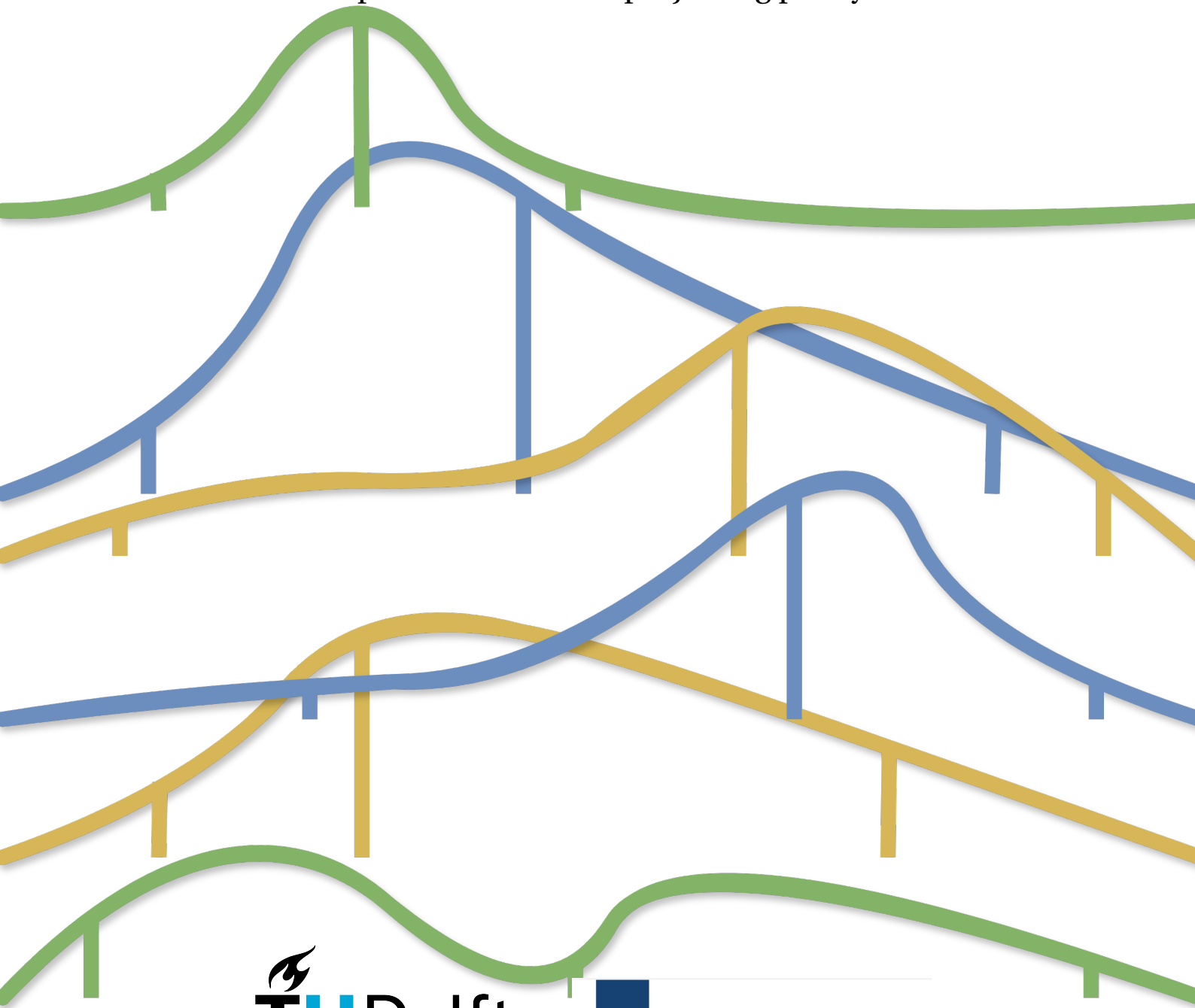
Supporting the expert judgement process for quantifying uncertainties in policy research

A case study of the unique challenges with implementing structured expert elicitation for projecting policy effects



TU Delft

Planbureau voor de Leefomgeving

SUPPORTING THE EXPERT JUDGEMENT PROCESS FOR QUANTIFYING UNCERTAINTIES IN POLICY RESEARCH

A CASE STUDY OF THE UNIQUE CHALLENGES WITH IMPLEMENTING STRUCTURED EXPERT ELICITATION FOR PROJECTING POLICY EFFECTS

Masters thesis submitted to Delft University of Technology in partial fulfillment of the requirements for the degree of

Master of Science

in Communication Design for Innovation

Thirza BOLHUIS
Student Number: 4660552

To be defended in public on 9 April 2026.

Graduation Committee
Chairperson and first supervisor: Drs. C. Wehrmann
Second supervisor: Dr. É. Kalmàr
Third supervisor: Dr. T. Nane
External supervisor: Dr. S. Setty

An electronic version of this thesis is available at
http://repository.tudelft.nl/.

TU Delft

Planbureau voor de Leefomgeving

PREFACE

As I'm becoming a professional, I'll be faced with making decisions based on my expertise. Though much of my studies (Mathematics, modelling, policy analysis...) has taught me valuable methods to calculate my way to a solution, I've also encountered many micro decisions that add up to a whole lot of assumptions and uncertainty. How do people in the real world deal with making the right decisions based on their expertise? Luckily we're not on our own. Therefore I'd like to thank several important people whose input supported me in creating a thesis I can be proud of.

Thank you Esther, for your expertise in design. I am very glad you would spend some time with me to bring my ideas to life.

Caroline, I want to thank you for your devotion to your students and the study program. At every point of my thesis journey, colleagues, experts, and others shared how much they appreciate that the TUDelft has CDI as a program, and were very interested in what I contributed precisely because it's so different from the other beta studies. Though the program will not continue, the time and attention you had for us and the lessons we learned have made us the students, professionals, and people we are today.

Éva thank you for your critical feedback and creative input. The research design was greatly improved because of your expertise on cocreation. And together with Steven, I'd like to thank you as teachers at CDI. Though some may view us students as vessels to be filled with knowledge, you have always supported us in growing as individuals, whose intuition and creativity matter, also in a research environment.

Tina, thank you for joining the project, first as a EJ expert and later as a committee member. Lisa and Floortje, thank you for your earlier guidance and input from the EPA perspective. I hope we can cross paths again sometime.

Also thanks to my external supervisors, Dr. Astrid Martens and Dr. Shruti Setty, for your time, expertise, and guidance at PBL. Fortunately for me, this project was exactly the topic of your interest. I hope you've had fun as well.

Lastly, I want to thank my colleagues and friends for our time together next to my studies for the past 8.5 years. Laura for being interested in my thesis topic and getting me to become her colleague. The 'Hot Chocolate Gang' at EPA for finding each other on day one, and being just infinitely interesting and fun to hang out with. Student association 'De Bolk' and the friends I've made there for being 'prettig gestoord' and a constant source of support in a constantly challenging study landscape.

Cheers! Here's to finishing the second of two master's theses (and hopefully never getting a concussion again).

SUMMARY

CONTEXT

Expert opinions are crucial to quantify unknown variables when it is not feasible to use traditional research methods (limited resources or questions about future values). Implementing structured expert elicitation (SEE) is important for the quality and credibility of assumptions made by these experts or the researchers who enlist them. However, application in practice has been limited.

RESEARCH GAP

When research is performed to project policy effects, the context is so volatile that the accuracy of experts' judgements cannot be validated. Moreover, experts may focus on quantifying uncertainty while ignoring extreme uncertainties (war, future technological inventions), operating from a status quo point of view, in order to support the policy relevance of their research. This leads to a unique context of the expert judgement process. It is unclear how experts address expert judgement in this practice. Moreover, it is not yet known what problems may limit application of SEE protocols in this context.

METHODOLOGY

I worked with the PBL Netherlands Environmental Assessment Agency to evaluate their expert judgement process for the KEV (The Climate and Energy Outlook). On a yearly basis their (sector) experts are required to judge dozens of uncertain variables for the projections of the progress on climate and energy goals. This includes policy effects, external factors like weather, and also scientific uncertainties like the inherent uncertainty of the models that are used for the projections.

How can practitioners in the policy research field be supported in implementing structured expert elicitation to improve their expert judgement?

To answer this research question and address it through communication design, I followed the Double Diamond method for design-based research (Discover-Define-Develop-Deliver). For the duration of a year, I was an intern at PBL, allowing me to observe and discuss their methodology, while inquiring after what problems they face. I helped organise a pilot for implementation of the IDEA protocol for structured expert elicitation. To evaluate this pilot I designed a focus group session with all different actors involved in the KEV project. Next, to discuss challenges and develop possible solutions I organised a co-design workshop with (sector) experts from different teams that work on the KEV.

RESULTS

The results show that experts face several unique challenges in their expert judgement process, which are not directly addressed by SEE protocols. The IDEA protocol shows promise to improve quality of judgements and transparency through increased documentation. However, the protocol lacks support experts say they require to select what uncertainties to

address in their research, as well as adequate support to quantify their uncertainty interval.

Siloed teams lead to diverging methodology and definitions, as well as missing overall problem demarcation. Because there is limited opportunity for feedback on the accuracy of experts' judgements, there is no universal measure of what makes a judgement 'good'.

COMMUNICATION DESIGN

To address the problems faced by the practitioners at PBL, I designed an instructional video. It informs experts about SEE and provides tangible examples and actionable steps the practitioners can take on the short and long term, on the individual and organisational level. This provides the researchers with the knowledge and control to improve their own practice. At the same time, the video also acknowledges the unique challenges the experts face and reassures them of the confidence they can have in their practice to make decisions. By targeting both teamleaders and sector experts, the facilitating party and the practicing party are involved and given agency in this process.

CONCLUSION

Therefore, research requiring expert judgement on policy uncertainties presents some unique challenges in the expert judgement process. When accuracy cannot be used as a measure of quality, it is important that the researchers invest in the credibility of the methodology. SEE can provide structure and transparency, but additional instructions and methods are needed to support experts to confidently make judgements that balance accuracy and information.

Though this case study focuses on a specific organisation and group of experts, the lessons learned from this study indicate a variety of challenges when applying SEE in different research domains and contexts. This could possible indicate the stagnated implementation in practice. Additionally, this research and the development of structured expert elicitation for policy assessment support credibility of expert judgement in this political environment.

SAMENVATTING

CONTEXT

Deskundigenadviezen zijn van cruciaal belang om onbekende variabelen te kwantificeren wanneer het niet haalbaar is om traditionele onderzoeksmethoden toe te passen (bijvoorbeeld vanwege beperkte middelen of onzekerheden over toekomstige waarden). Het toepassen van gestructureerde deskundigenraadpleging (structured expert elicitation, SEE) is belangrijk voor de kwaliteit en geloofwaardigheid van de aannames die door deze deskundigen of de onderzoekers die hen inschakelen, worden gedaan. In de praktijk wordt deze methode echter nog maar beperkt toegepast.

ONDERZOEKS LACUNE

Wanneer onderzoek wordt uitgevoerd om de effecten van beleid te ramen, is de context zo veranderlijk dat de nauwkeurigheid van de beoordelingen van deskundigen niet kan worden geverifieerd. Bovendien kunnen deskundigen zich richten op het kwantificeren van onzekerheid, terwijl ze extreme onzekerheden (oorlog, toekomstige technologische ontwikkelingen) buiten beschouwing laten en uitgaan van een status-quo-perspectief, om de beleidsrelevantie van hun onderzoek te onderbouwen. Dit leidt tot een unieke context voor het proces van deskundigenbeoordeling. Het is onduidelijk hoe deskundigen in de praktijk omgaan met deskundigenbeoordeling. Bovendien is nog niet bekend welke problemen de toepassing van SEE-protocollen in deze context kunnen beperken.

METHODOLOGIE

Ik werk samen met het PBL (Nederlands Milieuagentschap) aan de evaluatie van hun proces voor deskundigenbeoordeling ten behoeve van de KEV (Klimaat- en Energieverkenning). Jaarlijks moeten hun (sector)deskundigen tientallen onzekere variabelen beoordelen voor de ramingen van de voortgang op het gebied van klimaat- en energiedoelstellingen. Dit omvat beleidseffecten, externe factoren zoals het weer, en ook wetenschappelijke onzekerheden, zoals de inherente onzekerheid van de modellen die voor de prognoses worden gebruikt.

Hoe kunnen professionals in het beleidsonderzoek worden ondersteund bij het implementeren van gestructureerde expertbevraging om hun deskundige beoordeling te verbeteren?

Om deze onderzoeksvraag te beantwoorden en een oplossing aan te pakken via communicatieontwerp, volgde ik de Double Diamond-methode voor ontwerpgericht onderzoek (Ontdekken-Definiëren-Ontwikkelen-Leveren). Gedurende een jaar liep ik stage bij PBL, waardoor ik hun methodologie kon observeren en bespreken, terwijl ik onderzocht met welke problemen zij te maken hebben. Ik hielp bij het organiseren van een proefproject voor de implementatie van het IDEA-protocol voor gestructureerde expertbevraging. Om dit proefproject te evalueren, ontwierp ik een focusgroepsessie met alle verschillende actoren die bij het KEV-project betrokken zijn. Vervolgens organiseerde ik, om uitdagin-

gen te bespreken en mogelijke oplossingen te ontwikkelen, een co-designworkshop met (sector)experts uit verschillende teams die aan de KEV werken.

RESULTATEN

Uit de resultaten blijkt dat deskundigen bij hun beoordelingsproces met een aantal specifieke uitdagingen worden geconfronteerd, die niet direct door SEE-protocollen worden aangepakt. Het IDEA-protocol lijkt veelbelovend voor het verbeteren van de kwaliteit van beoordelingen en de transparantie door middel van uitgebreidere documentatie. Het protocol biedt echter onvoldoende ondersteuning om te bepalen welke onzekerheden in hun onderzoek aan de orde moeten komen, noch biedt het voldoende ondersteuning om hun onzekerheidsinterval te kwantificeren.

Gescheiden teams leiden tot uiteenlopende methodologieën en definities, evenals tot een ontbrekende algemene afbakening van het probleem. Omdat er beperkte mogelijkheden zijn voor feedback over de nauwkeurigheid van de beoordelingen van deskundigen, bestaat er geen universele maatstaf voor wat een beoordeling 'goed' maakt.

COMMUNICATIE ONTWERP

Om de problemen aan te pakken waarmee de praktijkbeoefenaars bij PBL te maken hebben, heb ik een instructievideo ontworpen. Deze video informeert deskundigen over SEE en biedt concrete voorbeelden en praktische stappen die de praktijkbeoefenaars op korte en lange termijn kunnen nemen, zowel op individueel als op organisatieniveau. Dit geeft de onderzoekers de kennis en het vertrouwen om hun eigen praktijk te verbeteren. Tegelijkertijd erkent de video ook de unieke uitdagingen waarmee de experts worden geconfronteerd en versterkt deze hun vertrouwen dat zij hebben om beslissingen te nemen. Door de communicatie zowel op teamleiders als op sectorexperts te richten, worden zowel de faciliterende partij als de praktiserende partij bij dit proces betrokken en krijgen zij zeggenschap.

CONCLUSIE

Daarom brengt onderzoek waarbij deskundig oordeel over beleidsmatige onzekerheden vereist is, een aantal unieke uitdagingen met zich mee in het proces van deskundig oordeel. Wanneer nauwkeurigheid niet als maatstaf voor kwaliteit kan dienen, is het van belang dat de onderzoekers de geloofwaardigheid van de methodologie versterken. SEE kan structuur en transparantie bieden, maar er zijn aanvullende richtlijnen en methoden nodig om deskundigen te ondersteunen bij het met vertrouwen vellen van oordelen waarin nauwkeurigheid en informatie met elkaar in evenwicht zijn.

Hoewel deze casestudy zich richt op een specifieke organisatie en groep deskundigen, wijzen de lessen die uit dit onderzoek zijn getrokken op een verscheidenheid aan uitdagingen bij de toepassing van SEE in verschillende onderzoeksdomeinen en contexten. Dit zou de stagnerende implementatie in de praktijk kunnen helpen verklaren. Bovendien ondersteunen dit onderzoek en de ontwikkeling van gestructureerde deskundigenraadpleging voor beleidsbeoordeling de geloofwaardigheid van deskundig oordeel in deze politieke context.

CONTENTS

1

INTRODUCTION

1.1. EXPERT JUDGEMENT TO QUANTIFY UNCERTAINTY

WHAT IS EXPERT JUDGEMENT?

Uncertainty is an inescapable part of research and decision-making. In many cases it is simply unfeasible to set out and collect empirical data to find an answer. This includes situations of limited time and resources to research a specific topic (A. M. Hanea et al., 2022, p.254), but it is certainly true for projections of the future.

> For example, say you are investigating the development of energy usage in the Netherlands from 2030 to 2040 and the impact on CO2 equivalent emissions. You may have to estimate how many households make use of heatpumps by 2040, decreasing their emissions impact. Historic data on the number of heatpumps sold is available. However, it is not feasible to know for certain what the future number of heatpumps is.
>
> Theoretically, this number or 'unknown variable' could be much less and much more than it is currently. There are many sources of uncertainty, like unknown effect of policies (heatpump subsidies), weather developments (particularly warm or cold winters), and technical developments (more efficient heatpumps). This leads to a large uncertainty of what the actual value is.

One way to address this uncertainty in a feasible manner is to consult experts. In academic literature, "expert knowledge", "expert judgement" and "expert opinion" seem to be used interchangeably. They all refer to consulting experts for their best-guess based on their experience. Expertise can be defined broadly based on the relevance of experience to the topic of interest (Krueger et al., 2012). However, this general definition puts the relevance of expertise up for debate. The inclusion of a diverse set of individuals, and more radical democratisation of expertise, can improve the quality of expertise as no one expert has a complete understanding of the world. Moreover, it improves inclusion of less conventional types of expertise (A. M. Hanea et al., 2022; Krueger et al., 2012).

Though there are many problems that require experts to make judgements (e.g. choice of research methods or assumptions for model structure), in this research I focus on expert

1

judgement to *quantify uncertain variables* when there is no feasible way of knowing something through traditional science and statistics (Krueger et al., 2012, p.8). In these instances experts are presented with questions about the true value of an unknown variable.

> For example, to continue the example on heatpumps, researchers could be faced with the question: *"What will be the number of households that make use of a heatpump, in the Netherlands, registered by the end of 2040?"*
>
> The answer will most likely be presented as an interval, *e.g. [20.000, 50.000]*, that represents the 90% most probable values, with the lower and upper 5% excluded from the interval. In some cases a median value, *e.g. 40.000 (the 50% quantile)* can be provided, that represents the best-guess.

In these cases, expert judgement could provide a reasonably accurate alternative that requires much less resources or time compared to an entire study. Consulting experts to provide their opinion happens all the time: in talkshows, news reports, and educational settings, but also to support decision-making by hiring professional consultants. An expert on heatpump or energy consumption may provide a valuable insight to deal with the uncertainty faced when estimating usage of heatpumps in the Netherlands.

WHAT IS EXPERT ELICITATION?

Following a formalized, documented procedure for obtaining and combining expert judgements is called expert elicitation (Colson & Cooke, 2018, p.113). Informal, unstructured protocols are considered unscientific (A. M. Hanea et al., 2022, p.256), for example by making an assumption without documenting the rationale. Formal standards for expert elicitation are the Classical Model (CM) (Colson & Cooke, 2018), IDEA (A. Hanea et al., 2017), and SHELF (Williams et al., 2021). They are grounded in research and practice in decision theory, statistics, social sciences, psychology, and engineering.

The Delphi method by the RAND corporation (Dalkey & Helmer-Hirschberg, 1962) is also a popular strategy for developing group consensus about the value of unknown parameters or various normative or policy issues, but is critiqued: “A Delphi is extremely efficient in obtaining consensus, but this consensus is not based on genuine agreement; rather, it is the result of ... strong group pressure to conformity.” (Woudenberg, as cited in Morgan, 2014, p. 7183)

There seems to be a general structure to the different expert elicitation procedures (A. M. Hanea et al., 2022). Generally, there is an *elicitation team* that facilitates the elicitation. Then there are the *experts*, which ideally is a diverse group of experts with expertise, but the selection can be based on convenience. The facilitation team develops a *list of questions* for the elicitation. Next, either point estimates or subjective distributions are asked from the experts through either probabilities or estimating percentiles. More questions may be asked about the reasoning. This all happens first individually to avoid groupthink and dominating personalities. Depending on the project, this can happen in person, or digitally over a longer timespan.

Most protocols recommend the experts to *practice assessing uncertainty* prior to elicitation to: "(1) reduce experts' apprehension, (2) increase the understanding of the process by which expert judgments are collected and aggregated,(3) motivate the experts,(4) identify common biases and the extent to which experts are predisposed to these, and (5) provide the facilitators and team with guidelines for working with the expert group" (Hodge, Evans, Marshall, Quigley, & Walls, 2001; Keeney & von Winterfeldt, 1991; Revie, Bedford, & Walls, 2011), as cited in A. M. Hanea et al. (2022, p.258). This practice can be done within the domain, or with neutral questions.

After the individual judgements have been collected, many protocols have a *discussion round* for identifying and debating evidence that underpin the judgements. The results can either be mathematically aggregated or a conclusion can be found through consensus. "However, if different experts base their judgments on very different models of the way in which the world works, or if they produce quite different judgments that will be used as the input to a nonlinear model, then combining judgments does not make sense." (Morgan, 2014, p.7183)

Lastly, preferably the *estimations are validated.* This can either happen when data becomes available, or experts are tested on calibration questions that are representative of the target questions.

QUALITY OF EXPERT JUDGEMENT

Structured elicitation protocols intend "to help reduce the influence of biases and to enhance the transparency, accuracy, and defensibility of the resulting judgements" (Hemming et al., 2018). Thorough documentation of the process supports transparency, as well as documenting expert rationales. Accuracy can be measured by collecting empirical data after the fact and applying a statistical test (usually by calculating a P-value). Defensibility can be supported by the scientific quality of the data collection (Colson & Cooke, 2018; Krueger et al., 2012, p.113, p.12). Expert elicitation must therefore be: reproducible; accountable; empirically controllable; neutral (in relation to the conduct of the elicitation process); and fair (equal treatment of experts).

However, the success of expert judgement is often focused on the accuracy. Firstly, the performance of experts and their training is focused on accuracy. For example, the use of test or calibration questions is encouraged to later inform which judgements are weighted as likely to be more accurate (A. Hanea et al., 2017; Hemming et al., 2020). Not only are experts compared on accuracy, but also protocols are compared based on how they influence accuracy scores (Hemming et al., 2020; Nemet et al., 2017; Welsh et al., n.d.).

Another measure of performance, that is less commonly addressed, is information (Colson & Cooke, 2018). The information score is higher when a narrower interval is provided; The information is more densely concentrated. This may therefore conflict with accuracy. The wider the interval, the more likely the true value is captured, however the information score is lower.

1

THE ROLE OF EXPERT JUDGEMENT IN POLICY ASSESSMENT MODELLING

Decision-making, especially in the policy field, requires dealing with uncertainty in complex problems. Decision-makers seek to reduce or tame uncertainty, to more easily make decisions (Raadgever et al., 2011). Presenting and analysing uncertainty lets decision-makers "quantify the confidence in a decision" and "evaluate the risk associated with taking a decision based on the [research] output" (Baustert et al., 2018). Uncertainty can however also be used strategically to delay action or reduce responsibility (Dewulf & Biesbroek, 2018; Moore, n.d.). Expert judgement plays an important role to support analysts and decision-makers, as it can be the only feasible option to gauge future effects of policy.

> To continue the example of the heatpumps, policy-makers could want to gauge how much they should invest in cleaner energy sources. It can be an important point on the political agenda how to do this exactly. The policy-makers present a new subsidy to incentivise households to buy heatpumps, leading to the following question:
>
> *"What will be the additional number of households that make use of a heatpumps because of the subsidy, in the Netherlands, registered by the end of 2040, compared to this year?"*
>
> Answering the question requires experts to extrapolate from incomplete *limited historical data*) or imperfect (*unspecific subsidy plan*) evidence, and account for complex socio-economic (*what type of households use heatpumps*), institutional (*ease of subsidy request*), and behavioural (*is the subsidy enough incentive*) dynamics.

Expert judgement is required to identify relevant causal mechanisms, judge how these mechanisms interact, and make credible assumptions. Likely, there are many other uncertainties that experts need to quantify. Experts cannot completely reduce uncertainty. The uncertainty they express influences the way decision-makers interpret their contribution (Bolhuis, 2025). Because of the impact of scientific uncertainty on decision-making, a key methodological challenge is how to ensure that expert judgement is elicited and incorporated in a credible, transparent, and reproducible manner.

Although expert judgement theory has advanced considerably, its practical uptake in organisations is lagging. Expert elicitation is conducted often in unstructured ad hoc manners with untransparent methods, even when structured methods are available. For example, modelling for covid policy showed clear signs of missing perspectives in elicitation that could be helped by SEE protocols (Burgman et al., 2021).

1.2. KNOWLEDGE GAP AND RESEARCH CONTRIBUTION

SEE protocols could support the quality and credibility of expert judgement in the context of policy assessments. Though the IDEA protocol was developed to help predict geopolitical events (A. Hanea et al., 2017), structured elicitation protocols have mostly been applied in contexts of environmental science.

To estimate the population of a specific species of starfish, an example question is formulated by Hemming et al. (2018):

'What will be the average density of Crown of Thorns Starfish recorded by the Australian Institute of Marine Science during their two-minute manta tow surveys of Rib Reef, on the Great Barrier Reef, in March 2016?' (At the time of that study, 2016 was in the future).

When comparing this question to the heatpump example, we can note some unique differences (not exhaustive):

- The policy domain allows little empirical control. By the time you would want to verify whether the number of households using heatpumps in 2040 was estimated correctly, the context (political will, economic development, etc.) has likely changed so much that the true value in 2040 can be much different than initially estimated. Though there are also many uncertainties when estimating the population survival of starfish, I think less erratic changes can be expected in comparison.
- Practical restrictions like time, resources and available expertise create a distinctly different research context for expert judgement. Whereas one study is focused on the isolated question of the starfish population, which is answered by recruited experts, the heatpump question is one of many estimations that are made by policy researchers.

The unique features of the context of projecting and assessing policy effects warrents a study focused on expert judgement from this different perspective. As adoption of SEE is low, likely it will be new to many groups of researchers in this context. To my knowledge so far, no studies looked into the specific challenges faced in applying SEE methods in this context. However, implementing SEE requires resources, training, and organisational support, and it must compete with established workflows, institutional norms, and time constraints. Furthermore, experts may be unfamiliar with probabilistic elicitation or resistant to formalisation, while facilitators may lack the skills needed to manage group dynamics and mitigate cognitive biases.

These observations highlight a significant research gap: there is limited empirical understanding of how expert-judgement principles are actually applied, or adapted, in policy assessments, the challenges therein. In extension, we can question whether researchers in this context require a different type of support in applying structured expert elicitation.

This study can explore ways to support expert judgement in practice, such as developing context specific instructions. Strengthening these supports will help organisations bridge the gap between theory and practice, ensuring that expert judgement is both methodologically rigorous and operationally feasible.

1

1.3. RESEARCH QUESTION FORMULATION

To address the identified gap, I want to study a case example of policy assessment to examine how expert judgement is mobilised in practice. Such a case can reveal how experts negotiate uncertainty, how elicitation procedures are adapted to organisational contexts, and where tensions arise between theoretical guidance and practical constraints. Consulting practitioners also enables identification of concrete bottlenecks and opportunities for improvement.

How can practitioners in the policy research field be supported in implementing structured expert elicitation to improve their expert judgement?

First, I established and explored the research context. The case I examined is the practice of expert judgement at PBL Netherlands Environmental Assessment Agency, specifically the KEV project (Climate and Energy Outlook). My first sub-question is:

SQ1: What is currently known about expert judgement in theory and in practice at PBL (for the KEV project)?

Next, I wanted to identify what challenges are faced by these practitioners, focussing on the process of expert judgement and perhaps what barriers they experience to implementing SEE. I expected there will be several theoretical and practial challenges. Therefore I used the insights into the organisational values and research context from SQ1 to derive criteria to prioritise what problems should be addressed. This leads to the following subquestion:

SQ2: What challenges are faced by the practitioners in the KEV project in the process of expert judgement? What problems should be prioritised considering the organisational context?

Through communication design, I intended to present a prototype solution to one of the problems identified in SQ2. I examined what could be effective in the context of this case study, and what is feasible in the scope of this research. This informs the last subquestion:

SQ3: What (combination of) solution(s) could effectively and feasibly support the practitioners in the KEV project in the challenges they face in the process of expert judgement?

1.4. SCIENTIFIC AND SOCIETAL RELEVANCE

Expert judgement plays a pivotal role in decision-making, and for policy assessments it is often the most feasible option to quantify important sources of uncertainty. The context and assumptions matter for the interpretation of the results. It's therefore a problem when the assumptions are not transparent or structured, reducing the replicability and scrutability of the research. Research into the practice of expert judgement can help practitioners increase the structure and transparency of their practice. In turn this may lead to increased trust in and understanding of their research.

Though theory on structured expert elicitation (SEE) is quite advanced, it is important to develop tools to support researchers with implementing SEE into practice as well. Policy assessments presents unique challenges for expert judgement. At the same time, the organisational context brings its own structures and limitations that shape the process of expert judgement. Research focused on the implementation of SEE in practice in new research domains can support researchers to tackle challenges for translating theory to practice that are unique to their context. Analysing this proces together with practitioners also allows to explore the diversity of experiences.

1.5. RESEARCH APPROACH AND REPORT OUTLINE

My approach to answering the research questions is based on the double diamond approach to design based research, a method developed by the Council (2004). The method is built on the following phases:

Discover the problem, rather than simply assuming what the problem is: I will detail in Chapter 2 how an internship at PBL allowed me to gain a unique opportunity to obtain insights into the practice of expert judgement from different perspectives.

Define the challenge based on the insights gained from both theory and practice: Together with different actors at PBL, the implementation of SEE in a pilot project was evaluated by means of a focus group in Chapter 3, indicating challenges and pathways for improvement.

Though this design process is conceived as a double diamond, the practice of my research was a more dynamic iteration between the diverging and converging phases. A co-design workshop with several researcers (experts in different domains), provided insight into the practice in different teams, which is presented in Chapter 4. Initially, this would function to develop solutions, but instead it brought new challenges to light. To finally combine the insights from both the focus group and the co-design workshop, a problem essence is defined in Chapter 5 which functioned as the focus for the following steps.

Develop solutions through codesign including a range of perspectives: The research results provided an important perspective from expert judgement practice. Together with insights from SEE and communication theory, a strategy to address the problem essence through was developed with input from practitioners at PBL as well as an expert in SEE. This is presented in Chapter 6

Iterative design finally lead to the *Delivery* of a prototype instructional video in Chapter 6. At different steps the designs were tested whether they effectively address the problem essence.

Chapter 7 concludes with the outcomes of the research and communication design process, together with recommendations for future research. Chapter 8 discusses the reliability and limitations of this research. In Chapter 9 I reflect on my personal experience in this research, as it relates to my journey to a master's degree and development as a professional.

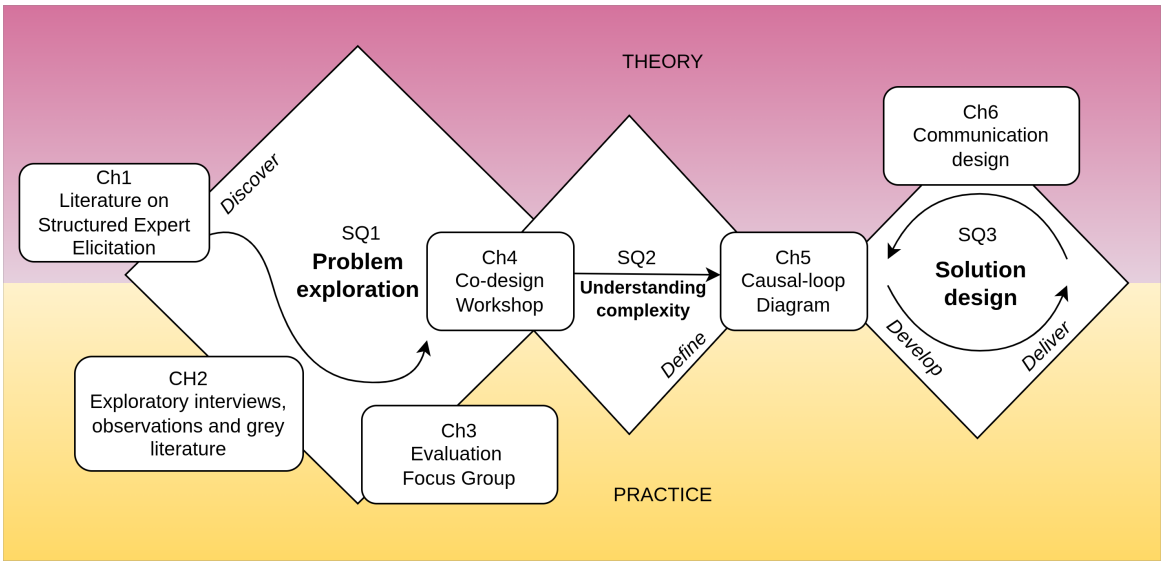


Figure 1.1: Visualisation of the Double Diamond structure of this thesis. The first two diamonds cover the problem analysis section of the research. Where the first attempt at defining a problem failed, I later succeeded at converging to a proble essence, leading to two parts of what was supposed to be one diamond. The last diamond covers the communication design process. I indicated how the research questions and methodological steps relate to the interation between insights from theory and practice.

