *advice takers* (the latter are experts who take advice at least once). We analyzed this and found that whether judges decide to ignore advice on all occasions (i.e., decide to be a holdout) is partly due to structural differences caused by the notions we described in our hypotheses. On average, holdouts conflict with their advisors more ($t(964) = 6.1$, $p < 0.001$) and have stronger differential desires ($t(964) = 4.6$, $p < 0.001$), but do not obtain more or less dissenting group advice ($t(964) = 0.2$, $p > 0.5$). Hence, we do not expect that confirmation of our hypotheses depends critically on including or excluding holdouts. In fact, we obtained the same structural results when using the full sample for our analysis as well as when estimating a sample selection model (Heckman 1979). Third, although judgments about long-term developments reflect important dimensions of real-world settings, the absence of a demonstrably correct answer implies that no accuracy-based incentives can be offered. Hence, one might suggest that some participants did not take the task seriously enough, but rather "clicked through it." In that case, excluding holdouts is likely to limit the proportion of none effortful experts.[4]

In sum, we believe that working with the subsample of "advice takers" enhances the rigor of our hypotheses tests while leaving a sizable 491 experts and 7,897 corresponding observations for analysis.

4. Estimation and Results

4.1. Descriptive Statistics

Table 1 depicts descriptive statistics and correlations for our data. First, we find that willingness to take advice is relatively low. On average, judges take advice only 13% of the time (full sample) and 27% of the time for judges that take advice at least once (advice takers). Hence, the experts in our study primarily ignore the advice of their peers. Second, we do

[4] The following should be noted. First, because no financial reward was offered for participation, participants who would have just "clicked through the survey" would have probably not made the effort to respond in the first place or even finish the survey. Second, we conducted several analyses to identify outliers and participants that did not make an effort to answer based on careful consideration (e.g., no variance over judgments). All conclusions indicated that these participants were among the holdouts.

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 20:38 . For personal use only, all rights reserved.

not observe very high correlations between the independent variables, which suggests that multicollinearity is not an issue. This was confirmed by the variance inflation factors (VIFs) with values below the commonly accepted threshold value of five (mean VIF = 1.33 in Model 7).

4.2. Hierarchical Generalized Linear Model

For our analysis, we use HGLMs, which take nesting in our data into account and allow for a nonlinear response variable (Raudenbush and Bryk 2002). Nesting results from the fact that a single expert obtains advice on more than one future event and her decisions to take advice are not independent; a simple hint to this is the variation in *ADVTAKE* between experts (see *ADVTAKE* at the subject level in Table 1). Here, advice-taking decisions (level 1) are nested within experts (level 2). Because of this data structure, there are two components of variance corresponding to each level: variance *within* each expert making several advice-taking decisions (level 1) and variance *between* different experts (level 2). We use random effects to explicitly model the variance between experts' willingness to take advice and to account for unobserved subject-level effects. Our analysis focuses on the within variance, i.e., the advice-taking decisions of an individual judge, who receives advice from the same advisor group over several judgments. This approach is advantageous for two reasons: First, it controls for judge/advisor-specific effects because the judge and the group of advisors are held constant at this level. Second, because our interpretation does not rely on between-subjects comparisons at a single future event, we can make meaningful inferences although advice was not manipulated across experts. Such idiographic regression-based approaches are common in the decision-making literature (Cooksey 1996). More specifically, previous advice-taking studies have used HGLMs when advice was provided by other participants as in our study (e.g., Gino and Moore 2007, Mannes 2009, Minson et al. 2011).

To make unbiased inferences from the within variance, we group mean center all covariates at level 1. This removes the between cluster variance in the level 1 predictors and allows their unbiased interpretation as individual level effects (Enders and Tofighi 2007, Hofmann and Gavin 1998). To avoid model misspecification, we capture the between-cluster variation by including the level 2 means of independent level 1 variables (Raudenbush and Bryk 2002). Note that our approach specifically avoids a bias in the estimation of level 1 coefficients that could occur when level 1 covariates are correlated with the random effects—a concern that the Hausman test intends to reveal and that is often taken as an argument to prefer fixed over random effect models in econometrics (Hausman 1978, Snijders and Berkhof 2008). Furthermore, all level 2 variables are grand mean centered, and thus we obtain meaningful cross-level interactions between *DISCREPA* and *DISTANCE* (Hypothesis 2). We estimate our models using the software package Stata 12, which provides maximum likelihood estimates for HGLMs. This estimation method remedies an underestimation bias of penalized quasi-likelihood (PQL) methods (Rodríguez and Goldman 2001) and provides the opportunity to compare models based on their log-likelihoods.[5] Because robust standard errors for our models are not available with Stata algorithms, we checked that an estimation with robust standard errors (provided by the Stata program GLLAMM; Rabe-Hesketh and Skrondal 2012) was in line with our results. Specifically, the robust standard errors provided no evidence for a misspecification of the model since they hardly deviated from the nonrobust ones (Raudenbush and Bryk 2002).

4.3. Hypotheses Testing

We regress our dependent binary measure $ADVTAKE_{ij}$, which indicates whether expert i decided to revise her prior judgment at future event j on the discussed predictors. We denote the vector of predictors by $\mathbf{X}_{ij}$, and, making the usual assumption for $ADVTAKE_{ij}$ to have a Bernoulli distribution, we define by φ_{ij} the expected probability that expert i takes advice at future event j, i.e., $\varphi_{ij} = \Pr(ADVTAKE_{ij} = 1 \mid \mathbf{X}_{ij})$. Therewith, we specify the log-odds of taking advice as $\eta_{ij} = \log(\varphi_{ij}/(1-\varphi_{ij}))$, which is predicted by the following equations:

$$\begin{aligned}\eta_{ij} = {} & \beta_{0i} + \beta_{1i} DISTANCE_{ij} + \beta_{2i} DISTANCE_{ij}^2 \\ & + \beta_{3i} DISSENT_{ij} + \beta_{4i} DIFFDESI_{ij} \\ & + \beta_{5i} CHNGDESI_{ij} + \beta_{6i} PROGRESS_{ij} \\ & + \beta_{7i} WITHINGR_{ij}, \end{aligned} \tag{1}$$

[5] A disadvantage of the maximum likelihood estimation is that it is computationally demanding, specifically when several random effects are included. To test whether our conclusions are robust when including random slopes for independent variables—in addition to a random intercept—we used the software package HLM 7, which provides PQL estimates (Raudenbush et al. 2011). We found that only the positive effect of *DISTANCE* showed significant additional variation between subjects. However, our conclusions were unaffected, and given the potential underestimation bias of PQL (Rodríguez and Goldman 2001), we refrain from including random slopes in our models. Furthermore, the (non)significance of random slopes requires a careful interpretation because it appeared that it was dependent on whether extrabinomial variation was specified (it is still up to discussion whether extrabinomial variation is recommendable; Fielding and Yang 2005, Skrodal and Rabe-Hesketh 2007). Importantly, our conclusions were robust even with random slopes and when allowing for extrabinomial variation.

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 20:38 . For personal use only, all rights reserved.

