Evaluating Qualitative Predictions:

A Novel Method Applied to the Automotive Powertrain Transition (1970–2009)

ABSTRACT

In times of high uncertainty and low data, expert judgment is the preferred foresight approach to guide strategic decision-making. We have developed a method for the retrospective assessment of qualitative predictions made by experts. This method is applied to the automotive powertrain transition. With our novel method, we analyzed the historical expert judgment from 1970 to 2009 to explore how accurate and consensual the experts have predicted the relative potential of alternative powertrain across this period. We collected 89 qualitative expert predictions and transformed them into quantitative probabilistic forecasts to measure their accuracy and the wisdom of the expert "crowd". To compare the accuracy of expert judgment with those of top managers, we coded 85 press releases of a leading car company related to alternative powertrains. The novel method shows that experts predicted significantly better than random guessing but the car company did not. Interestingly, we can show that a consensus forecast already emerged in the mid-1990s that battery electric powertrains would dominate fuel-cell powertrains. The leading car company did not see this for at least another 13 years.

Keywords: *Foresight, crowd wisdom, expert judgment, case study, automotive*

INTRODUCTION

In the face of ever-growing uncertainty in the business environment, foresight research has gradually gained importance (Marinkovic et al., 2022; Schoemaker et al., 2018). This is because foresight allows companies to reduce uncertainty about the future and gain competitive advantages (Fergnani, 2022). Supporting this claim, Rohrbeck and Kum (2018) showed that more extensive foresight capabilities of a firm lead to superior firm performance (corporate earnings and market capitalization) five years in the future. Foresight can be described as *"[…] identifying, observing, and interpreting factors that induce change […]*" (Gordon et al., 2020, p. 1) and by that, anticipating the future state of the world.

Multiple approaches to foresight exist such as scenario planning, quantitative modelling, or technology roadmapping. In cases of high uncertainty and low data, expert judgment of a group of experts frequently dubbed as "the wisdom of the crowd effect" is the most associated foresight approach. The effect describes that a crowd of individuals is on average more accurate than a single individual. This is due to the number of signals being processed and the elimination of systematic errors in judgment, called biases, at the group level (Petropoulos et al., 2022; Tetlock, 2005).

In the current age of digital media, many qualitative statements of scientific and practice experts are documented and accessible. Therefore, the retrieval of expert opinions that are publicly available for the reduction of uncertainty is intriguing.

A key theme in forecasting literature is the accuracy measurement (Petropoulos et al., 2022; Tetlock, 2005). The accuracy measurement verifies individual foresight as well as the crowd wisdom effect. For example, forecasting tournaments allow researchers to measure how accurately individuals have foreseen the future states of the world by applying variants of the Brier score in a structured environment (Chang et al., 2016; Kapoor & Wilde, 2023; Schuler et al., Forthcoming). To be able to draw on the vast documented predictions of experts, one needs a

score to pass judgment on the accuracy of qualitative predictions. For example, Stekler and Symington (2016) created a tailored scoring model for their study of qualitative economic predictions which unfortunately does not apply to other cases. A generic approach would help researchers assess qualitative predictions of individuals across research endeavors and retrieve the crowd wisdom to form a perspective on the future. Arb Research, a consulting firm specializing in forecasting, took a more general approach to qualitative judgment, using technology predictions of science fiction authors (Leech, 2023). However, the firm acknowledges that its method does not account for the vagueness of qualitative statements. Overall, there is no generalized measure for the accuracy and sentiment of qualitative predictions in science. For these reasons, we develop and test a general method to systematically assess qualitative predictions. With this method, researchers can measure the presence, accuracy, and limits of expert judgments and crowd wisdom across cases without setting up expensive other scientific interventions such as forecasting tournaments.

The method is tested in the context of the alternative powertrain transition in the automotive industry that is presently in full swing as countries are trying to phase out international combustion engine cars to reduce CO2 emissions. In 1970s, when car companies (OEMs) began to research alternative powertrains again, uncertainty about which powertrain technology would be the best alternative to the internal combustion engine (ICE) was omnipresent. Battery electric vehicles (BEVs) and fuel cell electric vehicles (FCEVs) were the two prominent non-combustion alternatives. While both drive with electric energy, the first relies on an externally charged battery, the latter converts hydrogen onboard. Due to the technological differences, the OEMs were forced to trade off their investment opportunities – BEV vs. FCEV. Tesla launched its successful Roadster model in 2008, while Mercedes-Benz's FCEV flopped in 2018. The underlying uncertainty about the better alternative imposed barriers to the strategic decisions

of automotive managers. As mentioned before, expert group judgment might have been valuable in this uncertain situation. We therefore use the developed method to assess how accurately experts as a group predicted the dominant technology of the future, what the consensus was, and whether companies would have benefited from the wisdom of the crowd. This is reflected in three research questions: 1) "*How accurate have experts been in foreseeing the future of alternative power trains?*", 2) "*What was the consensus of the experts?*" and 3) "*How did OEMs act?*". In Section 2, the previous efforts and our developed accuracy measure are detailed. In Section 3, the data collection is explained. In Section 4, the results for the expert crowd are laid out and, in Section 5, the results for Mercedes-Benz as representative for OEMs. Finally, the study concludes with the key findings, limitations, and future research trajectories in Section 6.

A NEW MEASURE FOR THE ACCURACAY OF QUALITATIVE PREDICTIONS

Previous efforts

As mentioned, scholars have made efforts to evaluate qualitative predictions. Goldfarb et al. (2005) developed a method for qualitative economic forecasts. It was further adjusted and re-applied by Lundquist and Stekler (2012) and Stekler and Symington (2016). Their scoring is based on the mindset of economic actors regarding the state of the economy. They paraphrased qualitative predictions of actors such as the Wall Street Journal and assigned scores from 1 to -1 with steps of 1/4 (e.g. 1 = "*the economy is strong […]*", -1 = "*Recession conditions are here and imminent […]*"). Hence, the scoring does not apply to other qualitative predictions as it is designed for a special case assessing the state of the economy. Additionally, the score captures sentiment, not accuracy. The paraphrased qualitative predictions are transformed purely based on sentiment into quantitative forecasts: e.g. "*[…] if the individual […] predicted that the economy would recover gradually over the next several months, that forecast would be scored as*

[+0.5]." (Goldfarb et al., 2005, p. 519). Further, they did not account for vague verbiage, using words instead of numbers to indicate one's belief in the stated future. For example, one prediction is: "*we conclude that a business upturn ... probably will be in progress by February or March 1931*" (Goldfarb et al., 2005, p. 522). The forecaster signals with "*probably*" that he or she believes less in the statement to come true than if they left the word "*probably*" out and thereby indicated absolute certainty. Hence, the forecaster would be less inaccurate in the prediction if it does not materialize. But the sentiment-based score does not address this issue. When the authors analyze the accuracy of the predictions, they infer it but do not calculate the accuracy of forecasts. For instance, Lundquist and Stekler (2012) compare their forecasts to an actual but do not apply any systematic forecasting error measure. Essentially, the different publications created a solid method to quantify qualitative economic outlooks but did not develop a method to score them or other qualitative predictions for accuracy.

Arb Research developed a thorough method that goes beyond economic forecasting, dealing particularly with technological developments. The objective was to measure the accuracy of futurologists such as Issac Asimov (Leech, 2023), focusing on their technological predictions and hence giving technological forecasts more weight than economic, techno-social, or cultural ones. In contrast to the described effort for economic outlooks, Arb directly integrated the correctness of the qualitative forecast in their final score. Next to correctness, the score is computed by the variables *category weight, difficulty*, and the difference between prediction and resolution year, called *years out*. A simple formula then calculates the score:

$$Final\ score = correctness * (\frac{difficulty}{10}) * category\ weight * \log(years\ out)$$

Correctness is scored with either 0 = "*unambiguously wrong",* 1 = "*ambiguous or near miss*", 2 = "*unambiguously right*". *Category weight* captures that the tech-oriented score weighs pure tech predictions (weight: *1.00*), e.g. "*a technology will be available",* higher than techno-economic (weight: *0.66*), e.g. "*the cost of new technology will diminish"*, or techno-social

predictions (weight: *0.33*), e.g. "*people will accept a technology",* and dismisses other cultural predictions (weight: *0.00*). Also, they define *difficulty* on a scale from "*1 - was already generally known*" to "*5 - prescient, no trend to go off*".