2

CURRENT THEORY AND PRACTICE OF EXPERT JUDGEMENT

In order to explore the implementation of SEE for research in the context of policy assessment, I contacted PBL Netherlands Environmental Assessment Agency (short: PBL). It is a research institute on the science-policy interface and performs (un)solicited research into environmentally and policy relevant topics (circular economy, energy use, climate impact, biodiversity, and more). This leads me to specify my main research question:

How can experts at PBL be supported in implementing structured expert elicitation in the KEV project to improve their expert judgement process?

I got to work with the researchers at PBL from Februari to December 2025 as part of an internship to research their expert judgement process. This section presents the details of this particular case study, based on publications from PBL, grey literature, and my observations on the expert judgement process.

2.1. METHODOLOGY

At PBL I worked with experts on uncertainty (analysis) who wanted to address the expert judgement process. Most of my work involved facilitating a pilot project with implementing, observing, and evaluating the IDEA protocol in one of the teams that work on the KEV project (the Climate and Energy Outlook, in Dutch it's abbreviated as KEV).

In order to learn about the practice of expert judgement at PBL, I looked into official pulications, internal (grey) literature and documents, and I conversed with different actors that are part of the KEV project. There are few publications on the uncertainty analysis for the KEV (Petersen et al., 2014; Stremler & Boot, 2024; van der Welle et al., 2017) of which some details are outdated, and there are no official publications on the expert judgement process. This internship provided me with access, time, and involvement of participants and experts that provide crucial insights into the actual practice of expert judgement. Moreover, I got into contact with experts at related organisations (RIVM, WUR, TNO, CPB) to inquire after their expert judgement processes.

Based on the literature and observations I will present the organisational and research context of PBL and their KEV project. To understand the expert judgement process I have also developed an overview of the process steps that show the relationship between modelling, expert judgement, and research outcomes. Lastly, I present an actor analysis to relate the research to the organisational structure. I will show how the insights from these steps informed the further research design.

In addition to this research, I also conducted research during the internship into perceptions of uncertainty on the science-policy interface (Bolhuis, 2025). Insights from that study are also incorporated in this report.

2.2. CASE STUDY: ENVIRONMENTAL ASSESSMENT AGENCY

POLICY ASSESSMENT

In the climate law of the Netherlands, it is established that regular monitoring of climate policy is obligatory. PBL reports on the achieved and expected effects of climate and energy policy in relation to the european, national, and domain specific goals (Stremler & Boot, 2024). The main national climate goals are stated in the Dutch law, requiring a minimal reduction of greenhouse gas emissions of 55% compared to 1990 levels by 2030.

To fulfil their task of assessing the climate and energy policy effects, PBL publishes projections based on a snapshot of the current Dutch national policies every year: The Climate and Energy Outlook ("Klimaat- en Energieverkenning" in Dutch) (short: KEV) (PBL et al., 2024). The expected effects of those plans are modelled for several policy domains as defined in the agreements with the government in the Climate Law (Electricity, Industry, Built environment, Mobility, Agriculture, and Land use). The 'sector teams' that work on the KEV are based on these policy domains.

From modelling the energy use and production, some key indicators for the national and sector energy and climate goals are reported, amongst which are the greenhousegas emissions and the share of renewable energy. The national goal and emissions forecast of KEV 2024 can be seen in Figure 2.1. Similar figures are generated for the relevant policy domains that were listed earlier.

The projection that is reported shows a calculated point-value every year, and an uncertainty bandwidth in reference years (here 2030 and 2035) which is the result of an adapted Monte-Carlo analysis: The variation in model result is simulated for 1000 random combinations of variations across all the uncertain factors, which are varied within their own respective uncertainty intervals.

MODELLING PROCESS

A yearly inventory is taken of all Dutch national policies. all policy is gathered and their effects modelled to project the impact on the key indicators for energy and climate goals. There are few official documents on this process (Stremler & Boot, 2024; van der Welle et al., 2017). Therefore the following information is largely based on conversations and observations during my internship at PBL.

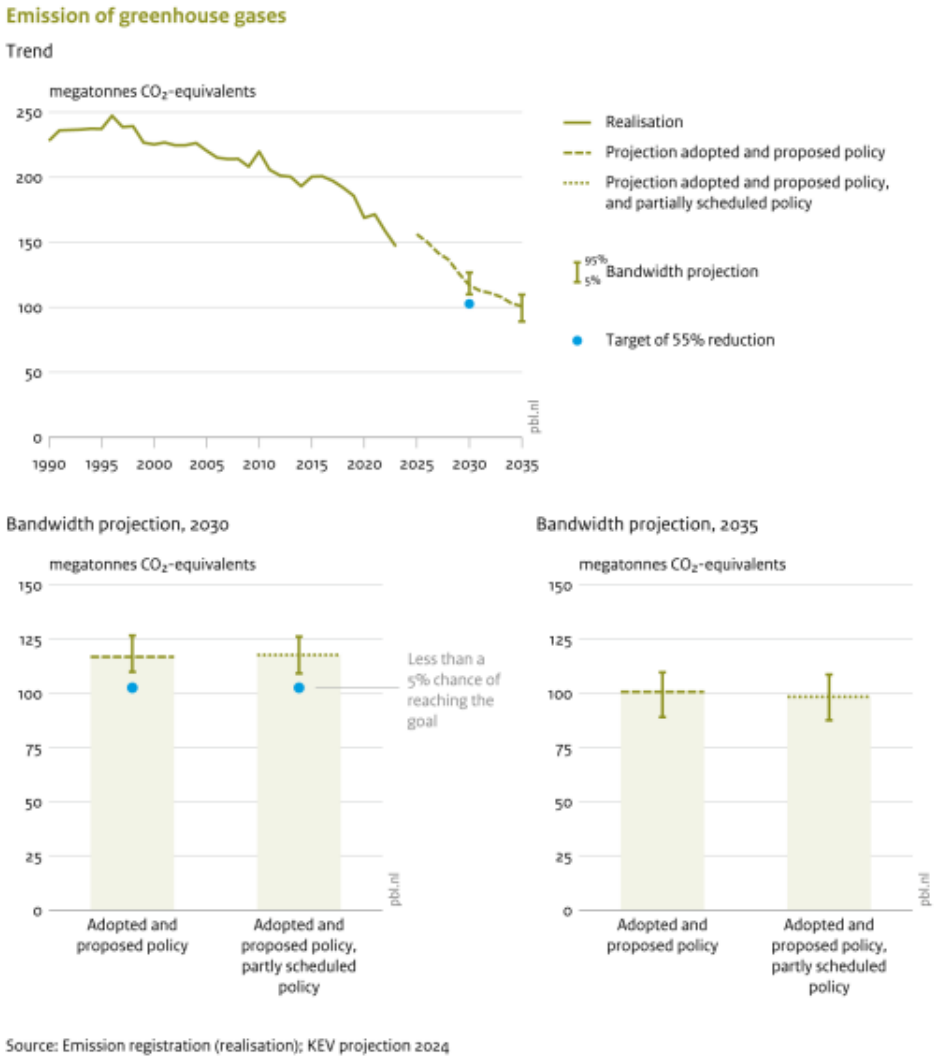


Figure 2.1: KEV 2024, pbl. National emissions forecast.

For the energy production and use across these domains, PBL uses around 10 different models that model everything from industry and agriculture energy use to emissions from vehicles. Some of these models are maintained at other organisations and require collaborations with researchers at WUR and RIVM. Some models currently used started development in the organisations preceding PBL. Through mergers and new mandates PBL was formed in 2008, and their models are still improved and adapted today.

The relevant policy officers of each policy domain are interviewed on the biggest changes. Many policies reoccur every year, and projecting their effects is therefore already baked into the modelling process. However, every year also new policies can be added, or old ones can be removed and new influences that lead to uncertainty can be taken into account.

The effects of these policies on the energy demand and supply are estimated and modelled for several years, leading to the best estimate and an uncertainty bandwidth. This requires model input on energy- and CO2 pricing; economic, demographic, and technological developments; weather; and relevant foreign policy. Statistics are largely derived

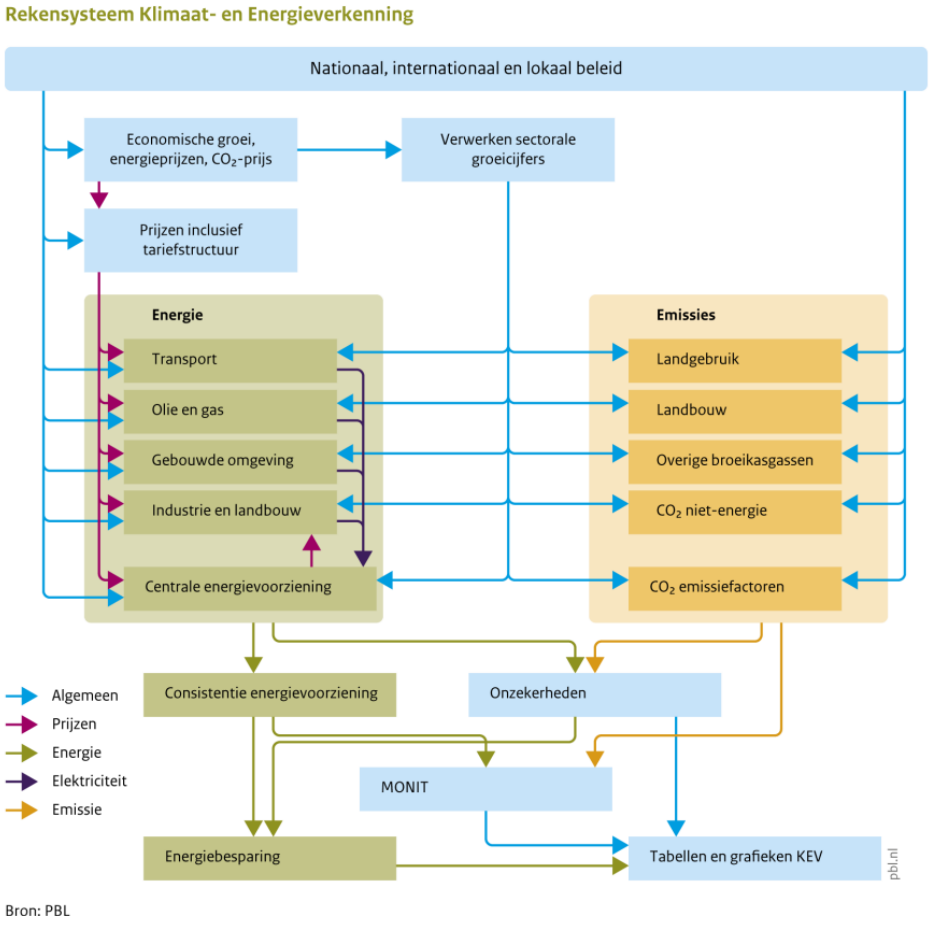


Figure 2.2: Illustration of the data sources and modelling sequence for the KEV (previously NEV). Source: Stremler and Boot (2024).

from CBS, and are consistent with other PBL studies. The researchers must judge the input on quality and relevance, and in case of insufficient data or in case of multiple scenarios, researchers must make assumptions based on the project demarcation.

During meetings of the sector teams, I observed that newer policies that are delivered by the ministries do not always provide a detailed execution plan and precise expected effects. Sometimes a general positive (+) or negative (-) effect on CO2 equivalent emissions is given, and other times a specific amount of expected decrease in Megatonnes is indicated. However, this does not show what mechanism leads to the effect. This requires back-and-forth between the researchers and the relevant policy officers to specify the intended policy effects.

Once all model input (parameters, policies, external factors) are gathered, the energy use is modelled (across several models for different sectors), after which the energy production is modelled to match the demand (also across several models for different sectors). There are additional models that support this supply-demand balance and the consistency across the different models, as can be seen in Figure 2.2.

Next, an uncertainty bandwidth is modeled for the projections in one or several tar-

get years. This uncertainty bandwidth is informed by variation in expected policy effects, varying expected demography changes, model uncertainty, and the effects of any relevant external factors that lead to uncertainty in the projection. In addition, the bandwidth also covers any new only partly scheduled policies or less developed policies that are submitted during a later stage of the project.

Important to indicate is that the projection and the uncertainty bandwidth are limited to a 90% interval, leaving out the top and bottom outliers of the projections. The presented results are the most plausible projections, thus leaving room for a slim possibility of the true value lying outside of this bandwidth. In addition, the projection is limited to a relatively normal policy future: when it comes judging for example what external factors are relevant, extreme uncertainties and changes are considered out-of-scope for this project (van der Welle et al., 2017). Factors like future policy development, changing monitoring practice and definitions, extreme (not necessarily unlikely) events like war and disasters, and game-changing technologies are not taken into account.

INCORPORATING EXPERT JUDGEMENT FOR UNCERTAINTY ANALYSIS

There is no official documentation available on the role of expert judgement in the KEV, but the relation between the modelling and expert judgement steps in this project is sketched in Figure 2.3.

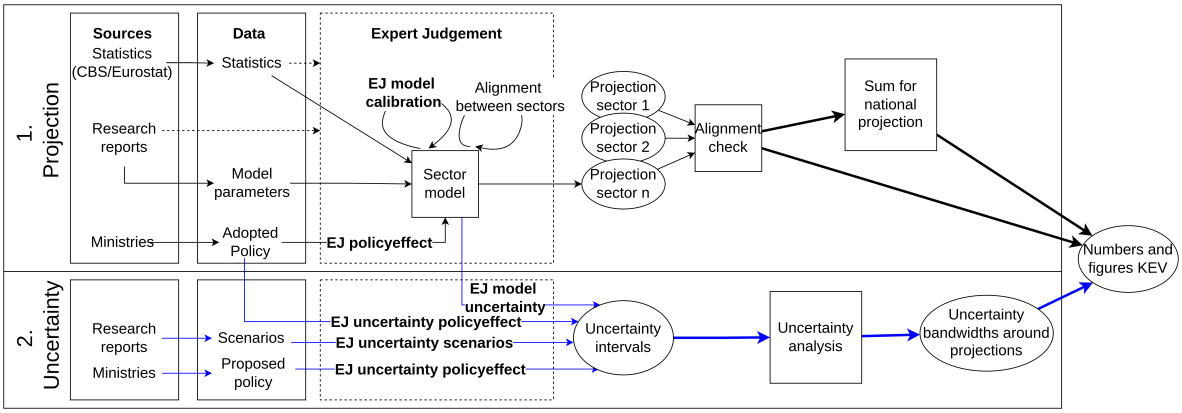


Figure 2.3: Illustrative representation of the KEV modelling process including indications of use of expert judgement. Estimating uncertainty intervals happens at a later stage in modelling than the central projection. Though Expert judgement is often needed for modelling, experts can also use models to support their expert judgement.

The uncertainty analysis is helmed by an uncertainty analysis team. They gather input from all sector experts on what uncertain factors are relevant, and what uncertainty can be expected in the (accepted and planned) policy effects. The sector experts are required to estimate the middle 90% most plausible values of these uncertain factors and policy effects (the 5th and 95th percentiles are elicited). Specifically, they must indicate the effect in target years, requiring assumptions about how quick policies have effect.

Sector experts approach judging these uncertainty bandwidths in two ways: 1) some bandwidths can be calculated using the models they have access to. 2) Other judgements

must be made based solely on their own expert judgement.

For example estimating the effect of a new subsidy for heatpumps requires looking into relevant studies on the use of these pumps, the expected machine life, behaviour of consumers, monitoring practice of correct policy implementation, and many more underlying mechanisms that impact the effect of policy.

The expertise of the researchers plays an important role, even when data (statistics, other independent studies, reports from industry or other sector relevant data) is available because the quality of the data needs to be judged. They have their own research as experience, but also stay in contact with policy officers, industry representatives and other experts from the relevant domains.

2.3. EXPERT JUDGEMENT PROCESS AT PBL

2.3.1. OBSERVATIONS CURRENT PRACTICE

In all the domains of the KEV, there are relevant unknown variables that influence the model outcomes. Deciding which factors are relevant (do we take into account foreign policy?) and quantifying these unknowns (how many windmills will have been built by 2035?) fall to the expert judgement of the sector experts. These experts have learned their own customs based on their own studies, experiences and values.

Organisations and teams work on various different projects and therefore develop different ways of approaching the task of quantifying unknown variables. I got to shortly discuss with some professionals at CPB, RIVM, and PBL. These conversations show that different settings also require different expert judgement processes, however these are not often formalised. Therefore, formalising the expert judgement process is not as easy as taking one protocol and standardising it, each domain and project presents its own limitations of available time and expertise and requirements on the outcomes.

The projects and organisations aim to present independent, credible, salient and legitimate research (Martens et al., 2021). This in some ways overlaps with views from expert judgement theory, where the scientific standard is persued (Colson & Cooke, 2018; A. M. Hanea et al., 2022; Krueger et al., 2012; O'Hagan, 2019). However, these seem to miss some practical values, that became apparent in the conversations I had: There are limited resources, diverse expert selection is not always possible, the protocol needs to work with the model workflow, and there is a wish to support sharing and growing knowledge within the organisation.

2.3.2. MOTIVATION FOR IMPROVING THE EXPERT JUDGEMENT PROCESS

Though the modelling process for the KEV (and previously the NEV) has been developed and improved over more than a decade, the current approach to expert judgement for the KEV could still be improved. It is sometimes experienced as unclear (as different expert teams have grown to approach this task in different ways), untransparent (as experts may understand their own process, but it may not be clear to others), and unstructured (as there is no structured protocol and little documentation that protects the reproducibility of the

process).

To continue to improve the methodology of their policy assessment, PBL formed a team that would run a pilot on improving expert judgement. This team, existing of researchers from the uncertainty team and the methodologists at PBL, would develop a proposal for an expert judgement protocol. They chose to base their protocol on the IDEA protocol (which will be discussed later). The intention is to improve the quality of the expert judgement process, making is more structured and transparent, but also increasing sharing knowledge between sector experts and reducing dependence on a small number of experts.

2.3.3. ACTORS

The main actors in this project are identified and presented in Figure 2.4.

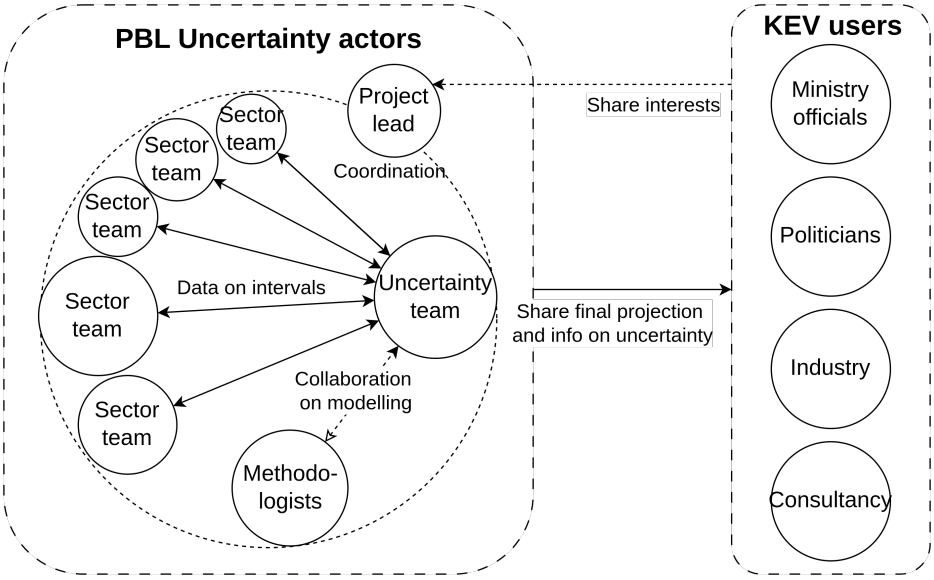


Figure 2.4: Actor map based on observations.

- Sector experts - There are 6 different policy domains (sectors) that each have a dedicated team with modellers and experts (Electricity, Industry, Built environment, Mobility, Agriculture and Land use). The sector experts perform expert judgement on the domain they have expertise in;
- Uncertainty team - They perform uncertainty analysis and require input from sector experts;
- Methodologists - Research and develop methodology at PBL. They work with all sectors;
- Project lead - Coordinates collaboration between the other actors and responsible for the final product;
- KEV users - These actors are interested in the KEV research results and reasoning. Examples are relevant ministries, political actors, but also decision-makers in the industry, consultants, and lastly the general public. The information interests of these

actors shape the relevance of the project and should therefore be considered by the project lead.

Initial conversations with these actors indicate a broader view on challenges and improvements in the expert judgement process:

Across different sectors, working methods vary, resulting in inconsistent levels of quality and transparency. Implementing a structured protocol could promote standardization and clarity. Currently, there is limited documentation of the considerations and assumptions underlying expert judgments. While it is assumed that experts work thoroughly, the risk of knowledge loss—due to illness, retirement, or turnover—remains a concern.

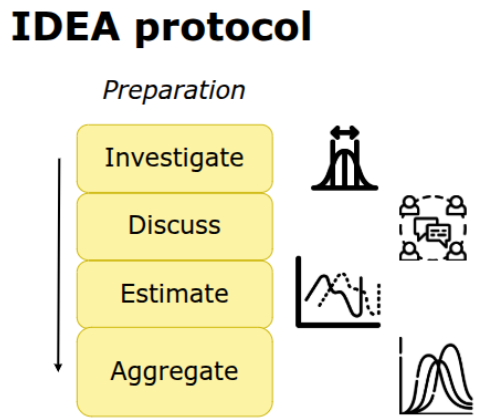
Experts are limited (in number and diversity). Would it be possible to incorporate a broader range of expertise and perspectives into the process? Finally, under time pressure, quality can be compromised, so any new protocol should ideally both save time and enhance the quality of outcomes.

For those reviewing expert judgments, it is often unclear what falls outside the specified bandwidth and whether the chosen percentiles (such as the 5th and 95th) are accurately applied or if they deviate significantly (Bolhuis, 2025). Judging minimum and maximum values might be more straightforward.

2.3.4. A PILOT WITH THE IDEA PROTOCOL FOR EXPERT JUDGEMENT

IDEA PROTOCOL

The IDEA protocol, and its supplementary guides (Hemming et al., 2018), is one of the existing protocols for Structured Expert Exlicitation (SEE). It was initially developed for predicting the outcomes of political events. The goal of the protocol is to improve accuracy of judgements and to increase transparency. It is meant to be applied to a few questions with a group of around 6-12 experts (Hemming et al., 2018). The protocol outlines what steps facilitators of expert elicitation should follow.



The facilitators compile background information on the problem and formulate relevant specific questions for the experts. They contact and brief experts on the elicitation process. Next, experts individually answer the questions and provide reasons for their judgements. A discussion is held where the individual answers are anonymously shared. This is followed by a 2nd round of final and private estimates that allow the experts to incorporate any new insights. Finally, the responses are aggregated. This last step is done mathematically based on experts' previous accuracy, though at PBL consensus is preferred.

The questions formulated by the facilitators will be answered in four steps. Only after considering the bounds will the experts make a best-guess. Lastly, experts are required to consider their confidence, which could help them iterate on their ealier answers.

What will be the additional number of households with heatpumps in 2035 compared to now because of the new subsidy?
Realistically, what do you think the lowest plausible value will be?
Realistically, what do you think the highest plausible value will be?
Realistically, what is your best guess?
How confident are you that your interval, from lowest to highest, could capture the true value? (Answer between 50% and 100%)

PILOT

Four experts from one sector team were participants to this pilot project (specific sector will not be mentioned to protect anonimity). Though ideally their whole process would be supplemented with the IDEA protocol, in practice the meetings for this protocol happened in parallel to their usual workflow. Before the meeting they identified together the list of uncertainties, which at the time consisted of the effect of 8 proposed policies. Next, all participants were required to quantify the uncertainties and document their decisions.

Usually the sector team would have responsible experts who would discuss their findings during their weekly meetings where other experts would check the work. The team and the variables they estimate are similar year after year. Having multiple researchers per variable, the explicit instructions to specify their considerations in more detail, and explicitly note their confidence in their own judgement were new to the participants. This led to difference in interpretation, questions and uncertainty. This process showed that though other experts can check others' work, a different level of engagement is required when they themselves have to make a separate individual judgement.

The group discussion where the individual judgements were discussed also did not go without challenges. The list of variables to discuss were formulated in simple terms, and not specific questions. Though all experts had prediscussed these variables, their interpretation of the questions were different. This led to answers in different units (number of houses, amount of CO2) and styles (ratio, absolute growth). This process showed where usually outsiders are uncertain how to interpret the contributions of the sector experts, even the experts themselves can differ in their interpretations.

Lastly, a common point of discussion was if the policy they are estimating the effect of was formulated specifically enough to judge. If it is too interpretable, the experts can't make a judgement without making political assumptions. It is important that a good dialogue is held with policymakers at the ministry to formulate the specifics of the policy.

Because of the extended discussion on the interpretation of the question and eachother's reasoning, it was no surprise that there was little discussion about the exact boundaries of the interval, though I expected that to be discussed. The sector experts finally decided that they would at a later point finalise their estimates, but this discussion was valuable input.

3

EVALUATION OF EXPERT JUDGEMENT PRACTICE

The experience and challenge of expert judgement (EJ) in practice can be contrasted with the theoretical principles. We've established the current approach to expert judgement in the KEV project. The current intention is to improve the documentation and decrease reliability on specific experts. In the context of the double-diamond, I'm looking to diverge on possible challenges that are faced by the teams in the expert judgement process.

3.1. METHODOLOGY

APPROACH

In order to explore challenges from a wide variety of perspectives, I wanted to evaluate the pilot project with representatives of all different actors 2.4. I designed a focus group to gather in-depth insights, where the actors can also compare and contrast their perspectives, allowing the evaluation to involve different values (Babbie, 2016). A secondary goal was to ideate solutions to improve the expert judgement process.

Evaluation criteria will be based on theory, organisational values, and participants' experiences. Participants will then discuss their experiences, evaluate (what is going well and what should change), and ideate solutions (ideally what should the process look like).

Because of the nature of the project there are restrictions to the method of evaluation: the participants have little time so I limited the meeting to 1 hour. Because the meeting should be easy to attend, it was a digital meeting so people working from home can join. Actor groups consist of multiple people. To have a large chance of participation I invited several individuals from each group, which meant the meeting would need to allow a large group of participants. To work around these restrictions, I checked with the participants pre- and post-participation what their perspective is and whether I presented it accurately.

The focus group session was conducted on June 25th 2025 online via MSTeams (Microsoft, 2020). The session lasted approximately 60 minutes. The meeting was not transcribed, but notes were taken by me and the contributions of participants were collected on the digital whiteboard platform Miro (Miro, 2011).

EVALUATION CRITERIA

To analyse the evaluation, I want to know what the actors actually find important. This can help indicate what problems and solutions should be prioritised. From the exploration of the theory and practice of expert judgement, 10 different concepts are highlighted to examine what trade-offs and challenges are faced in expert judgement at PBL. These are used as a basis for the questions during the focus group.

Standardisation	Independent judgements	Proportional effort
Transparency (external)	Realistic judgements	Transparency (internal)
Scientific basis	Priority for expertise	
Inclusivity	Low workpressure	

SESSION GUIDELINES

I designed a protocol for the focus group for me to guide the participants through discussing and evaluating their experience with the new expert judgement protocol. On the one hand there are several themes that should be discussed. On the other hand, having a more loose structure allows the participants to discuss and also add previously unconsidered themes.

The questions were based on preliminary research into expert judgement protocols and consultations with the participants. The goal was to get the participants to think about their experience individually beforehand so that they can discuss in a shorter amount of time. Participants were stimulated to note any thoughts on a digital whiteboard. The intention is for the participants to more easily share contributions during an online meeting.

What are the experiences with and feelings about the new protocol? (10 min)

- How much time did the protocol take?
- How was the experience with the process?
- How did you experience the outcomes?

After the initial discussion of the experiences, I would like to bring all actors to focus on what they find important for expert elicitation. Therefore before moving on to the evaluation, the participants discuss the most important values for the protocol design.

Values (10 min) An initial list of values was compiled from theory and through personal conversations and presented as a start of the discussion. At the end of the discussion there should ideally be a list of about 10 values. To avoid unnecessary discussion time the list needs to be clearly interpretable in one go. I will proceed to ask all participants to score the values by distributing 100 points.

The outcome of this distribution will support the actors to reflect on their values in rela-

tion to the others' because they are now presented in a complete list, but will also allow the researchers to discuss any potential trade-offs that have to be made or are controversial. These will be collected for everyone to see. This exercise allows to include everyone's values, also people who are less likely to speak up.

At this point we want to dive into the evaluation of the IDEA protocol. Therefore we need to ask the participants to reflect on their current experience. I will prompt with these questions for the participants to discuss. They can add on the Miro board their own ideas for what can be improved.

Evaluation discussion (10 min)

- What parts worked?
- What parts could be improved?

Next I would like to make space to look outside of the current experiences and imagine what the protocol could look like unrestricted by the current experiences. The following questions are posed to the participants to consider or to further prompt participants to speak up:

Design requirements (20 min)

- Ideally, how would you personally like to work?
- Ideally, what would the collaboration with others look like?
- What (support) is required to really make this work?

This section concludes with a list of design requirements, listing those that have consensus from the group but also those that may not be widely supported. These will also be collected on the Miro board.

MODERATION AND FACILITATION

The session was moderated by the primary researcher, who has some previous experience with social scientific interviews. The moderator ensured that all participants had the opportunity to speak and that the discussion remained focused on the research objectives. They also kept the time to ensure transitions to new topics.

PARTICIPANTS

The focus group consisted of 13 participants, reflecting the diversity in the relevant actor groups (Figure 2.4): sector experts (from the pilot detailed in Chapter 2), uncertainty team, project lead and coordinators, and methodologists. The participants were selected through a purposive sample with the following selection criterion: the participants must be involved in the pilot of expert judgement or must be responsible for the expert judgement process and project results.

The initial sample of participants was recruited through available contacts the researcher had access to. The intended number of participants requires balancing between what a fea-

Figure 3.1: Participants could contribute by using these post-its on the online whiteboard. Each actor group had their own colour assigned. The names for the individuals have been erased to keep them anonymous.

sible number of participants is for a focus group, and the diversity in perspectives: Ideally, at least two actors per group would participate (so 8 total), but in the end more participated. To identify the contributions of the individuals, they were assigned their own colour (Figure 3.1).

HUMAN RESOURCE ETHICS AND RISKS

I applied for ethical approval at the human resource ethics committee (HREC) at TU Delft (Application number: 5520). To comply with human resource ethics guidelines, potential participants were fully informed about the study's purpose, methods, and their rights regarding withdrawal from the study. A copy of the informed consent form is attached in Appendix A. In that form, participants are also made aware of potential risks of participating. Measures were taken to minimise these risks:

- Risk of re-identification: Quotes are only presented anonymised.
- Risk of reputational damage: Anonymisation of contributions which limits re-identification should mitigate risk of reputational damage. The written materials used in this research are shared with the participants so they have insight into what data is used for the research.
- Risk of data breach and re-identification: Personal identifiable information is stored securely and deleted before the end of the research.

DATA COLLECTION

The participants were asked to note their contributions on a digital whiteboard. The participants were prompted to take time to write out their ideas at the start of each topic, before they would discuss, ensuring that each actor group's perspective would be collected. However, the participants were free to adjust and change their contributions during the meeting.

DATA ANALYSIS

The written contributions were analysed based on common themes. Analysing the contributions per theme allows to identify emerging themes and compare different perspectives under each theme.

RELIABILITY AND VALIDITY

The participants could write out their contributions on their own. Time was allocated per question, but participants were allowed to double-back and adjust their contributions. The digital platform allowed less outspoken participants to contribute. In contrast, more outspoken individuals were reminded to also document their answers so they can be seen later.

The evaluation meeting was not tested beforehand, though a stress test of the digital platform was performed to rule out any technical challenges during the focus group.

3.2. EVALUATION RESULTS

During the evaluation meeting, several key topics were discussed to address important aspects of the expert elicitation protocol and its implementation. The contributions of the participants are presented in Table 3.1 and Figures 3.2 and 3.3 (originals in Appendix B).

WHAT DO WE FIND IMPORTANT?

A large theme during the discussion was practical requirements like time and other resource allocations. A new way of working shouldn't increase the requirements. The participants discussed that a structured protocol could potentially save time. Many values that underline credibility were valued, like the need for a scientific protocol, transparency, required 4-eye principle. If anything should be focused on, it is the standard of credibility of the work they do.

The participants discussed that an expert judgement process needs to be doable, but also credible, and that given these different pressures the question arises: given the limited time and resources, can they keep up with a high standard of credibility? Additionally the complicated relationship between trust and transparency was discussed: does transparency about uncertainty and assumptions actually increase trust? No conclusion was arrived at.

EVALUATION OF THE EXPERT ELICITATION PROTOCOL

The pilot project mainly required participants to document and discuss their assumptions and considerations more extensively, first individually and next in a group.

Figure 3.2: Contributions from the participants in the focus group. The illustration is a translation of the contributions that can be found in Appendix B.

Table 3.1: Participant responses to the question: "What do we find important?". The values were derived from previous interviews and literature. Participants were asked to distribute 100 points over the values. Each column is the response of one participant.