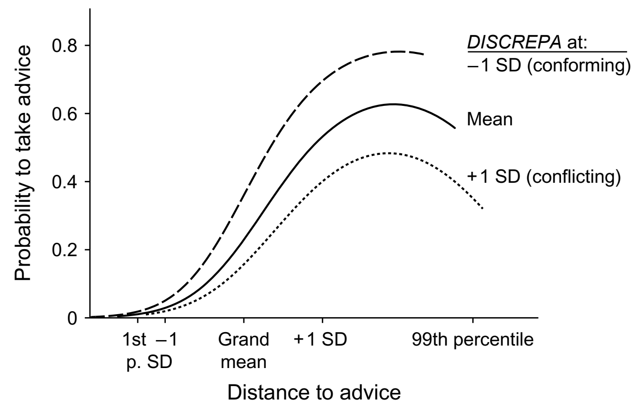
with level 2 units

$$\beta_{ki} = \begin{cases} \gamma_{00} + \gamma_{01} DISCREPA_i & \text{if } k=0, \\ \quad + \gamma_{02} DISSENT_i^{\sim} & \\ \quad + \gamma_{03} DIFFDESI_i^{\sim} & \\ \quad + \gamma_{04} GROUPSIZ_i + \varsigma_{0i} & \\ \gamma_{10} + \gamma_{11} DISCREPA_i & \text{if } k=1, \\ \gamma_{k0} & \text{if } 1<k\leq 7, \end{cases} \tag{2}$$

where $k \in \{0, \ldots, 7\}$ indexes the coefficients, and $\sigma_u = \text{stdev}(\varsigma_{0i})$ captures how the random intercept varies across experts. The model allows for an inverted U-shaped relationship with distance (Hypothesis 1) by including *DISTANCE* and its squared term $DISTANCE^2$ with corresponding coefficients γ_{10} and γ_{20}. The first level 2 expert equation includes the mean intercept γ_{00} plus the random effect ς_{0i} that captures expert i's specific difference from the mean intercept. This is conditional on experts' discrepancy with the advisor, the level 2 means of group dissent and differential desirability, and the size of the advisor group captured by γ_{01}, γ_{02}, γ_{03}, and γ_{04}, respectively. The second level 2 equation captures the mean slope γ_{10} of *DISTANCE*, which is conditional on the effect of experts' *DISCREPA* with the advisor (captured by γ_{11}). The resulting cross level interaction between *DISTANCE* and *DISCREPA* allows for the moderating effect of discrepancy as postulated in Hypothesis 2. Hypothesis 3 is tested by *DISSENT* and its corresponding coefficient γ_{30}. Finally, Hypothesis 4 is integrated in the model by *DIFFDESI* and its coefficient γ_{40}. As discussed, *CHNGDESI*, *PROGRESS*, *WITHINGR*, *GROUPSIZ*, $DISSENT^{\sim}$, and $DIFFDESI^{\sim}$ were included as control variables. Results are reported in Table 2.

Hypothesis 1 is consistently supported across all relevant models. Whereas the probability to take advice increases with *DISTANCE* ($\gamma_{10} > 0$, $p < 0.001$), its negative squared term ($\gamma_{20} < 0$, $p < 0.001$) penalizes large distances and affects willingness to take advice negatively. The resulting inverted U-shaped relationship is depicted by the solid line in Figure 2.[6] Willingness to take advice increases at first, but this effect starts to marginally decrease and eventually reaches an inflection point upon which larger distances decrease advice taking. More rigorous econometric tests confirmed this inverted U-shaped relationship (see Lind and Mehlum 2010; cf. Karim 2009): First, for conservatively chosen extreme points of (group mean centered) *DISTANCE*, we find a positive and statistically significant slope at the left-side extreme point (i.e., at the 1st percentile; $z = 8.46$, $p < 0.001$) whereas the slope at the right-side extreme point is negative and significant (i.e., at the 99th percentile; $z = -4.34$, $p < 0.001$). Second, as recommended by Lind and Mehlum (2010), we use the Fieller method (Fieller 1954) to obtain the 95% confidence interval $[23.6, 29.7]$ around the inflection point 26.3. The confidence interval lies inside the range of our conservatively chosen extreme points (1st and 99th percentiles, $[-21.6, 36.2]$) and thus clearly inside the data range of (group mean centered) *DISTANCE*. Both tests are therefore in support of Hypothesis 1 and confirm that there is a true inverted U-shaped relationship between distance to advice and willingness to take advice.

Figure 2 **Willingness to Take Advice and Distance**



Notes. The figure shows an inverted U-shaped relationship between distance and willingness to take advice and how this relationship is moderated by discrepancy with the advisor. The graph is based on Model 7. All level 1 and level 2 variables not displayed are held at their group or grand means, respectively. The dashed lines are aligned on the common grand mean to make them comparable.

The results in Table 2 also provide support for Hypothesis 2. Whereas discrepancy with the advisor has an insignificant main effect ($p > 0.1$ for γ_{01} across all models but Model 5), it moderates the positive effect of distance to advice on the probability to take it. The cross-level interaction between *DISCREPA* and *DISTANCE* is statistically significant across all models. Importantly, the interaction works into the opposite direction of the main effect; that is, higher *DISCREPA* lowers the positive effect of *DISTANCE* ($\gamma_{11} < 0$, $p < 0.001$). Hence, the curvilinear increase of probability to take advice with distance to advice becomes flatter as discrepancy increases. This supports Hypothesis 2; the positive effect of distance to advice on willingness to take advice is smaller for judges that have a higher discrepancy with their advisor. Note that interactions in logistic regression models depend on all model variables, and their signs can

[6] Note that the probability to take advice is calculated from the logistic function, $\Pr(ADVTAKE = 1 \mid v(\mathbf{X})) = [1 + \exp(-v(\mathbf{X}))]^{-1}$, where $v(\mathbf{X}) = \mathbf{X}\boldsymbol{\gamma}'$ is defined as the linear combination of independent variables (row vector $\mathbf{X}$) and corresponding model coefficients (column vector $\boldsymbol{\gamma}'$).

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 20:38 . For personal use only, all rights reserved.

Table 2 Regression Results of Hypotheses Tests

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Fixed effects							
Intercept	−1.468***	−1.223***	−1.243***	−1.204***	−1.284***	−1.277***	−1.299***
	(0.068)	(0.071)	(0.075)	(0.074)	(0.077)	(0.077)	(0.079)
DISCREPA			−0.013	0.018	0.027	0.019	0.020
			(0.014)	(0.014)	(0.015)	(0.016)	(0.016)
DISTANCE	0.089***	0.128***	0.128***	0.126***	0.118***	0.118***	0.120***
	(0.003)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)
× *DISCREPA*				−0.006***	−0.005***	−0.005***	−0.005***
				(0.001)	(0.001)	(0.001)	(0.001)
DISTANCE^2		−0.002***	−0.002***	−0.002***	−0.002***	−0.002***	−0.002***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
DISSENT					−0.026***	−0.026***	−0.025***
					(0.004)	(0.004)	(0.004)
DIFFDESI						−0.037	−0.351***
						(0.044)	(0.052)
Controls							
CHNGDESI							1.376***
							(0.104)
PROGRESS	−0.004	−0.007	−0.007	−0.005	−0.011	−0.011	−0.014*
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
WITHINGR					−0.779***	−0.778***	−0.784***
					(0.118)	(0.118)	(0.121)
DISSENT~					−0.024*	−0.024*	−0.024*
					(0.011)	(0.011)	(0.011)
DIFFDESI~						0.394	0.381
						(0.283)	(0.289)
GROUPSIZ	0.353***	0.333***	0.328***	0.340***	0.349***	0.359***	0.364***
	(0.066)	(0.068)	(0.068)	(0.068)	(0.069)	(0.070)	(0.071)
Random effect (σ_u)	1.243***	1.293***	1.291***	1.289***	1.318***	1.314***	1.346***
	(0.058)	(0.060)	(0.060)	(0.060)	(0.061)	(0.061)	(0.062)
Wald χ^2	983.4***	1,018.8***	1,019.4***	1,099.3***	1,123.7***	1,124.4***	1,186.8***
Log-likelihood	−3,589.6	−3,477.9***	−3,477.5	−3,431.3***	−3,353.4***	−3,352.1	−3,261.9***