The main message is that the higher the score, the better the prediction. For example, we illustrate the method using a prediction from Isaac Asimov out of Arb's database: "*Robots for gardening work will also have made their appearance."* The prediction was correct, pure tech, received a difficulty of 4 ("*above trend*") from the author. It was made in 1964 and the resolution year is 2014. Hence, the *years out* are 50. Therefore, the final score is as follows:

$$Final\ score = 2 * \left(\frac{4}{10}\right) * 1 * \log(50) = 1.359$$

This score is easily transferable to other qualitative predictions by adjusting or eliminating the category weighting, depending on the case. However, it exhibits two fundamental flaws. First, the variables *difficulty* and *years out* strongly correlate. Events that are already known tend to be short-term. Events without a trend at the time of prediction are either spontaneous or the trend is to be emerging. Therefore, the scale of difficulty is influenced by the resolution horizon. Yet, this is not mentioned as an intention of the authors. Second, and more important, Arb's method does not account for vague verbiage which is very common in human communication as mentioned before. The firm states in its limitation section: "*The predictions are usually very vague*." (Leech, 2023, p. 10). Looking at the paper's data, we also found that phrases like "*might be*", "*it is to be hoped*" are used in the collected predictions. As the objective of Arb's publication is the comparison of individuals' accuracy, it would have been required to assess whether the forecasters fully believed in their prediction or if they watered them down to be less accountable in case of misjudgment.

Following the limitations of prior efforts, we conclude that a more general method is required for measuring qualitative predictions. This method needs to be agnostic about the examined

domain and importantly be able to deal with vague verbiage to account for verbal/textual communication that existing methods have not tackled.

Addressing vague verbiage

As seen, qualitative forecasters hold different beliefs in their forecasted outcome which is reflected in their wording. Evaluating qualitative forecasts needs to account for this. Let us assume two persons predicting the same USA presidential election and both correctly predicting a Republican win as happened a few months ago. Person A states: "*The Republicans will likely win the next election*", while person B states: "*The Republicans will definitely win the next election*". If we measure accuracy, and thereby foresight, we are confronted with the same outcome but different levels of certainty. The example comparatively illustrates how foresight differs based on wording. A measure of accuracy needs to address this because person B provided a better forecast by implicitly assigning a higher likelihood.

The phenomenon of differed levels of linguistic certainty in communication is referred to as "*vague verbiage*" (Irwin & Mandel, 2020). Directly comparing the accuracy of statements using different words is difficult. For example, let us add a third person to the forecast about the Republican election victory and person C states: "*The Republicans probably win the election*". Vague verbiage complicates an assessment of the relative certainty of the three forecasters. While "*probably*" ranges between 55% and 80%, "*likely*" is ranging between 40% and 90% (Irwin & Mandel, 2020). Since people have different perceptions of the likelihood of these words, this extends to an inconsistent relative assessment (Friedman et al., 2018; Irwin & Mandel, 2020).

But there is a way out of these difficulties. Certainty stands for "*an individual's subjective sense of confidence or conviction*" (Rocklage et al., 2023, p. 871). It is translatable to the

context of probability forecasts in individual foresight because the assigned probability from individuals symbolizes their subjective belief of the future state of the world. Individuals assign probabilities from 0% to 100%, and the higher the probability, the more likely the forecaster considers an event to be (Kapoor & Wilde, 2023; Schuler et al., Forthcoming). Probability can therefore be equated with certainty. Our main contribution is the transformation of vague verbiage into probabilities and applying existing measures for probabilistic forecasts.

A scoring method for qualitative predictions

Recently, the time-weighted Brier score (TWBS) was developed to improve scoring quantitative probabilistic forecasts made at different time points (Schuler et al., Forthcoming). We build on this new score to transform the qualitative predictions into quantitative probabilistic forecasts and measure them according to the TWBS.

The basis of the score lies within the conceptualization of individual foresight as the ability to predict the future state of the world accurately and early. "*Accurately*" and "*early*" are key concepts of the TWBS. The score accounts for the (in)accuracy of the prediction and the time horizon involved in a single prediction made by an individual (Schuler et al., Forthcoming). The TWBS addresses that the signals indicating future events change over time. One assumption is that the closer a prediction is made to its resolution year, the easier it is. With more and stronger signals emerging toward the resolution, foresight is eased because uncertainty naturally diminishes. Therefore, errors weigh heavier. This is accounted for by integrating the time weights in the accuracy measure (Schuler et al., Forthcoming). The time weights address the problem of the correlation of difficulty and years out to resolution in Arb Research's method. Both dimensions are conceptually represented by one single variable, the time weights. Schuler et al. (Forthcoming) mathematically define the TBWS as follows:

$$TWBS_{iq} = \frac{\sum_{t=1}^{T_q} w_t \sum_{c}^{C} \left(f_{iqtc} - o_{qc}\right)^2}{\sum_{t=1}^{T_q} w_t}$$

where "$TWBS_{iq}$ is the TWBS of forecaster i on question q; T_q is the number of time-units (usually days) $t = 1, \ldots, T_q$ question q was forecastable; and $w_t > 0$ is any weight at time t that is strictly monotonically increasing over time $t = 1, \ldots, T_q$." (Schuler et al., Forthcoming, p. 13) and "C is the number of possible disjunct classes $c = 1, \ldots, C$ the event can fall into; f_{iqtc} is the probability forecast of forecaster i on question q at time t for class c; and o_{qc} indicates whether the class c is the true state of the world at the time of the resolution of question q ($o_{qc} = 1$ if yes, $o_{qc} = 0$ if not)." (Schuler et al., Forthcoming, p. 12). The TWBS measures the error in forecasting. Hence, the better the ability to foresee, the lower the score in the interval [0;2] (Schuler et al., Forthcoming).

However, the TWBS was developed for the closed environment of forecasting tournaments. Forecasting tournaments take place in environments where the questions and answer space are defined by the researcher. Since our method analyzes qualitative statements, the predictions will be coded to allow for the analysis and calculation of the TWBS. This requires deciding on handling vague verbiage and (unprecise) time horizons. While the first hinders the measurement of the forecasted probabilities (f), the second impacts the calculation of time weights (w). The resolution of the event itself (c, o) will be taken from the prediction itself and compared to the state of the world as is the case of quantitative questions in forecasting tournaments. The explanations and remedies of the issues follow in the next paragraphs.

We integrate vague verbiage in a measure of qualitative forecasts' accuracy by deriving the forecasted probability (f) from the words of certainty. The Lexical Suite of Rocklage et al. (2023) can process phrases and assign an average certainty score of 0 to 9 to them. The resulting certainty is transformed into probabilities by translating the certainty scale into a probability of 0 to 100 [%]. In that way, the forecasted probabilities for the respective event are obtained.

Because the Lexical Suite relies on a defined translation dictionary in the background, predictions without identifying words in the Lexical Suite are not converted into a certainty score. In such cases, the certainty is coded manually on the same interval based on the "*DIA Expressing Analytic Certainty*" framework (Irwin & Mandel, 2020). Mathematically, the probability (*f*) is derived in the following way:

$$f\ [\%] = \frac{(Average\ certainty\ score - 0\ )}{(9 - 0)}$$

For instance, the prediction "*EVs probably will be introduced with relatively little government involvement*" (Sperling & Deluchi, 1989, p. 414) receives a certainty score of 4.785. This is low due to the words "*probably*" and "*relatively*". The resulting probability (f) is 53.2%.

The disjunct classes (c = 1, …, C) are the future states of the world that are restricted to the predicted event and its mutual exclusive counter event. For example, the statement published in SAE Transactions in 1981 "*This study concludes that hydrogen will not be used as an automotive fuel during this century. It is unlikely to be given serious consideration […]*" (Kukkonen, 1981, p. 1433) with a derived probability of 71.4% has two disjunct classes:

- c_1 = H*ydrogen is not used automotive fuel by 2000*
- c_2 = *Hydrogen is used automotive fuel by 2000*

Based on the analytical reasoning in the article and the public information on market launches and fuel infrastructure, the conclusion is that c_1 is correct ($o_1 = 1$) and c_2 is incorrect ($o_2 = 0$). According to the formula of the TWBS, c_1 is then assigned $f_{iqtc,}$ and c_2 is assigned 1- f_{iqtc}. Then, the accuracy dimension of the TWBS is calculated as follows:

$$\sum_{c}^{C}\left(f_{iqtc} - o_{qc}\right)^2 = (0.714 - 1)^2 + (0.286 - 0)^2 = 0.164$$

Some readers might recognize this formula as the Brier Score (Brier, 1950). The TWBS extends its idea by incorporating the time dimension. Linear time weights were found to be the

best solution. Linearity means that for each day the question was forecastable, or the resolution has come closer, forecasting errors gradually weigh heavier (Schuler et al., Forthcoming). However, the translation of the concept to qualitative predictions that do not necessarily include a resolution time is not trivial. Our practical solution is measuring how much of the predictable time horizon is left out. From this, we can derive linear time weights according to the TWBS. The time weights are positive, linear, and strictly monotonously increase the closer a prediction is made toward its resolution. To account for this, we calculate the time weights *w* of prediction based on the relation of "*how much of the of the predictable time horizon is left out*" to "*how much would have been possible*":

$$w = \frac{how\ much\ of\ the\ of\ the\ predictable\ time\ horizon\ is\ left\ out}{how\ much\ would\ have\ been\ possible}$$