Value	Sub-topic														SUM
Credibility	Standardising definitions and process	10	20	20	20	30	40	10			15	15	25	20	225
Credibility	Transparancy in external communication	10	20	30	30	20		10	25	20	15	15	15	10	220
Credibility	Scientific standard	10		10	10	10		15	15	10	10	15	20	15	140
Legitimacy	Inclusivity	5		10				10			5				30
Independence	4-eye principle	15	20	20	30	10	30	15	25	20	15	10	17	10	237
Relevance	alignment of assumptions with practice	15	20					10	20		5	20		30	116
Workable	Judgements based on the right expertise	10	20			10	30	9			15	5			99
Workable	Low workpressure	5						5	15	10	5	10		15	70
Workable	Proportionality	5		10	10	10		6		40	7		3		91
Workable	Transparency in internal communication	15				10		10			7	10	20		72

The participants emphasized the value of enhanced documentation and an improved quality of discussions: new insights were gained. However, the question was posed to what extent this may improve results. An idea was offered that this protocol should be applied to the most impactful assumptions and not as extensively to smaller uncertainties to limit additional workload.

An important discussion was conducted on the confidence in quantifying uncertainties. Experts may feel that they are showing false certainty by specifying specifically 5 and 95 percentiles and assuming probability distributions. It is specifically the lack of data on these aspects that requires expert judgement, but which complicates the practice.

Lastly, it was considered important to view how such a protocol interacts with the planning over the whole research project as there are several moments of expert elicitation and the steps of the process could be distributed.

Participants noted it is key for all experts themselves to support the new protocol and this should be uniformly applied across teams. Perhaps an early introduction of the first steps of the protocol can help. No specific comments were raised regarding dependencies between different sectors.

An idea is offered to have experts estimate 0-100% instead of 5-95 as they may feel more confident in their assumptions. Other aspects that may help is additional examples and instructions for comprehensible and correct judgements and documentation. Lastly, it is offered to consider also more qualitative ways of communication assumptions and uncer-

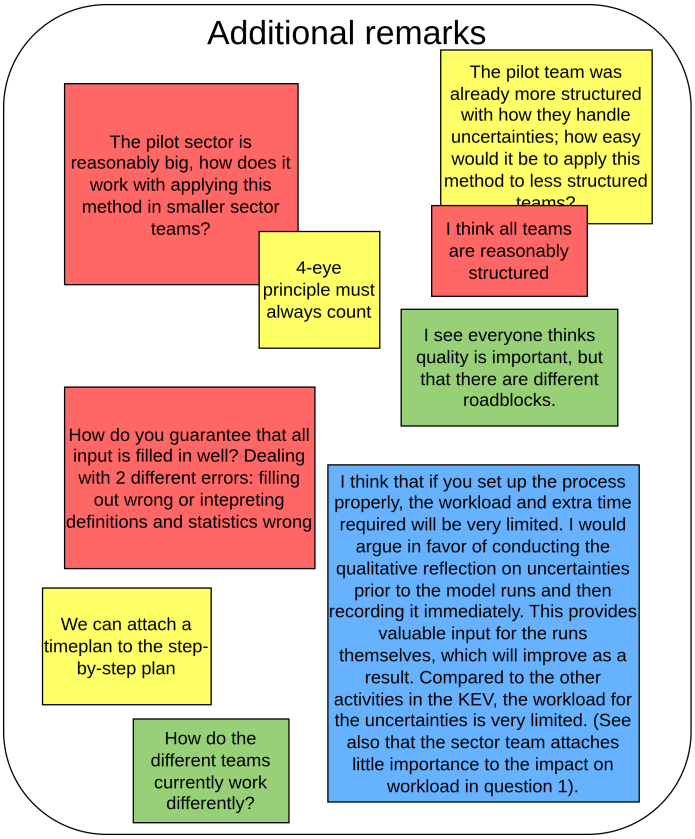


Figure 3.3: Additional remarks made by the participants in the focus group separate from the main discussion points. The illustration is a translation of the contributions that can be found in Appendix B.

tainties next to the quantitative model assumptions.

Participants agreed on the necessity for universal team support while recognizing the need for sector-specific adaptations within each team or sector. This nuanced approach ensures broader support while allowing for tailored adjustments to meet specific sector requirements.

3.3. DISCUSSION OF FOCUS GROUP RESULTS

Several themes are identified based on the participants' contributions. They are overarching themes across the topics that were discussed during the evaluation, showing that some themes are relevant in both the discussions of the problems and solutions.

METHODOLOGICAL RIGOR AND FEASIBILITY

Chapter 1 denoted that protocols are often evaluated on their impact on transparency, defensibility, and accuracy. The results from this evaluation show that other and mostly practical restrictions play an important role in evaluating the IDEA protocol in this case study.

Sector experts and project leads highlight practical challenges in integrating the new protocol into existing workflows, emphasizing the need for flexibility for different teams. They also express concern about increased workload, and call for workload measurement.

Additionally, widespread acceptance among all experts and actors is essential. While methodologists and the uncertainty team support the protocol's improved scientific basis and standardization, they acknowledge the practical obstacles but advocate for maintaining scientific standards.

Part of easy implementation is clear instructions and planning. A. Hanea et al. (2017) provide extensive instructions for implementing the IDEA protocol. Not only do they provide a written guide, a digital platform "IDEAcology" was developed (Courtney Jones et al., 2023) which can be used for elicitation. The pilot has shown these do not translate one-to-one in the research setting for the KEV. Adapting the method and instructions to this context takes time and dedication, also from teamleaders who have to learn to lead the elicitation process.

Time investment is for many a barrier to implementation (Courtney Jones et al., 2023). Though the starting up cost may be higher, time can be saved through efficient selection of which questions require more extensive elicitation and which do not. Furthermore, as one of the objections to the current method is that time is needed for back-and-forth discussion and checking between teams, time can also be saved in these aspects.

As practical restrictions play an important role in the ease and readiness to implement, it is understandable the project coordinators seek to understand what a proportional amount of resources is to invest in improving the expert judgement. However, I think this also highly depends on what is considered 'good expert judgement'. Sticking to a scientifically researched protocol underlines the methodological rigor, especially as studies show the IDEA protocol can improve accuracy. But in this case study, the accuracy of judgements cannot be verified. An important measure for the quality of the expert judgement could therefore be the process credibility, underlined by for example reproducibility or the steps taken to mitigate common biases.

PROTOCOL STEPS ARE AN IMPROVEMENT BUT INSUFFICIENT

The implementation of structured elicitation protocols are, to my knowledge so far, not studied in the context of modelling for policy assessment. These results provide a new perspective, indicating what may limit the applicability in this case study.

The protocol increases transparency, but clear instructions for practical implementation are still needed. Sector experts note that the IDEA protocol pilot does not cover the full expert judgement process. They want steps added for listing all factors to be estimated and determining which require expert judgement.

Extensive instructions for all practical matters (formulating questions, selecting experts, making a planning) were developed for the IDEA protocol(Hemming et al., 2018) to overcome such practical barriers such as the lack of instructions for implementation. We observed during the pilot that the focus lay on implementing the individual judgement and group discussion. Therefore, it is understandable the teammembers feel that not all necessary steps were covered. It is therefore important for future pilots that the full process is attended.

However, even considering the limitations of the pilot, it is interesting to note that there is a distinct difference in the contexts. The IDEA protocol is developed to support facilitators and experts who must address the uncertainty of a pre-defined problem. However, the sector experts in the KEV project also decide what uncertainty to address: together with the project leads they define the problem and what uncertain factors influence it. This gives them direct control over the list of uncertain variables they must estimate, in turn requiring a different type of expert judgement: "What sources of uncertainty are relevant?". To my knowledge this is not directly addressed in the IDEA protocol, but perhaps the criteria for scientific data listed in Chapter 1 (reproducibility, validity, etc.) can provide inspiration for instructions.

INDIVIDUAL ESTIMATES

Though the pilot indicates the elicitation process is more structured and transparent, the experts experience little support to improve the actual expert judgement step.

Both sector experts and the uncertainty team call for more detailed, sector-specific instructions and examples on expert judgement. The uncertainty and methodology teams stress the need for clearer implementation guidance to achieve the intended input. A major challenge is eliciting 5–95% confidence intervals, as experts prefer 0–100% ranges, which risks producing absolute rather than probabilistic intervals. Methodologists attribute this to experts' lack of confidence in precise estimation. Another issue is balancing confidence and false certainty: while methodologists encourage precision, some experts fear projecting unwarranted certainty, highlighting the need for reflection on uncertainty.

Elicitation protocols take different approaches to eliciting individual judgements. Both the quantiles (5-50-95 or 25-50-75) and methods of elicitation (iteratively approaching the best guess or direct elicitation) can differ (MOLE (Welsh et al., n.d.), IDEA(A. Hanea et al., 2017), Shelf (Williams et al., 2021)). Regardless of what approach is used, experts can be trained in assessing uncertainties and describing probabilities to gain experience and improve accuracy and validity, which most protocols recommend (A. M. Hanea et al., 2022). It could also reduce experts' apprehension and help them identify common biases.

Learning to do expert judgement through training is generally supported with a training set of questions of which the true values are known, or empirical data is collected at a later moment to validate the experts' judgements and score their performance. This type of feedback is however not feasible in the context of projecting policy effects. For example, in the heatpump example used earlier, new policies can be introduced in the mean time, geopolitical changes can have rapid influence on energy pricing, and human behaviour may not be as straightforward as expected (perhaps the subsidies are hard to request). This volatile context therefore limits the available feedback experts can receive.

Perhaps the limited training and feedback can help explain the feeling of false certainty some experts experience. They are required to provide a best-guess, but can never validate the accuracy of their judgement. Could this be a case of under-confidence? To my knowledge underconfidence and insecurity in experts is not researched. It seems these concepts may cover something not captured in the 'uncertainty' that the experts must quantify. Additionally, protocols are designed to mitigate overconfidence as this is a common pitfall of

experts, but these design choices have not been tested on whether they also address 'underconfidence'.

IMPROVED COLLECTIVE LEARNING

New insights and the involvement of new experts are positive outcomes of the pilot. The pilot provided valuable new insights into expert judgement and offered other actors a window into the sector experts' work. Experts appreciated the process of individual judgement followed by discussion. Methodologists see potential in reducing reliance on single experts, as structured elicitation and clear instructions could help integrate new colleagues more easily.

Not every SEE protocol includes a group discussion phase (A. M. Hanea et al., 2022; O'Hagan, 2019). The experts' judgements must be combined, which can be done mathematically or through consensus. Mathematical aggregation is a quick and often used method to combine the experts' judgements. Currently discussion is part of the pilot, but where the IDEA protocol instructs weighting and aggregation based on experts' performance, the pilot allowed the experts to find consensus. The experts also expressed they prefer to have control over the outcome, instead of mathematical aggregation.

So far it seems that it is appreciated that this new method creates an expert community that discusses, questions, and learns with eachother, though this is not commonly remarked upon as a benefit of such protocols. The benefits of the discussion phase specifically do seem to depend on the structure of the rest of the protocol. For example, it was observed during the pilot that much time of the discussion was spent on clarifying the question and context. This was a limit of the pilot. The ambiguity can be remedied by having the experts define the quetsions and problem demarcation together before the individual judgements, similarly to the role of the facilitators in the IDEA protocol.

The explicit inclusion of more experts is remarked to decrease dependence on single experts (which was one of the goals of introducing a protocol. Next to documentation, this may support knowledge rentention in the project on the long-term, when senior researchers leave. However, protocols instruct expert selection based on diverse individual perspectives. Therefore, being part of the same discipline and even the same team can be burdened by groupthink and diminish the independence of the individual judgements. (A. M. Hanea et al., 2022; Sutherland, 2022).

3.4. CONCLUSIONS AND NEXT STEPS

The evaluation of the pilot with the IDEA protocol shows participants appreciated the potential benefits of the protocol, as it grounds the judgement process in a scientific basis, makes the 4-eye principle explicit and leads to new insights and transparent expert judgement process amongst colleagues.

However, implementing the expert elicitation protocol requires dealing with some barriers, some of which have previously been noted (Courtney Jones et al., 2023; A. M. Hanea et al., 2022; Hemming et al., 2018; Morgan, 2014). Work pressure, proportional effort, and support for change should be taken into account in the implementation. Some barriers are

also more related to this specific judgement context. Projections of policy effects are hardly feasible to validate, leaving little feedback options for experts other than discussing with other experts. Moreover, this may explain feelings of false certainty.

Even though the actors rate credibility and independance higher than more practical values like workpressure, there are several barriers to succesful implementation that are of a more practical nature. Synergy with available resources (time planning and available expertise), investigating what experts need to confidently make judgements, and involving more and different experts could all be useful directions to improve the expert judgement process. The next step will be to narrow down and choose one problem to focus on in this research.

4

EXPERT WORKSHOP

In Chapter 3, several problems were identified that could be addressed to improve the implementation of the expert judgement protocol. The next step in the double diamond research plan is to use the insights gained so far to narrow down and define a problem to focus on in this research.

However, converging to one problem would turn out not to be so easy. Surprisingly, new themes emerged. In the coming section, the workshop and the results will be presented. The newly emerging problems and the complex interconnections will be discussed, after which there will be a new attempt at converging to a problem statement.

4.1. METHODOLOGY

APPROACH

The insights from the previous sections can be used to inform the problem definition. However, those results were based on the experience from one sector team using the IDEA protocol. To support generalisability, I want to work on the problem definition together with sector experts from other sectors. Moreover, I think it is important (as mentioned by actors in the previous section) that any solutions are supported by the sector experts. Therefore I think it is valuable to involve a diverse selction of sector experts in this stage of the research.

Instead of focussing on the organisational challenges, I want to dive deeper into the sector expert perspective. The evaluation showed the experts may struggle with confidence to make specific judgements. The goal is to find out how different sector experts experience these challenges, and design solutions based on their different experiences. Specifically, I want to know what influences their confidence in their judgement and what their experiences are with their current methods. This can provide important criteria for adapting the elicitation protocol to their practice.

I organised a codesign workshop for a diverse group of sector experts who work on the KEV at PBL. An interactive group discussion where participants work together allows collecting and comparing the experiences and ideas from the different sector experts by having the experts themselves discuss and question each other and stimulates creative think-

ing for solutions to address their problems.

Previous discussions have shown few sector experts are familiar with expert judgement protocols and confidence scores. Because there is limited time for discussion it may be important to introduce these important concepts to the experts beforehand. Therefore, the participants also filled out a survey during a short interview, where they will think about and describe their experiences with expert judgement. They can take these along to the workshop.

The workshop session was conducted on 10 november 2025 in the Netherlands in a hybrid meeting via MSTeams (Microsoft, 2020). The session lasted approximately 60 minutes.

PRE-SESSION INTERVIEW SURVEY

In preparation for the workshop, participants were surveyed on their individual experience and ideas. Each participant therefore got the space to talk about their own ideas before discussing them with the group.

The added value of this step is that more participants had already thought about and shared their own opinion on the topics in the workshop beforehand. This means that individual opinions are collected which may otherwise not have been highlighted during the discussion. Potentially time is saved because individuals because the topic is not new.

The survey took about 20 minutes to fill out, and participants were given the opportunity to iterate on their answers before the start of the workshop. The survey contained the following questions (which can be found in appendix C):

- Think of a moment where you had to apply your expert judgement. What made this judgement credible to you? And to others?
- Are there obstacles to making realistic judgements?
- What influences the credibility of a judgement made by a group of experts according to you? What is the role of bias here?
- How would you communicate your confidence in your own judgement? (Draw or write how you would do so)
- What should a protocol for expert judgement absolutely (not) do?

The survey was conducted either online (via MSTeams (Microsoft, 2020)) or in person, and guided by the researcher. The results were not transcribed, but collected in writing. This allowed the participants to be prompted for more explanation about their answers and explaining the questions. The interview was semi-structured, allowing the participants to answer the questions out of order.

SESSION GUIDELINES

The workshop was guided by a semi-structured discussion guide designed to explore the experiences and ideas of the participants on what could support the credibility and quality of their own expert judgement process.

To inspire creative thinking, the participants were invited to reflect on what they think would be the worst way to conduct expert judgement. The participants were encouraged to think towards more extreme and more realistic bad practices through questions like:

- What would be something that would really decrease the credibility of the expert judgement for you?
- What would be the worst way to approach expert judgement?

Next, the session followed three steps:

1. brainstorming initial ideas
2. making a sub-selection based on topics with support of the moderator
3. working out and evaluating specific ideas in more detail.

The guide included open-ended questions to initiate discussion:

- What support do you use or need for your expert judgement?
- What supports the credibility of your expert judgement?
- What would help you in applying the IDEA protocol?

To also specifically address problems that resulted from the evaluation focus group, the guide also included more targeted questions:

- How would you communicate your confidence?
- How would you use the confidence score?
- How would you approach the discussion phase of expert judgement?
- Is there support missing in this IDEA protocol? Topics it does not cover?

Participants were encouraged to share their thoughts freely and interact with each other to stimulate discussion and generate diverse perspectives.

MODERATION AND FACILITATION

I moderated the workshop. Having organised the evaluation focus group, I had now gotten experience with the facilitation of large group discussions. As a moderator, my role was to structure the meeting, keep the time, prompt elaboration on ideas, segue to the next topic on the agenda, and to monitor participation of all participants.

I tried to ensure that all participants had the opportunity to speak, by asking participants who had been quiet whether they would like to elaborate on their own ideas.

New topics were welcomed as this supports the goal of this research, but due to limited time most discussions had to be cut short. Participants were requested to write down any remaining thoughts, whether they were already discussed or not.

PARTICIPANTS

The workshop consisted of 6 participants, one of whom only did the pre-workshop survey. The participants were selected through who were selected through a purposive sample with the following selection criterion: the participants need to have experience in expert judgement at PBL in the KEV project and currently work at PBL. Because there was already a pilot project with the IDEA protocol, it was considered important that at least one participant of this workshop was part of that pilot.

The initial sample of participants was recruited based on who I was able to contact. To ensure diversity, participants were contacted across all different sector-teams in the KEV project. This means all participants have worked on the KEV project, but all using different models and topic perspectives (Electricity, Industry, Built environment, Mobility, Agriculture, Land use).

The intended number of participants required balancing between what a feasible number of participants is for a workshop, and how many teams there are for the KEV, concluding at 5-7.

HUMAN RESOURCE ETHICS AND RISKS

I applied for ethical approval at the human resource ethics committee (HREC) at TU Delft (Application number: 5520). To comply with human resource ethics guidelines, potential participants were fully informed about the study's purpose, methods, and their rights regarding withdrawal from the study. A copy of the informed consent form is attached in Appendix A. In that form, participants are also made aware of potential risks of participating. Measures were taken to minimise these risks:

- Risk of re-identification: Quotes are only presented anonymised.
- Risk of reputational damage: Anonymisation of contributions which limits re-identification should mitigate risk of reputational damage. The written materials used in this research are shared with the participants so they have insight into what data is used for the research.
- Risk of data breach and re-identification: Personal identifiable information is stored securely and deleted before the end of the research.

DATA COLLECTION

Notes of contributions were taken by the participants' themselves. Additional notes on the meeting's procedure and discussion were taken by the moderator to capture points of conflict, additional explanations by the participants, and reflect on the conduct of the participants and moderator.

The contributions were collected both in notes in person and digital notes (of those participants joining online). These notes were collected and grouped by the moderator based on similar topics, which supported deeper discussion on specific topics and led to more focused data collection in the second phase of the workshop.

Unfortunately, due to time constraints, not all topics were addressed in as much time as necessary to collect everyone's contributions. Therefore, additional data collection was necessary ex post via email.

DATA ANALYSIS

Again, the collected data is analysed according to emergent themes and themes of interest. During this workshop this allows the participants to pitch ideas outside the scope of the workshop. Participants are actively involved in recognising overarching themes, in both problems and solutions. Their ideas for solutions can later be assessed based on design criteria. This step will be discussed in the next chapter.

RELIABILITY AND VALIDITY

Participants were encouraged to discuss their interpretations of key topics like "confidence" and "expertise". This allows to gain a more shared understanding of these terms which are usually subject to interpretation.

Due to time limitations, it was considered important that participants had ample time to discuss while also writing out their ideas. Participants were given the opportunity to add any additional necessary written explanation of their ideas after the focus group via email. This helped balance the pressure to write everything out during the meeting, and time for exploring new ideas.

Because the connections and grouping of ideas in themes was done by the moderator, it is important the participants could validate any conclusions made by the moderator. The results of the focus group were shared afterwards, and participants were given the opportunity to comment on the conclusions that were drawn by the moderator and make any necessary corrections.

4.2. RESULTS

The results of the surveys the participants filled out as preparation for the workshop can be seen in Appendix C. The participants, guided by me, identified several themes along which they discussed their experience with expert judgement. Each theme encapsulates challenges faced by the sector experts as well as ideas for improvement.

WHAT WOULD LEAVE YOU WITH NO CONFIDENCE IN THE EJ PROCESS?

Confidence in the EJ process could be undermined by both methodological and practical shortcomings. The omission of significant uncertainties and the use of physically impossible or unexplained assumptions would cast doubt on the validity of the results. Similarly, results that cannot be reproduced or justified erode trust in the process. Conducting the process in isolation, without engaging other experts for discussion or validation, further reduces confidence in its robustness.

Practical limitations, such as insufficient time or inadequate documentation, also weaken confidence, as they may lead to oversights or incomplete analyses. Furthermore, the process risks losing credibility if judgments appear ideologically driven or if unclear sources, such as AI tools like Chat-GPT, are relied upon.

WHAT COULD HELP SUPPORT INDIVIDUAL JUDGEMENT?

Supporting individual judgment in the EJ process can be achieved through several strategies. Experts often find it challenging to estimate precisely at the 5th–95th percentiles, and would benefit from clearer guidance on when to use these ranges—as opposed to the full 0–100 percentiles—and what situations to include or exclude. While using the extremes of "no effect" or "full effect" as benchmarks for lower and upper bounds can be helpful, this approach may sometimes result in an overly broad range.

Judgment is supported by drawing on precedents and comparable sources, which provide valuable reference points. Confidence in judgment is strengthened when assumptions are underpinned by a logical narrative and a clear mechanism. Many experts emphasize that developing a deep understanding of sector dynamics, through ongoing engagement and tracking changes, enhances their expertise and the quality of their judgments.

For greater accuracy, individual judgments should be complemented by aggregating multiple expert opinions, as the average of several judgments tends to yield more reliable results than relying on a single perspective. However, experts do prefer control over the outcome and would therefore opt consensus over mathematical aggregation. Additionally, incorporating input from laypeople or external experts can introduce fresh insights and broaden the scope of perspectives.

WHAT HELPS TO REFLECT ON CONFIDENCE DURING DISCUSSION WITH EXPERTS?

Reflecting on confidence during expert discussions can be facilitated by several key practices. Standardizing judgments—such as using a confidence score—makes them more comparable and easier to evaluate. Employing a shared tool or scale (for example, a simple plus/minus system) can help ensure that all participants interpret and discuss confidence in the same way. Additionally, having a method to recalculate bandwidths into 5th–95th percentiles can provide a clearer, more flexible framework for assessment. It allows experts to make judgements of different intervals which can then be standardised for discussion.

The experts are aware that common biases like groupthink should be mitigated. In mixed expert groups, it is valuable to reflect on the range of expertise present and to actively seek out new perspectives during discussions. Even when individuals feel confident in their own judgments, engaging with colleagues can uncover blind spots and improve

collective understanding. While consulting more experienced colleagues can be helpful, it is important to be mindful of the potential for junior experts to unconsciously align their judgments with those of seniors.

Though teams and individuals can work differently, sharing similar methodologies (like a SEE protocol) can help to have more productive and consistent collaboration. This can additionally be supported by ensuring that the rationale behind each judgment is clear to everyone, even colleagues that are not directly involved but may need to use the outcome.

WHAT CREATES (MORE OR LESS) CONFIDENCE IN EXPERT JUDGEMENT?

See Table 4.1

High confidence	**Medium confidence**	**Low confidence**
Multiple sources that agree	Mainly expert judgement, few references	Considered unacceptable
Comparable situations (for example similar projects abroad)	Ad hoc	Only expert judgement, barely any references
There is precedent, can learn from the past dynamics	Getting to a conclusion that is acceptable to everyone	Usually in very new situations
Few extreme options are physically possible		Very volatile situations (e.g. an economic crisis could completely change the outcome)
Combining insights from both science and governance		Unclear behaviour / development
Including multiple perspectives (through discussion with colleagues)		Generic policy/ intervention (e.g. subsidy for e-boilers, hard to gauge how much effect this will actually have)
Explainability of rationale		Gap theory / practice
Contact with and research from external parties		

Table 4.1: Caption

INCLUDING AND HANDLING UNCERTAINTIES COULD BE SUPPORTED BY:

Better handling of uncertainties starts with aligning teams, which currently use different definitions (like “policy uncertainty” and “model uncertainties”). More interaction between teams would promote collective learning and help standardize methods.

Next there is little guidance on what uncertainties to include. Currently, uncertainty lists often reuse past years’ judgments, and time pressure doesn’t motivate adding new uncertainties. The experts provide a few criteria: 1. High impact variables (to decrease time spent on estimating factors that are of little impact) and 2. plausible impact (leaving out extreme scenarios that are of little relevance to the policy assessment that assumes a business as usual development).

The experts do stress that they would appreciate examples and instructions on which uncertain factors to include, and how, which would prevent oversight and standardize approaches. What could also help is upfront agreements on core assumptions across all teams. This would clarify what to include and exclude, boosting confidence and consistency. While sensitivity analysis could help experts to gauge magnitude and impact of uncertain variables, it is often impractical with current model infrastructure.

DISCUSSING CONFIDENCE IN EXPERT JUDGEMENT SHOULD:

The process should be simple, practical, and designed to aid reflection rather than become an administrative burden. Ultimately, it should help generate new insights. What could help is a clear qualitative scale, optionally linked to numerical values, that defines what each confidence category means. This ensures everyone shares the same understanding and limits differing interpretations.

4.3. DISCUSSION

The workshop allowed the experts to meet, and surprisingly for the first time discuss across teams how they approach expert judgement. This led to several insights: 1. The experts face an opportunity every year to include new uncertainties, but see too little guidance to do so. 2. They face the same challenges when it comes to individual judgements, though some teams can make use of computer models more than others, but they all find rationale and defensibility of their judgements the most important criterion of a good judgement; 3. They are aware of bias influenced by the expert selection, and the negative effects of the siloed nature of the sector teams.

CREATING LIST OF UNCERTAIN VARIABLES

The KEV project these sector experts participate in, reoccurs every year. The experts are required to choose what sources of uncertainty to quantify for the models and uncertainty analysis. However, these models and this project has been repeating for years. New experts are introduced to the project and can use the list of uncertainties from last year, which gives them a foothold in this giant project. Experts who have been working on the project for several years have gained experience with the existing sources of uncertainty, but therefore indicate to lack experience with analysing new uncertainties.

The IDEA protocol (A. Hanea et al., 2017) and its guides to implementation (Hemming et al., 2018) instruct facilitators on formulating specific and clear questions, but it generally assumes the researchers have already defined the problem demarcation: What is it that experts must estimate? Other protocols have similar limitations. What if you know the topic of interest but still have to decide on what uncertainties to estimate?

The experts themselves indicate that a selection will likely be limited by feasibility: there are so many uncertainties that could be taken into account, it is not feasible to address them all and certainly not all through structured expert elicitation. A selection could be made by focussing on 1. which have significant effect and 2. which are relevant to the policy assessment.

Significant effect is easier to assess for uncertainties that have previously been included, because the experts have more understanding of the nature of their impact which becomes

clear through modelling and by tracking real developments. This provides some feedback (though a true sensitivity analysis is unfeasible with the model structure of the KEV project). Significant effect is much harder to gauge for new uncertainties. Experts are likely inexperienced with these new uncertainties of interest and only have their expert judgement to rely on.

Relevance for policy analysts and decision-makers is also a good criterion. Even if an uncertainty may not end up very impactful, it is functional to show that it is included in the analysis to inform decision-making: no matter what this source of uncertainty may do, it probably has little effect on the outcome. On the other hand, relevance could be too strict a criterion. The research should be independent and may miss impactful uncertainties which at first may not seem relevant.

So, how to formulate a list of uncertainties? Including many very uncertain factors can lead to broad uncertainty bandwidths. In my previous project (Bolhuis, 2025) I showed politically driven decision-makers have an interest in many sources of uncertainty because wide uncertainty intervals can be ambiguous and support different political arguments. So, also formulating this list will be dependent on the expert judgement of the researchers.

RATIONALE AND DEFENSIBILITY

Current research on SEE implementation primarily evaluates its impact on the accuracy of judgments (A. Hanea et al., 2017; Welsh et al., n.d.). Additionally, several studies introduce tools designed to guide researchers through concrete steps for applying SEE (Courtney Jones et al., 2023; Hemming et al., 2018). However, a key gap remains: experts lack clear guidance on how to determine appropriate intervals. This shortcoming stems from the volatile research context which leads to two main issues: unclear underlying assumptions and limited feedback.

Underlying assumptions are particularly critical. Policy effects are judged based on a status quo development. This is standard practice in different policy assessment techniques, in contrast to assessing policy under different scenarios. It is assumed that contextual factors like economic and demographic development continue steadily based on the current policy. Specifically this means no extremely impactful events like war are considered to be of influence on these developments (van der Welle et al., 2017). This introduces complex conditional probabilities. However, this creates problems when the teams are siloed. Though sector experts work in differen teams, sources of uncertainty can be relevant to several sectors. A potential solution would be to standardize definitions and assumptions across problems and teams.

Limited feedback is another challenge caused by the volatile context. There is little feedback on the accuracy of past judgments, because empirical data cannot be gathered to validate the judgements as the context will have changed (new policies introduced and previously unknown uncertainties likely appeared). To address this, example questions could be tailored to sector-specific problems, or more general questions could be developed. These could form the basis of a feedback system, allowing experts to test their judgments. Though such exercises would only indirectly reflect their performance on real-world problems, it could support them to get experience with judging and reflect on their accuracy.

BIAS AND CONFIDENCE

The responses of the sector experts indicate a prevalence but also an awareness of sources of bias. 1. Social bias; 2. Underconfidence.

The sector experts noted how they worked in siloes. Though the researchers bring different backgrounds to the team, they may be inclined to converge in their opinions and be prone to groupthink after working together so often. This could be addressed through expert selection. Though the selection of experts is currently based on responsibility and sector expertise, perhaps teams can look into the flexibility to involve experts from outside their team when applying the IDEA protocol.

The sector experts are also aware of possible social bias in teams and warn about how junior researchers may view the opinion of senior researchers. This bias is frequently observed and can be addressed through expert selection and elicitation design. On the one hand a diverse selection of multiple experts should allow discussion. Preferably the experts discuss anonimised judgements so they cannot see who made what judgement. Lastly, the second round of individual judgements in the IDEA protocol preserves diversity of opinions. However, the KEV experts prefer to aggregate the judgements through consensus which does not protect anonimity nor preserves diversity of opinions.

Lastly, expanding on the results from the evaluation focus group from chapter 3, the sector experts provide insight into what influences their confidence. Expert judgement is required mostly when there is a lack of comparable resources, so it is interesting that the experts' confidence in their judgement largely depends on the availability of references and data. To my knowledge this has not been researched before, and therefore there is little proof for effective measures to take to limit underconfidence. However, I think it could be addressed through clear instructions about: how the protocol supports the scientific data collection; how experts can reflect on their confidence; how bias can be mitigated. This could reassure experts of the credibility of their process with the lack of feedback on accuracy.

The confidence score is presented in several papers as an important measure that decreases overconfidence as it requires critical reflection by the expert. Yet, in practice also Hanea leaves out the confidence score as it is experienced to be hard to interpret ("Anca Hanea: The IDEA protocol for SEJ", 2022). It has also not yet been shown to address underconfidence, a phenomenon that is indicated to also be important: balancing information and confidence is important when decisions are based on your uncertainty bandwidth: a wide bandwidth is very accurate but not very informative and therefore adds little value to salience of the research.

4.4. CONCLUSIONS AND NEXT STEPS

Though the intention of this section was to narrow down on a problem, the sectors experts found they had much to discuss as they had not previously compared their experiences. They came up with several concrete ideas to help them in their expert judgement process, there is no indication what can help them in situations that have no precedent. The next step will be for me to define what problem essence lies under this complex case.

5

PROBLEM DEFINITION

Earlier sections presented several avenues of challenges in implementing structured expert judgement, in policy research but also specifically at PBL. Following my double diamond research design, this chapter restructures the results of the focus group (Chapter 3) and sector expert workshop (Chapter 4) in order to find any underlying common sources of friction or identify larger structures that point to overarching problems. At the end of the chapter I define a problem essence I can address in the final chapters of this report.

5.1. APPROACH

The goal of this section is to capture the intricacies of the challenges faced by the actors, but also find some underlying commonalities that might be addressed. A causal loop diagram (CLD) is useful because it helps you lay out all the important factors in a problem and show how they influence each other. Instead of looking at things separately, you can see the chain of cause-and-effect relationships and the feedback loops that make a situation grow, stabilize, or decrease.

At the same time, drawing the diagram naturally pushes you to simplify. You have to decide which variables really matter and which links are most important. By doing that, the messy complexity of the problem becomes clearer, and you start to see the main feedback loops driving the situation. Those loops often reveal the core mechanism, or the essence of the problem.