Notes. Standard errors are in parentheses. Level 2 variables are indented. Advice taker sample, $n(N) = 7{,}897(491)$. The likelihood ratio test against the preceding model is indicated by the significance of log-likelihoods.
$^*p < 0.05$; $^{***}p < 0.001$.

change over the range of the predicted probabilities (to take advice). To confirm our finding, we followed Ai and Norton (2003) and calculated cross derivatives. In support of Hypothesis 2, we found that the average change in the slope of the distance trajectories with discrepancy was negative for 98.9% of all cases (mean $= -0.74 \times 10^{-3}$; SD $= 0.32 \times 10^{-3}$; $n = 491$).[7]

Figure 2 plots the interaction effect between distance and discrepancy. The solid line depicts the probability to take advice with *DISTANCE* at mean *DISCREPA*. The short and long dashed lines show this relationship for a "conflicting" and a "conforming" judge–advisor interaction, i.e., at plus and minus one standard deviation of *DISCREPA*, respectively. In the former case, judges conflict with their advisors more on average and demonstrate a less steep increase of probability to take advice with distance, and vice versa in the latter case. Furthermore, as a result of being less sensitive to the learning signal provided by distance, judges in conflicting interactions with their advisors generally take less advice than judges in conforming ones for any given distance. Indeed, the three curves differ significantly at their grand mean (pairwise Wald tests of linear combinations, all $z > 10$; $p < 0.001$).

Hypothesis 3 is also supported. Model 5 in Table 2 introduces *DISSENT* into the regression to account for the effect of group dissent. As suspected, dissent of the advisor group negatively relates to judges' probability to take advice ($\gamma_{30} < 0$, $p < 0.001$). It should be noted that dissent has a negative effect even though we control for an inside/outside group effect, which also has a significant influence ($\gamma_{70} < 0$, $p < 0.001$). Likewise, Hypothesis 4 is strongly supported; the more advisors' outcome desirability for far-future events differs from judges', the less likely advice is taken ($\gamma_{40} < 0$, $p < 0.001$ in Model 7). As

[7] It is worth mentioning that although it is possible to compute the statistical significance of individual cross derivatives, these should be interpreted with caution (Greene 2010; cf. Vissa 2011). As recommended by Greene (2010), we refer to the significance tests at the model level (see Table 2).

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 20:38 . For personal use only, all rights reserved.

expected, this effect is conditional on judges having constant desires. In Model 6, we did not control for *CHNGDESI*, and consequently *DIFFDESI* did not yield a significant effect ($\gamma_{40} < 0$, $p = 0.40$). Noticeably, judges who change their desires should find it harder to compute a differential desirability.

4.4. (Conditional) Weight on Advice

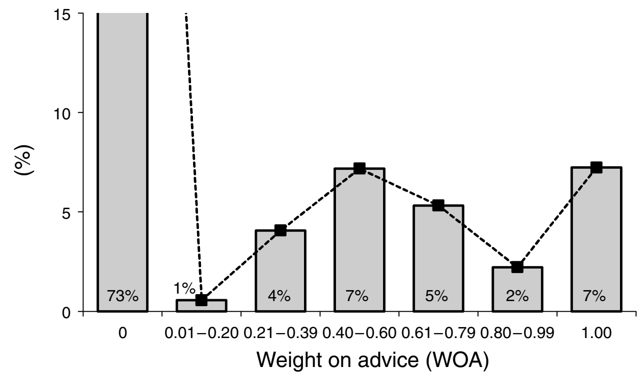
Thus far, our analysis has focused on judges' willingness to take advice, i.e., whether or not they take advice. This conceptualization of advice taking is different from that in previous studies that commonly used (some form of) "weight on advice" (WOA) as the dependent variable. WOA indicates how influential advice is in terms of how far the judge moves from her prior judgment toward the advice:

$$\begin{aligned} \textit{final judgment} &= \textit{WOA} \times \textit{advice} + (1 - \textit{WOA}) \\ &\quad \times \textit{prior judgment} \\ \Leftrightarrow \quad \textit{WOA} &= \frac{\textit{final judgment} - \textit{prior judgment}}{\textit{advice} - \textit{prior judgment}}. \end{aligned} \tag{3}$$

To better understand our results in relation to previous research, we conduct additional analyses with WOA as the dependent variable. Following previous studies (e.g., Soll and Larrick 2009), we calculate WOA as shown in Equation (3) and truncate it to $[0, 1]$ for the rare cases in which a revised judgment fell outside the range of the prior judgment and the advice (1.3% of all cases). Figure 3 plots the distribution of WOA for our data. We observe a similar W-shaped distribution as previous advice-taking studies reported for intellective tasks (Soll and Larrick 2009, Soll and Mannes 2011). Notably, the tendency to ignore advice is more pronounced in our study (more than 70% of the time) than in studies with intellective tasks (30%–50%; see Soll and Larrick 2009, Soll and Mannes 2011).

Model 8 in Table 3 reports the regression results when using WOA as the dependent variable in Model 7. Formally, we substitute η_{ij} in Equation (1) by $\mathbb{E}(WOA_{ij} \mid \mathbf{X}_{ij})$ and thus estimate an HGLM with a linear link function. The results are in accordance with our previous findings: the signs of the relevant coefficients and the levels of significance remain unchanged (see Table 3). At first glance, it seems that our cues to advice quality have the same effects on both how much weight judges give to advice and whether or not judges take advice. However, the regression results with WOA as the dependent variable have to be interpreted with caution. Soll and Larrick (2009) point out that inferences based on WOA can be misleading if the specific characteristics of the underlying distribution are neglected. In fact, when judges mainly ignore advice, as in our study, WOA contains very little additional information beyond willingness

Figure 3 **Distribution of Weight on Advice**



Notes. The distribution of weight on advice is shown for the advice taker sample ($n = 7{,}897$). Average WOA = 17.5%.

to take advice. Hence, we could (falsely) draw conclusions about *how much* judges weight advice, although we primarily measure *whether or not* they take advice. To account for this subtle but important difference, we can think of WOA as measuring (1) judges' willingness to take advice, $\Pr(WOA > 0)$, and (2) the weight judges give to advice if they take advice, $\mathbb{E}(WOA \mid WOA > 0)$. The latter represents the additional information that WOA contains beyond willingness to take advice (we call this "conditional WOA"). Model 9 in Table 3 shows the regression results when using conditional WOA as dependent variable. Formally, we now substitute η_{ij} in Equation (1) by $\mathbb{E}(WOA \mid WOA > 0, \mathbf{X}_{ij})$ and estimate another HGLM with a linear link function. We observe that all distance-related effects, including discrepancy and its interaction with distance, change their direction compared to Model 7. Figure 4 displays the relationship between distance and the dependent variables conditional WOA, willingness to take advice, and WOA. The conditional weight on advice exhibits a U-shaped relationship with distance, whereas willingness to take advice has an inverted U-shaped relationship with distance. WOA aggregates these two effects and, as a result, has a concave relationship with distance. We do not find such structural differences for our other explanatory variables dissent and differential desirability. Both have a negative relationship with conditional WOA and willingness to take advice. Their effect on conditional WOA is significant for dissent and marginally significant for differential desirability.