Mathematically, the weights will be calculated as follows:

$$w_t = \frac{t_i - T_i}{1 - T_i}$$

where w_t is the weight of a forecast given the predicted resolution time t_i and the predictable time horizon T_i of a forecaster i for a future state of the world. The predictable time horizon refers to the interval of the publication year to the year of the analysis plus one. This is mathematically required to calculate a weight for predictions whose resolution can be observed in this year. For example, if a prediction was made in 1980, the predictable time horizon in our analysis (in 2024) is 45 years. Then, we use the forecast horizon and subtract the predictable time horizon from it to obtain the numerator, "*how much of the of the predictable time horizon is left*". In our analysis in 2024, if a prediction from 2000 had its resolution in 2004, the predictable time horizon would have been 25 years and the forecast horizon is 4 years. Hence, the numerator is: "$4 - 25$". Then, we look at "*how much would have been possible*". Here, we subtract the predictable time horizon from one year. A prediction cannot be shorter than one

year in this analysis. Taking the last example, we then have "$1 - 25$". Finally, we then calculate the time weight *w* with our formula:

$$w = \frac{how\ much\ of\ the\ of\ the\ predictable\ time\ horizon\ is\ left\ out}{how\ much\ would\ have\ been\ possible} = \frac{4 - 25}{1 - 25} = 0.88$$

It is necessary to subtract the year of analysis plus one from the prediction horizon as well as one year to obtain always positive weights that are linear and strictly increase for shorter prediction horizons. For instance, if two predictions were made in the same year but one prediction covers a longer time horizon (5 years) than the other (10 years), the first receives a heavier time weight. Thereby, forecasting errors weigh linearly heavier the shorter to the resolution because it is easier to get such a forecast correct. This approach is flexible regarding the underlying time unit and can be altered for other analysis, e.g. to months.

For the time weights *w*, the problem of vague language is as prevalent as it is for the probability. Often, individuals do not state a specific resolution date, e.g. "*by the year 2010*" (explicit reference). Rather, they rely on time expressions that are only understood if one knows the date and time of the statement, e.g. "*end of the decade*" (indexical reference), or they state vague expressions, e.g. "*in the foreseeable future*" (vague reference), which are interpreted individually (Schilder & Habel, 2001; Tissot et al., 2019). While the first two reference types are directly codable from the prediction's metadata, the last type is not. For a manual approach, no applicable summary to clear the uncertainty surrounding vague temporal expressions was found. Therefore, an individual coding scheme is set up (see Table 1). The meaning is conditioned on two aspects: First, the goal is to be strict on vague expressions compared to concreteness. The rationale is that concreteness is more valuable for decision-making since "*in 10 years*" provides more informational value than "*in the near future*". Strictness is achieved by assigning low but reasonable forecasted years because then, errors weigh heavier. Secondly, the meanings have been conditioned on each other. Short-term, near-term, and long-term are interrelated. Additionally, the same adverbs are used in temporal expressions, e.g. "*long term*"

and "*long time*", which postulate similar meanings. For predicted time intervals, e.g. "*in the next 20 to 30 years*", the lower bound was used to maintain strictness. If the time expression is missing, the value is set to "*1*" to punish the forecaster, e.g. "*XY will happen*". Finally, the denominator of the original TBWS is set to one to obtain the TBWS of a qualitative forecaster. The denominator serves to calculate the weighted average of a forecaster who might update his or her belief for the future state of the world and thus, creates multiple forecasts at different points in time. This will not be the case in this study because the forecasters will be evaluated on single predictions.

TABLE 1
Coding of vague temporal expressions

Vague temporal expression	**Coding**
Several /few *time unit*	2
Short term	5 years
near term	10 years
foreseeable/near future	10 years
Far / medium term	15 years
future	20 years
long time / long term / long way	20 years

The coded temporal expressions are used to obtain the resolution year of the prediction and the time weights. We illustrate it using the following prediction of Chan (2002, p. 266): "*At the present time and in the foreseeable future, batteries have been agreed to be the major energy source for EVs*". The term "*foreseeable future*" is coded as "*10 years*". Therefore, the resolution year is 2012, despite not explicitly stated. Thus, the time weights are calculated as follows:

$$w_t = \frac{t_i - T_i}{1 - T_i} = \frac{10 - (2025 - 2002)}{1 - (2025 - 2002)} = \frac{10 - 23}{1 - 23} = 0.59$$

In 2012, the vast majority of EVs on the market have been powered by batteries. Therefore, the prediction was correct. The final TWBS of this prediction with an assigned probability (*f*) of 73.1% is then:

$$TWBS_{iq} = \frac{\sum_{t=1}^{T_q} w_t \sum_{c}^{C} \left(f_{iqtc} - o_{qc}\right)^2}{1} = 0.59 * ((0.731 - 1)^2 + (0.269 - 0)^2) = \ 0.085$$

Since the TWBS is set on an interval from 0 to 2, the accuracy measure allows us to argue that the forecaster showed strong foresight in this example. This accuracy measure is agnostic to the underlying topic and predictions because the coding is independent of the content. Hence it applies to equally our historical case study of alternative automotive powertrain developments that we conduct in the next sections as well as to any other conceivable topic that includes qualitative predictions.

DATA COLLECTION AND CODING

Data collection procedure

The observation period is set from 1970 to 2009. The lower bound of the time interval was set to 1970 because the 1970s were a "*very active period in the history of electric car development*" (Høyer, 2008, p. 66). Geopolitical events drove the interest in alternative powertrains in multiple domains (Høyer, 2008). The upper bound of the interval, 2009, is suitable for finding sufficient predictions that are resolved from today's perspective, while the BEV market has not yet taken up in 2009 and uncertainty was present (IEA, 2024).

For expert predictions, we targeted researchers, consultants, journalists, and employees in the automotive industry and adjacent industries (e.g. politics, energy, and materials). The identity of an expert is most often tied to superior knowledge and judgment, compared to lay people. This superior knowledge stems from training and practical exposure to a domain (Bolger, 2018). The knowledge is the reason why experts are valuable in the face of low data availability and high uncertainty as mentioned before (Martin et al., 2012; Petropoulos et al., 2022; Tetlock,

2005). The four expert types are exposed to practical activity in the field and also may have insights into (hidden) developments that are crucial for the future of alternative powertrains. This can arise through (cross-organizational) research projects, government interaction, upper management dialogues, or undisclosed firm activities.

The main sources for journal articles are relevant academic databases such as Science Direct, JStor, and Web of Science. Additionally, we scanned the libraries of renowned institutions in the field of automotive technology, for instance, the Fraunhofer Institute in Germany or the Argonne National Laboratory in the USA. Hence, the database contains mainly researchers. Nonetheless, we also retrieve data from digitalized media sources (i.e. Web.Archive, Factiva, and ProQuest). In the first step, the key terms are "*electric vehicle**", "*fuel cell* AND automotive*", "*alternative fuel**", and "*alternative powertrain**" without a focused research field. This allows for the inclusion of cross-sector expertise (e.g. automotive, energy, politics). Increasing diversity in the crowd significantly enhances its wisdom (Larrick et al., 2012). If the abstract offers the chance that relevant predictions may be stated, the full text is accessed. The focus is on predictions from which the future of alternative powertrains can be inferred. The second step is to search the citations backward and forward. This allows for the inclusion of updates in the belief about the potential of technologies.

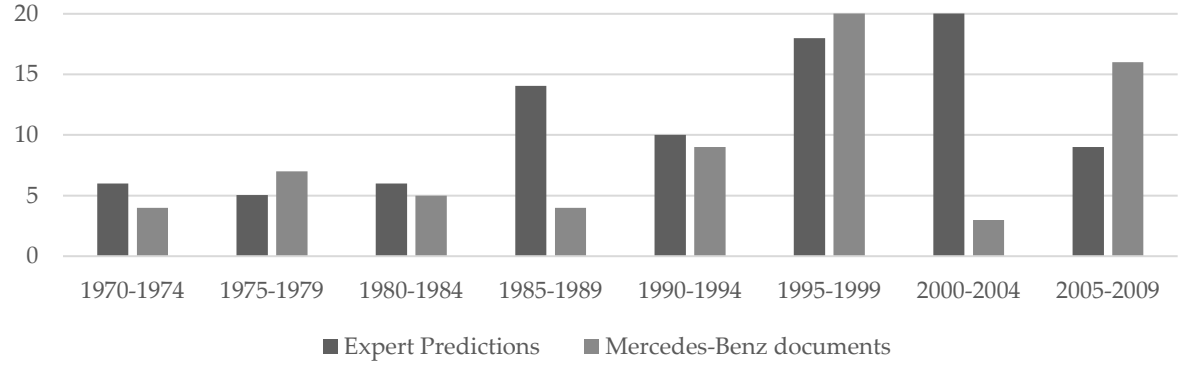
The activities of the OEMs in the period from 1970 to 2009 are based on research into the archives of Mercedes-Benz. Its corporate archive publicly offers news articles and official press releases regarding alternative powertrains. Mercedes-Benz is of special interest because it was seen as a technology leader in the past, but it is falling behind in the EV market (Kazeem, 2024). Its corporate archive is accessed online (https://mercedes-benz-publicarchive.com). The key search terms are the German words for "*electric vehicle**", "*fuel cell**", "*battery**", similar to the expert predictions. The documents are scanned for information about current activities, perceptions, and announcements regarding BEVs or FCEVs. This covers research, marketing,

partnering, production, and more. We address the intentionality of corporate communication in the discussion section.