CLD's must be interpreted in the following way: each node is a variable that can increase of decrease, represented by an oval. Whenever an incoming arrow has a +, the variable is positively correlated with the previous node (when the previous node increases/ decreases, the current variable also increases/ decreases). Whenever an incoming arrow has a -, the variable is negatively correlated with the previous node (when the previous node increases/ decreases, the current node inversely decreases/ increases). I personally also added arrow with both +/- to indicate that both were indicated by the earlier results, or a ? to incate that the results suggest a connection but not what type.

Whenever a concept was considered an intervention it is included as a 'lever', in a box. These nodes only have outgoing arrows and can be turned on/off.

5.1.1. SELECTION

As noted, the CLD pushes to select and simplify earlier results and find connections between the concepts. The concepts I select to include as the nodes in the CLD are all ideas noted down by the participants of the focus group (Chapter 3) and workshop (Chapter 4). The causal connections between these concepts were either noted or inferred from the discussion.

I decided not to add all ideas and concepts to the CLD, based on 2 criteria:

1. **Simplifying through aggregation.** I decided to combine some ideas onto a higher level of aggregation. For example, *transparency* and *sharing information with other teams* are both included in sharing knowledge. For the sake of simplicity, similar concepts will not be presented multiple times in the diagram.

2. **Focussing on research relevance.** Some contributions of participants focused not on their current experience, but were rhetorical questions: *How can we apply this in smaller sectors?* Other contributions focused on very specific practical issues: "We need to make sure the data storage allows external partners to also access the documentation". Though these are valid concerns and perhaps problems, the participants may already have a clear way of addressing them (data storage) or are a limitation of the current pilot but not related to the experience of the current team (no experiences from other teams with the protocol were collected).

In addition to the contributions from the participants, I also added concepts from the context as sketched in Chapter 2 that I think are of relevant influence. To clearly delineate my contributions from those made by the participants these additions are indicated with dashed lines in contrast to the other content with filled lines.

5.1.2. RELIABILITY AND VALIDITY

Restructuring the results with a CLD provides a simplified overview of the results. However, this may erase details that are important for interpretation such as who the contributions were from and in what context they were elicited. Therefore it is important to view these diagrams in the context of the previous chapters.

Moreover, as I mentioned, I made a selection of what concepts to include and exclude. Though this is a necesarry function of creating this diagram, I may have left out concepts that the participants or the reader may consider relevant. The participants could be contacted to reflect on the CLDs and whether they reflect their experience. However I did not do so. I made these selections by myself to based on the focus of this research. The unaltered contributions can be found in Chapters 3 and 4 if the reader would like to see those and make their own interpretation.

5.2. EVALUATION FOCUS GROUP

WHAT ARE THE IMPORTANT CONCEPTS

The question at the centre of the evaluation is the value of investing in the new protocol. This is therefore represented as a lever that impacts the CLD (Figure 5.1). It increases the number of included perspectives, standardises the method across experts, structures the

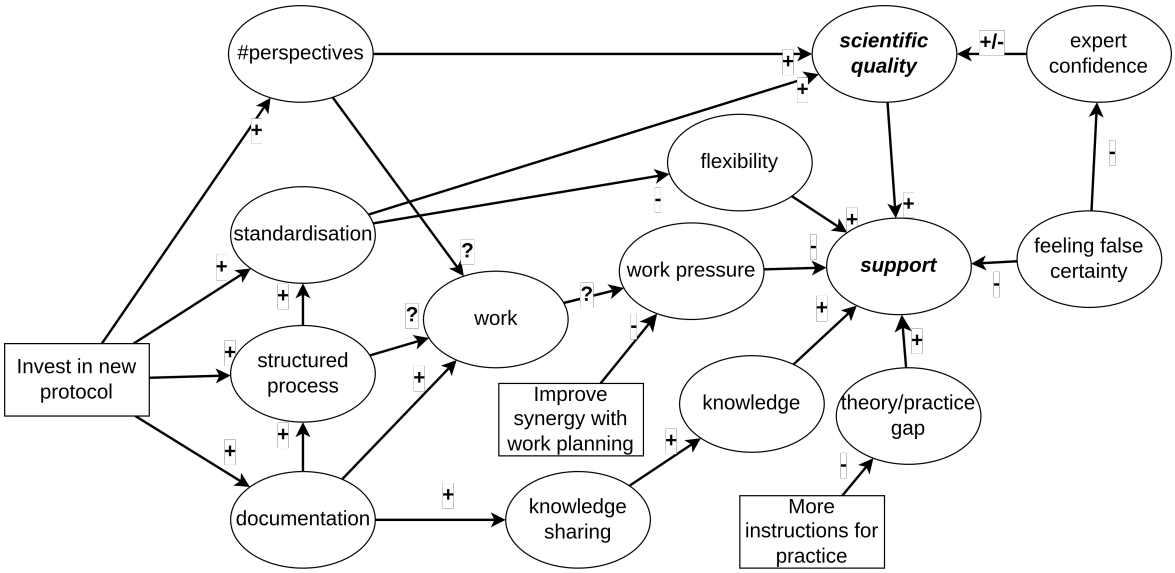


Figure 5.1: Causal Loop Diagram based on the outcomes of the Evaluation focus group

process with specific steps to follow and increases documentation.

However, this is not the only intervention mentioned. There are several specific comments that focus on practical specifics like planning which influence work pressure and their support for the protocol. For example, the exact format and timing of collection of expert input. Though such practical aspects are very important for good implementation of the protocol, they indicate a gap between the theory of the protocol and the practical implementation. The participants also suggest an intervention: 'Improve synergy with work planning' and 'More instructions for practice'.

These levers influence several important nodes that are mentioned to be effects: work pressure, flexibility, knowledge sharing and eventually also new insights due to the knowledge sharing. Though everyone values increasing these concepts, there are also still questions to what extent the new protocol actually influences work pressure. Therefore introducing other interventions are important to address work pressure and the theory/practice gap.

Lastly, we can identify two sinks, nodes that are effected by other nodes, but do not lead to other nodes: scientific quality and support. (In this case I make an exception as one node does affect the other, but no other nodes.)

WHAT STRUCTURES SUPPORT OR CONFLICT EACHOTHER

No loops are identified, indicating no reinforcing or balancing loops. When we relate the discussion presented in this CLD to the values mentioned earlier we can note that there are connected concepts that support eachother. Credibility through standardisation and transparency(as documentation and knowledge sharing) are clearly present and support eachother and scientific quality. Independence also is mentioned to support scientific quality. We recognise that these concepts reflect the highest rated values and it is therefore consistent that they are of great influence on the scientific quality and support variables in

this CLD.

The sinks of the CLD are part of the dilemma facing the participants. On the one hand, there are many interventions that could increase scientific quality (many also excluded of this evaluation because this evaluation focused on the current new protocol, but many other interventions could be possible.) However, many would not be feasible or would not be supported by the colleagues having to implement them based on flexibility, work pressure, and implementation issues (theory/practice gap).

WHAT PROBLEMS DO WE IDENTIFY?

We recognise a conflict in the flexibility of the old/current way of working and the dilemma of the new protocol that has as of yet unknown impact on work pressure. There are several questions left:

- Is there actually a conflict in the impact of a new protocol on the work pressure and scientific quality?
- How can agency of different teams and experts who have trained their skill in expert judgement be supported in a protocol that drives standardisation?
- How can the theory of the IDEA protocol and similar expert elicitation protocols translate to the practice of research in the policy context?
- What is the role of expert confidence and feeling false certainty in the implementation of an expert judgement protocol and can they be influenced?

5.3. SECTOR EXPERT WORKSHOP

WHAT ARE THE NEW IMPORTANT CONCEPTS

The question at the centre of the workshop is how experts experience the credibility and confidence in their judgement, and what can help them. This is reflected in the interventions that act as levers in this CLD (Figure 5.2): defining and scoring confidence, and 'more instructions' for judging new and different types of unknowns. I have also included the intervention 'investing in new protocol' as this was presented during the workshop, but all experts present hadn't worked with this protocol yet [indicated with dashed line]. The sinks identified in this CLD are related to the focus of the workshop: what is credible expert judgement and what is (the role of) expert confidence?

It is clear from these results that experts think more instructions on how to handle expert judgement would support them in getting a grip on their task as experts. However, the question is whose task is it to create these instructions. The methodologists and uncertainty team can play a role (as they do now) in developing a protocol for expert judgement, but such protocols do not specify how to tackle specific types of uncertainty. This is where the experience and knowledge of the policy context of the experts play an important role. This is where theory and practice are balanced.

At the same time, we recognise the confidence and credibility the experts experience are influenced by several other (many similar) variables, like the explainability of judgements which relates to the availability of data and as discussed in the previous paragraph,

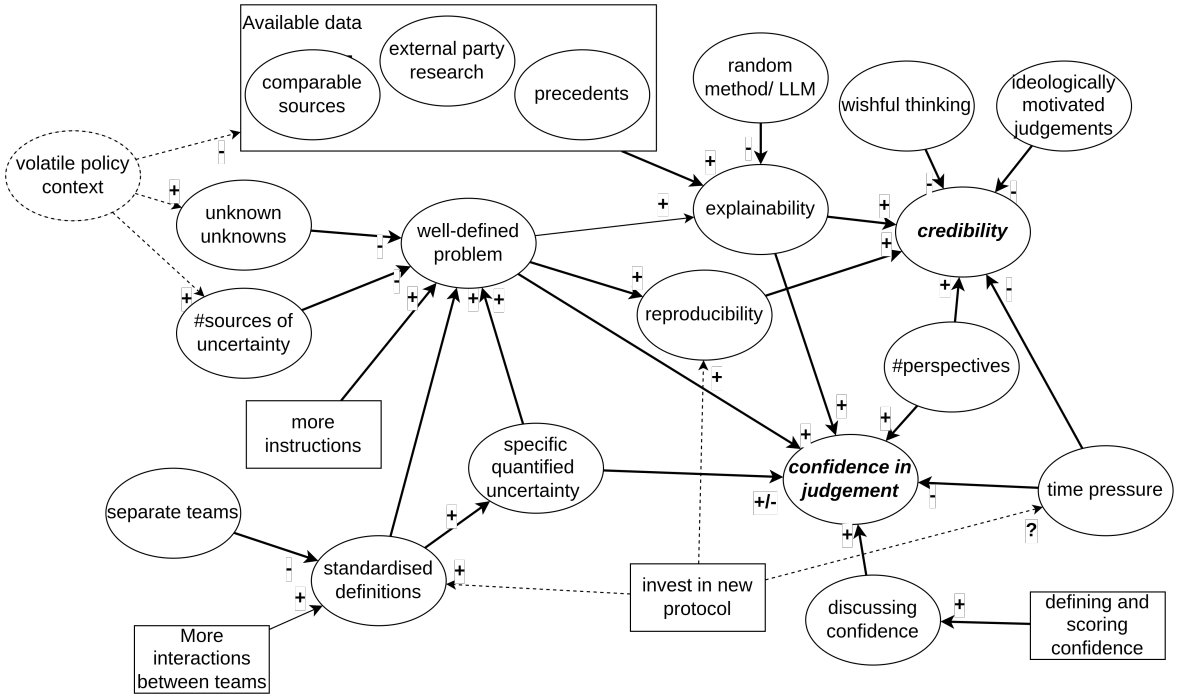


Figure 5.2: Causal Loop Diagram based on the outcomes of the expert workshop

the grasp experts have on defining the problem they must tackle (or specifically, defining the unknown variables they must estimate). However, due to the influence of the volatile policy context (indicated with dashed lines), more specific instructions may not help if experts' confidence and experience of credibility relies on data and well-definedness.

Discussing their judgements and their confidence play an important part in their judgement. It allows to check with other experts, using the 4-eye principle, and reflecting on their judgements. However, there is a lack of definition. On the one hand there is the definition of what their uncertainty interval is which complicates comparing judgements, which instructions could tackle. The participants also suggest defining and scoring confidence to support the discussion of their judgements.

The participants also suggest an intervention that supports cohesion and standardisation across teams. During the workshop the participants noticed that their teams have different methods and perspectives. On the one hand, it is important to instruct and standardise the methods, but a wealth of perspectives is very valuable as teams tend to homogenize. Therefore cohesion between the teams is a dilemma as it has both positive and negative effects on the sinks, our concepts of interest

WHAT STRUCTURES SUPPORT OR CONFLICT EACHOTHER

There are no loops identified, so no obvious reinforcing structures. Next, we also see that the concepts supporting credibility do not per se support confidence.

The conflict that separate teams and perspectives leads to is how standardisation has positive impact on credibility, but requires mitigating homogeneity. In contrast to the eval-

uation focus group, there were no remarks on what may possibly impact time pressure. Being required to specify and quantify uncertainty both positively and negatively impacts confidence, as on the one hand experts experience the uncertainty bandwidth to not be well-defined (which could lead to feelings of false certainty as the focus group reflected). Whereas if they could get more support in this aspect, explicitly communicating uncertainty could support their confidence.

WHAT PROBLEMS DO WE IDENTIFY?

To reiterate, this CLD reflects the experiences shared by the participants during the 1-hour workshop. It is therefore realistic that this CLD may not present a complete picture. For example, some connections between concepts were not explicitly discussed, but I infer they are connected and clearly show this distinction with the dashed lines.

The participants share largely a need for more connection to share experiences between teams (which can create a more shared understanding of their task through practice), but also a need for instruction from an outside perspective to define and specify (which can create more grip on the problem at hand through outside intervention). However, focussing too much on either the confidence of experts or the credibility of the process can lead to conflicting results.

5.4. EXPERT INTERVIEW

There are several practical hurdles that still need to be addressed, however because I cannot tackle practical challenges like planning, a feasible solution would require me to focus on what information I can provide and discussions I can facilitate.

My first instinct was to tackle the challenge of expert confidence. The confidence score of the IDEA protocol that is supposed to support experts in reflecting on their confidence and reduce overconfidence, was hard to use. However, at the same time there are indications of underconfidence: a tendency to focus on physical bounds when there is a lack of indication to choose anything different and a worry for presenting false certainty. There are no steps in the IDEA protocol tackling this as a serious bias. I therefore wanted to consult an expert in expert judgement to discuss my interpretation.

Tina Nane is an expert in applied probability theory and is specialised in uncertainty quantification through data and structured expert judgement.

I explained the case study and my research goal to Nane. I highlighted the specifics of the number of experts (less than ideal for the IDEA protocol) and the number of variables to estimate (could be dozens, but many in a short amount of time).

In the case study at hand, few experts are responsible for the estimations and they have this responsibility every year. Nane shared that an initial step should ideally be to define the question you want answered first, and then ask what expertise is needed. This can indicate gaps in knowledge and motivate to include more experts. Additionally, the anonimity and pressure to conform can be large when there are just 2 experts. Nane pitches that a second round of revising individual judgements and then aggregating them would be valuable to

support a difference in expert opinion.

A challenge in the application of the IDEA expert judgement protocol on policy assessment is that most likely the experts' judgements will turn out to be incorrect (as new policy is developed to transform the context). This makes it hard to calibrate and give feedback on the performance of expert judgement. Nane shares that some relevant questions could still be formulated on a topic in the experts' domain of expertise. As the experts have to learn to work with the new protocol, a capsule training of for example 5 questions could already provide some important feedback.

The confidence score is not interpreted as intended by the experts the researcher discussed with. Perhaps categories of confidence would be interpreted better? Nane indicates she has had similar experiences and prefers not to use the confidence score as it is generally experienced as unintuitive and requires a lot of instruction and training. They share that for understanding underlying use of data and expertise, having experts write down their rationales would be beneficial.

The confidence score also functions to rescale the contributions of the experts for comparability, but if the numeric score isn't used this can't be done. Nane shares that in practice it is very difficult to estimate exactly 5-95. So having the judgements of plausible lower and upper bounds be comparable could mostly be improved by letting the experts have access to the same data and having them establish the same definition of 'plausible'.

The application of expert judgement in policy assessments could be prone to underconfidence as well as overconfidence (but it is rarely researched in contrast to overconfidence). Even though expert judgement is needed when there is a lack of data, experts seem to have little confidence when there is a lack of data. Perhaps instructions about what expert judgement is and the value of it would support expert confidence. Nane was surprised by the observation of underconfidence, and is more weary of overconfidence. They agree that some context could be sketched, that even in the absence of data someone can be confident in their expert judgement.

This discussion led me to double-back and revise my problem definition. Though the idea of underconfidence is an interesting area of further research, it is clear support for implementation runs in too many different problems. Focusing on an underlying problem could therefore help provide a good foundation.

5.5. PROBLEM ESSENCE DEFINITION

There is a need to implement structured expert elicitation because the documentation and standardisation improves independence and credibility. Apart from practical restrictions, implementation is a challenge because of the lack of support for experts, both regarding the appropriate (scientific) tools to confidently face new uncertainties as well as the connections with other experts.

Sector experts are faced with recurring but also unfamiliar unknowns they must judge. The policy context is volatile and confident judgements should not rely on the availability of

concrete data. However, they require more support in tackling these unknown unknowns. It is however in the nature of unknown unknowns that there are no instructions yet. Therefore principles of best-practice for dealing with uncertainty lack.

Best practices are developed in learning communities, but experts miss opportunities to learn together with and from their colleagues. This would also increase encounters with new perspectives, as teams are prone to fall into groupthink due to social biases and dependence on senior experts. Though someone is the responsible expert on a topic, this does not mean they are alone. Structured expert elicitation provides the foundation principles for including multiple experts and other pillars of credible expert judgement, even if only one expert is ultimately responsible.

6

COMMUNICATION DESIGN

In Chapter 5 I presented the following problem essence: The policy context is volatile. There is a need to implement structured elicitation, but this does not support the experts in confidently estimating new uncertainties nor does the organisational context promote interacting with new and different experts. Probable directions for solutions are to communicate best-practices for dealing with uncertainties and promoting interaction between experts so they can learn from and reflect on eachothers' examples.

Following the double diamond research design, this section will present a prototype communication design to tackle this challenge. First, I will present perspectives and inspiration that informed my plan for the communication design, after which I will present a prototype.

6.1. APPROACH

In communication between sender and receiver, there are several levels of analysis. In this context, the Laswell model of communication (Figure 6.1) helps us separate and understand different aspects of a linear communication process. This linear simplification fits the context of this project because I would like to provide the uncertainty team with a communication product they can use to address other actors in the KEV project. I will create an outline of the communication plan based on this model.

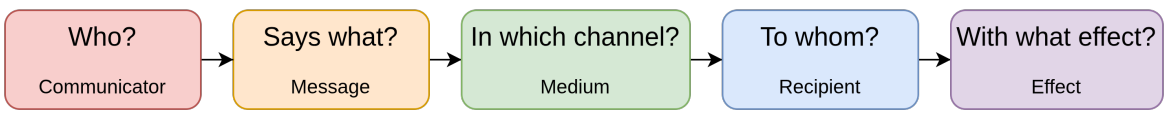


Figure 6.1: Laswell's model of communication. Source: Lasswell, 1948

Effective communication in this model would mean that the communicator sends their message, through an appropriate medium that reaches the intended recipient with the intended effect. The communicator in our case will be the uncertainty team at PBL, who are in charge of the pilot on structured expert judgement. To create a main outline of the communication we need to identify what their intended communication effect is. Then we can identify who the best recipients are, what message(s) should be send to them and what

medium would be best to use.

Figure 6.2 summarises the steps of my design process. After the overview of the communication plan, I drafted a storyboard with text and accompanying visual design to try out how to communicate the core concepts with the right tone (Appendix D). Someone from the uncertainty team (the communicator in our model) is contacted to provide feedback on the the storyboard (and therefore the communication plan). For the next iteration of the storyboards, I will draw up a list of design criteria and decide on the aesthetic of the communication design. At that time I will seek feedback on the aesthetic of the communication design from a design expert. The last steps will include making a prototype, validating the prototype with representatives from the intended audience, and concluding any revisions that can be made for a final product.

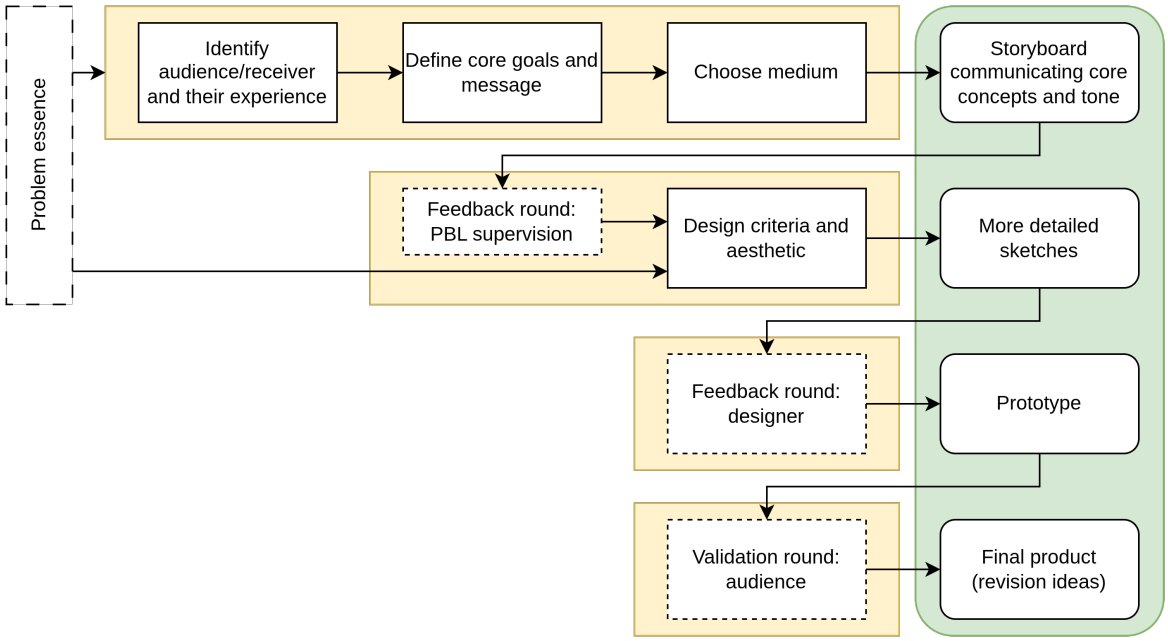


Figure 6.2: Overview of the design process

6.2. TARGET AUDIENCE

Sector experts, team leaders, project leaders, uncertainty team and methodologists all play an important role as described in the actor analysis. Targeting the problem essence could be done from any of those angles. Communication design could target how colleagues define and discuss confidence in judgements, though this is partially covered in protocols like the IDEA protocol (A. Hanea et al., 2017) it could be expanded upon to ease implementation. Perhaps it's this implementation perspective that could be communicated to the methodologists, to create more common understanding across teams. Team and projectleaders could also become more involved in instruction and leading of the expert judgement protocol, so understanding the challenges the experts face could be useful. However, it's the experts themselves who have the most experience with their tasks. Supporting their autonomy, informing on the principles and benefits of structured expert elicitation and ac-

tivating them to reflect and work on their own practice could lead to change from bottom-up, rather than strengthening instructions and standardisation top-down.

The main audience is therefore the sectors experts (old and new), and the secondary audience are the team and project leaders who play a supportive role. Expert judgement however is not isolated to only the sector experts of the KEV-project at PBL. Implementing structured expert judgement is a challenge other (policy) research teams and organisations also face. Designing a message that is more widely applicable could easily be shared and help other researchers also reflect. This is especially important as researchers often work together with researchers from other teams and organisations. The scientific standard and implementation of such protocols could benefit if this design is applicable to a wider audience. This would mean a more general message.

Sector experts and other researchers with similar tasks most likely have their own working protocol or habits in their team and organisation. Deriving from the research results in this thesis, they will be groups of experts of different backgrounds and have the responsibility to estimate unknowns to support research and modelling. They may be familiar with protocols like the Delphi method that is commonly used in research to reach consensus in groups of experts, but it can not be assumed they are familiar with protocols for quantifying unknown variables like the IDEA protocol.

Assuming based on the evaluation focus group and expert workshop, there are constraints of resources, and there may be pre-conceived ideas about the bureaucracy involved with a new protocol which may be misconceptions. They are however generally motivated to improve their work, especially if it could save them time in the future and if it makes their work more explainable. They value the flexibility they have to work in a manner that fits the constraints of their team, but they do see value in learning from eachother and standardising their methods as long as it synergises with their planning and workload.

6.3. COMMUNICATION GOAL

The problem essence shows support to bring theory into policy practice in needed. Communication can be supportive by informing, but communication can also have the intention to influence behaviour.

I specify the communication goal with the help of the hierarchy of communication effects (Figure 6.3). As most experts are aware of changes happening, the most logical next steps of communication should focus on increasing knowledge and sharing experiences, which requires a different strategy than if no experts would be aware of anything related to expert judgement protocols (Blythe, 2005).

The primary goal of communication will therefore be to inform sector actors about the principles of structured expert elicitation (SEE), and spark conversation on how to tackle challenges of implementation in this specific policy research context.

However, to actually achieve change and improve the expert judgement process, information is not enough. Motivation to act and change of social norms in the organisation are

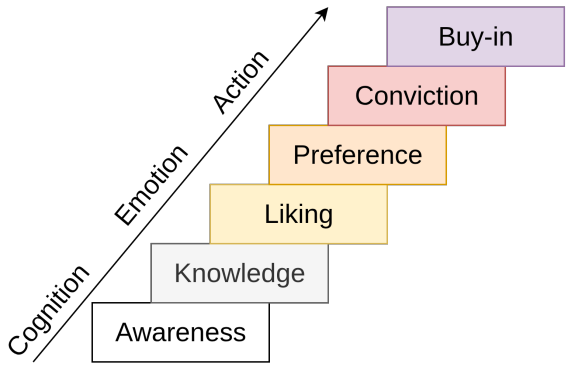


Figure 6.3: Blythe hierarchy of communication effects. Source: Blythe, 2005.

also important.

The secondary goal requires addressing emotions, which is the next step in the hierarchy of effects. Sector experts must be reassured of the importance of applying their expertise without data to back it up. The earlier results indicated

Lastly, the tertiary goal is to inspire them to take action on their working method and built a learning community with their colleagues where they can share new insights with eachother.

6.4. CORE MESSAGE

The primary communication goal is to inform old and new sector experts about SEE. Chapters 1 and 2 already addressed what aspects SEE protocols have in common and what they do differently. The choice between the protocols therefore depends on for example restrictions (how many experts you have access to) or goals (consensus or mathematical aggregation). Because PBL is focused on the IDEA protocol, that protocol will be used as the main example, detailing how it supports the scientific standard of expert judgement.

Additionally, there are some challenges that the sector experts at PBL face that are unique to their context, which were identified in Chapters 3 and 4. I would therefore like to address the following points:

- Choosing what uncertainties to include in the expert judgement process.
- Choosing experts and how to work together with colleagues
- Assumptions that inform the problem demarcation and the uncertainty intervals.

6.5. MEDIUM AND TIMING OF COMMUNICATION

With any request for attention, it is important to respect the audience's time. As the goal is to inform old and new sector experts and motivate reflection, no long interactive session is required. A concise message can also be conveyed repeatedly and remembered better than

a long lecture.

A short message could be conveyed during a live presentation or preferably an interactive workshop, but this requires a presenter and availability of the target audience. Repetition is also expensive in such methods. A message does not have to be provided verbally. Information could be provided through writing or a combination of means. Though writing can be shared and accessed easily, it is hard to convey tone. Additionally it is up to the reader to choose what information they stress.

A video message could easily convey the same dynamic visual and textual information as a presentation and similarly can communicate a tone. It is also as easily sharable as a written text. Moreover, whereas the communicators have control over the dosing of the information (speed and amount of information on screen), receivers also have control and can go back and forth through the video at any time.

Information about expert judgement would best be shared in the preparation period for the project. The KEV project for example, follows a yearly cycle and planning of expert judgement and other steps is an important part of the preparation period. Any changes and reflection would best be discussed in this period while the planning is being made.

However, there are many other challenges that require attention. Expert judgement is only a small part of the entire research process and even the KEV-project is not the only project the experts must spend time on. It would be in the interest of the receivers of the communication that the information is more widely applicable.

6.6. SUMMARY OF COMMUNICATION STRATEGY

The uncertainty team at PBL is positioned within an organisation with its own structures, culture, and practical constraints. Communication will target both sector experts and supportive actors, including project and team leaders, who play a crucial role (see Table 6.1).

The goals are to inform, reassure, and inspire, driving both immediate and fundamental long-term change (see Figure 6.4). To achieve this, communication should emphasize how to maintain a scientific standard (to support confidence in expert judgement) and foster a learning community that shares knowledge and questions organisational practices (to address bias and challenge the status quo).

6.7. COMMUNICATION DESIGN CRITERIA

To inform the communication design, I will develop a list of design criteria that will support reaching the desired communication effects. This section presents a theoretical framework based on two dimensions: the personal-organisational learning dimension, as well as the short-long term change dimension. In addition, (grey) literature on effective instructions are consulted.

	1st Sector Experts	2nd Project leaders
Desired behaviour	Reflect and act	Connect and support
Communication aim	Informing and reassure that Structured expert elicitation (SEE) can support experts with their unique challenges Inspire to implement in own work	Spark conversation on challenges for expert judgement (EJ) in this policy context and how project/ team leads can support
Communication effect	Experts can be reassured that they can make quality informed decisions, even in very uncertain cases	Project leaders have a better understanding of what is needed to support the quality of expert judgement
Key message	This is a way you (the expert) can make judgements with confidence and be transparent about them.	These are principles that can supports experts in their work and you can help them with it.

Table 6.1: Addressing two types of recipients as part of a larger strategy

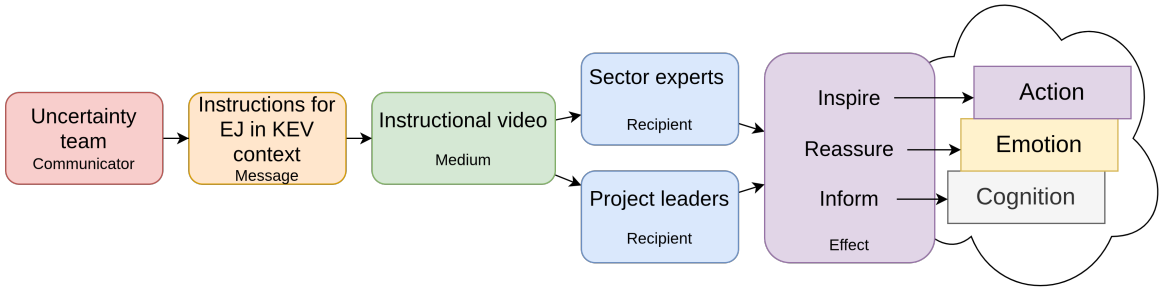


Figure 6.4: Laswell's model of my communication strategy. The effects refer back to Blythe's hierarchy of effects.

PERSUASIVE MESSAGING AND INDIVIDUAL BEHAVIOUR CHANGE

Persuasive messaging (Dainton & Zelley, 2018, Ch. 7) builds on the appeal to ethos, pathos, and logos. Not only appealing with elaborate arguments but also appealing to emotions is consistent with the dual-process model of persuasion in the elaboration likelihood model. Processing messages through the *central route* requires personal motivation and cognitive capacity for analysing arguments. However, the *peripheral route* processes cues that are influenced by: pre-commitment, liking, and bandwagon effects.

A persuasive message may however not lead to long-term behavioural change. This can be explained by the Theory of Planned Behaviour: Autonomy (believed capacity to take control and perceived power) is important for the intention to perform a behaviour (change). Together with the social norm and the receiver's personal attitude, these are predictors for behavioural change.

Perceived autonomy and personal attitude can be partially addressed at the individual. But a social norm that supports the behaviour (change) in the long-term requires a look at the organisational context.

ORGANIZATIONAL LEARNING AND LONG-TERM CHANGE

For an organisation to address its social norms and underlying assumptions requires a deeper form of (organisational) learning (Georges L. Romme & Van Witteloostuijn, 1999):

- Single-loop learning: Am I doing things right?
- Double-loop learning: Am I doing the right things?
- Triple-loop learning: Questioning underlying assumptions about what the right thing is.

Deeper forms of learning are stimulated by questioning assumptions, instigating reflection, and initiating dialogue on problem framing. Embedding these activities in teams supports deeper reflection. Therefore communication should target not only individuals, but acitvating reflexive learning communities in the organisation could actually address the social norms.

The organisational structure is an important aspect of the communication route. Addressing individuals in different teams and levels of management activates learning in different points of the organisation. Addressing gate-keepers in different levels of management (team leaders) but also first-movers researchers supports creating a learning network from multiple directions. Alternatively, if a message is sent to only a project leader who should diffuse the information top-down, information travels less fast and may diminish perceived autonomy by other team-members.

6.7.1. DESIGN PRINCIPLES

In addition to communication and learning theory, the communication design can built on best-design-practices from both academic literature (multimedia learning research: Mayer (2005); instructional video design: van der Meij and Hopfner (2022)), and grey literature on practical video guidelines from "Video-toolkit" (n.d.) by Leiden University, and memorability research as articulated by Heath and Heath (2007) in 'Made to Stick'.

VIDEO FORMAT

On firsthand, a video format provides flexibility of reaching the audience when at different times with a visual and spoken message. According to the Modality Principle (Mayer's Cognitive Theory of Multimedia Learning) learners process information more effectively when graphics are paired with narration rather than with on-screen text. Moreover, conveyeying procedural knowledge is also supported by animation (van der Meij and van der Meij) because it externalises the actions.

FAMILIAR CONTEXT

Situating information within the familiar context of the company, both in the sequencing of content and in the visual design, aligns with several principles.