5. Discussion

Our study provides first empirical insights into the advice-taking behavior for long-term judgments. We find that it is structurally similar to what past research—in other domains—has found. In fact, the expert judges in our study used the typical advice-taking strategies identified by Soll and Larrick (2009):

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 20:38 . For personal use only, all rights reserved.

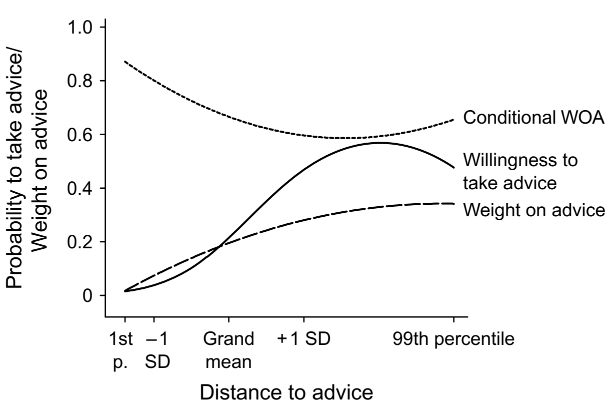
Table 3 Regression Results of Additional Analysis

	Model 7 Willingness to take advice	Model 8 Weight on advice[a]	Model 9 Conditional WOA[a]
Fixed effects			
Intercept	−1.299*** (0.079)	1.955*** (0.079)	6.659*** (0.096)
DISCREPA	0.020 (0.016)	−0.011 (0.016)	−0.076*** (0.020)
DISTANCE	0.120*** (0.005)	0.079*** (0.003)	−0.079*** (0.007)
× *DISCREPA*	−0.005*** (0.001)	−0.005*** (0.000)	0.002* (0.001)
DISTANCE^2	−0.002*** (0.000)	−0.001*** (0.000)	0.002*** (0.000)
DISSENT	−0.025*** (0.004)	−0.026*** (0.003)	−0.029*** (0.005)
DIFFDESI	−0.351*** (0.052)	−0.334*** (0.044)	−0.141† (0.080)
Controls			
CHNGDESI	1.376*** (0.104)	1.593*** (0.099)	0.609*** (0.137)
PROGRESS	−0.014* (0.006)	−0.014** (0.005)	−0.019* (0.009)
WITHINGR	−0.784*** (0.121)	−0.402*** (0.092)	0.599** (0.207)
DISSENT~	−0.024* (0.011)	−0.027** (0.011)	−0.014 (0.013)
DIFFDESI~	0.381 (0.289)	0.451 (0.293)	0.177 (0.327)
GROUPSIZ	0.364*** (0.071)	0.291*** (0.071)	−0.101 (0.082)
Random effect (σ_u)	1.346*** (0.062)	1.433*** (0.055)	1.230*** (0.075)
Wald χ^2	1,186.9***	1,582.8***	263.0***
Log-likelihood	−3,261.9	−1,057.9	70.6
n(*N*)	7,897(491)	7,897(491)	2,098(491)

Notes. Standard errors are in parentheses. Level 2 variables are indented.
[a]Coefficients, standard errors, and random effects are shown at ×10.
†$p < 0.1$; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

they primarily chose their prior judgment ($WOA = 0$), chose the advice ($WOA = 1$), or averaged the two ($WOA \approx 0.5$), resulting in the predicted W-shaped distribution of WOA (see Figure 3). There was also a clear preference of the judges to stick with their prior judgment rather than to update their judgment with the advice they received. We find that this tendency is particularly pronounced in our study of long-term judgments. Compared to the studies of Soll and Larrick (2009) and Soll and Mannes (2011), in which 30% to 50% chose not to update, our judges ignored advice more than 70% of the time. It is worth emphasizing this finding: experts ignored the opinions of their peers more than 7 out of 10 times in a domain that was professionally relevant. This occurred despite the fact that (1) the task was exceedingly difficult, which should have increased judges willingness to take advice (Gino and Moore 2007), and (2) advice in our study was the aggregated judgment of a group of other experts and, hence, should have represented a valid source of information to the judge (Mannes 2009).[8]

Figure 4 Measures of Advice Taking and Distance



Note. Differential relationships between distance to advice and measures of advice taking are shown. The graph is based on Models 7–9. All level 1 and level 2 variables not displayed are held at their group or grand means, respectively.

The relatively low rate of advice taking in our study seems counterintuitive at first. However, it is most likely a result of both sample and task characteristics. First, judges in our study were experienced professionals (not students) and had relevant expertise in the industries they made judgments for. Their experience could have been "conducive to overconfidence" because the feedback that they "receive from their failures in long-term judgments is delayed, sparse, and ambiguous" (Kahneman and Klein 2009, p. 523). Hence, experts in our study likely overestimated their own ability to make long-term judgments, leading

[8] Note that aggregated expert judgment is indeed an important source of information for long-term judgments (Duboff 2007). This is also reflected by many commercial studies as well as a multitude of studies commissioned by public institutions that try to predict important future developments based on some form of aggregation of expert opinions (e.g., Alfaro et al. 2005, Boston Consulting Group 2010).

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 20:38 . For personal use only, all rights reserved.

to a low willingness to take advice (Bonaccio and Dalal 2006, Soll and Larrick 2009). Furthermore, for long-term judgments there is no demonstrably correct solution in proximate time, and it is impossible to determine the quality of advice by objective measures (ex ante). A judge may easily argue, "if the advisor does not have an objectively better solution, why should my own belief be any less appropriate?" Without objective measures for the quality of advice, there is little to counteract basic tendencies such as egocentrism and overconfidence (Krueger 2003), and their negative effect on willingness to take advice will likely be more pronounced (Van Swol 2011).

Our results do, however, indicate that judges use the scarce information they have to evaluate the quality of advice for long-term judgments. We find empirical support for our initial premise that judges will resort to soft cues to advice quality when deciding whether or not to update their prior belief based on the advice they receive. In the following sections, we will provide a detailed discussion of the individual effects of these cues. It is interesting to observe that beyond the individual effects, a more general theme emerges from our results: judges are sensitive to any evidence that indicates some form of conflict. A conflicting relationship with the group of advisors (in terms of absolute or average distance to advice as well as differential desirability) and conflicting opinions within the group of advisors (i.e., dissent) reduce judges' willingness to take advice. Put differently, judges in our study prefer group advice from others that are (1) collectively similar to themselves and (2) similar among each other. This is generally in line with social comparison theory (Suls et al. 2002); when (1) and (2) are high, the group as a whole will be most similar and, at the same time, most influential. Recently, Mannes (2009, p. 1268) addressed the "gap in our understanding of group informational influence" on an individual's belief and demonstrated that group influence increases in group size. Our findings add to this and suggest that the informational influence of a group on an individual's belief increases (decreases) in two dimensions of similarity (conflict).