Data overview

The data collection retrieved 89 expert predictions and 85 press releases of Mercedes Benz between 1970 and 2009 (see Figure 1). A longitudinal analysis shows that interest in alternative fuels has increased over the decades. Alternative powertrains were a central issue in the 1990s. The years are clustered in half decades for readability and comprehensibility. The amount of data points is lower from 2000 forward than in the half-decades before, which could be mistakenly interpreted as a reversed trend. However, the decrease has three reasons. First, per default, fewer predictions could be included in the analysis as the resolution time frame is shorter than in the years before. From 2005 to 2009, 19 to 15 years to predict were left. Predictions beyond the time frame were excluded based on the study design. The second reason is connected to the first one. The focus of researchers shifted to hybrid electric vehicles as the most promising short- and medium-term alternative as it became known that the technology was ripe and a full transition to other powertrain solutions would take more time (Morita, 2003). Therefore, predictions regarding BEV and FCEV are less prevalent.

FIGURE 1
Sampled predictions and documents (1970-2009)



For the expert predictions, the biggest sources are the SAE (Society of Automotive Engineers) Transactions (11) and independent or governmental reports (11). They are followed by the Journal of Power Sources (9) and the International Journal of Hydrogen Energy (6). Overall, the data collection covers automotive (e.g. SAE Transactions), energy (e.g. Journal of Power Sources), politics (e.g. Transport Policy), chemistry (e.g. Journal of Electroanalytical Chemistry and Interfaces), sustainability (e.g. Journal of Cleaner Production), other journals (e.g. Technological Forecasting and Social Change), newspapers (e.g. Scientific America), and books (e.g. "*Innovationsprozess vom Vebrennungsmotor zur Brennstoffzelle*" (engl.: Innovation process from combustion engine to fuel cell)). Alongside, the forecasters were associated with 53 different institutions at the time of publication, including universities (e.g. University of California, University of Hong Kong), research institutes (e.g. Fraunhofer Institute), government institutions (e.g. Joint Research Centre EU, International Energy Agency), companies (e.g. Mercedes-Benz, Nissan, Ford), consultancies (e.g. Delco Systems Operations), and newspaper (e.g. Financial Times). Most predictions are made by researchers (75%). Corporate employees account for 12 % and the rest is split between consultants, journalists, analysts, and press notes. In this context, press notes are references or citations of corporate communication in the publications. For example, Wengel and Schirrmeister (2000, p. 100) feature the statement of an infrastructure company regarding hydrogen refueling stations: "*The company expects that a relevant demand for hydrogen will only become noticeable on the fuel market in around 20 years, provided that the necessary framework conditions have been clarified by then*". Such predictions were crucial indicators for automotive managers to understand how the infrastructure players perceive the alternatives and infer strategic decisions. Therefore, they are included.

Mercedes-Benz published 85 documents that address alternative powertrains. It consistently released information on their technological progress, upcoming events, and speeches of

managers. Within the documents, 32 predictions were found. In contrast to the experts, Mercedes-Benz provides much information from 2005 to 2009. This may have two reasons. First, Mercedes-Benz increased its marketing activities as it came closer to the actual market launch of an EV. Second, many documents covered Mercedes-Benz's actual activities. Then, this sample is less constrained. Overshooting the year 2024 with forecasts is not a problem, as it is with the expert's predictions.

Other coding

Besides the accuracy, two attributes of the predictions are coded: Powertrain technology (*BEV*, *FCEV*) and sentiment towards technology (-1 = *negative*, 0 = *neutral*, 1 = *positive*).

Technologies coded as either "*BEV"* or "*FCEV"* for the technologies in the focus of the prediction, or as the subject of the phrase if both were mentioned. For instance, a prediction in the form of "*BEV will outperform FCEV by 2010*" would be coded as "*BEV*". Sentiment describes an expert's positive or negative attitude toward the coded technology (Rocklage et al., 2023). Hence, a large discrepancy in the average sentiments towards BEVs and FCEVs signals that the majority favors one technology over the other. Based on this, the degree of consensus is inferred and applied in the sentiment analysis later. Negative sentiment is characterized by foreseen issues such as limited technology adoption, shortcomings in performance, or barriers to market entry. In contrast, positive sentiment is characterized by (commercial) successes, technological advancements, OEM push, or customer pull factors. The neutral sentiment describes mixed or ambiguous predictions, e.g. "*FCEVs will gain market share, but hydrogen infrastructure will be a barrier to their acceptance.*"

Combined with the meta-attributes of the prediction (author, institution, year, publication medium), this creates a longitudinal analysis of the prediction's correctness and consensus.

The documents of Mercedes-Benz are coded for technology, sentiment, and themes. Technology and sentiment coding are executed identically to the expert predictions. The themes arise through bottom-up coding. First, each statement related to alternative powertrains is tagged by a theme (e.g. range, charging speed). In the next step, these themes receive a higher-level code (e.g. range and charging speed refer to technical attributes). Thereby, the sentiment towards technology is connected to the themes Mercedes-Benz uses to reason. Additionally, predictions of Mercedes-Benz are retrieved and measured, identically to the expert predictions.

According to the third-eye principle, the explained codes for scoring, technology, and accuracy are assigned in a three-step process. In the first step, two researchers code independently. In the next step, differences will be discussed and eliminated by the two researchers. In the third step, if differences in judgment prevail, a third researcher is consulted which has the decision power about the respective codes.

RESULTS: EXPERT CROWD WISDOM

Experts' accuracy

For the expert predictions, the TWBS indicates the level of foresight the experts showcased in history. This is compared to "*guessing*". A guess is defined as the assignment of a probability of 50% to an event and its counter-event without a time horizon. Therefore, guessing constitutes a TBWS of 0.5. Comparing this to the expert crowd, the t-Test shows it was significantly better than guessing in the past (see Table 2). The average TWBS of all experts is 0.32.

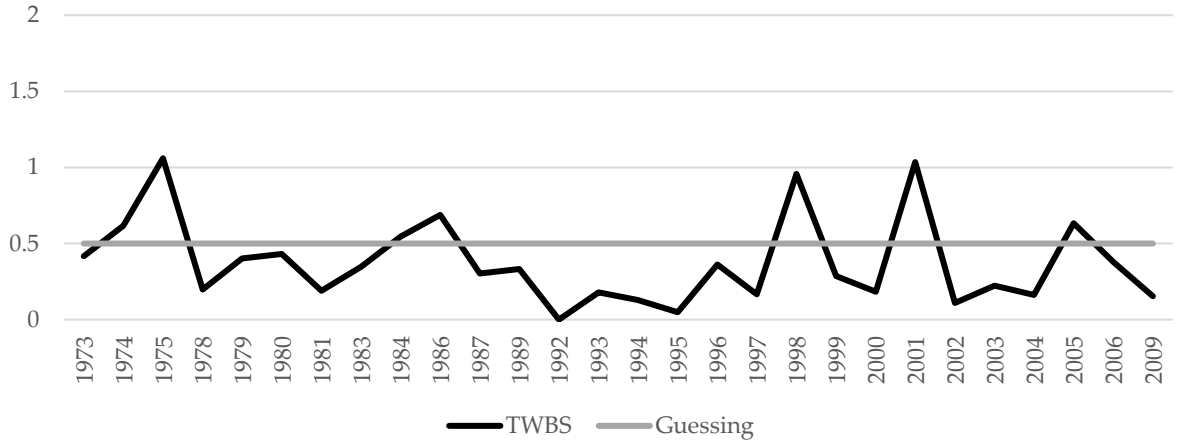
TABLE 2
One-sample t-Test of expert predictions (comparison value = 0.5)

Metric	Mean	*t*	*df*	95% confidence intervall	*p*-value
TWBS	0.3199	-4.3798	88	0.2382; 0.4016	0.0000**

N = 89
† p <0.10
* p <0.05
** p <0.01

The experts have been highly accurate and early (TBWS < 0.15) in 1992, 1994, 1995, and 2002 (see Figure 2) Kukkonen and Shelef (1994b, p. 146) predicted: "*The overall conclusion is that hydrogen will not be used as an automotive fuel before the year 2020 at the earliest.*" Tokuda (1995, p. 344) graphically displays electric vehicles with lithium batteries to become available in the next 20 years. Chan (2002, p. 266) gave this outlook: "*At the present time and in the foreseeable future, batteries have been agreed to be the major energy source for EVs.*" In the few years of 1975, 1998, and 2001, the experts' average was exceptionally worse than guessing (TWBS > 0.85). For instance, Parrish et al. (1975, p. 75) state in a U.S. government report "*There are no insurmountable technical problems that must be solved before hydrogen can be adopted as a synthetic fuel for use in transportation*." Yet, the hydrogen distribution infrastructure remains a major technical problem (Tashie-Lewis & Nnabuife, 2021). Clark et al. (2005, p. 46) predicted: "*The future roadway will have within a few years hundreds of regional hydrogen power stations for local energy needs, as more and more vehicles come into the marketplace.*" But the prediction has not materialized.

