Methods to Effectively Communicate Verbal Probability Expressions in Human-AI Teams

Christian Fleiner[1,*], Joost Vennekens[1,2]

[1]*KU Leuven, Department of Computer Science, 2860 Sint-Katelijne-Waver, Belgium*

[2]*Vrije Universiteit Brussel, Department of Informatics and Applied Informatics, 1050 Brussels, Belgium*

Abstract

In knowledge acquisition, the elicitation of probabilities is a challenging task because many domain experts prefer to communicate probability estimates with verbal probability expressions (VPEs; e.g., "likely") rather than precise numerical values. Since the 1960s, many methods and approaches were introduced to operationalize verbal probability expressions. Given the conclusion that an individual's intended meaning of an expressed verbal probability is at risk of being lost in means of group-based aggregations, a co-learning approach between a human individual and an AI agent has been proposed. In this paper, we summarize methods that are capable of contributing to the realization of such a co-learning process. The methods are translation tables, fuzzy sets, the Sheffield Elicitation Framework (SHELF), the Rational Speech Act (RSA) model, and large language models (LLMs).

Keywords

human-agent collaboration, hybrid intelligence, hybrid team, knowledge acquisition, preference paradox, subjective probability, uncertainty communication

1. Introduction

Knowledge acquisition is one of the major challenges in the field of knowledge representation and reasoning (KRR). There exist multiple methods and protocols for expert knowledge elicitation (EKE), that allow to acquire knowledge from domain experts. For instance, the European Food Safety Authority (EFSA) published a guide on EKE in 2014 [1]. Nevertheless, the problem remains far from solved [2]. One challenge is that the communication between domain expert and knowledge engineer is often inefficient and ambiguous. After all, an expert's opinion is just a "subjective assessment, evaluation, impression, or estimation of the quality or quantity of something of interest that seems true, valid, or probable to the expert's own mind."[3, p. 98].

To handle the inherent uncertainty and subjectivity, the elicitation of probabilities is an important aspect of EKE. While people typically prefer to *hear* probabilities expressed as numbers between 0.0 and 1.0, when *expressing* probabilities themselves, they often prefer words (this is known as the *preference paradox* [4, 5]). Words and word combinations that carry probabilistic meaning (e.g., "likely") are referred to as verbal probability expressions (VPEs), although many synonyms exist in literature like probabilistic phrases, probability terms, or judgment terms. Research on VPEs can be roughly divided into two waves [6]. Researchers of the first wave (1967 – 1996) concluded that VPEs do not translate to fixed numerical probabilities due to a between-subject variability in most studies. An extensive summary about this period is provided by Clark [7]. The second wave started around 2013 and is still ongoing. An overview of relevant work is provided by Dhami and Mandel [8].

Recently, Fleiner and Vennekens [6] have proposed a co-learning approach "to efficiently and effectively communicate (subjective) probabilities", in which a personalized translation table is developed by iterative communication between a human and an AI agent (who form a human-AI team). The

co-learning process consists of three phases which are illustratively depicted in Figure 1. The co-learning process and its phases are described more in detail in section 2.

With many studies indicating the high effort of eliciting probabilities from domain experts for knowledge engineers and the existence of the *preference paradox*, such a co-learning approach seems promising to increase the scalability of knowledge acquisition. However, Fleiner and Vennekens [6] provided only a high-level description of the co-learning concept for probability elicitation and identified two research questions which first have to be thoroughly answered before such a co-learning approach can be realized.

In this paper, we aim to answer the first research question from Fleiner and Vennekens [6]:

> *Which algorithms, mechanisms, and methods are most appropriate to establish co-learning processes for estimating uncertainty in the context of collaborative tasks in human-agent teams?*

To answer the research question, we first describe identified methods and their related work. Then, we provide explanations in which phase each method can be applied during the co-learning process. Additionally, we provide code examples online[1]. The identified methods are translation tables, fuzzy sets, the Sheffield Elicitation Framework (SHELF), the Rational Speech Act (RSA) model, and large language models (LLMs).