Taken together, our explanatory variables explain a substantial part of the total variation in willingness to take advice for long-term judgments. Following Snijders and Bosker (1999), we calculate their extension of McKelvey and Zavoina's (1975) measure for the explained proportion of variance. We find that our variables in Model 7 explain more than one-third (34.9%) of the variance. Besides this considerable explanatory power, it is interesting to note that unobserved subject-level characteristics account for 23.1% of the unexplained variance. Judging from Tetlock's (2005) extensive evidence, we expect that this is—to a certain extent—a result of experts' different "cognitive styles" in making long-term judgments (e.g., the degree of integrative complexity). Note that our advice/advisor characteristics in conjunction with the unobserved subject characteristics make up for more than 55% of the variance in willingness to take advice for long-term judgments.

5.1. Distance to Advice

Judges' willingness to take advice changes depending on how much the advice of the expert panel deviates from judges' prior judgments. Our data suggest that this relationship can best be characterized by a curvilinear function: At first, judges are more willing to take advice as distance increases. At a certain point, however, willingness to take advice grows at diminishing rates and eventually reaches an inflection point after which more distant advice is more likely to be ignored (see Figure 2). As argued previously, distance is the most obvious cue to the quality of advice, and we expected a strong influence on judges' willingness to take advice. This is confirmed by our results and its effect can be observed in Figure 2. For example, when distance changes from one standard deviation below to one standard deviation above the individual mean, willingness to take advice increases from less than 5% to approximately 50%.

The inverted U-shaped relationship we find supports our argument that judges indeed process distance both as learning and quality signals (Minson et al. 2011, Sharot et al. 2011). The judge perceives advice as increasingly beneficial with distance; beyond a certain point, however, she gradually questions the quality of advice and, as a consequence, tends to ignore advice more often.[9] This inverted U-shaped relationship is generally in line with findings from classic studies on attitude change (e.g., Bochner

[9] An anonymous reviewer correctly pointed toward another explanation for the relationship between distance and willingness to take advice: when judges are highly uncertain about an item, they may tend to give answers that are far away from the average answer and, thus, receive very distant advice. Because of their uncertainty—not because of distance per se—they may be more willing to take advice. Although such an effect cannot be strictly ruled out in the present study, we believe that in our particular setting uncertainty does not explain the (curvilinear) relationship between distance and willingness to take advice observable in our data. First, we cannot expect a positive relationship between judges' uncertainty and distance. Judges that are very uncertain will most likely provide a response close to 50%. In our study, the majority of the aggregate group advice was between 40% and 60%. Hence, we believe that experts in our study who were uncertain about an item would in most cases not have faced a very high distance to advice. Second, uncertainty as an explanation by itself would predict that willingness to take advice strictly increases in distance (uncertainty); this is not in line with the inverse U-shaped relationship for which we find strong empirical evidence.

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 20:38 . For personal use only, all rights reserved.

and Insko 1966). However, it is structurally different from previous findings of advice-taking studies, which suggest a negative relationship between advice taking and distance (Minson et al. 2011, Yaniv 2004). These studies focused on how much judges weight advice (WOA) rather than whether or not they take advice (willingness to take advice). In fact, our additional analysis provides evidence that the relationship between distance and advice taking may depend on how advice taking is operationalized (see Figure 4). In our study, WOA measured little beyond willingness to take advice because judges primarily ignored it. As a result, WOA also has an inverted U-shaped relationship with distance (although less pronounced). However, by considering the additional information in WOA beyond willingness to take advice, we were able to replicate part of the negative relationship between weight on advice and distance found in previous studies: if judges decide to take advice, their conditional weight on advice has a U-shaped relationship with distance, which indicates that—up to a certain point—the judge puts less weight on advice as distance increases. Interpreting this in light of the advice-taking strategies proposed by Soll and Larrick (2009), we can assume that in the rare cases where experts decide to take advice for small distances (i.e., lower than the mean distance in Figure 4), they tend to choose the advisor's judgment (i.e., $WOA = 1$). For example, if an expert decides to take advice when her initial judgment is 55% and the advice is 60%, she will tend to choose the advice rather than weighting it with her initial judgment. As distance increases, an expert that decides to take advice will more frequently average her initial judgment with the advice (i.e., $WOA \approx 0.5$).

It should be noted that the relationship between (conditional) WOA and distance must be interpreted with caution for two reasons: First, when (conditional) WOA is regressed on distance, a negative relationship could artificially be induced because distance is used as the denominator in calculating WOA (for a related argument, see Bonaccio and Dalal 2006). Second, advice was generally more distant for observations where $WOA > 0$ because the probability to take advice increased in distance. Hence, the negative association between conditional WOA and distance could also result from the fact that the majority of advice was perceived as conflicting.[10]

[10] In an effort to validate and to better understand the distance effect, we analyzed the experimental data reported in the four studies of Soll and Larrick (2009) (we thank Soll and Larrick for providing their data). In these four experiments, students completed intellective judgments and received advice from another student. Using our analytical approach, we find that both willingness to take advice and conditional WOA exhibit the same structural relationship with distance that we observe in our data.

Summarizing our discussion, judges' willingness to take advice has an inverted U-shaped relationship with distance. However, although judges are more willing to take more distant advice (up to a certain point), this does not necessarily imply that they also give a higher weight to advice. Hence, it is important to allow for the possibility that the effect of distance on advice-taking differs between (1) whether or not judges take advice and (2) how much weight judges give to advice (if they take it). Using WOA exclusively as a measure for advice taking may conceal these structural differences because it aggregates both effects.

5.2. Discrepancy with the Advisor

Our results for Hypothesis 2 demonstrate that the positive effect of distance is moderated by judges' discrepancy with the advisor, i.e., the average distance between the two. We find that the positive influence of distance to advice on judges' willingness to take advice decreases in discrepancy. As can be seen from Figure 2, this effect is substantial. Comparing willingness to take advice at grand mean distance, we predict that judges in a conflicting relationship with the advisor (i.e., discrepancy at plus one standard deviation) are willing to take advice approximately 15% of the time, whereas judges in a conforming relationship do so approximately 35% of the time. Moreover, the latter are more willing to take advice for any given distance than the former.

This supports our argument that frequent conflicts with the advisor cause a generally lower perception of the quality of the advisor's judgment, which asymmetrically looms to subsequent interactions (Yaniv and Kleinberger 2000). This argument implicitly hinges on the premise that past observations of distance determine how judges process future learning signals from distance. To provide further support for this notion, we also used the "current discrepancy" of judge i with the advisor at judgment $\tilde{j}$, defined as $(1/(\tilde{j}-1)) \sum_{j=1}^{\tilde{j}-1}(DISTANCE_{ij})$, as a moderating variable and obtained comparable results. Hence, our findings suggest that discrepancy reflects a general perception of the quality of advisor's judgments that builds from past distances of advice and impacts willingness to take advice negatively but indirectly through a change in how learning signals from subsequent distances are processed. We conclude that it is important to single out the effects of distance and discrepancy whenever judges interact with the same advisor on several occasions. To the best of our knowledge, this has not been considered by past research with the exception of Minson et al. (2011). They also did not find a significant direct effect of discrepancy on advice taking, but did not report whether there was a moderating effect on distance.

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 20:38 . For personal use only, all rights reserved.