FIGURE 2
Time-weighted Brier score of expert predictions



Experts' sentiment

Following accuracy, the sentiment towards the different powertrain technologies is analyzed. The objective is to examine whether the expert crowd had a stronger positive sentiment toward one technology compared to the other. This analysis focuses on the different sentiments, not filtered by correctness. Due to the scientific inquiry, we are eager to understand whether the crowd had a consensus that would have aided Mercedes-Benz with its asset allocation.

The sentiment of the expert predictions has been inconsistent from 1970 to 1994. Then, the positive sentiment consequently shifted to BEV as a credible alternative. The trendline indicates how the negative opinions about BEV turned positive (see Figure 3). For clarification, the average sentiment towards FCEV in 1985-1989 was exactly 0. From 1995 onwards, the consensus favors BEV over FCEV.

In the 1970s, it was the assumption that hydrogen would fuel ICEs (Winsche et al., 1973). Therefore, experts expected the technical transition to be smooth because the underlying powertrain attributes remain. However, hydrogen infrastructure criticism arose in the 1980s which deviated from the vehicle technology but impacted the viability of an alternative powertrain: "*Therefore a wide-ranging hydrogen infrastructure for motor vehicles will probably not be realized in the future due to economical reasons as well as specifically for the European*

area for political reasons" (Buchner, 1983, p. 374). The hydrogen fuel cell replaced hydrogen ICEs as the preferred technology to use hydrogen as energy source in the 1980s, but in addition to the problems of an early innovation, the infrastructure issues remained. An automotive employee, Carl A. Kukkonen from Ford, precisely predicts that "*hydrogen will not be used as an automotive fuel during this century*" (Kukkonen, 1981, p. 1433) based on fuel cell technology, hydrogen production, and onboard storage. In his analysis, he ranked powertrain technologies by economical, technological, and ecological aspects, concluding that hydrogen-fueled ICEs and FCEV rank behind BEV. Yet, BEVs' technical features were seen as the major downside throughout the years (i.e. low range, heavy battery, long recharging time): "*Electric cars can never match those with petrol or diesel engines because batteries weighing half a tone store only as much energy as a few litres of fuel*" (Marshall, 1989).

In the first half of the 1990s, the sentiment turned generally negative, especially for FCEV. Kukkonen updated his alternative fuel analysis and extended his negative view on the future of hydrogen: "*Therefore, it is concluded that hydrogen will not be used as an automotive fuel before the year 2020*" (Kukkonen & Shelef, 1994a, p. 385). A government report for the U.S. Department of Transportation concludes in a comparative view of the infrastructure for alternative fuels: "*LPG, CNG, electricity, ethanol, and methanol follow fairly closely together in that order, while hydrogen trails badly*" (Amann et al., 1994, p. 36). Nonetheless, the requirements for a BEV infrastructure have also been criticized: "*And even the most advanced EV designs currently envisioned will need a costly charging infrastructure to permit 'refueling' at each long-term parking space*" (Womack, 1994, p. 266).

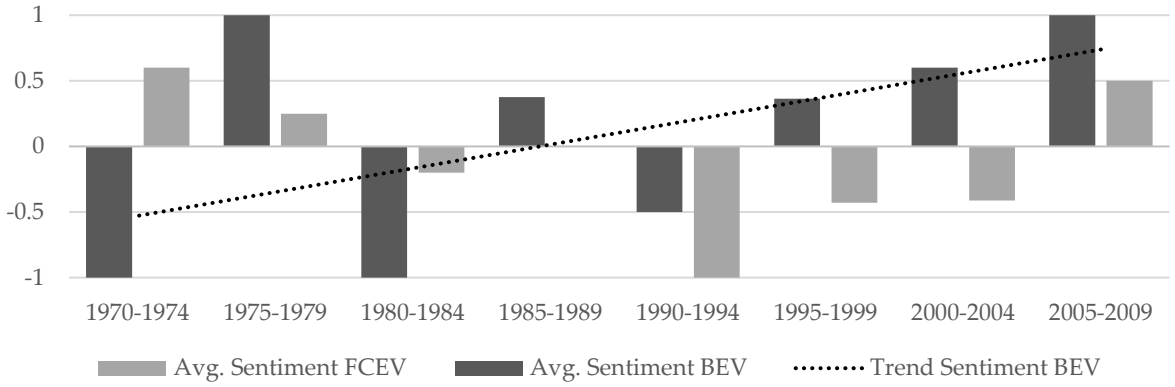
Until the millennium, the sentiments drifted apart, signaling a stronger consensus that BEV would be the dominating alternative. The experts foresaw the market launch of lithium batteries which was an early sign of technological advancement in battery technology that would put the BEV ahead as it promised increased driving range and power (Cowan & Hultén, 1996; Tokuda,

1995). BEVs were more likely to take market share, but they were still far away. At the time, only employees in the automotive industry favored FCEVs because they were expecting a series-ready FCEV by 2005-2006 (Panik, 1998; v. d. Weiden, 1998).

As major hurdles the viability of FCEVs became clear, the consensus solidified. In Wengel and Schirrmeister (2000), gas station operators are questioned to understand the requirements of the FCEV infrastructure. They did not see hydrogen as a relevant business opportunity for the next 20 years and pointed to the extremely high investment costs for hydrogen refuelling stations. Therefore, they were reluctant to invest a lot of capital. This limited the market acceptance of the FCEV. Carlsson and Johansson-Stenman (2003, p. 25) concluded the same disadvantage: "*The profitability of FCCs [Fuel Cell Cars] is largely reduced by the fact that a distribution system for hydrogen was not considered realistic in a 10-year perspective.*" Meanwhile, BEVs were seen as a more promising alternative and batteries were seen as the dominating energy source for EVs: "*MIT results imply that powertrain technology that has been in the market for over a century (ICEs and electric drive […]) will continue to improve from 2010-2020, while FCVs will not*" (An & Santini, 2003, p. 648).

Towards the end of the analyzed time frame (2005-2009), the gap between the sentiments was narrower. Despite the old arguments against FCEV, some automotive and governmental employees remained bullish on this technology, e.g. "*The transition to hydrogen-powered fuel cell vehicles will occur over the next 10–15 years*" (Chalk & Miller, 2006, p. 73). However, the sampling restrictions in the time frame 2005-2009 need to be kept in mind when assessing the informative value of this sub-sample.

FIGURE 3
Sentiment of experts towards FCEV and BEV



Beginning in the mid-1990s, the expert predictions point to BEV as the best future alternative powertrain technology. But how did Mercedes-Benz view the potential of BEVs and FCEVs since the 1970s?

RESULTS: MERCEDES-BENZ (OEM) ACTIONS

Mercedes-Benz accuracy

Mercedes-Benz[1] offered 32 predictions in our sample period. In contrast to the expert group, Mercedes' predictions are statistically not better than guessing (see Table 3). Its average accuracy score is 0.57.

TABLE 3
One-sample t-Test of Mercedes' predictions (comparison value = 0.5)

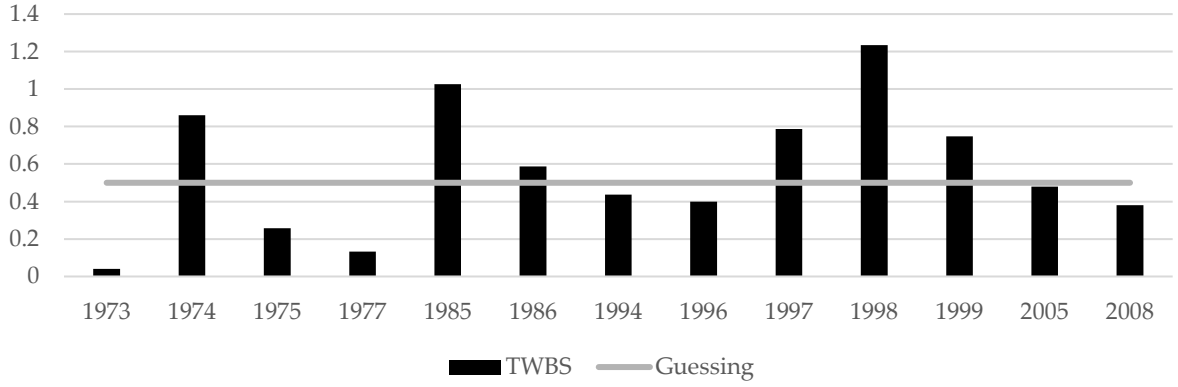
Metric	Mean	*t*	*df*	95% confidence intervall	*p*-value
TWBS	0.5707	0.7467	31	0.3777; 0.7636	0.4609

N=32
† p <0.10
* p <0.05
** p <0.01

[1] Mercedes-Benz is called «Mercedes» in the following text. The name will be used, although there have been several re-brandings in the company's history.