Constructivist learning theory assumes that individuals construct knowledge by linking new information to prior experiences. Emphasizing previous steps and existing experiences provides grounds for building new knowledge and skills.

The Pre-training Principle (Mayer's Cognitive Theory of Multimedia Learning) suggests that learning is enhanced when learners already understand key elements of the instructional environment. Van der Meij and van der Meij further argue that new concepts should be introduced in context and aligned with authentic tasks. Therefore introducing new information in the context of a worked out familiar example supports learning.

From the perspective of memorability, Heath and Heath emphasize the importance of concreteness and credibility. Grounding principles in the company's actual environment increases both.

ACTIONABLE INSTRUCTIONS

Though abstract concepts can sometimes provide illustrations, it is important for implementation that actionable instructions are provided. Van der Meij and van der Meij explicitly recommend focusing on procedural rather than conceptual information in software training contexts and following the learner's mental plan when describing action sequences. This approach aligns with Mayer's guidance on managing essential processing: abstract theoretical exposition increases intrinsic cognitive load and may impede comprehension when learners' primary need is task performance. By presenting principles in actionable form, the instruction remains concrete and directly applicable, thereby facilitating both immediate use and long-term retention. This approach also resonates with the argument in Made to Stick that concrete ideas are more likely to be understood and remembered than abstract generalities.

GRAPHICS AND NARRATION

Employing symbols and graphics in combination with narration, rather than text accompanied by narration, is grounded in Mayer's Redundancy and Modality Principles. Presenting identical verbal information in both spoken and written form can overload the visual channel and create unnecessary cognitive processing. By replacing extensive on-screen text with visual cues, icons, and highlighting, the design leverages dual-channel processing: visual resources are used for graphical representation while auditory resources process narration. Van der Meij and van der Meij further stress the importance of synchronizing action and voice and using highlighting to guide attention.

CONVERSATIONAL STYLE AND MOTIVATING TONE

Adoption of a conversational, motivational tone rather than a detached and purely informational style is supported by Mayer's Personalization and Voice Principles. Learners demonstrate deeper processing when instructional language is conversational rather than formal, and when narration is delivered in a friendly human voice. Direct address ("you," "we") fosters social presence. Van der Meij and van der Meij also recommend promoting the goal of the task and enhancing perceptions of relevance, both of which are strengthened through conversational framing. Heath and Heath further argue that ideas are more likely to endure when they evoke emotion and connect to meaningful concerns. A motivational tone therefore does not merely increase engagement superficially, it contributes to improved retention.

6.7.2. DESIGN CRITERIA

The theoretical framework and communication design principles are distilled in the following design criteria, which inform the design choices detailed in the next section (see Table 6.2).

Criterion	Communication Theory	Design principle
Guide attention at information by showing the benefits for the viewer and building on their familiar knowledge first.	Liking	Familiar context
Minimize use of abstract concepts. Show implementation through use of examples.		Actionable instructions
Balance communicator's authority to instruct with receiver's intrinsic motivation to change behaviour	Authority; Elaboration Likelihood;	Conversational style and motivating tone
Provide concrete examples of actions that are in control of the viewer	Theory of planned behaviour	
Strike a reassuring tone and address the viewer directly		Conversational style and motivating tone
Pose rhetorical questions	Organisational learning	Actionable instructions
Combine verbal and graphical information		Video format; Graphics and narration

Table 6.2: Design criteria based on the communication theories and design principles discussed in this section.

Lastly, the project also poses constraints on the communication design. Any deliverables would have to be handed off. This would mean that any prototype or design should be easy to transfer and not need any continued maintanance or attention from me. Practically, either I or someone at PBL needs to have the skills and resources to use the communication design.

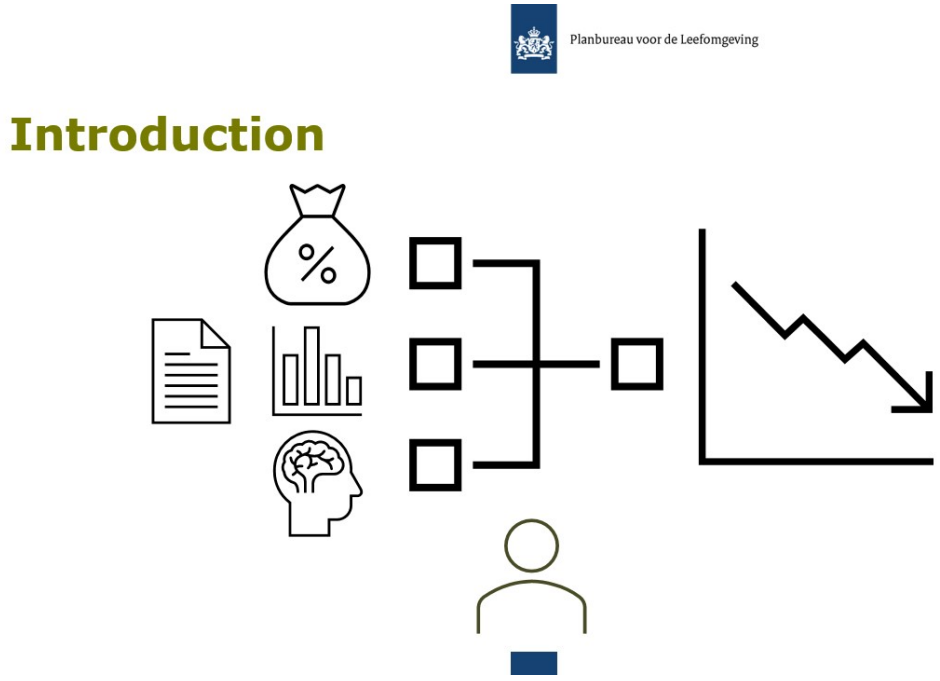
6.8. DESIGN DELIVERY AND JUSTIFICATION

The final communication product is an instructional video.[1] The video was created by recording my own voice and screen while presenting illustrations in MSPowerpoint (Microsoft Corporation, 2025). The rest of this section will detail the content of the video while presenting the design criteria and how they were met. Animations will therefore not be clearly visible in this document, but are visible in the video.

Voiceover: *This video addresses what principles you can use to support your judgements for policy analysis when facing uncertainty.*

Grab Attention and inform	The content of the video and the intended audience is made clear through the image, titles and voiceover. The viewer is addressed directly. This allows the viewer to gauge the value of watching the video.
Claim to authority	The link to the research institute and the university are made clear, and are specifically connected to the uncertainty team as the source of the research, information, and communication.

[1]The final instructional video is available for research and educational purposes at 4TU.ResearchData. DOI:10.4121/14223f51-d909-4be5-8eca-22b1720570f5

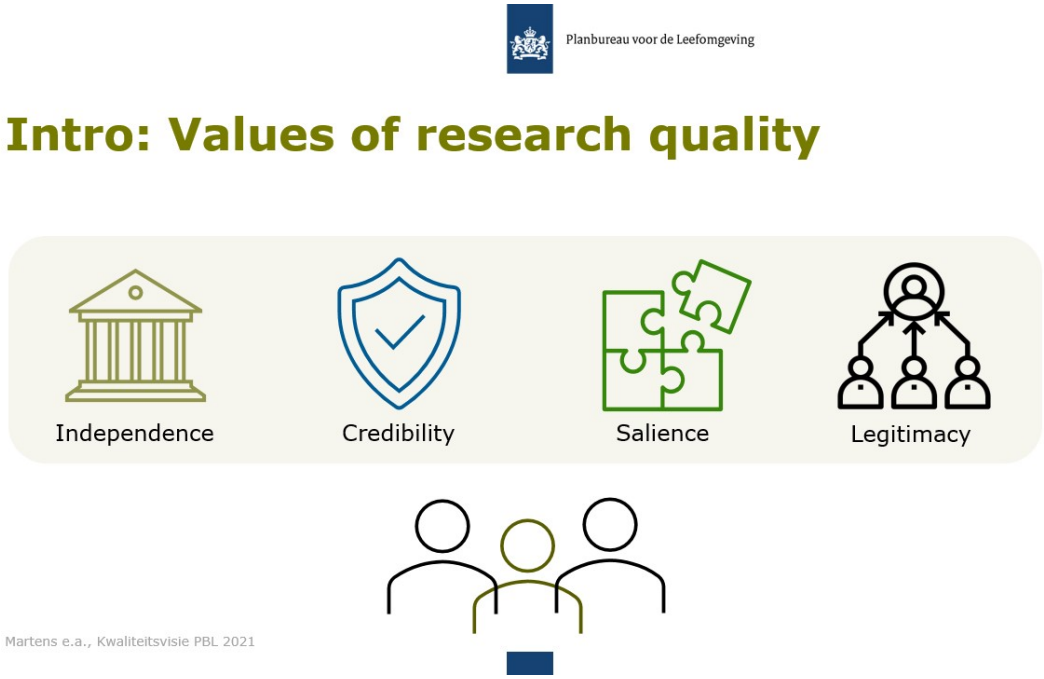


Voiceover: *When we perform research in the policy field, we will have to deal with unknown variables. It is our responsibility to judge the input data like a new policy, and any sources of evidence, like historical data and comparable cases. When there is insufficient data, we need to use our own knowledge and experience to make a best guess.*

Create shared perspective	The policy research context is established. In an extension of the introduction slide, the viewer is made aware of the context in which the video discusses expert judgement. It is important to do so because expert judgement can be addressed in different contexts, so this creates a shared understanding of the context.
Personal connection	The viewer is included in the narration through the use of 'we'. This makes the message more personal. It invites the viewer to imagine themselves as the subject that is discussed and connect any new information to their existing experience.

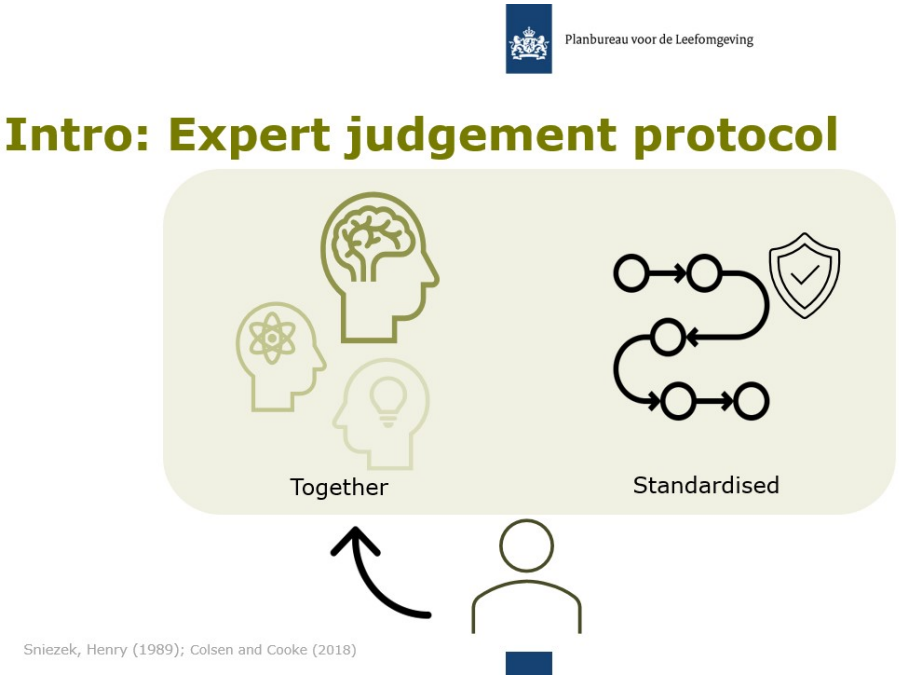
Voiceover: *Consider what influences your judgement in a volatile policy context. Perhaps you feel insecure because new uncertainties arise, or you may feel more confident when studies by other researchers are available.*

Personal connection	The viewer is now directly addressed and invited to reflect on their experience.
Introduce concepts	Sources of uncertainty and confidence are presented early as important concepts that relate to the information in the rest of the video.



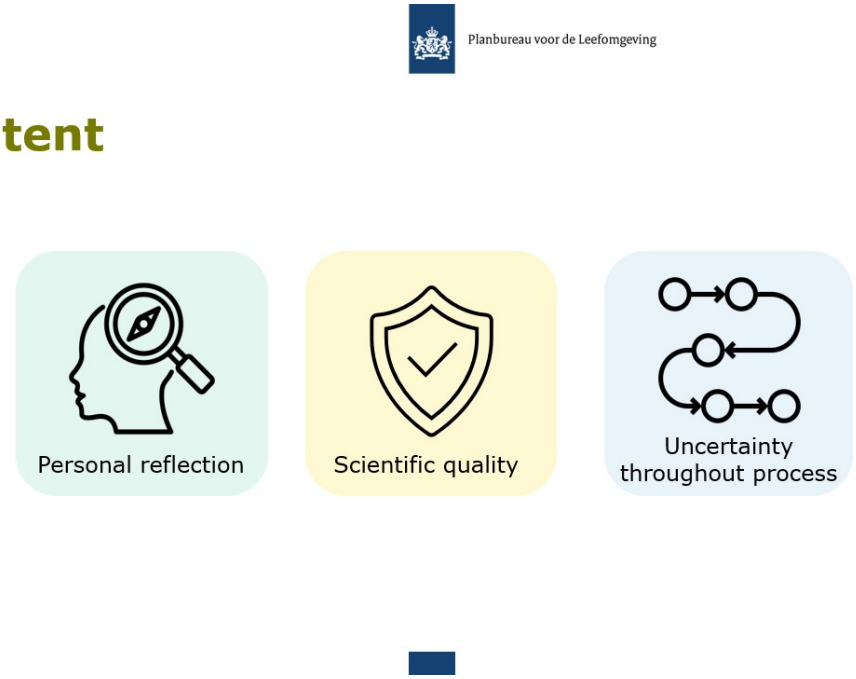
Voiceover: *The quality of our research at PBL is shown by our independence, credibility, salience, and legitimacy. Anyone using our reports also has expectations. How do we approach expert judgment in a way that ensures quality by standards expected internally and externally?*

Organisational and professional values	The value of the information in the video is directly connected to the organisational context. Through the use of 'we' in rhetorical questions reassures the viewer they are not alone in this challenge.
Grab attention	Posing a rhetorical question the viewer wants answered reastablishes the value of watching the video.



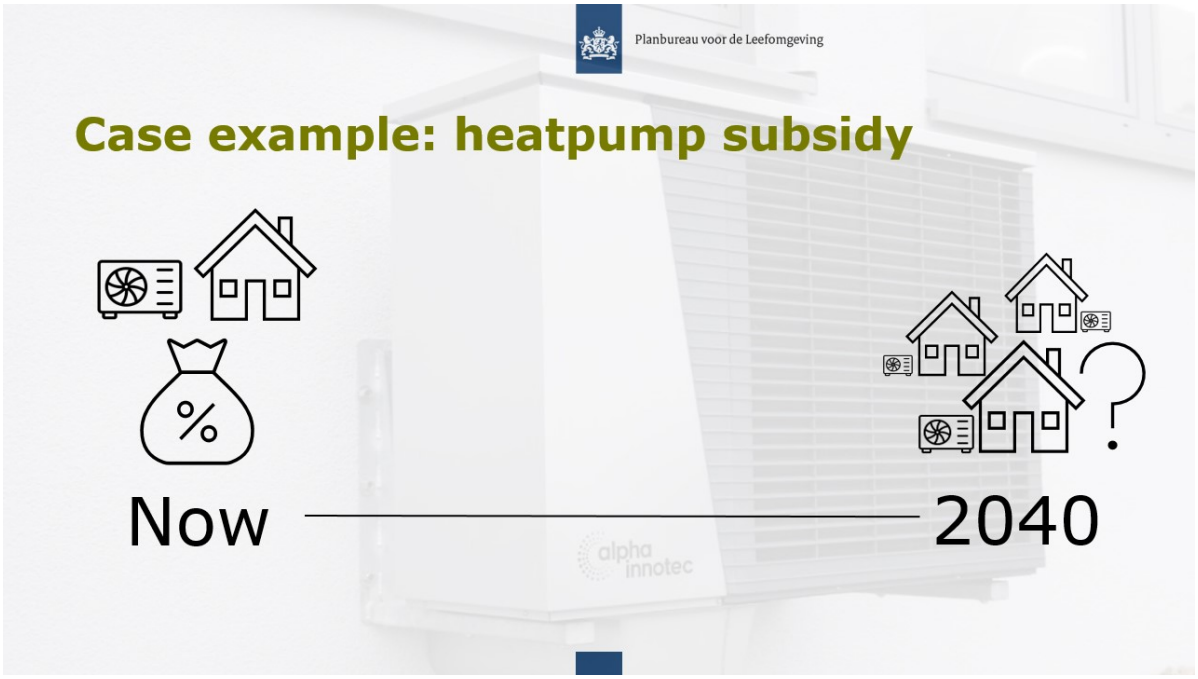
Voiceover: *Studies show that groups of experts with relevant knowledge often surpass single experts in the accuracy of their judgments. Researchers across disciplines developed tested and accepted protocols that support you in this task. You may recognise the Delphi method by the Rand Corporation. Or the Classical Model by mathematician Roger Cooke.*

Reassure and claim to authority	The information is connected to theory and reassures the viewer there is research that supports the content of this video. Specifically that information can address challenges the viewer faces.



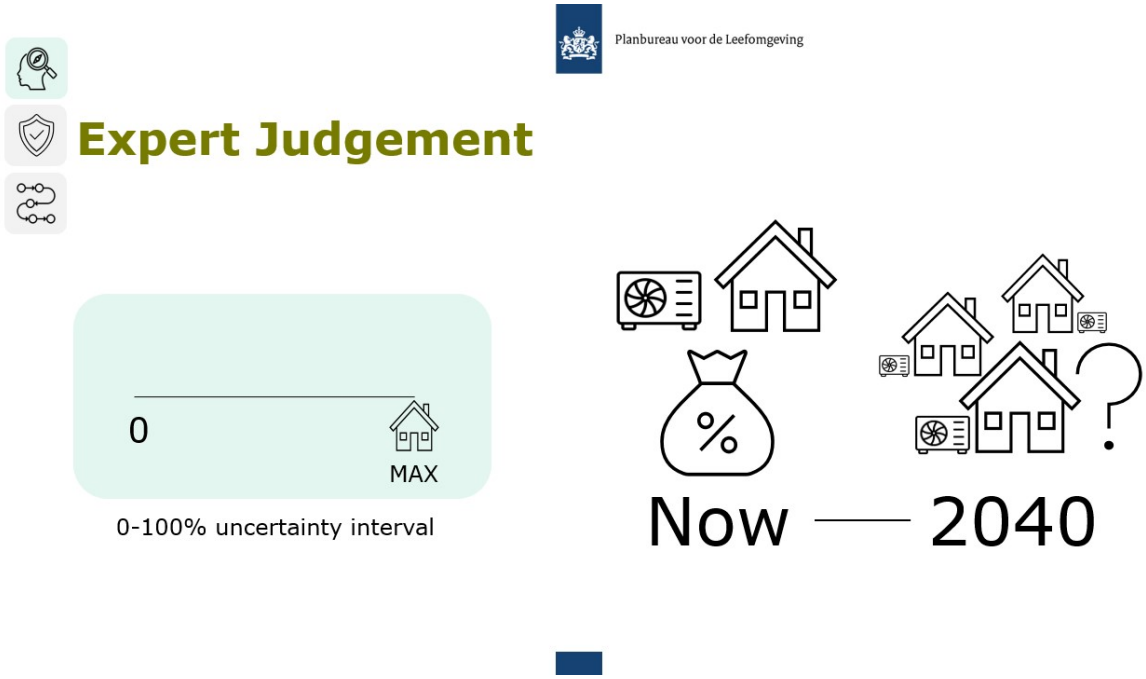
Voiceover: *How can you credibly use your expert judgments as a source of data? This video will discuss the following principles of expert judgement protocols: How you reflect on your judgement under uncertainty; How you can scientifically use your judgment as a source of data; And how you can explicitly treat uncertainty throughout the entire process.*

Grab attention	The introduction has grounded the message in a personal, organisational and theoretical context. Now this has build to the rhetorical question the viewer can expect an answer to in this video.
Manage expectations	By presenting the structure of the message, the viewer is guided to what they can expect the rest of the video to discuss.
Build on existing knowledge	The structure is not based on the procedure of expert elicitation, but favours building on the most familiar to less familiar perspectives: starting with making a judgement, building to a protocol (some are already familiar, others not, but the concept of such protocols has at least been presented in the introduction) and ending with actionable steps to implement theory on uncertainty that are relevant throughout the entire process.
Visual cues	The structure of the information is not only presented in the voice-over, but the visual structure is also introduced. The logo's and colours are continued throughout the next slides, providing the viewer with visual assistance to track the structure and support memory.



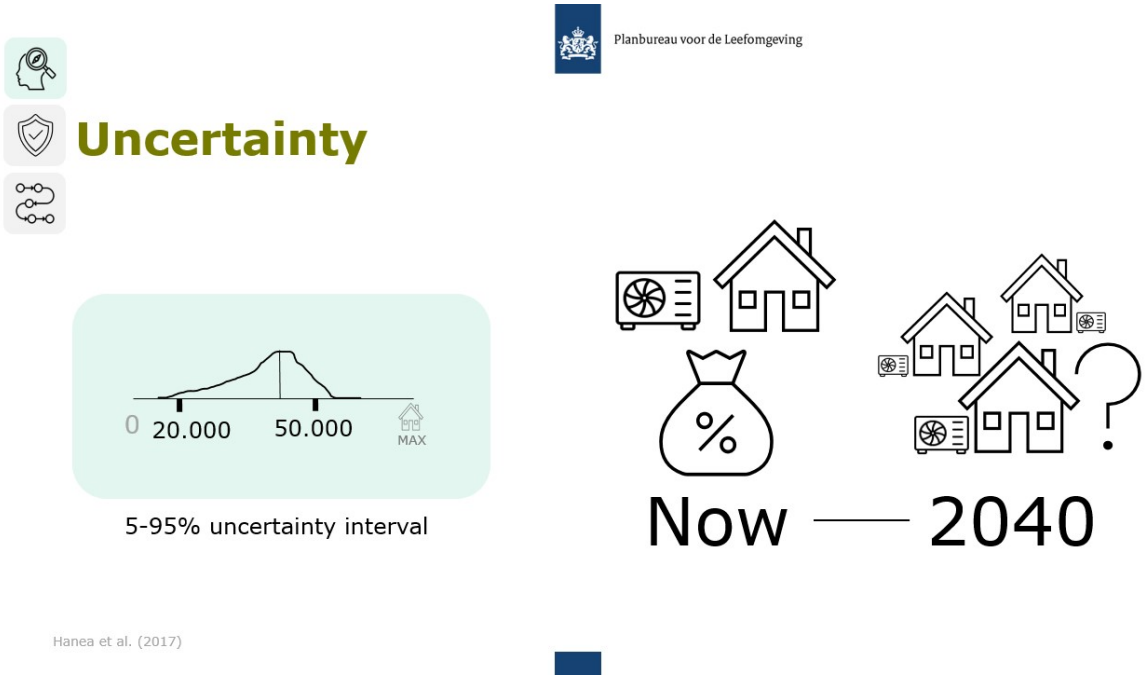
Voiceover: *Now, imagine you need to estimate the emission effect in 2040 of a new heatpump subsidy scheme for households. We need to assume a unit to express our answer. We will assume the expected effect is the number of additional households getting heat pumps because of the subsidy that otherwise wouldn't.*

Tangible example	Instructions are better understood when they are grounded in tangible examples. This example is based on a real uncertainty, but the numbers are hypothetical.
Engage	The viewer is addressed directly and prompted to 'imagine', improving the engagement in learning.



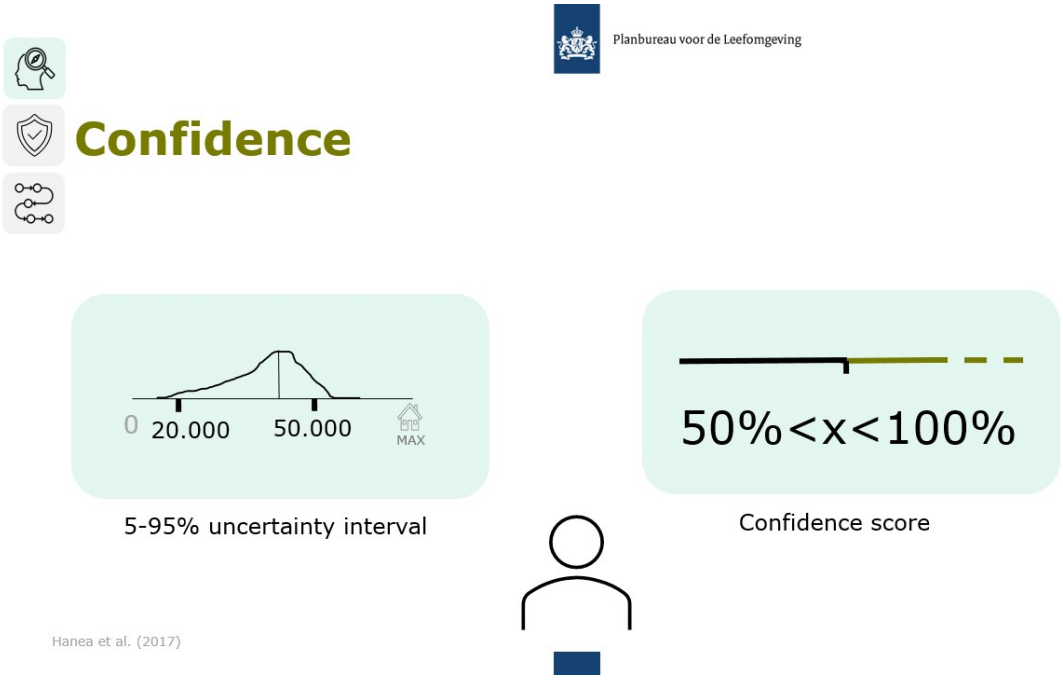
Voiceover: *By 2040 so much could change. A simple approach would be to say the effect lies somewhere between 0 households and the maximum that the budget allows for. This is our 0 to 100% uncertainty interval, taking into account all uncertainty, but what information can readers and policymakers gain from this judgment?*

Actionable instructions	The viewer is guided step-by-step through the thought process.
Rhetorical question	The viewer is inspired to question the process as presented, which might in some ways be similar to their own process but at least invites a critical look.



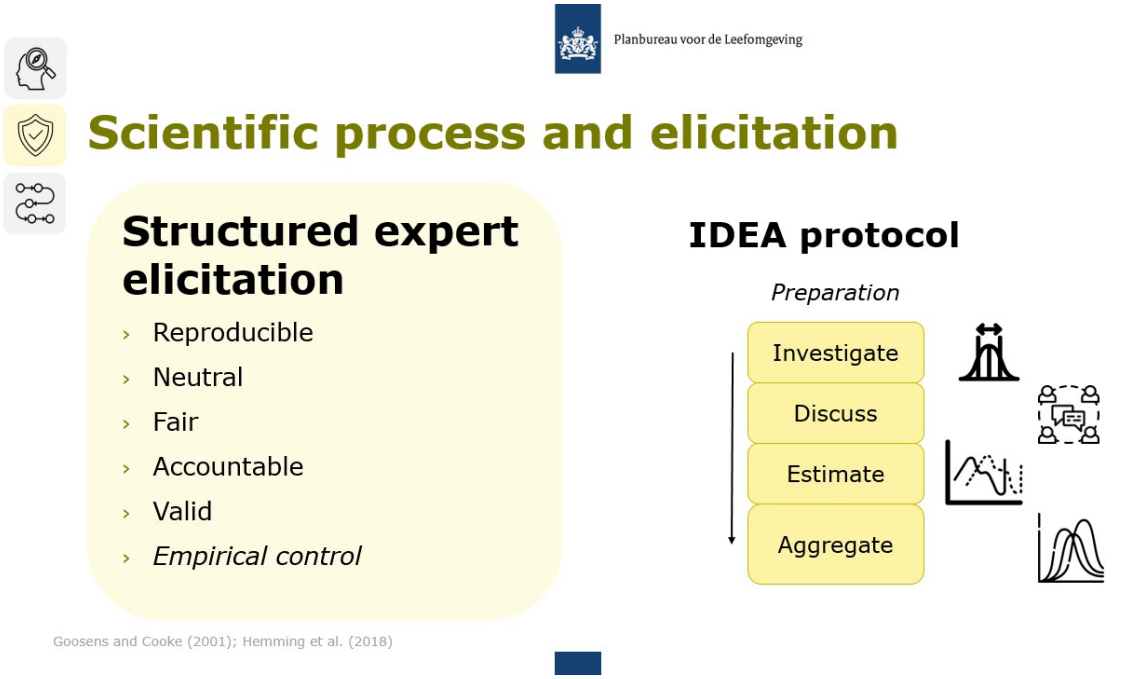
Voiceover: *To leave out the most unlikely values, you are expected to look at the most plausible lower and upper value. This will result in a 5 to 95% interval. As an expert you might assume based on data and your judgement a realistic lower bound is 20.000 households. You might similarly conclude a realistic upperbound of 50.000. After expressing your uncertainty, what distribution do you expect, and what could be a realistic best-estimate of the true value?*

Actionable instructions	The viewer is guided step-by-step through the thought process. The steps are introduced visually in sequence along with the voiceover.
Engage	The viewer is not only following along, butlearning the process is eased by thinking about the answers to the process questions.



Voiceover: *Reflecting on your confidence improves your judgement. For example, you conclude you are less than 50% confident that the true number of households in 2040 lies in your interval, that means you are more confident the true value lies outside your interval. You can expand your bandwidth to adjust. Should you conclude you are 100% confident, you may decide to choose a slimmer interval that you are still confident about. Your final interval will thus be a balance between confidence and how much information it can provide for policy.*

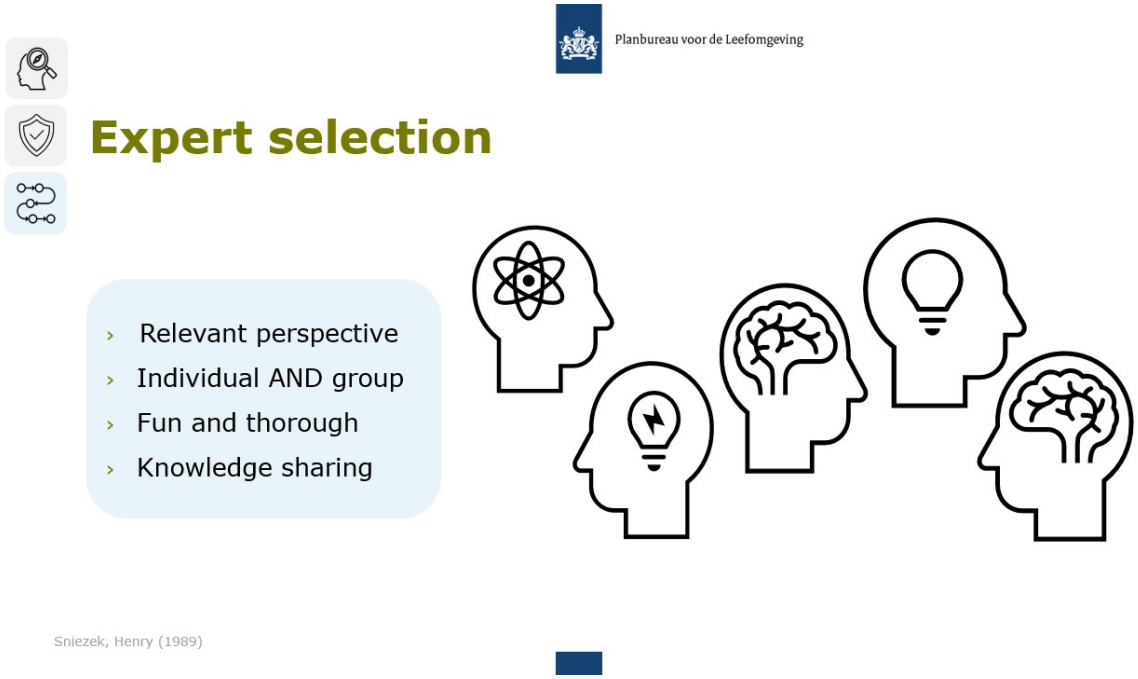
Actionable instructions	The viewer is guided step-by-step through the thought process. The steps are introduced visually in sequence along with the voiceover.
Multimodel communication	By combining voiceover with visual representation of the distribution and confidence score, the information is presented in multiple ways. The visual representation also stresses the distinction between the two concepts (uncertainty and confidence).