5.3. Dissent of the Advisor Group

The willingness to take advice of the judges in our study decreases as dissent within the group of advisors increases. They perceive dissent as a negative signal to the quality of (aggregated) advice. Our results suggest that this effect is again substantial. For example, willingness to take advice almost halves (from above 30% to below 20%) when the group dissent observed by an individual judge increases from minus to plus one standard deviation. In accordance, judges' conditional weight on advice decreases in dissent.

The negative impact of advisor dissent on willingness to take advice for long-term judgments is generally in line with prior results, e.g., for trivia questions (Yaniv et al. 2009). Our judges also adhere to consensus and perceive agreement within the group of advisors as a signal of lower uncertainty of (aggregated) advice and, as a consequence, higher reliability. It appears that judges do not fear that consensus may be a result of interdependent and collectively biased expert beliefs. Likewise, dissent seems to be an indicator of higher uncertainty of the advice rather than valuable "group wisdom" stemming from independent expert judgments (Lorenz et al. 2011). Mannes (2009) concludes that people misappreciate the wisdom of crowds by giving too much weight to their own judgment when faced with collective group advice. Our results suggest that this effect may be even more pronounced when judges know about the distribution of (independent) group judgments under conditions of high uncertainty. This extends insights gained from the yet small number of advice-taking studies that consider collective group advice. With the exception of Mannes (2009), other advice-taking studies typically present group advice as a set of individual opinions rather than an aggregated group opinion. The benefit of presenting the judge with aggregated advice and a measure of advisor dissent is that the effects of distance and dissent can be addressed individually. This is not the case when the judge is confronted with a set of opinions from individual advisors; when facing dissenting advice, for example, a judge may tend to give more weight to individual advisors whose advice is less distant (Yaniv and Milyavsky 2007). This raises the question of whether dissent and distance are independent cues to advice quality. To further explore this issue, we included the interaction between distance and dissent (in Model 7). Our results indicate that there is no statistically significant interaction effect between the two ($\gamma > 0$, $p = 0.76$). Hence, the negative association between dissent and judges' willingness to take advice does not change depending on the distance to the collective group advice. Vice versa, distance to advice does not have a different effect in light of varying dissent among the advisors. We therefore conclude that judges perceive dissent of the advisor group as a strong negative signal to the quality of advice and that this signal functions independently of how distant the collective group judgment is.

5.4. Differential Desirability

Judges perceive differential desirability (i.e., the advisor states a desirability that deviates from the judge's) as a negative signal for advice quality. Willingness to take advice decreases considerably as the difference between the judge's and advisor's desirability increases. For instance, willingness to take advice decreases from approximately 25% to 15% when the differential desirability observed by the judge increases from minus to plus one standard deviation of the individual mean. In accordance, judges' conditional weight on advice has a negative, though marginally significant, relationship with differential desirability.

These results support our argument that the judge may perceive the existence of differential desirability as a signal for a biased advisor and, consequently, low advice quality. For example, the judge could suspect a bias from dissimilar self-interests and question the advisor's trustworthiness. The judge may also believe that the advisor is optimistically biased (e.g., Armor et al. 2008). More precisely, the judge could presume a bias from wishful thinking: the advisor's higher (lower) probability judgment is driven by the advisor's higher (lower) desirability. It is important to note that our results are conditional on the judge being confident of her desirability and retaining it after seeing the desirability of the advisor (i.e., the expert group). When judges change their desires in light of this social influence, they have no basis from which to compute a differential desirability, and there is no significant effect (Yaniv et al. 2011).

6. Conclusions and Implications

Our results indicate that a tendency to *not* update prior beliefs with additional relevant information is not a phenomenon that is observed merely in lab settings and for trivia questions, but also occurs in a professionally relevant context. In our domain of long-term judgments, experienced professionals show a pronounced tendency to ignore the collective wisdom of other experts. This information, we considered as *advice*, was free of charge in our study, and it would be unreasonable to assume that an individual already had sufficient information given the high uncertainty associated with long-term judgments (e.g., the market penetration of electric cars in 10 years). Although expert advice may not always be free of charge and paying for it could increase advice taking (Gino 2008), most relevant information for long-term judgments becomes publicly available over time, and decision makers could err in relying too heavily on their own

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 20:38 . For personal use only, all rights reserved.

prior beliefs. In any case, it is a particular characteristic of long-term judgments that one can hardly assess judgment performance—simply because there are no hard measures to it. This is true both for the specific judgments made in our study and for those that are an integral part of crucial strategic decisions in the corporate world (e.g., investments in alternative power train technologies). For this reason, it is virtually impossible to determine what leads to good or bad long-term judgments, ex ante, and to infer optimal levels of advice taking. There is, however, broad consensus that the behavior we find in our study can lead to both inaccurate long-term judgments and poor strategic decision making. More specifically, past research has shown that other's conflicting opinions are particularly valuable, and that ignoring them can have a strong negative impact (e.g., McDonald and Westphal 2003, Tetlock 2005). Interestingly, the results of our study suggest that conflict is in fact an important reason for judges to ignore advice. Conflicts in long-term judgments can occur in multiple dimensions: between the judge and the advisor (i.e., in terms of distance, discrepancy, desires) and within the group of advisors (i.e., dissent). They all have in common that their occurrence substantially lowers judges' willingness to take advice. Our study thus sheds light on what leads experts to take advice in the process of developing long-term judgments for practically relevant settings. In that way, our findings foster decision makers' understanding of their own advice-taking behavior as well as the behavior of others. On the one hand, knowing about one's own determinants to update prior beliefs can lead to better-informed strategic decisions. On the other hand, understanding what drives others' long-term judgments can be an important basis for managers and policy makers to evaluate the judgments of others, i.e., of their peers, team members, and competitors.

Acknowledgments
The authors are most grateful to Jack Soll, Ansgar Richter, Alexander Rothkopf, the participants of the OB-Lab at Harvard Business School, and three anonymous referees for their invaluable advice, comments, and support. The authors thank Heiko von der Gracht and his team at EBS Business School as well as Jack Soll and Richard Larrick for making their data available to the authors.

Appendix. List of Future Studies and Events

Table A.1 List of Events in Future Study of Automotive Industry

No.	Future event
1	2030: In many cities, only emission-free vehicles are allowed to be driven.
2	2030: Customers mainly use an optimally coordinated network of multimodal mobility services (bus, car sharing, train, airplane, etc.).
3	2030: The acceptance of new combustion engines varies significantly in different regions.
4	2030: Customers demonstrate a high willingness to pay for vehicles with new power train technologies.
5	2030: Government interventions have primarily shaped the provision and use of different power train technologies.
6	2030: The risks (political, economic, ecological) associated with oil- and gas-based fuels have increased the pressure to implement new power train technologies.
7	2030: Energy for new power train technologies is derived mainly from renewable sources.
8	2030: The use of electric power train technologies has led to a significant reduction of original equipment manufacturers' value added activities in vehicle manufacturing.
9	2030: The structure of the after-sales market has changed significantly with the implementation of new power train technologies.
10	2030: Automobile manufacturers have each specialized in one power train technology.
11	2030: New entrants significantly shape the automotive industry.
12	2030: Regardless of the progress in power train concepts, the significance of owning a car has lost its meaning.
13	2030: New power train technologies have succeeded in the market for commercial vehicles.
14	2030: The proliferation of electric drive vehicles has been limited by bottlenecks in the supply of raw materials.
15	2030: The widespread availability of effective battery charging stations has contributed to the proliferation of electric vehicles.
16	2030: The global market for electric vehicles is dominated by Asian producers.
17	2030: The recycling problems of electric vehicles (e.g., batteries) have been solved.
18	2030: Power train technologies that exclusively run on electricity (without hybrid or range extender technologies) have established themselves as the sole model for the future.
19	2030: Cars powered by internal combustion engines (gasoline and diesel engines, including efficiency technologies and alternative fuels) still dominate the number of new registrations.
20	2030: A newly discovered power train concept will shape the future of mobility.