The years 1973 and 1977 stand out positively (see Figure 4). In 1973, the director of commercial vehicles powertrain stated: "*In future, special requirements could be placed on the environmental friendliness of vehicles travelling in certain urban areas,*" 20 years before California's Zero-Emission Act, the first of its kind, was announced. Otherwise, Mercedes made many inaccurate predictions, especially in 1974, 1985 and 1998. In 1985, they praised the potential of hydrogen: "*Corresponding technologies will therefore play an important role in energy supply in the future.*" Today, hydrogen still does not live up to this claim. In 1998, the company foresaw its fuel cell breakthrough looming: "*Our strategic alliance with Ballard Power Systems Inc. in Canada and the Ford Motor Company in the USA will lay the foundations for the introduction of fuel cell vehicles in 2004.*" The first FCEV full series launched on the market 14 years later than predicted.

FIGURE 4
Time weighted Brier score of Mercedes-Benz' predictions



Mercedes-Benz activities

Bottom-up qualitative coding revealed that Mercedes was focused on hydrogen – first combusted in an ICE, then converted in the fuel cell. In the 1970s, Mercedes acted in line with the expert judgment. Hydrogen combustion was considered a valid alternative: "*In the long term, hydrogen combustion therefore represents a viable alternative to petroleum-based fuels"* (*WASSERSTOFF ALS ALTERNATIV-KRAFTSTOFF*, 1974, p. 3). This view was later reflected in headlines such as: "*Hydrogen - energy supply of the future*" (Molter, 1978, p. 2). BEVs

received attention because they were perceived as an inferior technology due to range disadvantage and recharging requirements. Still, Mercedes pursued a dual research strategy at that time, also examining the potential of battery powertrains (*Sondermappe zur IAA 77*, 1977). As a result, Mercedes presented their T-model-based prototypes of a hydrogen combustion vehicle and a BEV at the beginning of the 1980s. Surprisingly, BEVs became scarce in Mercedes' communication in the remainder of the decade. The company focused its activities on hydrogen: "*In the search for alternative energies for road transport, hydrogen is one of the few promising alternative fuels"* (*MERCEDES-BENZ-FLOTTENERPROBUNG IN BERLIN*, 1984, p. 1). Mercedes' tests showed the inferiority of contemporary battery technology due to the technical features (i.e. recharging time, speed, range, weight). Hydrogen was framed as sustainable and praised for its gasoline-like refueling time. Yet, the onboard storage of hydrogen remained a critical issue. The mid-1980s were relatively calm in the alternative powertrain transition of Mercedes. By the end of the decade, Mercedes showed "*[…] proof of the technical applicability of hydrogen for vehicle operation […]*" (Hörnig, 1989, p. 9). This took an interesting twist in the 1990s. While batteries were under research without satisfying results: "*However, there are still some problems to be solved: The lifespan and recycling of the batteries, for example, and above all the costs"* (*Mercedes-Benz auf dem Genfer Automobilsalon*, 1991, p. 6*),* Mercedes stayed intriguingly shallow about hydrogen combustion engines at the beginning of the decade. Concluding from its communication, Mercedes quietly pivoted from hydrogen combustion to the fuel cell because it is more efficient than ICEs: "*The fuel cell vehicle thus achieves significantly better efficiency than a car with an internal combustion engine*" (*Die Vorteile der Brennstoffzelle*, 1994, p. 1). Mercedes then confirmed it invested hundreds of millions in fuel cell technology and partnered with Ballard-Power Systems from Canada. Also, Mercedes claimed setting up hydrogen infrastructure was easy in 1994: "*It would be no problem to equip the existing petrol stations with a methanol dispenser*" (*Die Vorteile der*

Brennstoffzelle, 1994, p. 1). As mentioned earlier, the experts heavily criticized hydrogen as an energy carrier for its infrastructure requirements before and after that time. Nonetheless, Mercedes presented its first FCEV prototype, the NECAR ("*New Electric Car*"). What followed was systematic investment and development of fuel cells within the organization. Mercedes researched BEVs and achieved ranges of up to 220 km, but it declared the fuel cell drive as its strategic priority (Scheunemann, 1995). Compared to BEVs, Mercedes' FCEVs achieved a range of 250 km. Therefore, Mercedes' top management believed that "*From today's perspective, [it is] the alternative drive system with the greatest opportunities to seriously compete with the combustion engine*" (Schrempp, 1997, p. 2) and made more investments into fuel cell technology. First, it extended its partnership with Ballard and established a joint venture. Mercedes' board member, Jürgen Hubbert, announced an investment of 600 million DM at the IAA in 1997. The goal for an FCEV series was set for 2005. Then, Ford got on board. Mercedes and Ford intertwined their fuel cell research operations across multiple organizations. Mercedes invested heavily but also split the risk by partnering and outsourcing. In the following years, Mercedes continuously presented updated versions of its NECAR. This cumulated in the California Fuel Cell Partnership in 1999. The U.S. state, Mercedes, Ford, Shell, and more companies joined forces. One reason was that Mercedes acknowledged the infrastructure hurdle as *"one of the biggest challenges"* (Eaton, 1999, p. 1). But it remained hopeful that others would solve the issue. At that time, the NECAR 4 achieved a range of up to 450 km. After the millennium, news became relatively scarce again. Mercedes focused on the further development of the NECAR and planned to invest €1 billion until 2004, while BEV received little attention. In their drive strategy of 2004, BEVs were not mentioned, only hybrids and FCEVs. However, the announced market readiness of FCEVs was postponed by 7 to 10 years, signaling problems or prior misjudgment. Infrastructure was still seen as a major obstacle by Mercedes, but this did not affect its powertrain strategy. In 2007, the Bibendum challenge took place. In this

tournament, companies present their alternative mobility solutions of all kinds. Similarly to the 1998 Bibendum challenge, Mercedes showed up with FCEVs and hybrid Smarts, but no BEV prototype. The shift in Mercedes' narrow strategic focus on the fuel cell occurred one year later according to its "Road to the Future" publication. Here, Mercedes mentions BEV and FCEV in the same sentence as the solution for emission-free driving which was not the case in prior strategy papers. The following statement implicitly states Mercedes neglected BEVs for the sake of the FCEVs and has started to transfer the fuel cell electric knowledge to the battery domain: "*The company has been conducting research in the field of fuel cell technology since 1991 and therefore has extensive experience in electric driving. The knowledge gained from this can also be utilized profitably for the development of battery vehicles"* (Horn & Wiese, 2008, p. 6). In the same document, Mercedes refers to the application of the lithium-ion battery, which could solve the frequently mentioned problems of range, weight, and charging speed. This was more than one decade after the experts foresaw the potential of the lithium-ion battery. Along with the strategic change, Mercedes announced several partnerships for battery production and infrastructure, labeled as "*investment in the future*" (Müller et al., 2008, p. 1). For the next years, Mercedes again followed a dual drivetrain strategy approach – FCEV and BEV. Outside the period analyzed, this culminated in the late market launch of the GLC F-Cell in 2018, the first and last FCEV full series, and the introduction of the first Mercedes-branded BEV in 2019.

DISCUSSION AND CONCLUSION

Key findings

The transformation of qualitative predictions into quantitative forecasts allows measurement of accuracy of the expert crowd. The method measures expert judgments and crowd wisdom across cases without setting up expensive and effortful other scientific interventions such as forecasting tournaments. It enables inferences about the individual foresight of experts, the

expert crowd, and a firm and its top management (in our case Mercedes Benz). Nonetheless, it can be applied to different research endeavors because it is agnostic to the underlying case. In our study, we focused on Mercedes-Benz as research object. Going forward, we will add two other companies to see whether Mercedes-Benz is an outlier or representative of most legacy car makers in the period.

The analysis of the experts' predictions has shown, first, their crowd wisdom, and second, their consensus. The experts showcased accurate and early foresight on the future of alternative powertrains. The expert crowd anticipated today's BEV dominance. We unveiled that a consensus for BEV emerged within the expert crowd, starting in 1995. On the one hand, battery technology advanced. On the other hand, the awareness of the infrastructure as a critical barrier to hydrogen as an energy carrier rose. By 2005, BEV prototypes outperformed FCEVs (Hammerschlag & Mazza, 2005). Thus, an earlier market introduction for BEVs and conclusively, a win against FCEVs in the race for the future alternative came closer. Although BEVs also require infrastructural investments (e.g. generation capacity, grid, charging stations), the FCEV investment greatly surpasses it due to less electric efficiency and related additional generation capacity needs, machinery (electrolyzer, compression) to produce the fuel, distribution network, refuelling stations, and almost no potential for self-supply of vehicle owners (Eaves & Eaves, 2004).

In contrast, Mercedes-Benz did not show good foresight, because its predictions were not better than someone simply guessing. Contrary to the experts, Mercedes clung to its FCEVs. Mercedes acknowledged the problems raised by the experts. Nevertheless, the company stuck to its narrow fuel cell strategy, even when the key bottleneck for BEVs (economical high-performance batteries) was getting solved with lithium-ion battery technology.