Voiceover: *To elicit and collect such judgments according to the scientific standards, protocols like the ones mentioned earlier were developed. The protocol discussed in this video is the IDEA Protocol developed by Hanea et al. (2017). Multiple experts are recruited who independently* ***investigate*** *the problem and make an individual estimation of the 5-95 interval and best guess percentiles. Their anonymous estimations are shared and* ***discussed****, after which all participants can adjust their* ***estimates*** *based on new insights.All judgments and rationales are documented and* ***aggregated****.*

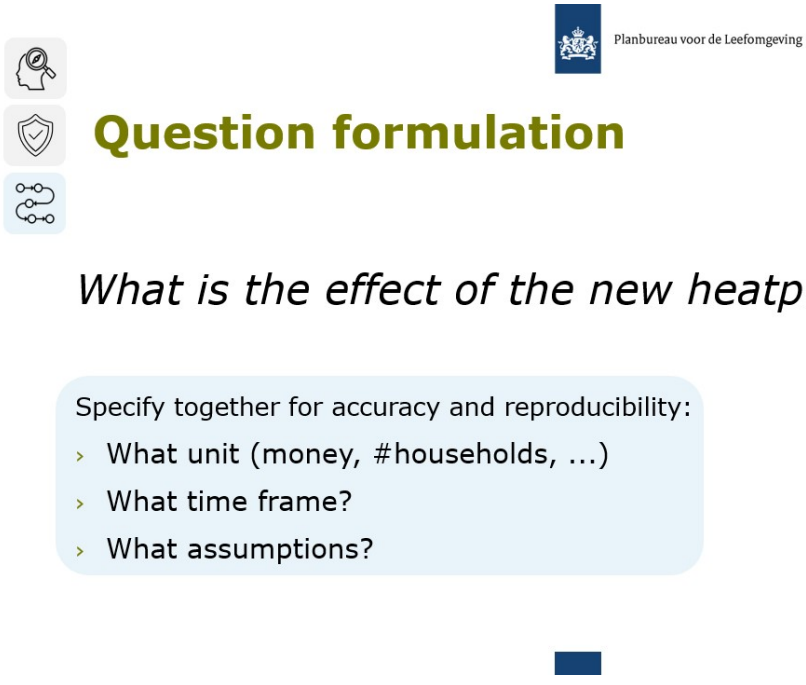
We see that it supports the scientific standard of data collection. The structure of the protocol supports ***reproducibility****. Independent estimates minimizes social bias, providing a* ***neutral*** *process. Incorporating diverse expert perspectives and disagreement preserves a* ***fair*** *process.* ***Accountability*** *is ensured by documenting assumptions and reasoning, allowing judgments to be scrutinized. By working together on problem statements, experts ensure the elicitation process is* ***valid****.* ***Empirical controllability*** *is the only step that may not be feasible. It is most likely new policies will influence the context in the future but projections are made from a status quo point of view.*

Inform	This slide connects two important pieces of information: the scientific standard and the protocol for expert elicitation. The different parts are introduced piece by piece (in voice-over and animation on the slide). It shows a protocol is more than a series of steps, but provides value to the scientific quality.
Visual support	The steps of the IDEA protocol are stressed by graphical representation.



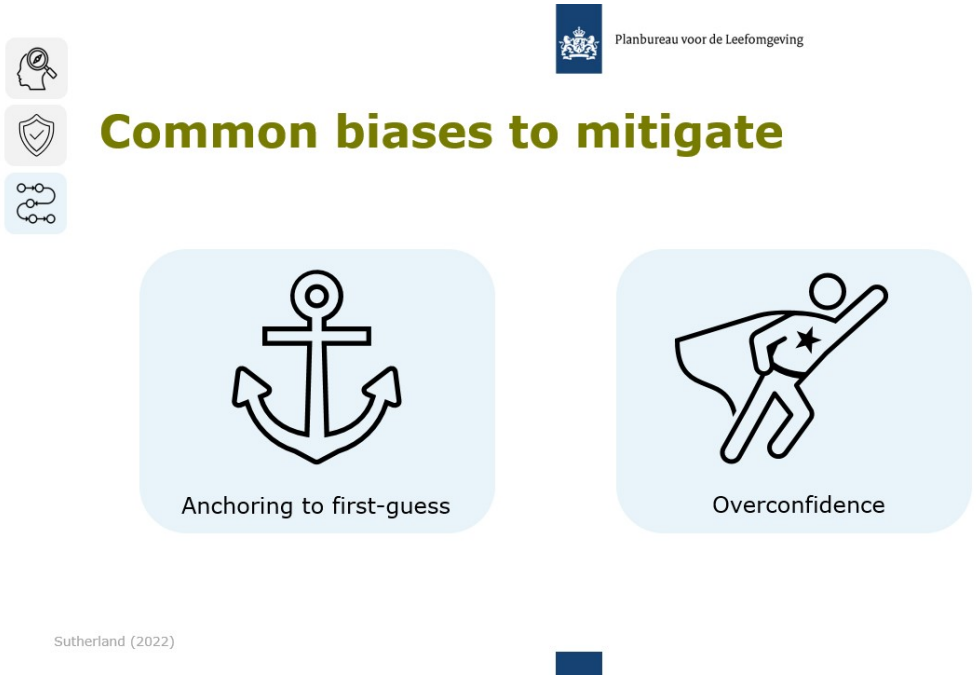
Voiceover: *As mentioned, studies show that groups of experts who do not have perfect knowledge often have fairly accurate results when compared to best performing individuals. So even if you would not consider yourself the best expert on heat pumps, you can contribute if you bring a relevant perspective, for example from behavior or energy markets. Combining individual and group processes mitigates groupthink. Working with colleagues also makes the process more fun and thorough as it allows knowledge sharing.*

Inform and Reassure	Though engaging more experts is important, the viewer is reassured of the value of their individual perspective.
Liking	An emotional appeal is made to the viewer by linking the information to information of a new approach to the idea of a more fun approach.



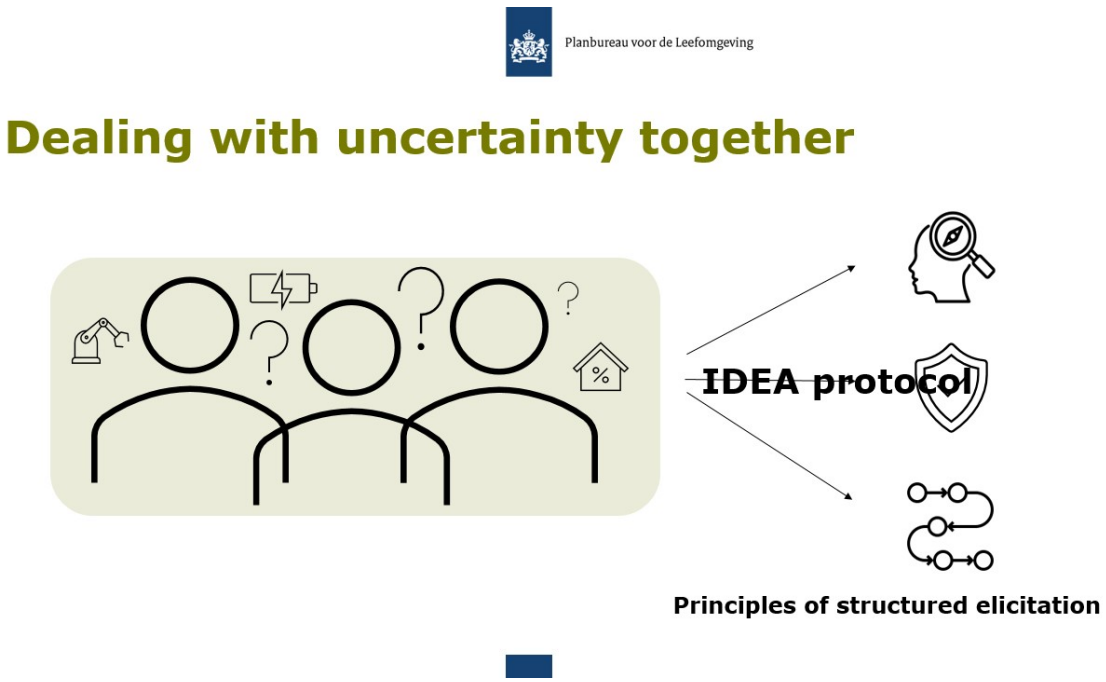
Voiceover: *Needing to judge the effect of the new heatpump policy, you may ask: What will the effect be? This lacks specificity. A well-defined question is very important for accuracy and ease of your judgement. Specifying the question with your colleagues creates a common understanding of the problem. A more general problem framing with assumptions across all problems could support standardisation and reproducibility.*

Deeper level of learning	The previous section focused on how the viewer applies their expertise, but now they are asked to reflect on whether they are even asking the right question.
Connecting to personal experience	When new uncertainties arise, experts look for guidance. However, here the idea is posed that the experts can look to themselves not on how to answer the question, but on how to ask the right question: structuring the problem, which also improves reproducibility. New situations rarely come with instructions, therefore it is important that new problems are first structured well.



Voiceover: *A structured protocol helps to mitigate some common biases, availability bias and groupthink are addressed by our earlier steps. Additionally there is anchoring, where experts do not explicitly consider new information as they are prone to stick to their first guess. Maybe you have to judge the same variable year after year. Therefore consider your uncertainty before concluding a best-guess. Experts are also often overconfident. Explicitly considering uncertainty upfront and expressing confidence as shown to as shown to decrease overconfidence.*

Inform	Common biases are introduced and explained.
Actionable examples	Building from theory to practice, tangible steps are presented that the viewer can take to address the biases.
Deeper level of learning	The viewer is prompted to question their own assummptions and biases, prompting a deeper level of reflection which builds on the earlier steps (first how they make a judgement, then whether they ask the right question, and now what assumptions and biases they bring to their approach).

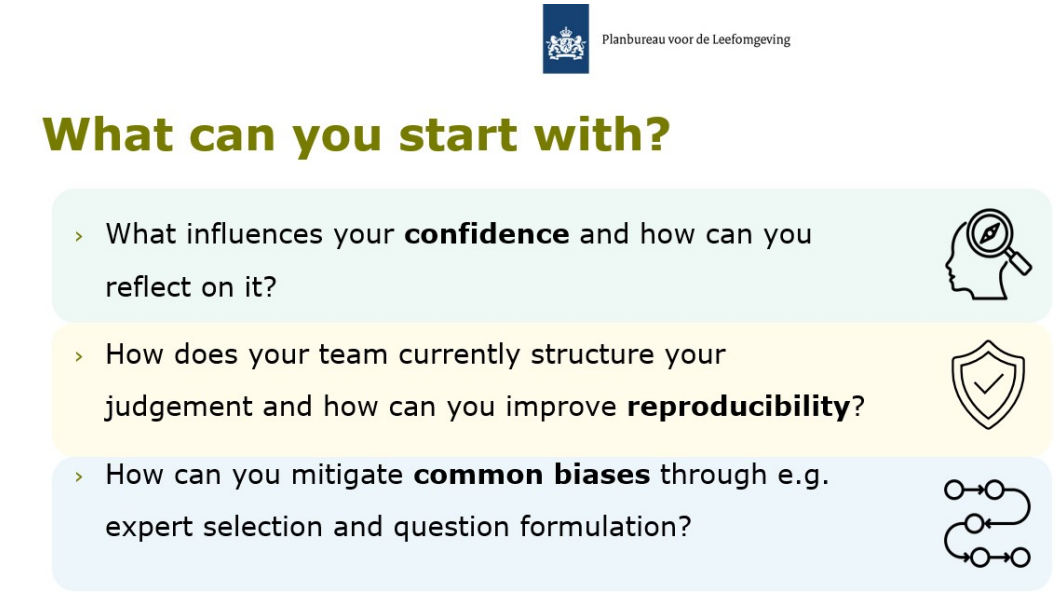


Voiceover: *Dealing with unknowns is part of your work. Most new situations will not come with instructions on how to deal with them. Together with your colleagues you can approach and reflect on your own uncertainty and improve the quality of your judgements. Compare assumptions with other teams and approach other expertises for discussion.*

We've observed in the pilot with the IDEA protocol in 2025, that structured expert elicitation strengthens knowledge sharing and can lead to new insights. Because of these positive outcomes, in time, more teams will be implementing the IDEA protocol.

However, following the principles of structured expert elicitation, you can take action and feel assured that you are addressing uncertainty. Together, we invest in the future quality of research at PBL. What can you start with today?

Connect concepts	A visual (animation on the slide) and verbal transition is made, mirroring the transition in the introduction, between the elicitation protocol and the principles that are applied.
Engage short-term	Showing the steps being taken by other teams and individuals in the organisation makes use of the bandwagon effect to engage the viewer in taking steps themselves.
Engage long-term	Bridging from the organisational perspective (company values) to the personal interest in learning and adapting (professional responsibility and challenges) supports motivation for long-term change.



No Voiceover

Mono-modal communication	The viewer is left without voiceover to internally read and ask themselves the questions on the screen.
Apply learning	Questions are posed that engage the viewer in applying the knowledge they've gained from the video immediately. This supports learning and retention.
Invite reflection	The last questions on screen send-off the viewer with questions they can discuss with their colleagues.
Visual continuity	The last questions are based on the earlier structure and principles, supporting retention of the information.

6.9. VALIDATION AND EVALUATION

The main intended receivers of the communication are firstly sectors experts in the KEV project at PBL, and secondarily the project and team leaders. I contacted one person from each category to review the video. I intend to validate the design by checking whether the communication goals were met and whether I did so effectively. I asked them:

- What are your initial ideas after watching the video?
- What is your take away from this video?
- The goals were ..., do you think that I'm able to achieve those?
- What could I have done differently? (e.g. are the examples clear? are the graphics clear?)

The participants selected to validate the communication design were contacted based on convenience are kept anonymous. They have however previously participated in this research and were therefore already familiar with this research and its results. Though on the one hand this provides an informed perspective on what may be missing, the information provided is not expected to be new to them. This validation could therefore be improved by also contacting new participants.

6.9.1. SECTOR EXPERT

First thoughts:

- Overall it's a nice video, though it has a few technical hiccups.
- The tone is quite corporate or educational, but that's good if you're going for that.
- The structure seems to be going back and forth randomly, maybe this can follow chronological sequence?

Great:

- Very informative
- The ending is great. Right tone, good info.

Improve:

- The overview of the content of the video.
- More clearly state who the video is for
- Can leave out scientific references in speech and text.
- The protocol structure and scientific bases can be moved to the beginning
- Though some things might be informative, they are not relevant to this case and could be left out, whereas other pieces of info could be made more relevant by contextualising them.

Missing:

- Some parts are missing that are very important in this context: formulating list of uncertainties for example
- Motivation for the project overall could be started with, something like: "the EJ process is experienced as unstructured..."

6.9.2. PROJECT LEAD

First thoughts:

- Explain empirical control
- Show the good examples on screen, not only the bad cases (e.g. give an example of good question formulation)

Great:

- It's informative
- Good reassuring tone at the end
- Good practical examples, but can also be more active at calling to action

Improve:

- It must be more clear for who the message is and with what goal
- If the intention is to inspire discussion, say so explicitly

Missing:

- What is the list of criteria that EJ protocol fulfills?
- Ask the reflection questions in the end out loud, with voiceover for clarity.

6.9.3. EVALUATION

Though the conclusion of my research comprises of what the challenges are of implementing SEE in the context of policy research, I'm suprised to say that these challenges and how to address them were ineffectively communicated. Looking back at the design process, I focused on communicating many research results and SEE instructions, leaving out the insights that were gained from my own research.

Though the ending of the video was viewed as a strong ending, the participants underlined that the ending could use voice-over, asking the questions to the viewer out-loud. For the sake of time, I chose to end the verbal information earlier than the visual information, allowing the viewer to pause the screen to read the questions. However, this may leave viewers confused because they had too little time to read.

What I have improved for the final design:

- Firstly, I changed the intro, starting with addressing the audience and goal of the video more specifically. Though I intended the video to be more widely applicable, I should state the primary audience and what's in it for them.
- Though the current sturcture is based on familiar to new information, based on the feedback I structured the content of the video in chronological sequence that the teams have to follow when performing expert judgement.
- The information on scientific basis is moved to the front. Though I opted to provide it as an example of a protocol for EJ, the context of the audience will primarily be users of the IDEA protocol. Moving this up front will provide a better overview of the structure.
- I emphasize the unique context of the KEV, based on the ideas of the experts in the codesign workshop in Chapter 4. Moreover, the specific challenge with limited feedback on the accuracy of expert judgement is highlighted when discussing the IDEA protocol in relation to the scientific quality.

7

Conclusion

This research aimed to address the question: *How can practitioners in the policy research field be supported in implementing structured expert elicitation to improve their expert judgement?* The focus lay on expert judgement for quantifying uncertain variables. Given that studies have already shown structured expert elicitation (SEE) improve accuracy of expert judgement, I chose to focus on the barriers to implementation in practice and how the process of expert judgement can be supported in a new context: modelling for policy assessment.

I approached answering the research question by using a design-based method: a double-diamond research design *Discover-Define-Develop-Deliver* (Figure 1.1). By means of an internship I studied the expert judgement process in the Climate and Energy Outlook (abbreviated as KEV in Dutch) at PBL Netherlands Envrionmental Assessment Agency. This led me to specify my research question:

How can experts at PBL be supported in implementing structured expert elicitation in the KEV project to improve their expert judgement process?

SQ1: What is currently known about expert judgement in theory and in practice at PBL (for the KEV project)?

Though unstructured approaches are generally considered unscientific, in practice at PBL, expert judgement is experienced as unstructured, with too little documentation and dependence on few experts. A focus group evaluated their current practice and a pilot with a structured expert elicitation protocol. The results indicate that a new protocol like the IDEA protocol can lead to new insights, more structured elicitation, more transparent communication and documentation of assumptions, and overall increase the credibility of the expert judgement process.

SQ2: What challenges are faced by the practitioners in the KEV project in the process of expert judgement? What problems should be prioritised considering the organisational context?

Focus groups with PBL actors and a co-design workshop with KEV sector experts of different teams revealed diverse perspectives on problem discovery and definition. While structured expert elicitation can enhance judgment, practitioners highlighted a persistent gap

between theory and practice in policy assessment, complicating implementation. PBL actors prioritize standardization and transparency, which could be addressed with a protocol. However, participants noted that the IDEA protocol for expert judgment fails to meet their practical needs.

- Practical restrictions like work pressure and a strict modelling process limits available time and flexibility to spend on the expert elicitation process. The organisational structure has also led to isolated teams that have diverging approaches to expert judgement.
- Additionally, sector experts experience feelings of false certainty. This may be explained by a lack of guidance on what should be included and excluded in uncertainty intervals. Specifically, SEE theory doesn't address what uncertainties to incorporate in research, and how to align the problem demarcation across teams.

The focus of the problem definition therefore lies in the lack of support for experts, both because of the limited connections across teams which inhibits organisational learning, as well as lack of appropriate tools to confidently address new uncertainties.

SQ3: WHAT (COMBINATION OF) SOLUTION(S) COULD EFFECTIVELY AND FEASIBLY SUPPORT THE PRACTITIONERS IN THE KEV PROJECT IN THE CHALLENGES THEY FACE IN THE PROCESS OF EXPERT JUDGEMENT?

A solution should help bridge the theory-practice gap by distilling actionable principles that recognises the challenges of implementation by giving the sector experts autonomy but also provides scientific basis for credibility.

The final communication design is an instructional video. It allows the uncertainty team to instruct and inspire sector experts in the short term on the adaptation of the theoretical principles of the IDEA protocol through tangible examples and actionable next steps that adapt to their unique context and challenges. Experts are reassured that they can make credible judgements bacause the methodology mitigates common biases. Challenges of implementation in policy context are recognised, but in the long term researchers should be inspired to discuss these challenges and learn together outside their siloed teams.

GENERAL CONCLUSION

The quality of expert judgement is often measured based on accuracy, but in the KEV project validating estimations is not possible. In such cases the process of expert elicitation could an provide important scientific basis. Overall, these findings show that though structured expert elicitation can support quality of expert judgement, these protocols require tailoring to the specific research domain and context.

This study provided a new perspective on the implementation of expert judgement protocols. Though several studies outline measures for practical implementation, this case study is unique when it comes to the policy research domain, the small number of experts, and the large number of judgements that are required. Moreover, this study explores the unique challenges that experts face in the context of projecting policy effects, like guidance in selecting sources of uncertainty that need estimation and support in making judgements

with confidence, and provides ideas for how these challenges could be addressed in the instruction of expert judgement.

By eliciting the experiences and ideas from multiple actors involved in the expert judgement proces, we gain a more complex perception of the challenges of expert judgement practice. Moreover, by embedding the research in an internship, I got to observe and experience the work structures and culture. This provides insights into day-to-day practice no (grey) literature can provide. This includes the complex reality of the human role in modelling, the interaction between teams, and the organisational structure.

Recommendations

The participants in this research indicated that they sometimes feel they present false certainty. Other participants indicate experts are very uncertain (choosing large uncertainty intervals) and are underconfident in making more specific judgements (slimmer, more informative intervals). Research so far has focused on overconfidence as a common bias, but future research could explore how underconfidence plays a role in policy assessment. This is especially relevant as decision-support is often targeted at decreasing uncertainty or gaining information despite uncertainty. Additionally, practical tools could be developed to provide experts with feedback or support in cases where they cannot get feedback on the accuracy on their judgements.

This research may provide interesting insights for decision-making under deep uncertainty (DMDU). Tools are developed to support decision-makers to grapple with uncertainty and make decisions that ackowledge uncertainty and are effective despite the uncertainty. As presented in Bolhuis (2025), in the end political actors are responsible for decision-making and make use of the information about uncertainty to gauge what policies could be effective. However, the participants in this study show that experts that are enlisted for their expert judgement may struggle similarly to balance acknowledging uncertainty with providing policy-relevant insights. It would be interesting to futher analyse what problem demarcation (and therewith all underlying assumptions) provide a policy-relevant perspective, and how a realistic view would differ.

This research has highlighted expert judgement in a policy context. More research can be done in this domain to look into whether the lessons from this study can be translated or generalised to other cases. Specifically, domains where experts cannot get feedback on their judgements, or projects where few experts are required to make many estimations are interesting because they look into contexts that the classic protocols do not address.

The experts indicate, and protocols stress, that the reasoning behind judgements are very important and should be documented. Though experts discuss these rationales, when having to discuss and document many judgements it would support these experts to be able to describe these rationales in a standardised way that increases transparency for outsiders. Future research could focus on developing practical tools to convey such rationales.

For practitioners, there is a need for a targeted discussion on what research quality can be achieved in projects like the KEV, particularly regarding the management of uncertainty and the value added by expert elicitation. While expert input can significantly enhance

credibility, extensive elicitation is resource-intensive and often contributes only a small fraction to the overall project. In practice, it is crucial to weigh the benefits of structured expert elicitation against the costs, especially when resources could be allocated to other critical projects.

Furthermore, the assumptions made during model creation may have a more substantial impact on outcomes than the expert judgments themselves. However, due to the complexity and scale of the model, conducting a comprehensive sensitivity analysis is often impractical. Developing methodology to understand the most impactful assumptions could help interpretation and transparency of similar projects.

8

Discussion

This study set out to explore the practice of expert judgement in the policy analysis in order to define and address key challenges with the implementation of structured expert judgement theory in this volatile context. The outcomes of this research show unique limitations to the applicability of structured expert elicitation protocols in the case of the KEV project at PBL. To address this gap, a communication product is developed to support the researchers to improve their expert judgement process.

The results from the focus group and codesign workshop that inform the problem definition were discussed in Chapters 3-5, where they are contrasted to other studies. This section focuses on discussing the limitations of the findings and generalisability of the results.

Limitations

The literature consulted for this research was focused on expert judgement from the perspective of quantifying uncertain variables. It is important to mention that though I point out the limited research into the implementation of support tools for expert judgement (like SEE protocols), there are other types of expert judgement in many areas of research. Prominently epidemiology research features expert judgement to estimate risks (Walker et al., 2001), and the healthcare sector addresses how medical experts combine clinical guidelines and expert judgement in medical decision-making (Kea & Sun, 2015). Perhaps these domains could provide more insight into how to support expert judgement and confidence in expertise. For example, Walsh et al. (2021) present a tool for medical students to reflect on their confidence in their expertise. Because of the different types of expert judgement and varying terminology it was hard to find inspiration that I can transfer to this study. But if anyone would like to expand on this study, this would be the next area of research I would say deserves more attention.

The first focus group was centered on evaluating a pilot project where one team of the KEV project at PBL implements the IDEA protocol. However, this implementation turned out to not be ideal. I supported the uncertainty team with instructing the sector team on how to address implementation of the individual expert judgement and the group discussion phase. This was very new for the team. Usually they would divide responsibilities for

expert judgement over the team members and discuss the estimations over multiple meetings. They would end up implementing some parts of the IDEA protocol, in parallel to their usual process. The results from the evaluation focus group would therefore address only a limited experience with implementing structured expert judgement. Ideally a second pilot would follow, where the same sector team would be instructed to implement the IDEA protocol in their workflow, not in parallel. However, though some limitations of the results could be addressed by better instructions in a second pilot, the participants of the focus group point out challenges that would be relevant, especially if the protocol would be followed more precisely. Therefore I think most of the results of the focus group would likely be replicated if the pilot would be executed more precisely.

Because the focus group highlighted challenges with the IDEA protocol, it was important to address the challenges currently faced without a protocol. The codesign workshop with sector experts from different teams was initially designed to specifically discuss what influences their confidence in expert judgement and involve them in designing a solution that would support them in their expert judgement process. I specifically did not focus on accuracy because, as this thesis discussed, the accuracy is not feasible to check. However, the workshop would raise more problems than solutions. It turned out that the experts had rarely talked to one another about their judgement methods. Therefore on the one hand this workshop did not go into enough detail to actually function as a codesign excercise, but rather functioned as an new perspective on what problems exist.

VALIDITY

The participants of both the evaluation focus group and the workshop were requested to write out their contributions. On the one hand hand this allows them to formulate their answers after thinking it through. However it also takes time and may not capture the nuance and additional remarks that come up throughout discussion of these contributions, these had to be noted separately as my observations and therefore more dependent on my interpretation. Recording these meetings could've provided a rigor to the research to capture specific remarks and show a clear distinction between the contributions of the participants and my interpretation as a researcher. However, because of practical reasons (many participants, lots of coding work) I opted not to record and code the meetings.

GENERALISABILITY

Quantifying uncertain variables using expert judgement is common practice in many fields. However, this study was tailored to discover the challenges this specific group of researchers face in this project context. This limits the generalisability of the results. The study's findings are shaped by the organisational structure, the available experts, team dynamics, and the timeplanning and available resources for the project that requires expert judgement.

Nevertheless, several insights offer lessons for other research projects that use expert judgement in projecting policy effects. For instance, the importance of problem demarcation as an integral step in preparing expert judgement can likely be meaningful in other contexts. Clear problem definition helps align teams, assumptions across different estimations, guide question formulation, and provides transparency.

Moreover, other research domains would probably also experience limited options to

verify the accuracy of expert judgements. In such cases, the quality of expert judgement might be better assessed through the credibility of the process instead of the accuracy of the outcomes. Ensuring the process is reproducible, clearly documenting rationales and including diverse perspectives are important for scientific quality.

9

REFLECTION

INTERNSHIP AND INDEPENDENT RESEARCH

For my master's theses, I set out with the goal for my research to mean something. I consider myself fortunate that I got to apply my research plan at an actual research institute by means of an internship at PBL Environmental Assessment Agency in 2025. For the duration of this internship I got to work in the office, converse with colleagues of different departments and positions, and get to experience the KEV project from the inside.

Most of my work was for the pilot project with the IDEA protocol. I prepared meetings, collected input from different actors, wrote instructions and created presentations, sat in with a team for the duration of their expert judgement process (observerd during several meetings), organised the evaluation and analysed the results. The internship provided me with access, time, and involvement of participants and experts that provided crucial input for this research.

I must however also critically reflect on the impact this position had on my research. The internship was paid, though not a full salary. I do not think this impacted my research. I have to do a master's research anyway and the pay does not depend om my research results.

What I do think impacted how I conducted my research is my position as a student. PBL is an agency that in my eyes is a prospective employer. I think this definitely impacted how I positioned myself at the organisation and the people I sought out to talk to. I do not think this impact was negative. If anything it caused me to spend more time on activities that were not mandatory for my thesis which allowed me to build rapport with my colleagues.

However, it's the rapport I built that has influenced my observations and how I present my opinion on the organisation and the research they do. I have the greatest respect for the researchers and staff at PBL. The work they do is so complex, especially when you get into the details of the computer models and how they are managed. Probably my written account of their work in this thesis still falls short of capturing the true complexity. I have caught myself at several moment thinking of not writing an observation, specifically of what I think are points for improvement, because I was afraid of being wrong.

To address this challenge I think it was valuable to discuss my ideas with supervisors and experts both within the organisation, and outside. This provided opportunity for correction from different perspectives, but also support to have the confidence to bring up criticism.

Action research is an important method in social science. Learning about the true practice of researchers by working along-side them provided insights I would not have gotten otherwise. The position of the researcher is challenging.

Iterative design and independence

The Council (2004) explicitly noted that including different perspectives, gathering inspiration, and testing designs with different audiences is part of the double diamond method.

When I was a bachelor student of mathematics, I mostly worked on my own, there were very few group projects. I valued my personal accomplishments, and have always been driven to present what I can contribute. During the CDI (Communication Design for Innovation) master as well as the EPA (Engineering and Poilcy Analysis) master I was part of groups. Because of my beta focused bachelor I spent a lot of time on programming in modelling projects, and still had a lot to learn about creative processes.

I felt like my master thesis was the time to show what I as an individual researcher can accomplish. Independence is one of the criteria we're graded on. The downside of this striving for independence is that I became perfectionistic from the first meeting. I wanted to present others with ideas and products that were finished. If I would've sought others' opinions earlier on in the process this could've saved time on iterations later on.

My first supervisor provided me with a valuable insight on this challenge. Independence does not mean every thought needs to be compeletely originally produced by me. Firstly, I'm still learning to perform research, and we develop opinions through discussion with other perspectives so it's especially not bad to incorporate ideas from someone who may be more of an expert on the topic. I can show independence by reaching out independently, by providing my own perspective and actively asking for input from people with interesting and relevant expertise. This shift clearly shows later in my research, where I became less afraid to use my intuition to inform choices, but sought input from others to reflect on these decisions.

In CDI we have a course called personal professional development. We learn to reflect on who we are as a person, as a professional, and as a part of a larger team and organisation. One of the learning goals I defined was to connect with others on eachother's expectations, and develop more confidence in having and presenting my opinion.

What I take away from this experience is that it shows I can work independently, though I can connect with peers and colleagues. I can be part of a team, but I can also offer critique on what isn't going well. I think it's good I can change my opinion and to not be afraid to offer it to be scrutinised.

THE VALUE OF THE DOUBLE DEGREE (EPA AND CDI)

When applying to perform a double degree to combine Engineering and Policy Analysis (EPA) and Communication Design for Innovation (CDI), I had to write a motivation letter about the value of combining these perspectives. I focused on the relevance and overlap of the two studies and how they would be an extension of eachother.

At the start of my master theses, I did not have planned out which project would be for which study. They both have a social science, but also a policy analysis aspect. My combined background led me to quickly integrate my knowledge on policy analysis practice, but also the methods I learned from a science communication perspective in both theses.

Looking at this thesis specifically, the idea for the research question came from my personal experience in policy analysis. I've often had to make assumptions for model parameters and extrapolate from input data. I wanted to know more about this practice by professionals, how this data informed by my estimations can be used as scientific data and what this means for the research results. However, diverging on what problems definitions were possible, I saw complex interactions between the modelling practice, and the social and organisational structures. Where I could've focused on developing a method connected to the modelling process, I ended up designing a communication product that addresses the experience of the practitioners, informs and inspires them to take action. This doesn't mean modelling methods are not also important, but my CDI lens allowed me to tackle this problem from another perspective has gotten little attention so far.

I incorporated my own intuition, insights from both theory and practice, and finally got to design a prototype communication product that I'm proud of. I've expanded my horisons as a researcher, developed capabilities across disciplines, but also learned to integrate the perspectives truly. Though the conclusion of my journey in this double degree did not end up with one integrated project, I think I instead ended up with two. Though I write about two perspectives, and I will have finished 2 studies, I have developed a transdisciplinary perspective. I look forward to further developing myself as a transdisciplinary researcher and value all the steps, challenges, and opportunities that got me there.

BIBLIOGRAPHY

Anca hanea: The idea protocol for sej. (2022, November 29). Retrieved March 28, 2026, from https://youtu.be/lbQ0J4key3U?si=h9Bcc740lGHnzT34

Babbie, E. R. (2016). *The practice of social research* (Fourteenth edition). Cengage Learning.

Baustert, P., Othoniel, B., Rugani, B., & Leopold, U. (2018). Uncertainty analysis in integrated environmental models for ecosystem service assessments: Frameworks, challenges and gaps. *Ecosystem Services*, *33*, 110–123. https://doi.org/10.1016/j.ecoser.2018.08.007

Blythe, J. (2005). *Essentials of marketing* (3. ed). Pearson Education.

Bolhuis, T. (2025). *Perceptions of uncertainty on the science-policy interface* [Master's thesis, Delft University of Technology]. https://resolver.tudelft.nl/uuid:c691f6b8-feb0-4d45-9126-7f458aab7b43

Burgman, M., Layman, H., & French, S. (2021). Eliciting model structures for multivariate probabilistic risk analysis. *Frontiers in Applied Mathematics and Statistics*, *7*, 668037. https://doi.org/10.3389/fams.2021.668037

Colson, A. R., & Cooke, R. M. (2018). Expert elicitation: Using the classical model to validate experts' judgments. *Review of Environmental Economics and Policy*, *12*(1), 113–132. https://doi.org/10.1093/reep/rex022

Council, D. (2004). *Framework for innovation*. Retrieved October 15, 2025, from https://www.designcouncil.org.uk/our-resources/framework-for-innovation/

Courtney Jones, S. K., Geange, S. R., Hanea, A., Camac, J., Hemming, V., Doobov, B., Leigh, A., & Nicotra, A. B. (2023). IDEAcology: An interface to streamline and facilitate efficient, rigorous expert elicitation in ecology. *Methods in Ecology and Evolution*, *14*(8), 2019–2028. https://doi.org/10.1111/2041-210X.14017

Dainton, M., & Zelley, E. D. (2018). *Applying communication theory for professional life: A practical introduction* (Fourth edition). SAGE.