2. Co-learning approach and its phases
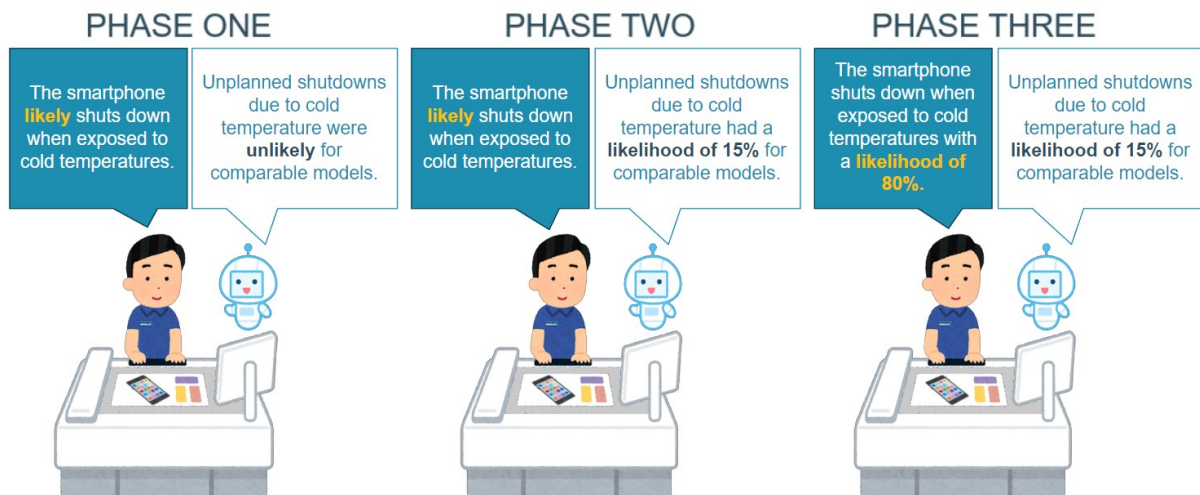
Figure 1: Outline of the co-learning process where a service technician analyzes a damaged smartphone with the aid of an AI agent (example from [9]). In the first phase, human and AI agent communicate using verbal probability expressions. In the second phase, the AI agent uses numeric estimates based on evidence. In the third phase, human and AI agent communicate using solely numerical estimates. [6, Table 2].

The co-learning approach described by Fleiner and Vennekens [6] applies in the context of a hybrid team. A hybrid team consists of a set of intelligent (human or software) agents who engage in joint task performance. Each agent develops and refines "a mental model containing knowledge of other agent's needs, goals, values, capabilities, resources, plans, and emotions" [10] through interaction and feedback. Van Zoelen et al. describe "co-learning" as an iterative cycle of co-adaptation and feedback [11].

The described co-learning process has as its final goal the bidirectional communication of numerical probabilities to remove the vagueness and misinterpretations of VPEs. The co-learning process consists of three phases, where the overall goal of the human-AI team is achieved when the team enters the third phase. Depending on the use case, it might be impossible or not necessarily desired to enter the

---

[1]https://gitlab.com/EAVISE/CFL/synergy25_codeExamples, last accessed on May 27th, 2025

third phase. Additionally, it is likely that a human-AI team switches between phases depending on the context.

Phase 1 In the first phase, the team is familiarized with a selected translation table that serves as probability reference and provides a small set of VPEs. The numerical translations are not important in the first phase as the team solely communicates probabilities by using VPEs. The co-adaptation consists of developing and using a VPE vocabulary that sufficiently and precisely enough describes the relevant probability scale which depends on the use case. For instance, a quality assurance operator who regularly tests material with a defect frequency with 0.05% will not necessarily need a vocabulary that ranges along the entire probability scale, but may require a detailed vocabulary for distinguishing various degrees of "unlikeliness". Thus, an essential element of the first phase is the introduction of new VPEs.

Phase 2 In the second phase, the AI agent starts to communicates numerical probabilities while the human team member still relies on using VPEs. A prerequisite for the second phase is that enough evidence has been collected to reliably map the VPEs to numerical probabilities. The required evidence does not necessarily need to be acquired by the human-AI team itself. For instance, a newly observed machine defect which was described using a VPE might become numerically translatable after the machine manufacturer analyzed the issue and updated customers about the results. The phase's co-adaptation consists of understanding the reasoning behind production and interpretation of the VPE vocabulary.

Phase 3 In the third phase, the human individual has gained enough experience to express probability estimates numerically. The co-adaptation refers to the granular adjustment to make more precise numeric estimates. While the overall goal (that both members use numeric probabilities) was already achieved by entering the third phase, an important element of this phase is to numerically translate already expressed estimates (using VPEs) to generate additional evidence.

3. Methods for the Co-learning Approach

3.1. Translation tables

Table 1
The EFSA scheme was inspired by the IPCC scheme and is a translation table using 9 VPEs to describe the entire probability scale. Optionally, experts can report an "inconclusive" estimate when they are not confident enough about a specific matter.