Our analysis of the expert crowd demonstrated the value of qualitative expert judgment for strategic decision-making. Before Ford invested in the partnership of Mercedes-Benz and

Ballard Power Systems, and long before Mercedes-Benz showed up with the fuel-cell-based F-Cell to the Michelin Bibendum challenge, the crowd wisdom of the experts pointed to a different direction, namely BEVs as the more promising alternative powertrain technology.

Limitations

These research results have their limitations. These arise through the translation of the TWBS and the sampling of the predictions.

First, our developed scoring rule is not strictly proper. Briefly, a strictly proper scoring rule encourages the forecaster to state his or her true belief (Gneiting & Raftery, 2007). The original Brier score and the TWBS are strictly proper scoring rules (Schuler et al., Forthcoming). Our application of the TWBS, however, would theoretically encourage forecasters to extend their prediction horizon to minimize their score, without it being their true belief. Hence, it is not strictly proper. Still, it is a rigorous representation of the time dimension of foresight in qualitative predictions because it follows the criteria of positive, linear, and strictly monotonously increasing time weights. Also, the scoring is based on indirect data collection. The forecasters are not directly in contact with the research. Hence, it is reasonable to assume that they are not incentivized in one way or another and, thus, they state their true beliefs.

Second, although the sampling follows a clear process, it has two problems. First, the approach is thorough but does not inevitably lead to an exhaustive database. More precisely, the data collection process does not capture every prediction of experts globally and yearly. Hence, there are under- and over-representation of author types, regions, and years in the sample. While the portion of the researchers' group is expected given the analytical focus on published journals due to their documentation trail, the geographical bias might distort the overall result. The overall sample is Western-centric. While the experts were affiliated with organizations

from 14 different countries, 51% of their predictions were made by experts affiliated with U.S. organizations. Germany contributed 14%, and the UK 10% to the sample. Two major car manufacturing nations, China and Japan, are represented by small sub-samples of 3% and 4%. Two reasons for the bias can be identified. Firstly, language barriers prevent the inclusion of native-language publications such as government reports or books. Secondly, the Chinese automobile industry slowly emerged in the 1990s and accelerated in the 2000s (Chen et al., 2020). We can reasonably assume that the scientific interest was small throughout large parts of the study time frame. Nonetheless, the sample of this study compiles various nations and institutions. The open approach to collecting predictions achieved a large set of individual opinions. But some specific years are not included, and others are covered by a only few predictions while some years have a lot predictions. However, from a holistic perspective, it is not necessary to represent every year. Journal articles are asked for revision and research for one publication might take significantly longer than for other articles. Therefore, a prediction in a single year does not precisely display the belief in that year, instead, it demonstrates a contemporary view.

Future Research

We developed a generic, replicable foresight measure for qualitative predictions. Currently, a large literature on foresight accuracy measures exists which we enrich with a new method of scoring systematically qualitative predictions. Going forward, foresight scholars are encouraged to confirm its validity beyond our case study. It needs to be replicated in other scenarios of high uncertainty impacting strategic decisions (e.g. energy transition and renewable energies). Multiple applications and testing of the method will establish its external validity.

Next to the foresight measure, the expert crowd wisdom and its value for foresight and strategic decision-making under high uncertainty has been shown in our study. Further research should explore how to automatize and further standardize the method. While the method is sound, the

procedure involves intense manual work. To further leverage the value of measuring qualitative predictions, automizing several steps (e.g. assessing correctness, coding temporal vagueness, coding technology) would support its application. Digital technologies such as natural language processing or generative large language models are suited for this.

For the field of strategic management, a second research trajectory emerges. The alternative powertrain transition is a fertile research ground for examining technology strategy and decision-making. Mercedes-Benz as an incumbent slept on the EV trend and is struggling to catch up with Chinese competitors. A follow-up research question is "*how do management theories explain the strategic failure?*" The strategic management & innovation management theories might offer different explanations for Mercedes' failure. According to the theory-based view of the firm and theory-based decision making under uncertainty, Mercedes would have lacked the right theory about the situation and did not appropriately experiment or react on experiments' results (Camuffo et al., 2024). According to managerial cognition, the cognitive abilities of Mercedes' top management hampered their managerial abilities (Helfat & Peteraf, 2015). This would have doubling down on the fuel cell despite old arguments against BEVs have vanished, instead of updating its beliefs. According to absorptive capacity theory, Mercedes' would not have successfully leveraged external knowledge for their decision making (Cohen & Levinthal, 1990). As seen, various stakeholders in the powertrain ecosystems stated their concerns regarding hydrogen. Despite Mercedes' partnerships, it still acted against the external knowledge. It is to be seen to what degree the theories can explain the failure and where differences in the explanations lie. By doing such an extensive case study of the failure of Mercedes Benz to see what the expert crowd already saw, we might be able to advance out theories about corporate foresight.

LITERATURE

Amann, C., Bonderson, L., & Herrett, R. (1994). *Precursor Systems Analyses of Automated Highway Systems - Activity Area M - Alternative Propulsion Systems Impact*. (FHWA-RD-95-149). Federal Highway Administration

An, F., & Santini, D. (2003). Assessing Tank-to-Wheel Efficiencies of Advanced Technology Vehicles. *SAE Transactions*, *112*, 636-654.

Bolger, F. (2018). The Selection of Experts for (Probabilistic) Expert Knowledge Elicitation. In *Elicitation - The Science and Art of Structuring Judgement* (Vol. 261, pp. 393-444). https://doi.org/https://doi.org/10.1007/978-3-319-65052-4_16

Brier, G. W. (1950). Verification of Forecasts expressed in terms of Probability. *Monthly Weather Review*, *78*(1), 1-3. https://doi.org/10.1175/1520-0493(1950)078

Buchner, H. (1983). The Question of the Hydrogen Infrastructure for Motor Vehicles. *International Journal of Hydrogen Energy*, *8*(5), 373-380. https://doi.org/10.1016/0360-3199(83)90053-8

Camuffo, A., Gambardella, A., & Pignataro, A. (2024). Theory-Driven Strategic Management Decisions. *Strategy Science*, *9*(4). https://doi.org/10.1287/stsc.2024.0173

Carlsson, F., & Johansson-Stenman, O. (2003). Costs and Benefits of Electric Vehicles: A 2010 Perspective. *Journal of Transport Economics and Policy*, *37*(1), 1-28.

Chalk, S. G., & Miller, J. F. (2006). Key challenges and recent progress in batteries, fuel cells, and hydrogen storage for clean energy systems. *Journal of Power Sources*, *159*(1), 73-80. https://doi.org/10.1016/j.jpowsour.2006.04.058

Chan, C. C. (2002). The State of the Art of Electric and Hydrid Vehicles. *Proceedings of the IEEE*, *90*(2), 247-275.

Chang, W., Chen, E., Mellers, B., & Tetlock, P. (2016). Developing expert political judgment: The impact of training and practice on judgmental accuracy in geopolitical forecasting tournaments. *Judgment and Decision Making*, *11*(5), 509-526.

Chen, Y., Lawell, C. Y. C. L., & Wang, Y. S. (2020). The Chinese automobile industry and government policy. *Research in Transportation Economics*, *84*. https://doi.org/10.1016/j.retrec.2020.100849

Clark, W. W., Rifkin, J., O'Connor, T., Swisher, J., Lipman, T., & Rambach, G. (2005). Hydrogen energy stations: along the roadside to the hydrogen economy. *Utilities Policy*, *13*(1), 41-50. https://doi.org/10.1016/j.jup.2004.07.005

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive-Capacity - a New Perspective on Learning and Innovation. *Administrative Science Quarterly*, *35*(1), 128-152. https://doi.org/Doi 10.2307/2393553

Cowan, R., & Hultén, S. (1996). Escaping Lock-in: The Case of the Electric Vehicle. *Technological Forecasting and Social Change*, *53*, 61-79.

Die Vorteile der Brennstoffzelle. (1994, 13.04.1994).