Dalkey, N. C., & Helmer-Hirschberg, O. (1962). *An experimental application of the delphi method to the use of experts*. RAND Corporation.

Dewulf, A., & Biesbroek, R. (2018). Nine lives of uncertainty in decision-making: Strategies for dealing with uncertainty in environmental governance. *Policy and Society*, *37*(4), 441–458. https://doi.org/10.1080/14494035.2018.1504484

Georges L. Romme, A., & Van Witteloostuijn, A. (1999). Circular organizing and triple loop learning. *Journal of Organizational Change Management*, *12*(5), 439–454. https://doi.org/10.1108/09534819910289110

Hanea, A., McBride, M., Burgman, M., Wintle, B., Fidler, F., Flander, L., Twardy, C., Manning, B., & Mascaro, S. (2017). I nvestigate d iscuss e stimate a ggregate for structured expert judgement. *International Journal of Forecasting*, *33*(1), 267–279. https://doi.org/10.1016/j.ijforecast.2016.02.008

Hanea, A. M., Hemming, V., & Nane, G. F. (2022). Uncertainty quantification with experts: Present status and research needs. *Risk Analysis*, *42*(2), 254–263. https://doi.org/10.1111/risa.13718

Heath, C., & Heath, D. (2007). *Made to stick: Why some ideas survive and others die*. Random House.

Hemming, V., Armstrong, N., Burgman, M. A., & Hanea, A. M. (2020). Improving expert forecasts in reliability: Application and evidence for structured elicitation protocols. *Quality and Reliability Engineering International*, *36*(2), 623–641. https://doi.org/10.1002/qre.2596

Hemming, V., Burgman, M. A., Hanea, A. M., McBride, M. F., & Wintle, B. C. (2018). A practical guide to structured expert elicitation using the IDEA protocol (B. Anderson, Ed.). *Methods in Ecology and Evolution*, *9*(1), 169–180. https://doi.org/10.1111/2041-210X.12857

Kea, B., & Sun, B. C.-A. (2015). Consensus development for healthcare professionals [Cited by: 67; All Open Access, Green Open Access]. *Internal and Emergency Medicine*, *10*(3), 373–383. https://doi.org/10.1007/s11739-014-1156-6

Krueger, T., Page, T., Hubacek, K., Smith, L., & Hiscock, K. (2012). The role of expert opinion in environmental modelling. *Environmental Modelling & Software*, *36*, 4–18. https://doi.org/10.1016/j.envsoft.2012.01.011

Lasswell, H. D. (1948). *The structure and function of communication in society. in l. bryson (ed.), communication of ideas*. New York: Harper & Row.

Martens, A., Arts, B., Bakema, A., Mylius, S., & Porsius, J. (2021). Kwaliteitsvisie PBL 2021.

Mayer, R. E. (2005). Cognitive theory of multimedia learning. *The Cambridge Handbook of Multimedia Learning*, 31–48. https://doi.org/10.1017/cbo9780511816819.004

Microsoft. (2020). *Microsoft teams*. Retrieved March 31, 2026, from https://www.microsoft.com/en-us/microsoft-teams

Microsoft Corporation. (2025). *Microsoft powerpoint online* [Web-based presentation software]. Microsoft. Retrieved April 1, 2026, from https://www.office.com/launch/powerpoint

Miro. (2011). *Miro: Online collaborative whiteboard platform*. Retrieved March 31, 2026, from https://miro.com

Moore, M. (n.d.). Political practice: Uncertainty, ethics and outcomes.

Morgan, M. G. (2014). Use (and abuse) of expert elicitation in support of decision making for public policy. *Proceedings of the National Academy of Sciences*, *111*(20), 7176–7184. https://doi.org/10.1073/pnas.1319946111

Nemet, G. F., Anadon, L. D., & Verdolini, E. (2017). Quantifying the effects of expert selection and elicitation design on experts' confidence in their judgments about future energy technologies. *Risk Analysis*, *37*(2), 315–330. https://doi.org/10.1111/risa.12604

O'Hagan, A. (2019). Expert knowledge elicitation: Subjective but scientific. *The American Statistician*, *73*, 69–81. https://doi.org/10.1080/00031305.2018.1518265

PBL, TNO, CBS, & RIVM. (2024). Klimaat- en Energieverkenning 2024.

Petersen, A., Janssen, P., van der Sluijs, J., Risbey, J., Ravetz, J., Wardekker, J., & Martinson Hughes, H. (2014). Leidraad voor omgaan met onzekerheden. *2*.

Raadgever, G., Dieperink, C., Driessen, P., Smit, A., & Van Rijswick, H. (2011). Uncertainty management strategies: Lessons from the regional implementation of the water framework directive in the netherlands. *Environmental Science & Policy*, *14*(1), 64–75. https://doi.org/10.1016/j.envsci.2010.11.001

Stremler, J., & Boot, P. (2024). Klimaat- en energieverkenningen door de jaren heen. tien jaar nev en kev en de invloed op het beleid.

Sutherland, W. J. (Ed.). (2022, December 6). *Transforming conservation: A practical guide to evidence and decision making* (1st ed.). Open Book Publishers. https://doi.org/10.11647/OBP.0321

van der Meij, H., & Hopfner, C. (2022). Eleven guidelines for the design of instructional videos for software training. *Technical Communication, 69*(3), 5–23. https://doi.org/10.55177/tc786532

van der Welle, A., Hekkenberg, M., Geilenkirchen (PBL), G., van Hout, M., Menkveld, M., Peek (RIVM), K., Plomp, A., van Schijndel (PBL), M., van der Sluis (PBL), S., Smekens, K., van Stralen, J., Traa (PBL), M., Tigchelaar, C., & Wetzels, W. (2017). Achtergronddocument onzekerheden Nationale Energieverkenning 2017.

Video-toolkit. (n.d.). https://www.medewerkers.universiteitleiden.nl/vr/videotoolkit?cf=wiskunde-en-natuurwetenschappen%5C&cd=leiden-institute-of-advanced-computer-science-liacs

Walker, K. D., Evans, J. S., & MacIntosh, D. (2001). Use of expert judgment in exposure assessment part i. characterization of personal exposure to benzene [Cited by: 45]. *Journal of Exposure Analysis and Environmental Epidemiology, 11*(4), 308–322. https://doi.org/10.1038/sj.jea.7500171

Walsh, P., Owen, P., & Mustafa, N. (2021). The creation of a confidence scale: The confidence in managing challenging situations scale. *Journal of Research in Nursing, 26*(6), 483–496. https://doi.org/10.1177/1744987120979272

Welsh, M., Lee, M. D., & Begg, S. (n.d.). More-or-less elicitation (MOLE): Testing a heuristic elicitation method.

Williams, C. J., Wilson, K. J., & Wilson, N. (2021). A comparison of prior elicitation aggregation using the classical method and SHELF. *Journal of the Royal Statistical Society Series A: Statistics in Society, 184*(3), 920–940. https://doi.org/10.1111/rssa.12691

A

INFORMED CONSENT FORM

Consent form

Research title: Supporting expert judgement of uncertain factors

Opening statement

You are being invited to participate in a research study titled "Supporting expert judgement of uncertain factors". This study is being done by Thirza Bolhuis from the TU Delft with supervisor Caroline Wehrmann (TU Delft). The research is part of an internship at PBL (Environmental Assessment Agency) with supervisors Shruti Setty and Astrid Martens.

The purpose of this research study is to explore and evaluate expert judgement elicitation processes, and will take approximately 60 minutes to complete. We will be asking you to elaborate on your experience with expert judgement.

Your answers will be collected in writing. Any data will be anonimised and will be stored on the TU Delft One Drive. The researcher and supervisors from TU Delft and PBL as named earlier will have access to the data.

As with any online activity, the risk of a breach is always possible. To the best of our ability, your answers in this study will remain confidential. We will minimize any risks by storing your personal information and any recordings of the interview (be it written or audio) on a password secured location for at most the duration of the study. Only the researcher and supervisors are allowed access to this data, as mentioned earlier.

The contributions will be used for a master thesis report, which will become publicly available in the university repository of TU Delft. In the report, only anonymised (and aggregate) data will be used, accompanied by a description of the expertise of the interviewees. Below, you can provide consent for this use of the aggregate anonymous data.

You must be made aware that a risk of re-identification is possible because there is only a low number of participants and you have a specialist position. In case of re-identification, personal risks that may arise can include reputational risk as you provide information on your expert opinion.

Your participation in this study is entirely voluntary, and you can withdraw until one month after the focusgroup, but not after the end of this research. You are free to omit any questions. A copy of the contributions will be shared for your inspection, and any answers can be left out until one week after inspection. If you withdraw, your answers will then be omitted from the research and publication.

Corresponding researcher: Thirza Bolhuis –

Responsible researcher: Assistant professor Caroline Wehrmann –

PLEASE TICK THE APPROPRIATE BOXES	Yes	No
1. I consent to the recording and processing of the data provided in this focusgroup as described above.		

By signing this form, you consent to the above.

Signatures

_____ _____ _____

Name of participant Signature Date

I, as researcher, have accurately read out the information sheet to the potential participant and, to the best of my ability, ensured that the participant understands to what they are freely consenting.

Thirza Bolhuis

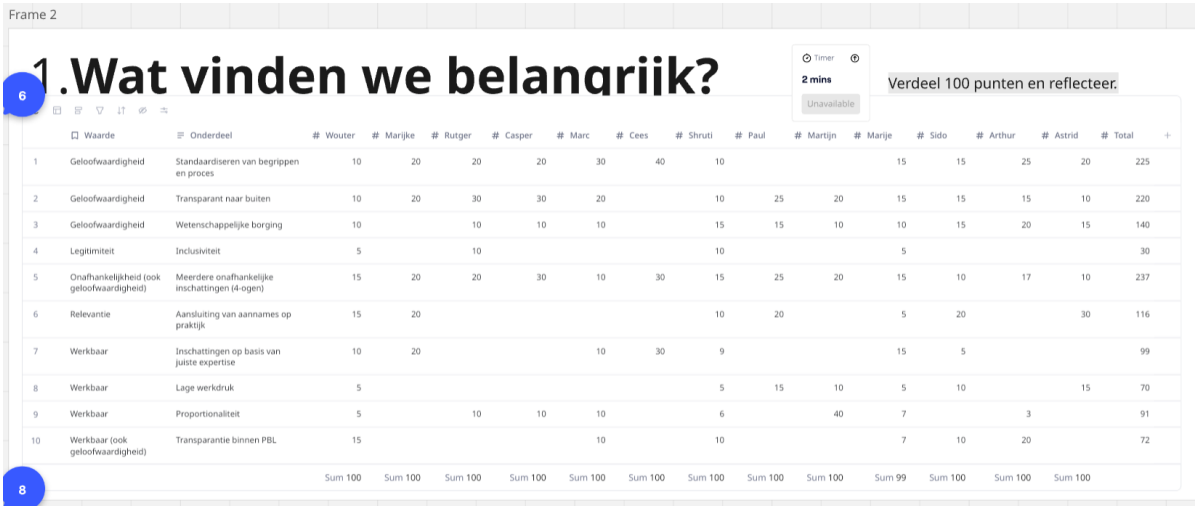
Name of researcher Signature Date

Study contact details for further information:

Thirza Bolhuis

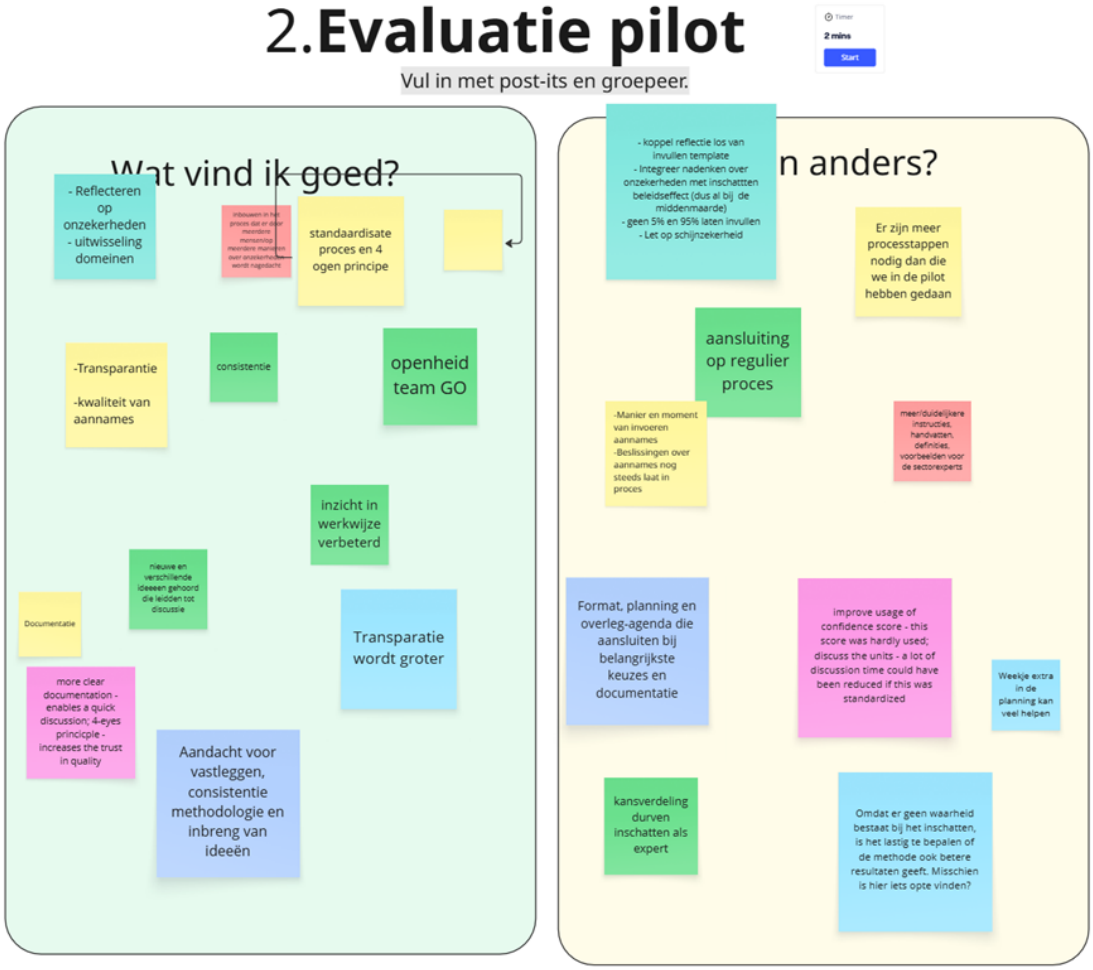
B

Results Evaluation



	Waarde	Onderdeel	# Wouter	# Marijke	# Rutger	# Casper	# Marc	# Cees	# Shruti	# Paul	# Martijn	# Marije	# Sido	# Arthur	# Astrid	# Total
1	Geloofwaardigheid	Standaardiseren van begrippen en proces	10	20	20	20	30	40	10			15	15	25	20	225
2	Geloofwaardigheid	Transparant naar buiten	10	20	30	30	20		10	25	20	15	15	15	10	220
3	Geloofwaardigheid	Wetenschappelijke borging	10		10	10	10		15	15	10	10	15	20	15	140
4	Legitimiteit	Inclusiviteit	5		10				10			5				30
5	Onafhankelijkheid (ook geloofwaardigheid)	Meerdere onafhankelijke inschattingen (4-ogen)	15	20	20	30	10	30	15	25	20	15	10	17	10	237
6	Relevantie	Aansluiting van aannames op praktijk	15	20					10	20		5	20		30	116
7	Werkbaar	Inschattingen op basis van juiste expertise	10	20			10	30	9			15	5			99
8	Werkbaar	Lage werkdruk	5						5	15	10	5	10		15	70
9	Werkbaar	Proportionaliteit	5		10	10	10		6		40	7		3		91
10	Werkbaar (ook geloofwaardigheid)	Transparantie binnen PBL	15				10		10			7	10	20		72
			Sum 100	Sum 100	Sum 100	Sum 100	Sum 100	Sum 100	Sum 100	Sum 100	Sum 100	Sum 99	Sum 100	Sum 100	Sum 100	

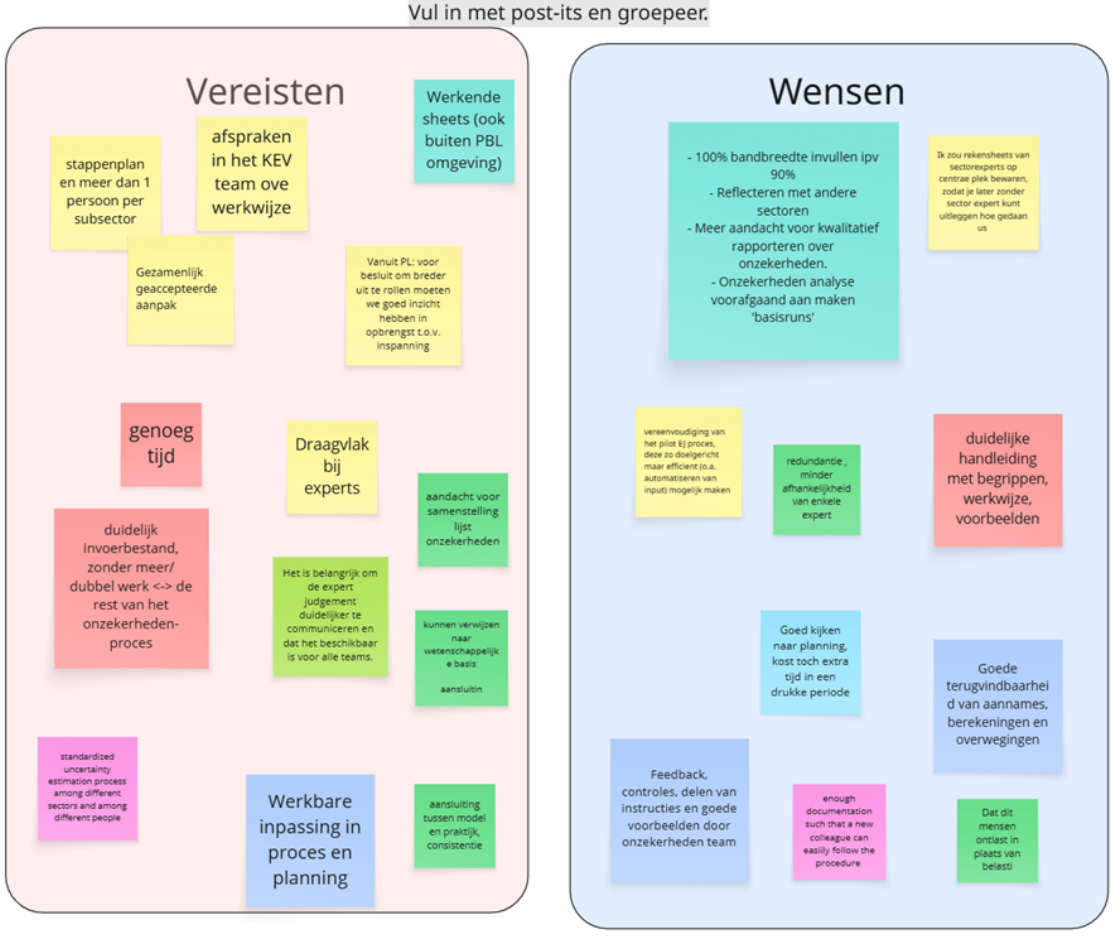
B


2.Evaluatie pilot
Vul in met post-its en groepeer.
2 mins
Start
Wat vind ik goed?
- Reflecteren op onzekerheden
- uitwisseling domeinen
standaardisate proces en 4 ogen principe
-Transparantie
-kwaliteit van aannames
consistentie
openheid team GO
inzicht in werkwijze verbeterd
Documentatie
Transparatie wordt groter
more clear documentation - enables a quick discussion; 4-eyes principle - increases the trust in quality
Aandacht voor vastleggen, consistentie methodologie en inbreng van ideeën
n anders?
- koppel reflectie los van invullen template
- Integreer nadenken over onzekerheden met inschattten beleidseffect (dus al bij de middenmaarde)
- geen 5% en 95% laten invullen
- Let op schijnzekerheid
Er zijn meer processtappen nodig dan die we in de pilot hebben gedaan
aansluiting op regulier proces
-Manier en moment van invoeren aannames
-Beslissingen over aannames nog steeds laat in proces
meerduidelijkere instructies, handvatten, definities, voorbeelden voor de sectorexperts
Format, planning en overleg-agenda die aansluiten bij belangrijkste keuzes en documentatie
improve usage of confidence score - this score was hardly used; discuss the units - a lot of discussion time could have been reduced if this was standardized
Weekje extra in de planning kan veel helpen
kansverdeling durven inschatten als expert
Omdat er geen waarheid bestaat bij het inschatten, is het lastig te bepalen of de methode ook betere resultaten geeft. Misschien is hier iets opte vinden?

3.Ontwerpen

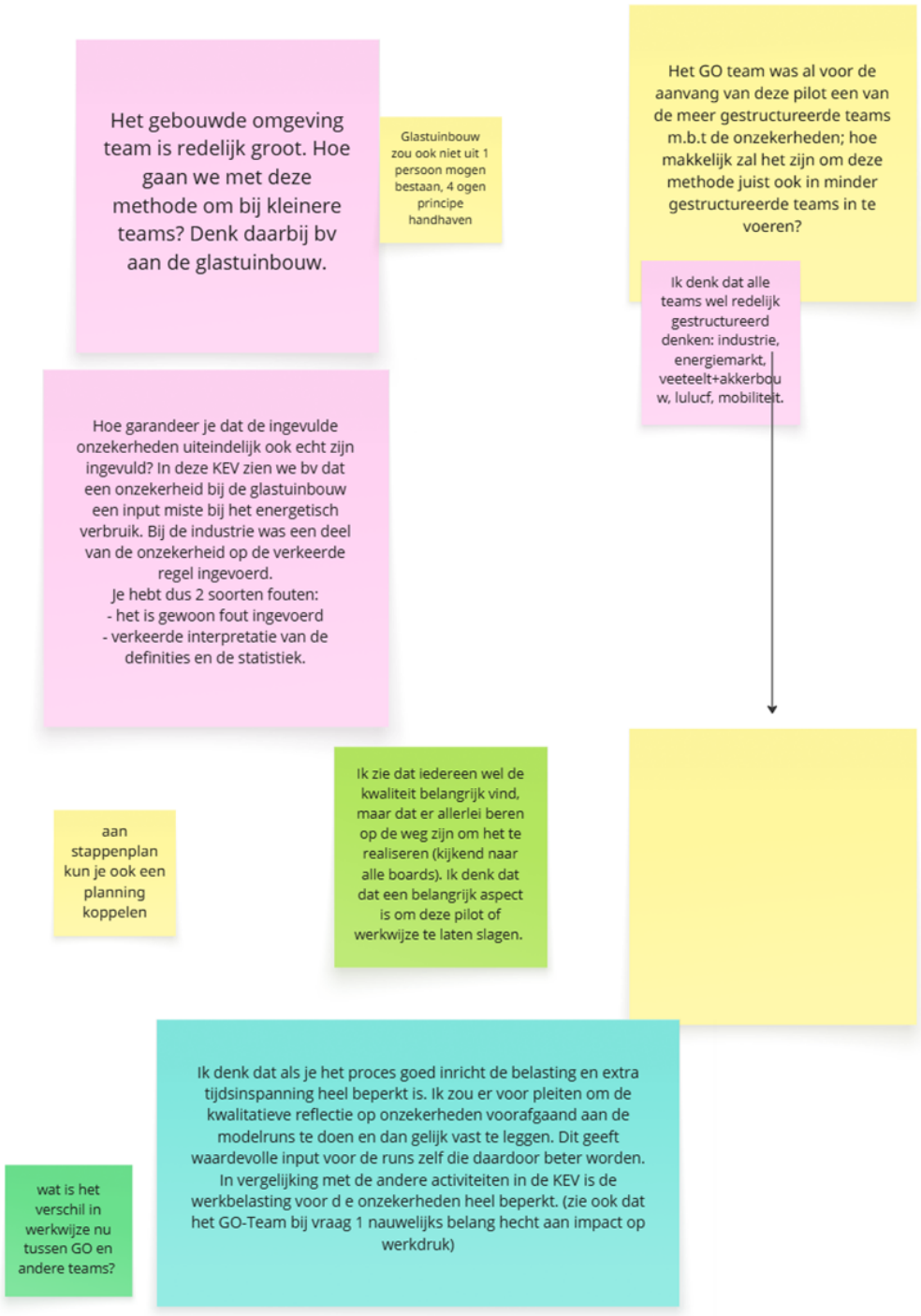
Timer
5 mins
Start

Vul in met post-its en groepeer.



B

Overige notities



C

RESULTS WORKSHOP

This appendix shows the raw contributions for the codesign focus group, including the pre-focus group contributions and post-focus group additional input. There were in total 6 participants, one of whom only joined the pre-focus group survey.

C.1. PRE-FOCUS GROUP SURVEY

6 1

Expert judgement – een protocol ontwerpen

Gebruik dit blad om te tekenen, schrijven of schetsen hoe jij denkt over geloofwaardige inschattingen, vertrouwen en mogelijke bias. Er zijn geen goede of foute antwoorden — het gaat om jouw ervaring en manier van denken.

A. Situatie & ervaring

Denk aan een moment waarop je een belangrijke inschatting moest maken onder onzekerheid.

Wat maakte het oordeel (minder) betrouwbaar voor jouw? En voor anderen?

Als ik een bron heb (die misschien niet alles deelt)
als de redenatie goed uit te leggen is
Naarmate verder in de toekomst/ samenhangt met onzekere ontwikkelingen.

Wat maakt het juist lastiger om **realistische** inschattingen te maken?

Geen vergelijkingsmateriaal
Wat helpt is als je kan doorredeneren naar iets waar je wel inschattingen van hebt

B. Bias & geloofwaardige inschattingen in een groep

Wat beïnvloedt volgens jou de geloofwaardigheid van een inschatting die experts samen moeten maken?

- Denk aan ervaring, groepsdynamiek, data, overtuigingen, tijdsdruk, geloofwaardigheid van expertise ...
- Wat helpt bias te verminderen?

als een ervaren medewerker met een junior ~~[illegible]~~ werkt: risico dat ervaren collega toch alleen een inschatting maakt.

redenatie moet wel te snappen zijn voor alle ervaringsniveaus

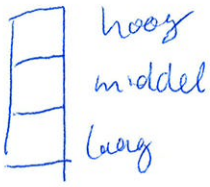
checken bij een ervaren collega, zeker als je zelf ervaring hebt.

C. Vertrouwen en geloofwaardigheid communiceren

Hoe zou jij laten zien hoeveel vertrouwen (confidence) je hebt in je eigen inschatting?
Teken of markeer hoe jij dat zou uitdrukken, bijvoorbeeld:

- een schaal van laag ↔ hoog vertrouwen
- een thermometer
- cirkels van vertrouwen
- verschillende dimensies
- of iets anders wat voor jou werkt.

discrete intervallen



bronnen, meer relevante leads, meer vergelijkings materiaal

totaal nieuw heel ver in de toekomst mechanisme onduidelijk.

D. Richtlijnen voor een protocol

Wat zou een goed protocol moeten doen om geloofwaardige inschattingen te ondersteunen? (individueel en in een groep)

Een goed protocol moet: werkbaar, 4-ogen moet gewaarborgd, reflectie zekerheidsmarge

Het mag vooral niet:

2

Expert judgement – een protocol ontwerpen

Gebruik dit blad om te tekenen, schrijven of schetsen hoe jij denkt over geloofwaardige inschattingen, vertrouwen en mogelijke bias. Er zijn geen goede of foute antwoorden — het gaat om jouw ervaring en manier van denken.

A. Situatie & ervaring

Denk aan een moment waarop je een belangrijke inschatting moest maken onder onzekerheid.

Wat maakte het oordeel (minder) betrouwbaar voor jouw? En voor anderen?

- Weinig onderbouwing uit berekeningen en/of literatuur. (-)
- Ontbreken van ledere onderbouwing/situatieschets. In welke situatie of omstandigheden zou bb gelden? ontbreekt of bovenkant (-)
- Inzicht/experts uit meer wetenschappelijk en meer beleidsperspectief (+) Beren.

Wat maakt het juist lastiger om **realistische** inschattingen te maken?

- Weinig onderbouwing uit berekeningen en/of literatuur.
- Niet volledig inhoudelijk inzicht (bij landgebruik is PBL meer beleid, WUR berekeningen → ieder met eigen perspectief)
- Niet volledig systeeminzicht → niet alle variatie in landgebruik is (eenvoudig) verklaarbaar. Bijv. bijv. vaak interactie-effecten.

B. Bias & geloofwaardige inschattingen in een groep

Wat beïnvloedt volgens jou de geloofwaardigheid van een inschatting die experts samen moeten maken?

- Denk aan ervaring, groepsdynamiek, data, overtuigingen, tijdsdruk, geloofwaardigheid van expertise ...
- Wat helpt bias te verminderen?

Telkenmale! Soms experts iets te veel 'uit de losse pols'
↓
PBL: onderbouwing/transparantie.
+ brede blik

- Duurzigheid gedeelde benchmark verhaallijn (+)
- Voldoende aandacht voor vragen (+ transparantie) → bijv. een keer nieuwe achtergronddoc over onzeker leden? (+)
  ↳ onderwijging: Vbn de plank.

ervaring in lg: te nauwe bb.

Idee: realistieve onzekerheid onderlijg grote berekeningen. Daarom niet altijd de meeste aandacht/investering. Makkelijk om te zeggen: doen zoals vorige keer.

C. Vertrouwen en geloofwaardigheid communiceren

Hoe zou jij laten zien hoeveel vertrouwen (confidence) je hebt in je eigen inschatting?
Teken of markeer hoe jij dat zou uitdrukken, bijvoorbeeld:

- een schaal van laag ↔ hoog vertrouwen
- een thermometer
- cirkels van vertrouwen
- verschillende dimensies
- of iets anders wat voor jou werkt.

→ 5 puntsschaal? of ++ etc.

↓

dingen echt op getalniveau is gedetailleerd, maar verwacht dat het op een puntschaal wel mogelijk is.

↓

wel goed om af te spreken / gedeeld beeld waar deze schaal voor staat. (verschillende interpretaties mogelijk!)

D. Richtlijnen voor een protocol

Wat zou een goed protocol moeten doen om geloofwaardige inschattingen te ondersteunen? (individueel en in een groep)

Een goed protocol moet:

- Zorgen voor uniformiteit over sectoren
- leidraad bij gesprek/proces inzakkosten, op een manier die erg gericht is op KEV proces.

Het mag vooral niet:

- voelen als een (vooral) grote administratieve last, zonder praktische uitkomst.
- "een moetje zijn"

3

Expert judgement – een protocol ontwerpen

Gebruik dit blad om te tekenen, schrijven of schetsen hoe jij denkt over geloofwaardige inschattingen, vertrouwen en mogelijke bias. Er zijn geen goede of foute antwoorden — het gaat om jouw ervaring en manier van denken.

A. Situatie & ervaring

Denk aan een moment waarop je een belangrijke inschatting moest maken onder onzekerheid.

Wat maakte het oordeel (minder) betrouwbaar voor jouw? En voor anderen?

- Unieke emissiereductie projecte bij individuele bedrijven → effecten bekend. Project kan wel/niet doorgaan.
- Modelonzekerheid oude jaard te bepalen → vergelijking historische jaren
- Bij generieke emissiereductie-projecten lastiger in te schatten. (bv. e-boilers, warmtepompen)

Wat maakt het juist lastiger om **realistische** inschattingen te maken?

- Productie volumes industrie ook moeilijk in te schatten.

B. Bias & geloofwaardige inschattingen in een groep

Wat beïnvloedt volgens jou de geloofwaardigheid van een inschatting die experts samen moeten maken?

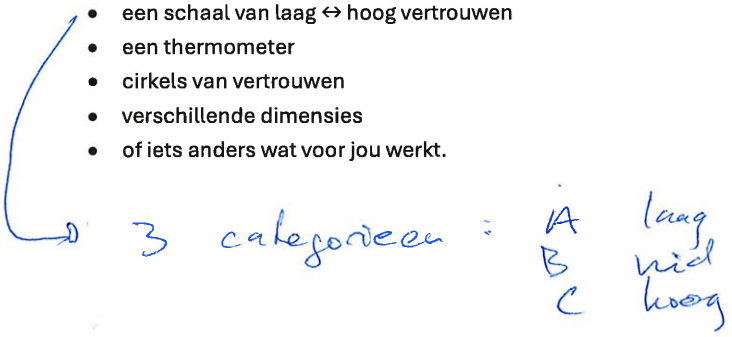
- Denk aan ervaring, groepsdynamiek, data, overtuigingen, tijdsdruk, geloofwaardigheid van expertise ...
- Wat helpt bias te verminderen?

Een semble gemiddelde is altijd nauwkeuriger dan een individuele inschatting.

C. Vertrouwen en geloofwaardigheid communiceren

Hoe zou jij laten zien hoeveel vertrouwen (confidence) je hebt in je eigen inschatting? Teken of markeer hoe jij dat zou uitdrukken, bijvoorbeeld:

- een schaal van laag ↔ hoog vertrouwen
- een thermometer
- cirkels van vertrouwen
- verschillende dimensies
- of iets anders wat voor jou werkt.