EFSA [12]			
Probability term	**Subjective probability range**	**Additional options**	
Almost certain	99-100%	More likely than not: >50%	Unable to give any probability: range is 0-100%
Extremely likely	95-99%		
Very likely	90-95%		
Likely	66-90%		
About as likely as not	33-66%		Report as 'inconclusive', 'cannot conclude', or 'unknown'
Unlikely	10-33%		
Very unlikely	5-10%		
Extremely unlikely	1-5%		
Almost impossible	0-1%		

Description and related work   In general, translation tables translate VPEs to crisp probability sets or thresholds (see Table 1). Translation tables are also known as numerically bounded linguistic probability (NBLP) schemes [13]. Common translation tables contain less than ten VPEs. While most translation tables aim to cover the entire probability scale (or at least between 1% and 99%), the Professional Head of Intelligence Assessment (PHIA) Probability Yardstick intentionally contains 5% gaps to avoid the conflation of VPEs [14]. A recent summary on translation tables was already provided by Fleiner and Vennekens [6]. Some translation tables are accompanied by a confidence scale. However, research indicates that non-experts and experts struggle to separate probability and confidence estimates [15]. Furthermore, translation tables are not effectively used as look-up tables and thus VPEs should be always reported together with their numerical translation [16]. Lastly, a major concern is the lack of empirical validation of translation tables [17].

Application in co-learning process   While the application of a translation table is still controversially discussed in the research community, translation tables represent a good initial reference within the co-learning process. A translation table provides a distinguishable VPE set that ranges along the entire probability scale which would be difficult (or at least strenuous) to achieve quickly for a human individual. As we share the concern of lacking empirical evidence of applied translation tables, we recommend to implement a mechanism to validate if individuals are really compliant with the ordinal order of the chosen translation table.

Besides serving as initial reference, an AI agent should be capable of showing the current VPE vocabulary as a translation table during all three phases for the human user to check. Depending on the current phase, more or less information might be shown in the translation table. As translation tables normally contain non-overlapping ranges for different VPEs, there might be always some information loss involved in favor of clarity when the AI agent generates a translation table from its current model. The limitation of crisp ranges can be addressed by fuzzy sets.
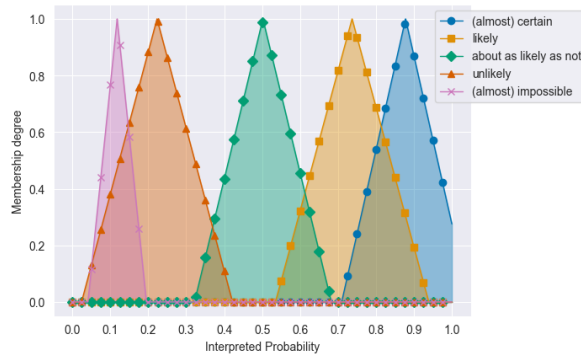
3.2. Fuzzy sets

Description and related work   Instead of describing a probability by a crisp value between 0 and 1, it is also possible to use a fuzzy set as a more imprecise but flexible representation. Fuzzy sets were especially popular in the first VPE research wave [3, 7]. Membership functions were mostly kept simple and were triangular or trapezoidal. For instance, Bonissone et al. [18] integrated fuzzy sets with trapezoidal memberships functions in an expert system to handle uncertainty. They defined the membership function $\mu_{L_i}(x)$ with L_i being part of the term set L as

$$\mu_{L_i}(x) = \begin{cases} 0 & \text{if } x < (a_i - \alpha_i) \\ (\alpha_i^{-1})(x - a_i + \alpha_i) & \text{if } x \in [(a_i - \alpha_i), a_i] \\ 1 & \text{if } x \in [a_i, b_i] \\ (\beta_i^{-1})(b_i + \beta_i - x) & \text{if } x \in [b_i, (b_i + \beta_i)] \\ 0 & \text{if } x > (b_i + \beta_i) \end{cases} \tag{1}$$

where a_i and b_i defined the interval where the membership function returns 1.0; α_i and β_i describe the left and right width of the probability density function. Figure 2 depicts the membership functions of the shortened EFSA scheme which were retrieved from Fleiner and Vennekens' dataset [19].

Application in co-learning process   Fuzzy sets are an easy and fast way to numerically represent VPEs which will be relevant in the second phase of the co-learning process where evidence is used by the AI agent to numerically translate VPEs. The primarily applied triangular or trapezoidal membership functions, however, are inappropriate to represent multi-peaked distributions (as summarized by Clark [7]) and are therefore not necessarily a good choice depending on the context. Nonetheless, the

parameters for a membership function are easy to elicit. For instance, a user could adjust parameters in an interactive graph to provide feedback to the AI agent in the third phase.