Notes. Events translated from German. Experts were asked to evaluate the probability of occurrence in year 2030 of each event on a scale from 0 to 100.

Table A.2 Description of Future Studies

No.	Industry focus and main objective	Regional focus	Time horizon	No. of experts	No. of events
1	Aerospace (in India)	India	2019	33	18
2	Automotive (alternative power train technologies)	Germany	2030	138	20
3	Aviation (business passengers)	Global	2025	13	30

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 20:38 . For personal use only, all rights reserved.

Table A.2 (Continued)

No.	Industry focus and main objective	Regional focus	Time horizon	No. of experts	No. of events
4	Aviation (cargo)	Global	2025	18	23
5	Aviation (passengers)	Global	2025	23	25
6	Banking (general)	India	2020	28	20
7	Consulting (general)	India	2020	25	20
8	Consumer goods (general)	India	2020	17	20
9	Consumer goods (general)	Global	2030	80	16
10	Defense (general)	India	2020	28	20
11	Insurance (general)	India	2020	31	20
12	Logisitcs (services)	Global	2030	66	20
13	Logistics (emerging markets)	Global	2030	87	16
14	Logistics (general)	Global	2025	60	20
15	Logistics (sustainability)	Global	2030	48	18
16	Logistics (transport infrastructure)	Global	2030	104	16
17	Mergers and acquisitions (general)	India	2020	17	20
18	Pharma (general)	India	2020	43	20
19	Private equity (general)	India	2020	16	20
20	Social business (general)	Germany	2030	69	16
21	Steel (general)	Global	2020	22	13
				966	411

References

Ai C, Norton EC (2003) Interaction terms in logit and probit models. *Econom. Lett.* 80(1):123–129.

Alfaro J, Delrio C, Dima G, Dondi C, Fischer T, Zuheros A (2005) Understanding change, adapting to change, shaping the future. MENON Network EEIG, Brussels, Belgium.

Armor DA, Massey C, Sackett AM (2008) Prescribed optimism: Is it right to be wrong about the future? *Psych. Sci.* 19(4):329–331.

Aronson E, Turner JA, Carlsmith JM (1963) Communicator credibility and communication discrepancy as determinants of opinion change. *J. Abnormal Soc. Psych.* 67(1):31–36.

Bochner S, Insko CA (1966) Communicator discrepancy, source credibility, and opinion change. *J. Personality Soc. Psych.* 4(6): 614–621.

Bonaccio S, Dalal RS (2006) Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organ. Behav. Human Decision Processes* 101(2):127–151.

Bonaccio S, Dalal RS (2010) Evaluating advisors: A policy-capturing study under conditions of complete and missing information. *J. Behav. Decision Making* 23(3):227–249.

Boston Consulting Group (2010) Batteries for electric cars: Challenges, opportunities, and the outlook to 2020 (Boston Consulting Group, Boston).

Budescu DV, Rantilla AK (2000) Confidence in aggregation of expert opinions. *Acta Psychologica* 104(3):371–398.

Cooksey RW (1996) *Judgment Analysis: Theory, Methods, and Applications* (Academic Press, San Diego).

Dalkey N, Helmer O (1963) An experimental application of the delphi method to the use of experts. *Management Sci.* 9(3):458–467.

Duboff RS (2007) The wisdom of (expert) crowds. *Harvard Bus. Rev.* 85(9):28–28.

Eisenhardt K (1989) Agency theory: An assessment and review. *Acad. Management Rev.* 14(1):57–74.

Enders C, Tofighi D (2007) Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psych. Methods* 12(2):121–138.

Festinger L (1954) A theory of social comparison processes. *Human Relations* 7(2):117–140.

Fielding A, Yang M (2005) Generalized linear mixed models for ordered responses in complex multilevel structures: Effects beneath the school or college in education. *J. Roy. Statist. Soc.: Ser. A* 168(1):159–183.

Fieller EC (1954) Some problems in interval estimation. *J. Roy. Statist. Soc.: Ser. B* 16(2):175–185.

Gino F (2008) Do we listen to advice just because we paid for it? The impact of advice cost on its use. *Organ. Behav. Human Decision Processes* 107(2):234–245.

Gino F, Moore DA (2007) Effects of task difficulty on use of advice. *J. Behav. Decision Making* 20(1):21–35.

Gino F, Schweitzer ME (2008) Blinded by anger or feeling the love: How emotions influence advice taking. *J. Appl. Psych.* 93(5):1165–1173.

Gino F, Shang J, Croson R (2009) The impact of information from similar or different advisors on judgment. *Organ. Behav. Human Decision Processes* 108(2):287–302.

Gordon T, Pease A (2006) RT delphi: An efficient, "round-less" almost real time delphi method. *Tech. Forecasting Soc. Change* 73(4):321–333.

Greene W (2010) Testing hypotheses about interaction terms in nonlinear models. *Econom. Lett.* 107(2):291–296.

Harvey N, Fischer I (1997) Taking advice: Accepting help, improving judgment, and sharing responsibility. *Organ. Behav. Human Decision Processes* 70(2):117–133.

Hausman JA (1978) Specification tests in econometrics. *Econometrica* 46(6):1251–1271.

Heckman JJ (1979) Sample selection bias as a specification error. *Econometrica* 47(1):153–161.

Hofmann DA, Gavin MB (1998) Centering decisions in hierarchical linear models: Implications for research in organizations. *J. Management* 24(5):623–641.

Hogarth RM (1978) A note on aggregating opinions. *Organ. Behav. Human Performance* 21(1):40–46.

Insko CA, Murashima F, Saiyadain M (1966) Communicator discrepancy, stimulus ambiguity, and influence. *J. Personality* 34(2):262–274.

Kahneman D, Klein G (2009) Conditions for intuitive expertise: A failure to disagree. *Amer. Psychologist* 64(6):515–526.

Karim S (2009) Business unit reorganization and innovation in new product markets. *Management Sci.* 55(7):1237–1254.

Kennedy KA, Pronin E (2008) When disagreement gets ugly: Perceptions of bias and the escalation of conflict. *Personality Soc. Psych. Bull.* 34(6):833–848.

Krizan Z, Windschitl PD (2007) The influence of outcome desirability on optimism. *Psych. Bull.* 133(1):95–121.

Krueger JI (2003) Return of the ego-self-referent information as a filter for social prediction: Comment on Karniol (2003). *Psych. Rev.* 110(3):585–590.

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 20:38 . For personal use only, all rights reserved.