D. Richtlijnen voor een protocol

Wat zou een goed protocol moeten doen om geloofwaardige inschattingen te ondersteunen? (individueel en in een groep)

Een goed protocol moet:

Het mag vooral niet:

4

Expert judgement – een protocol ontwerpen

Gebruik dit blad om te tekenen, schrijven of schetsen hoe jij denkt over geloofwaardige inschattingen, vertrouwen en mogelijke bias. Er zijn geen goede of foute antwoorden — het gaat om jouw ervaring en manier van denken.

A. Situatie & ervaring

Denk aan een moment waarop je een belangrijke inschatting moest maken onder onzekerheid.

Wat maakte het oordeel (minder) betrouwbaar voor jouw? En voor anderen?

kennengroe - voorbeelden buiten land (recent)
- gat theorie / praktijk

Wat maakt het juist lastiger om **realistische** inschattingen te maken?

geen referenties / niet toetsbaar

Algemene info, geschiedenis, dynamiek
Maakbaarheidsgeloof.

B. Bias & geloofwaardige inschattingen in een groep

Wat beïnvloedt volgens jou de geloofwaardigheid van een inschatting die experts samen moeten maken?

- Denk aan ervaring, groepsdynamiek, data, overtuigingen, tijdsdruk, geloofwaardigheid van expertise ...
- Wat helpt bias te verminderen?

- groupthink
+ input leken
+ input externen

C. Vertrouwen en geloofwaardigheid communiceren

Hoe zou jij laten zien hoeveel vertrouwen (confidence) je hebt in je eigen inschatting?
Teken of markeer hoe jij dat zou uitdrukken, bijvoorbeeld:

- een schaal van laag ↔ hoog vertrouwen ← 1 tot 5 sterren met gedefinieerde kansen of in % kansen
- een thermometer
- ~~cirkels van vertrouwen~~
- ~~verschillende dimensies~~
- of iets anders wat voor jou werkt.

D. Richtlijnen voor een protocol

Wat zou een goed protocol moeten doen om geloofwaardige inschattingen te ondersteunen? (individueel en in een groep)

Een goed protocol moet:

Het mag vooral niet:

5

Expert judgement – een protocol ontwerpen

Gebruik dit blad om te tekenen, schrijven of schetsen hoe jij denkt over geloofwaardige inschattingen, vertrouwen en mogelijke bias. Er zijn geen goede of foute antwoorden — het gaat om jouw ervaring en manier van denken.

A. Situatie & ervaring

Denk aan een moment waarop je een belangrijke inschatting moest maken onder onzekerheid.

Wat maakte het oordeel (minder) betrouwbaar voor jouw? En voor anderen?

→ onderzoeken externe partijen
→ discussie met collega's omtrent kanten te belichten

Wat maakt het juist lastiger om **realistische** inschattingen te maken?

→ wat helpt is dat je hele extreme scenario's kan negeren.
→ waar leg je precies de grens? Is dat als alles mee zit, versus alles tegen zit?

B. Bias & geloofwaardige inschattingen in een groep

Wat beïnvloedt volgens jou de geloofwaardigheid van een inschatting die experts samen moeten maken?

- Denk aan ervaring, groepsdynamiek, data, overtuigingen, tijdsdruk, geloofwaardigheid van expertise ...
- Wat helpt bias te verminderen?

→ op zoek gaan naar andere perspectieven.
→ ook kijken extreme, gekke situaties.
→ perspectief aan nemen van niet-innovatieve, niet-technische burgers/ondernemers.
→ junior ~~xxxx~~ collega's ook vragen naar hun blik.

C. Vertrouwen en geloofwaardigheid communiceren

Hoe zou jij laten zien hoeveel vertrouwen (confidence) je hebt in je eigen inschatting? Teken of markeer hoe jij dat zou uitdrukken, bijvoorbeeld:

- een schaal van laag ↔ hoog vertrouwen
- een thermometer
- cirkels van vertrouwen
- verschillende dimensies
- of iets anders wat voor jou werkt.

D. Richtlijnen voor een protocol

Wat zou een goed protocol moeten doen om geloofwaardige inschattingen te ondersteunen? (individueel en in een groep)

Een goed protocol moet:

Simpel, zodat iedereen op dezelfde lijn zit.
Moet behulpzaam zijn (ondersteunend).

Het mag vooral niet:

Expert judgement – een protocol ontwerpen

Gebruik dit blad om te tekenen, schrijven of schetsen hoe jij denkt over geloofwaardige inschattingen, vertrouwen en mogelijke bias. Er zijn geen goede of foute antwoorden — het gaat om jouw ervaring en manier van denken.

A. Situatie & ervaring

Denk aan een moment waarop je een belangrijke inschatting moest maken onder onzekerheid.

Wat maakte het oordeel betrouwbaar en wat waren obstakels voor betrouwbaarheid?

De betrouwbaarheid werd vergroot door op zoek te gaan naar actuele informatie, contact te leggen met externe organisaties, de redeneerlijn te bespreken met collega's en de uitgangspunten duidelijk vast te leggen. Obstakels waren dat betrouwbare informatie maar beperkt beschikbaar was en de tijd om tot een inschatting te komen beperkt was.

- Denk aan ervaring, groepsdynamiek, data, overtuigingen, tijdsdruk ...
- Wat helpt bias te verminderen?

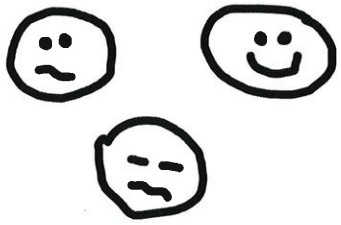
Wat maakt het juist lastiger om **realistische** inschattingen te maken?

Het is lastiger om realistische inschattingen te maken als data en kennis ontbreken en als de tijd om tot een inschatting te komen beperkt is.

C. Vertrouwen en geloofwaardigheid communiceren

Hoe zou jij laten zien hoeveel vertrouwen (confidence) je hebt in je eigen inschatting? Teken of markeer hoe jij dat zou uitdrukken, bijvoorbeeld:

- een schaal van laag ↔ hoog vertrouwen
- een thermometer
- cirkels van vertrouwen
- verschillende dimensies
- of iets anders wat voor jou werkt.



D. Richtlijnen voor een protocol

Wat zou een goed protocol moeten doen om geloofwaardige inschattingen te ondersteunen? (individueel en in een groep)

Een goed protocol moet: gebaseerd zijn op wetenschappelijke inzichten, de gebruiker ondersteunen, de gebruiker op nieuwe ideeën brengen, praktisch toepasbaar zijn

Het mag vooral niet: inspanning en tijd niet in goede verhouding met de voordelen, alleen gebruikt worden om de organisatie in te dekken

C.2. WORKSHOP CONTRIBUTION

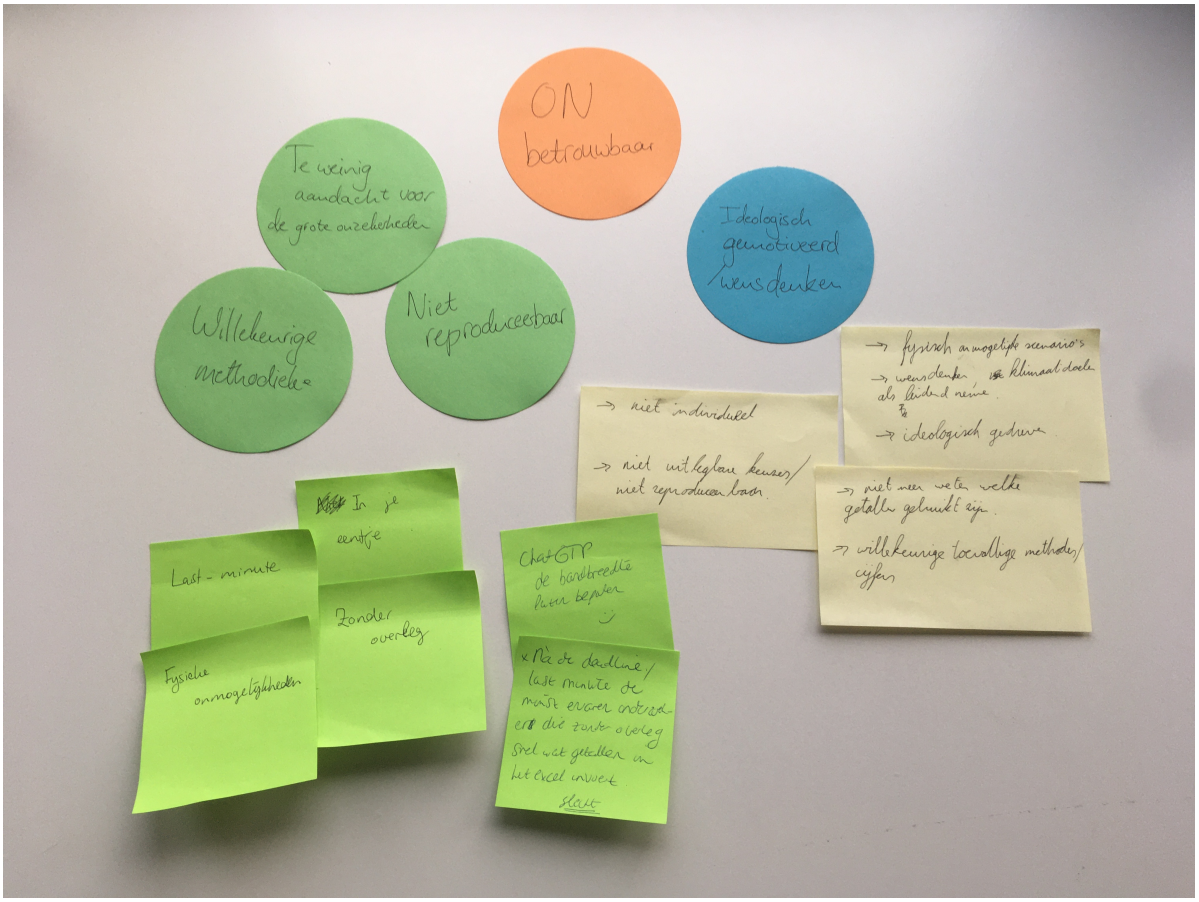


Figure C.1: Enter Caption

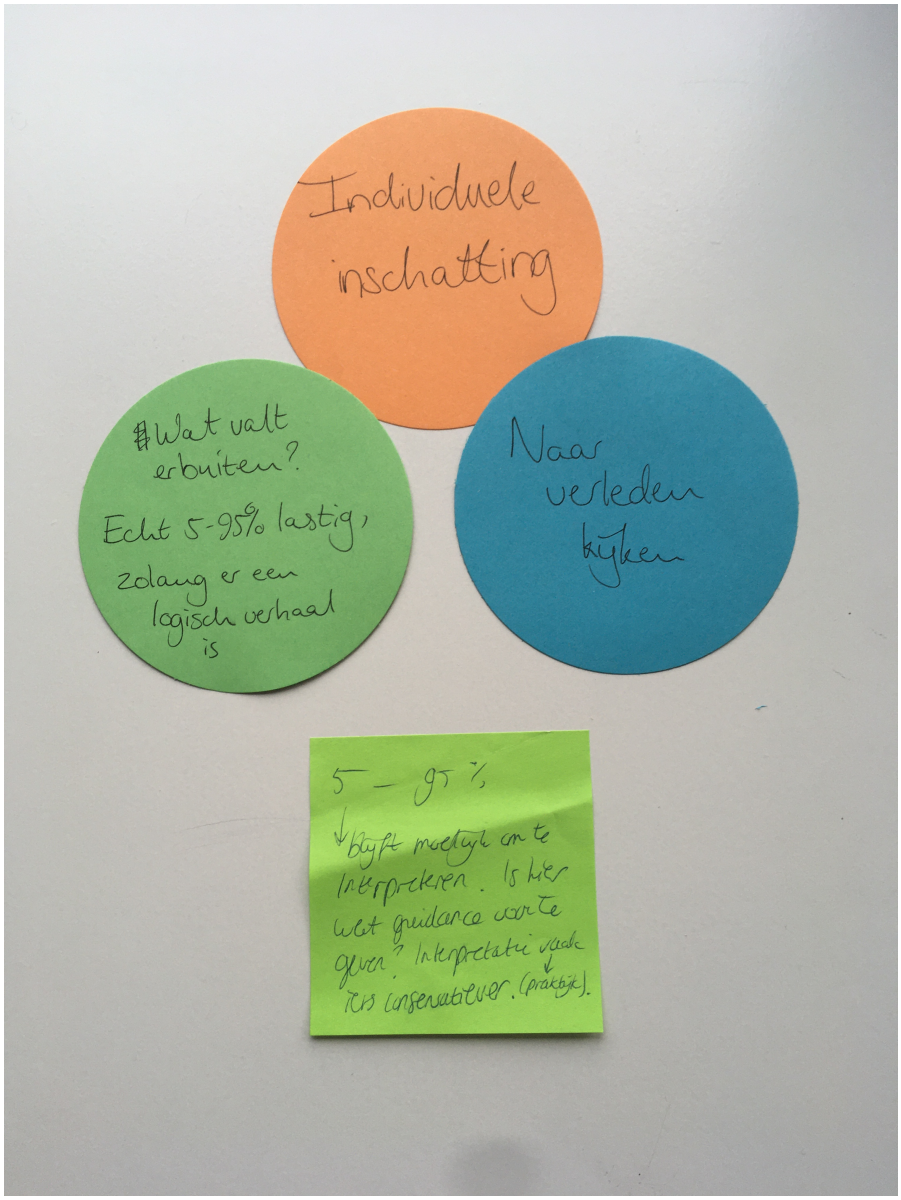


Figure C.2: Enter Caption

C

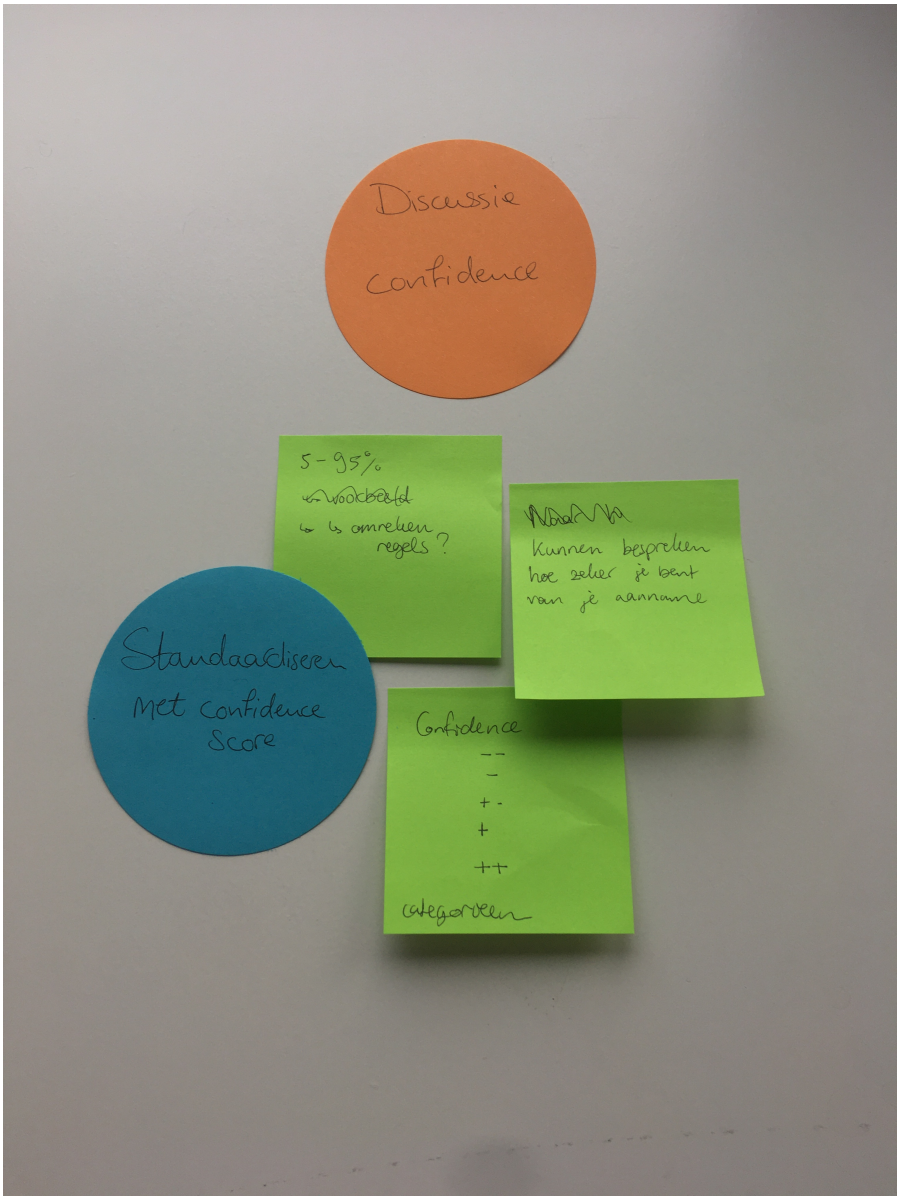


Figure C.3: Enter Caption

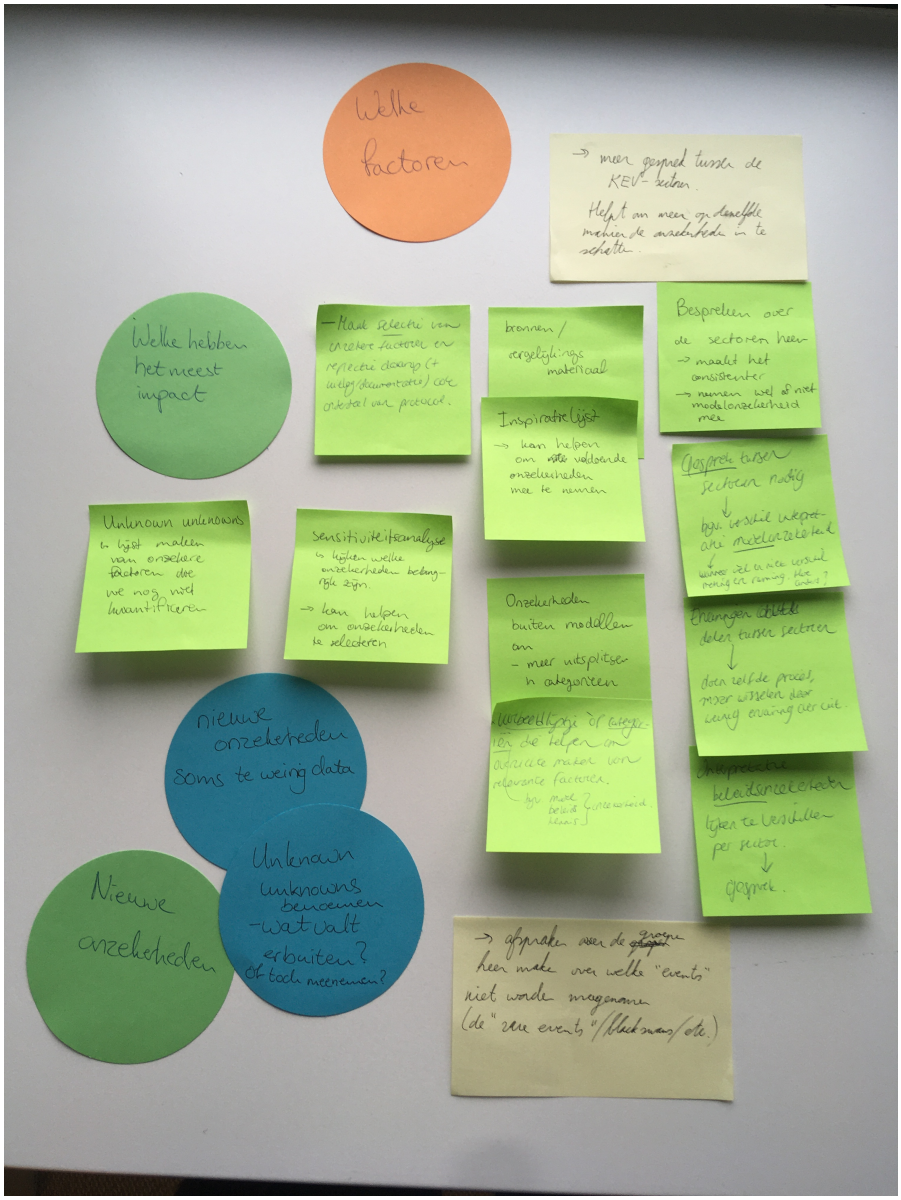


Figure C.4: Enter Caption

D

Storyboards

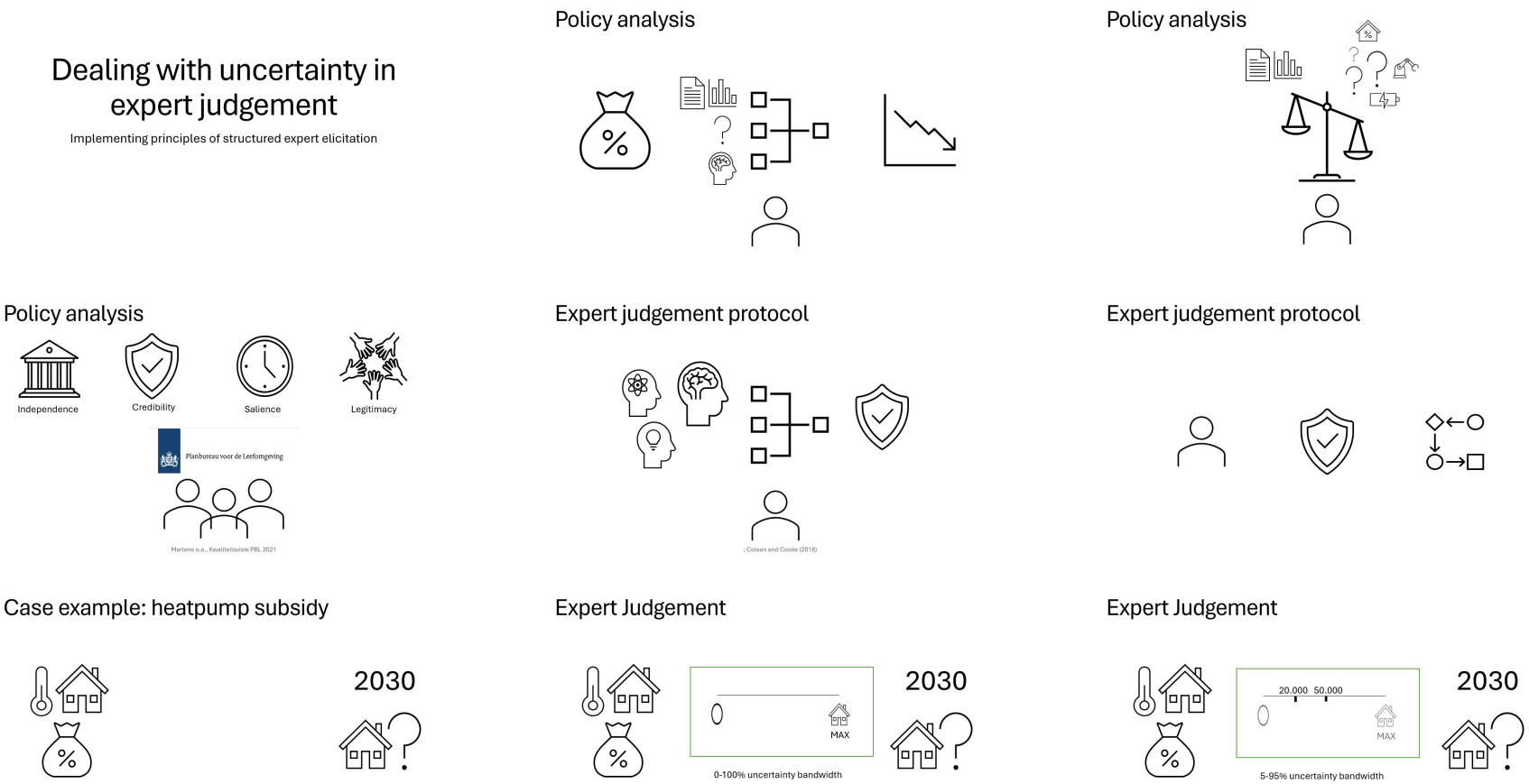
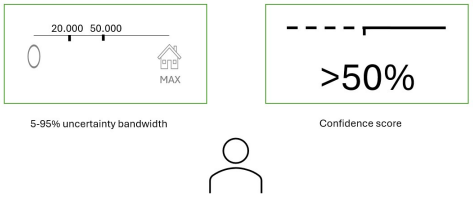


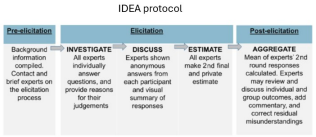
Figure D.1: Caption

Expert Judgement



Scientific process for expert judgement

Structured expert elicitation
- Reproducible
- Neutral
- Fair
- Accountable
- Valid

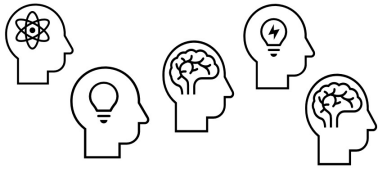


Hanea et al. (2017)

Explicitly addressing uncertainty

- Expert selection
- Question formulation
- Mitigating common biases

Explicitly addressing uncertainty: expert selection



Explicitly addressing uncertainty: Question formulation

What is the effect of the new heatpump subsidy?

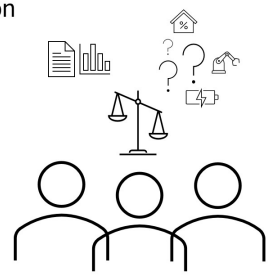
- What unit (money, #households)
- What time frame?
- What external factors?

Explicitly addressing uncertainty: common biases

Anchoring to first-guess

Overconfidence

Conclusion



What principles can you apply?

- What **assumptions** do you make in your estimations, and do others have the same view?
- What expertise is included in your judgments, and how could you address **missing perspectives**?
- How do you mitigate **common biases** like anchoring and overconfidence?

Structured expert elicitation at PBL

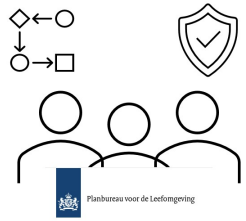


Figure D.2: Caption

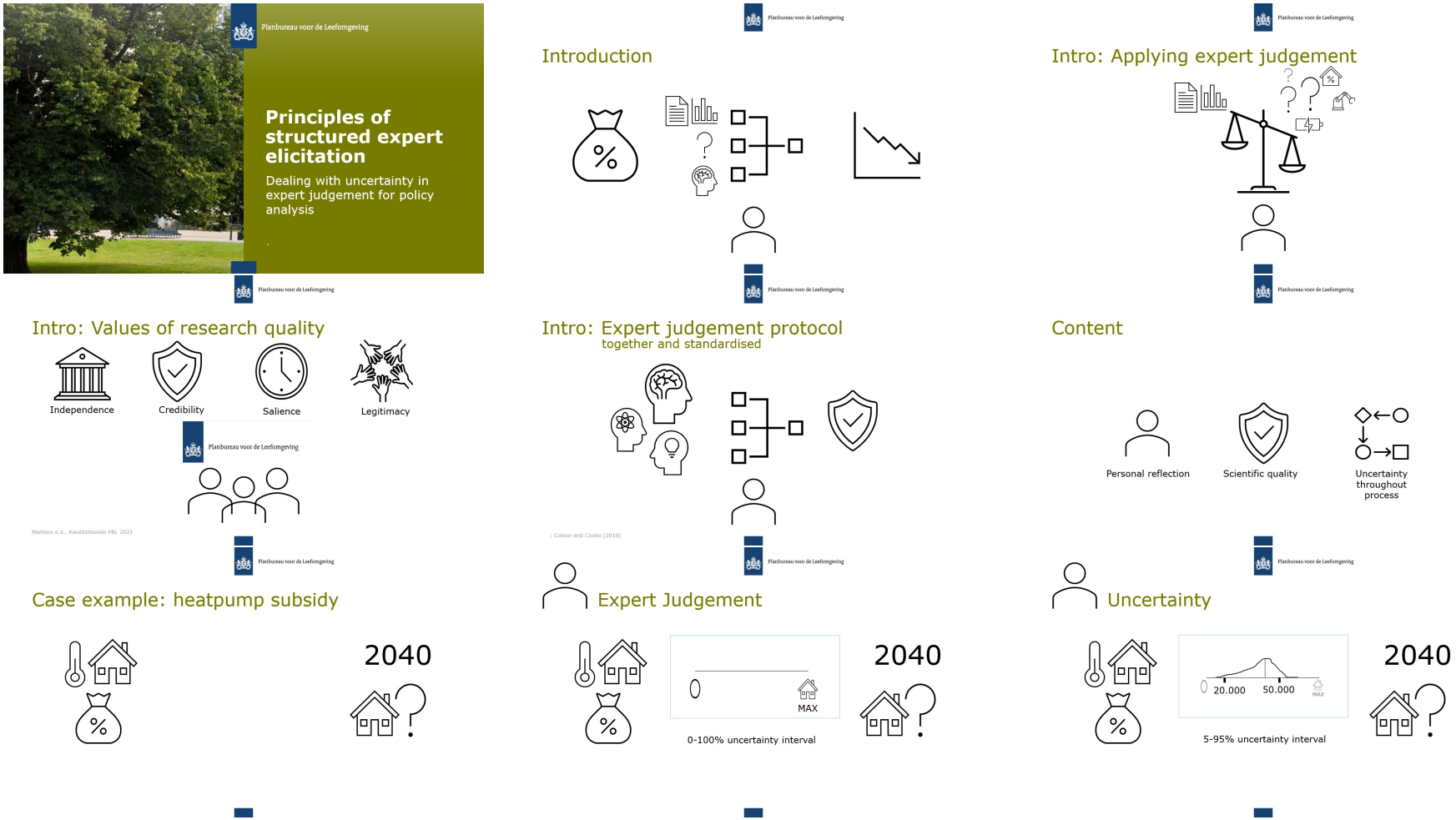
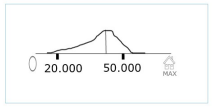
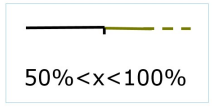


Figure D.3: Caption

Reflection on confidence



5-95% uncertainty interval, distribution, and best-estimate



Confidence score

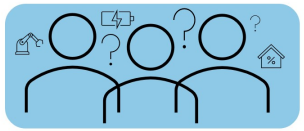
Question formulation

What is the effect of the new heatpump subsidy?

Specify together for accuracy and reproducibility:

- What unit (money, #households, ...)
- What time frame?
- What assumptions?

Together, we start today.

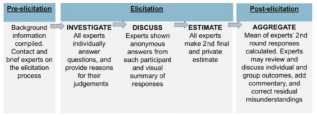


Scientific process and expert elicitation

Structured expert elicitation

- Reproducible
- Neutral
- Fair
- Accountable
- Valid
- *Empirical control*

IDEA protocol



Goosens and Cooke (2001); Hanea et al. (2017)

Common biases to mitigate

Anchoring to first-guess

Overconfidence

Sutherland (2022)

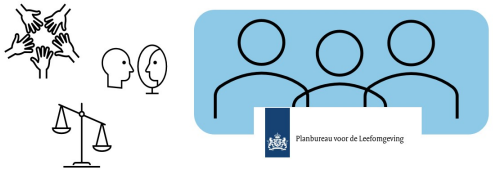
What can you start with?

- What influences your **confidence** and how can you reflect on it?
- What expertise is included in your judgments, and how could you address **missing perspectives**?
- How do you mitigate **common biases** like anchoring and overconfidence?

Expert selection

Sniezek, Henry (1989)

Conclusion: Dealing with uncertainty together



Sources (draft)

1. Hanea, A. M., Hemming, V., & Nane, G. F. (2022). Uncertainty Quantification with Experts: Present Status and Research Needs. Risk Analysis, 42(2), 254–263. https://doi.org/10.1111/risa.13718
2. Krueger, T., Page, T., Hubacek, K., Smith, L., & Hiscock, K. (2012). The role of expert opinion in environmental modelling. Environmental Modelling & Software, 36, 4–18. https://doi.org/10.1016/j.envsoft.2012.01.011
3. Hemming, V., Burgman, M. A., Hanea, A. M., McBride, M. F., & Wintle, B. C. (2018). A practical guide to structured expert elicitation using the IDEA protocol. Methods in Ecology and Evolution, 9(1), 169–180. https://doi.org/10.1111/2041-210X.12857
4. Martens, A., Arts, B., Bakema, A., Mylius, S., & Porsius, J. (n.d.). Kwaliteitsvisie PBL 2021.
5. Colson, A. R., & Cooke, R. M. (2018). Expert Elicitation: Using the Classical Model to Validate Experts' Judgments. Review of Environmental Economics and Policy, 12(1), 113–132. https://doi.org/10.1093/reep/rex022
6. Wintle, Bonnie C., Nibedita Mukherjee, Victoria Hemming, Stefano Canessa, en Marissa McBride. '5. Improving the Reliability of Judgements'. In Transforming Conservation, 1ste dr., onder redactie van William J. Sutherland. Open Book Publishers, 2022. https://doi.org/10.11647/OBP.0321.05.

Figure D.4: Caption