Eaton, R. (1999, 20.04.1999). *DaimlerChrysler gründet gemeinsam mit Partnern die "California Fuel Cell Partnership"*

Eaves, S., & Eaves, J. (2004). A cost comparison of fuel-cell and battery electric vehicles. *Journal of Power Sources*, *130*(1-2), 208-212. https://doi.org/10.1016/j.jpowsour.2003.12.016

Fergnani, A. (2022). Corporate Foresight: A New Frontier for Strategy and Management. *Academy of Management Perspectives*, *36*(2), 820-844. https://doi.org/10.5465/amp.2018.0178

Friedman, J. A., Baker, J. D., Mellers, B. A., Tetlock, P. E., & Zeckhauser, R. (2018). The Value of Precision in Probability Assessment: Evidence from a Large-Scale Geopolitical Forecasting Tournament. *International Studies Quarterly*, *62*(2), 410-422. https://doi.org/10.1093/isq/sqx078

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*(477), 359-378. https://doi.org/10.1198/016214506000001437

Goldfarb, R. S., Stekler, H. O., & David, J. (2005). Methodological issues in forecasting: Insights from the egregious business forecast errors of late 1930. *Journal of Economic Methodology*, *12*(4), 517-542. https://doi.org/10.1080/13501780500343524

Gordon, A. V., Ramic, M., Rohrbeck, R., & Spaniol, M. J. (2020). 50 Years of corporate and organizational foresight: Looking back and going forward. *Technological Forecasting and Social Change*, *154*. https://doi.org/10.1016/j.techfore.2020.119966

Hammerschlag, R., & Mazza, P. (2005). Questioning hydrogen. *Energy Policy*, *33*(16), 2039-2043. https://doi.org/10.1016/j.enpol.2004.04.010

Helfat, C. E., & Peteraf, M. A. (2015). Managerial Cognitive Capabilities and the Microfoundations of Dynamic Capabilities. *Strategic Management Journal*, *36*(6), 831-850. https://doi.org/10.1002/smj.2247

Horn, C., & Wiese, E. (2008, 13.06.2008). *Mercedes-Benz "Road to the Future"*

Hörnig, R. (1989). *Forschung und Technik im Daimler-Benz-Konzern*

Høyer, K. G. (2008). The history of alternative fuels in transportation: The case of electric and hybrid cars. *Utilities Policy*, *16*(2), 63-71. https://doi.org/10.1016/j.jup.2007.11.001

IEA. (2024). *Global EV Outlook 2024*. https://www.iea.org/reports/global-ev-outlook-2024

Irwin, D., & Mandel, D. R. (2020). *Variants of Vague Verbiage - Intelligence Community Method for Communicating Probability* (Assessment and Communication of Uncertainty in Intelligence to Support Decision-Making, Issue.

Kapoor, R., & Wilde, D. (2023). Peering into a crystal ball: Forecasting behavior and industry foresight. *Strategic Management Journal*, *44*(3), 704-736. https://doi.org/10.1002/smj.3450

Kazeem, A. O. (2024). *Electric vehicles: market data & analysis* (Market Insights, Issue.

Kukkonen, C. A. (1981). Hydrogen as an Automotive Fuel. *SAE Transactions*, *90*, 1425-1461.

Kukkonen, C. A., & Shelef, M. (1994a). Hydrogen as an Alternative Automotive Fuel: 1993 Update. *SAE Transactions*, *103*, 385-405.

Kukkonen, C. A., & Shelef, M. (1994b). Prospects of Hydrogen-Fueled Vehicles. *Progress in Energy and Combustion Science*, *20*, 139-148. https://doi.org/https://doi.org/10.1016/0360-1285(94)90008-6

Larrick, R. P., Mannes, A. E., & Soll, J. B. (2012). The social psychology of the wisdom of crowds. In J. I. Krueger (Ed.), *Social judgment and decision making* (pp. 227-242). Psychology Press.

Leech, G. (2023). *Scoring the Big 3's predictive performance - Asimov, Clarke, and Heinlein as forecasters*. https://arbresearch.com/files/big_three.pdf

Lundquist, K., & Stekler, H. O. (2012). Interpreting the Performance of Business Economists During the Great Recession. *Business Economics*, *47*(2), 148-154. https://doi.org/10.1057/be.2012.2

Marinkovic, M., Al-Tabbaa, O., Khan, Z., & Wu, J. (2022). Corporate foresight: A systematic literature review and future research trajectories. *Journal of Business Research*, *144*, 289-311. https://doi.org/10.1016/j.jbusres.2022.01.097

Marshall, S. (1989). Motoring: A future for current models? - Electric cars could be the urban runabouts of the 21st century. *Financial Times*.

Martin, T. G., Burgman, M. A., Fidler, F., Kuhnert, P. M., Low-Choy, S., McBride, M., & Mengersen, K. (2012). Eliciting Expert Knowledge in Conservation Science. *Conservation Biology*, *26*(1), 29-38. https://doi.org/10.1111/j.1523-1739.2011.01806.x

MERCEDES-BENZ-FLOTTENERPROBUNG IN BERLIN. (1984, 05.10.1984).

Mercedes-Benz auf dem Genfer Automobilsalon. (1991).

Molter, G. (1978). *DAIMLER-BENZ FORSCHUNG RUND UMS AUTO*

Morita, K. (2003). Automotive power source in 21st century. *JSAE Review*, *24*(1), 3-7. https://doi.org/10.1016/S0389-4304(02)00250-3

Müller, B., V., S., Horn, C., & Wiese, E. (2008, 15.12.2008). *Evonik und Daimler gründen strategische Allianz zur Entwicklung und Fertigung von Lithium-Ionen Batterien*

Panik, F. (1998). Fuel cells for vehicle applications in cars - bringing the future closer. *Journal of Power Sources*, *71*(1-2), 36-38. https://doi.org/10.1016/S0378-7753(97)02805-X

Parrish, W. R., Voth, R. O., Hust, J. G., Flynn, T. M., Sindt, C. F., & Olien, N. A. (1975). *Selected Topics on Hydrogen Fuel*. Institute for Basic Standards, National Bureau of Standards

Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., Ben Taieb, S., Bergmeir, C., Bessa, R. J., Bijak, J., Boylan, J. E., Browell, J., Carnevale, C., Castle, J. L., Cirillo, P., Clements,

M. P., Cordeiro, C., Oliveira, F. L. C., De Baets, S., Dokumentov, A.,…Ziel, F. (2022). Forecasting: theory and practice. *International Journal of Forecasting*, *38*(3), 705-871. https://doi.org/10.1016/j.ijforecast.2021.11.001

Rocklage, M. D., He, S., Rucker, D. D., & Nordgren, L. F. (2023). Beyond Sentiment: The Value and Measurement of Consumer Certainty in Language. *Journal of Marketing Research*, *60*(5), 870-888. https://doi.org/10.1177/00222437221134802

Rohrbeck, R., & Kum, M. E. (2018). Corporate foresight and its impact on firm performance: A longitudinal analysis. *Technological Forecasting and Social Change*, *129*, 105-116. https://doi.org/10.1016/j.techfore.2017.12.013

Scheunemann, W. (1995, 14.05.1995). *DAIMLER-BENZ STELLT REVOLUTIONÄRES FAHRZEUG VOR*

Schilder, F., & Habel, C. (2001). From Temporal Expressions to Temporal Information: Semantic Tagging of News Messages. Proceedings of the ACL 2001 Workshop on Temporal and Spatial Information Processing,

Schoemaker, P. J. H., Heaton, S., & Teece, D. (2018). Innovation, Dynamic Capabilities, and Leadership. *California Management Review*, *61*(1), 15-42. https://doi.org/10.1177/0008125618790246

Schrempp, J. E. (1997, 14.04.1997). *Daimler-Benz beteiligt sich an Ballard Power Systems*

Schuler, B. A., Murmann, P. J., Beisemann, M., & Satopää, V. (Forthcoming). INDIVIDUAL FORESIGHT: CONCEPT, OPERATIONALIZATION, AND CORRELATES. *International Journal of Forecasting*.

Sondermappe zur IAA 77. (1977).

Sperling, D., & Deluchi, M. A. (1989). Transportation Energy Futures. *Annual Review of Energy*, *14*, 375-424. https://doi.org/10.1146/annurev.eg.14.110189.002111

Stekler, H., & Symington, H. (2016). Evaluating qualitative forecasts: The FOMC minutes, 2006-2010. *International Journal of Forecasting*, *32*(2), 559-570. https://doi.org/10.1016/j.ijforecast.2015.02.003

Tashie-Lewis, B. C., & Nnabuife, S. G. (2021). Hydrogen Production, Distribution, Storage and Power Conversion in a Hydrogen Economy - A Technology Review. *Chemical Engineering Journal Advances*, *8*. https://doi.org/10.1016/j.ceja.2021.100172

Tetlock, P. E. (2005). Expert Political Judgment: How Good Is It, How Can We Know. *Expert Political Judgment: How Good Is It, How Can We Know*, 1-321. <Go to ISI>://WOS:000287913900012

Tissot, H., Del Fabro, M. D., Derczynski, L., & Roberts, A. (2019). Normalisation of imprecise temporal expressions extracted from text. *Knowledge and Information Systems*, *61*(3), 1361-1394. https://doi.org/10.1007/s10115-019-01338-1

Tokuda, T. (1995). Automobiles for sustainable mobility (FISTA 94, Beijing, 20 October 1994). *JSAE Review*, *16*, 341-348.

v. d. Weiden, S. (1998, 17.04.1988). Die Technologie der Brennstoffzelle hat grosse Fortschritte gemacht. *Die WELT*.

WASSERSTOFF ALS ALTERNATIV-KRAFTSTOFF. (1974).

Wengel, J., & Schirrmeister, E. (2000). *Innovationsprozess vom Verbrennungsmotor zur Brennstoffzelle - Chancen und Risiken für die baden-württembergische Industrie*.

Winsche, W. E., Hoffman, K. C., & Salzano, F. J. (1973). Hydrogen - Its Future Role in Nations Energy Economy. *Science*, *180*(4093), 1325-1332. https://doi.org/10.1126/science.180.4093.1325

Womack, J. P. (1994). The real EV challenge: reinventing an industry. *Transport Policy*, *1*(4), 266-270.