Lamberson PJ, Page SE (2012) Optimal forecasting groups. *Management Sci.* 58(4):805–810.
Larrick RP, Soll JB (2006) Intuitions about combining opinions: Misappreciation of the averaging principle. *Management Sci.* 52(1):111–127.
Liberman V, Minson JA, Bryan CJ, Ross L (2012) Naïve realism and capturing the "wisdom of dyads." *J. Experiment. Soc. Psych.* 48(2):507–512.
Lind JT, Mehlum H (2010) With or without U? The appropriate test for a U-shaped relationship. *Oxford Bull. Econom. Statist.* 72(1):109–118.
Lorenz JR, Rauhut H, Schweitzer F, Helbing D (2011) How social influence can undermine the wisdom of crowd effect. *Proc. Natl. Acad. Sci. USA* 108(22):9020–9025.
Mannes AE (2009) Are we wise about the wisdom of crowds? The use of group judgments in belief revision. *Management Sci.* 55(8):1267–1279.
McDonald ML, Westphal JD (2003) Getting by with the advice of their friends: CEOs' advice networks and firms' strategic responses to poor performance. *Admin. Sci. Quart.* 48(1):1–32.
McDonald ML, Khanna P, Westphal JD (2008) Getting them to think outside the circle: Corporate governance, CEOs' external advice networks, and firm performance. *Acad. Management J.* 51(3):453–475.
McKelvey RD, Zavoina W (1975) A statistical model for the analysis of ordinal level dependent variables. *J. Math. Sociol.* 4(1): 103–120.
Meloy MG, Russo JE, Miller EG (2006) Monetary incentives and mood. *J. Marketing Res.* 43(2):267–275.
Minson JA, Liberman V, Ross L (2011) Two to tango: Effects of collaboration and disagreement on dyadic judgment. *Personality Soc. Psych. Bull.* 37(10):1325–1338.
Pronin E (2004) Objectivity in the eye of the beholder divergent perceptions of bias in self versus others. *Psych. Rev.* 111(3): 781–799.
Pronin E (2007) Perception and misperception of bias in human judgment. *Trends in Cognitive Sci.* 11(1):37–43.
Pronin E (2008) How we see ourselves and how we see others. *Science* 320(5880):1177–1180.
Rabe-Hesketh S, Skrondal A (2012) *Multilevel and Longitudinal Modeling Using Stata* (Stata Press, College Station, TX).
Raudenbush SW, Bryk AS (2002) *Hierarchical Linear Models: Applications and Data Analysis Methods* (Sage Publications, Thousand Oaks, CA).
Raudenbush SW, Bryk AS, Cheong YF, Congdon RT, du Toilt M (2011) *HLM 7: Hierarchical Linear and Nonlinear Modeling* (Scientific Software International, Lincolnwood, IL).
Ravazzolo F, Røisland Ø (2011) Why do people place lower weight on advice far from their own initial opinion? *Econom. Lett.* 112(1):63–66.
Reed J (2012) Switch to hybrids and electric likely to be slow. *Financial Times Online* (March 5), http://www.ft.com/cms/s/0/57ecdfac-6203-11e1-807f-00144feabdc0.html.
Rodríguez G, Goldman N (2001) Improved estimation procedures for multilevel models with binary response: A case-study. *J. Roy. Statist. Soc.: Ser. A* 164(2):339–355.
Rowe G, Wright G (2001) Expert opinions in forecasting: The role of the delphi technique. Armstrong JS, ed. *Principles of Forecasting: A Handbook for Researchers and Practitioners* (Kluwer Academic Publishers, Boston), 125–144.
Russo JE, Yong K (2011) The distortion of information to support an emerging evaluation of risk. *J. Econometrics* 162(1):132–139.
See KE, Morrison EW, Rothman NB, Soll JB (2011) The detrimental effects of power on confidence, advice taking, and accuracy. *Organ. Behav. Human Decision Processes* 116(2):272–285.
Sharot T, Korn CW, Dolan RJ (2011) How unrealistic optimism is maintained in the face of reality. *Nat. Neurosci.* 14(11): 1475–1479.
Skrodal A, Rabe-Hesketh S (2007) Redundant overdispersion parameters in multilevel models for categorical responses. *J. Educational Behav. Statist.* 32(4):419–430.
Sniezek JA, Van Swol LM (2001) Trust, confidence, and expertise in a judge-advisor system. *Organ. Behav. Human Decision Processes* 84(2):288–307.
Snijders TAB, Berkhof J (2008) Diagnostic checks for multilevel models. de Leeuw J, Meijer E, eds. *Handbook of Multilevel Analysis* (Springer, New York), 141–175.
Snijders TAB, Bosker R (1999) *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling* (Sage Publications, London).
Soll JB, Larrick RP (2009) Strategies for revising judgment: How (and how well) people use others' opinions. *J. Experiment. Psych.: Learn., Memory, Cogn.* 35(3):780–805.
Soll JB, Mannes AE (2011) Judgmental aggregation strategies depend on whether the self is involved. *Internat. J. Forecasting* 27(1):81–102.
Suls J, Martin R, Wheeler L (2002) Social comparison: Why, with whom, and with what effect? *Current Directions Psych. Sci.* 11(5):159–163.
Tetlock PE (2005) *Expert Political Judgment: How Good Is It? How Can We Know?* (Princeton University Press, Princeton, NJ).
Tost LP, Gino F, Larrick RP (2012) Power, competitiveness, and advice taking: Why the powerful don't listen. *Organ. Behav. Human Decision Processes* 117(1):53–65.
Van Swol LM (2011) Forecasting another's enjoyment versus giving the right answer: Trust, shared values, task effects, and confidence in improving the acceptance of advice. *Internat. J. Forecasting* 27(1):103–120.
Vissa B (2011) A matching theory of entrepreneurs' tie formation intentions and initiation of economic exchange. *Acad. Management J.* 54(1):137–158.
von der Gracht HA, Gnatzy T, Darkow IL, Gordon TJ, Glenn J (2011) New frontiers in delphi research: Experiences with real-time delphi in foresight. *Conf. Volume of the World Future 2011, Vancouver*, July 2011.
Warth J, Zimmermann M, von der Gracht H, Darkow IL, Esch FR, Langenhan F, Georgi C (2011) AIM Automotive Zukunftsbarometer Antriebskonzepte der Zukunft. EBS Business School, Wiesbaden, Germany.
Yaniv I (2004) Receiving other people´s advice: Influence and benefit. *Organ. Behav. Human Decision Processes* 93(1):1–13.
Yaniv I (2011) Group diversity and decision quality: Amplification and attenuation of the framing effect. *Internat. J. Forecasting* 27(1):41–49.
Yaniv I, Kleinberger E (2000) Advice taking in decision making: Egocentric discounting and reputation formation. *Organ. Behav. Human Decision Processes* 83(2):260–281.
Yaniv I, Milyavsky M (2007) Using advice from multiple sources to revise and improve judgments. *Organ. Behav. Human Decision Processes* 103(1):104–120.
Yaniv I, Choshen-Hillel S, Milyavsky M (2009) Spurious consensus and opinion revision: Why might people be more confident in their less accurate judgments? *J. Experiment. Psych.: Learn., Memory, Cogn.* 35(2):558–563.
Yaniv I, Choshen-Hillel S, Milyavsky M (2011) Receiving advice on matters of taste: Similarity, majority influence, and taste discrimination. *Organ. Behav. Human Decision Processes* 115(1): 111–120.
Zarnoth P, Sniezek JA (1997) The social influence of confidence in group decision making. *J. Experiment. Soc. Psych.* 33(4):345–366.

RIGHTSLINK