Figure 2: Fuzzy sets of VPEs (Interpretation) with cardinality of 2.0 [19].
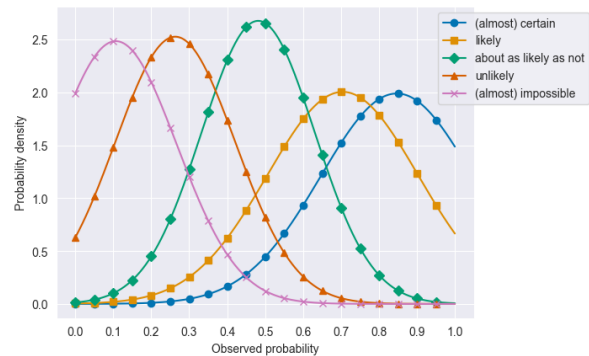
Figure 3: Normal distributions of VPEs (Production) [19].

3.3. Sheffield Elicitation Framework

Description and related work   The Sheffield Elicitation Framework (SHELF) is a collection of methods and materials for conducting expert knowledge elicitation (EKE) since 2008. Although SHELF materials are primarily used in workshops where a moderator supports the elicitation process of an expert group, a subset can be also used for remote knowledge elicitation of single experts. SHELF's *quartile method* is one of EFSA's recommended knowledge elicitation methods described as *Sheffield method* [1]. A more visual approach is SHELF's roulette method, where "experts are asked to build histographic representations of densities that reflect their beliefs about the quantities of interest." [20, p. 12]. The method's name is an analogy to the casino game because the elicitors "bet" on the true value being in a specific range by placing probability units in a column. We have transformed the production concerned dataset of Fleiner and Vennekens [19] to make use of the quartile method and the SHELF tool[2] to derive the optimal normal distribution parameters of the shortened EFSA scheme (see Figure 3).

Application in co-learning process   While the adjustment of membership function parameters (fuzzy sets) becomes more demanding with the complexity that was chosen for the membership function, SHELF's roulette method can be used instantly to elicit single-peaked and multi-peaked distributions. Together with the SHELF software tool, which outputs optimized parameters for advanced distributions (like skewnormal or beta distributions) from elicitation results, we see SHELF as an advanced successor to traditional fuzzy set elicitation for VPEs. That said, SHELF with its roulette method is especially useful in the third phase of the co-learning process to derive a good VPE distribution. Additionally, the roulette method might be also helpful in the first phase to emphasize the ordinal order of the initial VPE set and to identify synonymous VPEs.

3.4. Rational Speech Act Model

Description and related work   The Rational Speech Act (RSA) model was first introduced in 2012 [21] and is a Bayesian interpretation of the RSA theory which "predicts an interaction between (shared) knowledge about a speaker's knowledge state and a listener's interpretation of his utterance" [22]. The RSA theory is a probabilistic formalization of Grice's pragmatics theory [23] where the assumption is that an utterance must be informative to meet the speaker's goal.

The RSA model considers three actors (the literal listener, the rational speaker, and the rational listener) who reason about the state of affairs by means of a set of utterances. Applied to the Bayesian model,

---

Hat + Glasses   Glasses   Neither
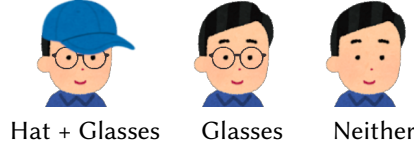
Figure 4: Illustrative example by Goodman and Franke [24]. The world consists of three faces which can be described by the speaker via the utterances: *"hat"*, *"glasses"*, and *"neither"*. The speaker uses the utterance *"glasses"* for the listener to identify the speaker's friend.

the internal reasoning of the literal listener (P_{Lit}) is represented in the prior probability distribution, the likelihood function (P_S) depends on the rational speaker's utility (U), and the posterior probability distribution (P_L) is the consequence of the rational listener's reasoning. We adopt the equations from Goodman and Frank[24]:

$$P_L(w|u) \propto P_S(u|w)P(w), \tag{2}$$

$$P_S(u|w) \propto exp\left(\alpha U(u;w)\right), \tag{3}$$

$$P_S(u|w) \propto exp\left(\alpha U(u;w)\right), \tag{4}$$

$$U(u;w) = log\ P_{Lit}(w|u), \tag{5}$$

$$P_{Lit}(w|u) \propto \delta_{[[u]](w)}P(w) \tag{6}$$

where u is the speaker's chosen utterance out of a set of utterances intended to describe the state of the world w. The α coefficient adds a simple mechanism to adjust the speaker's (assumable) rationality, with $\alpha = 1$ representing a purely rational speaker. P_{Lit} consists of the prior probability distribution and an indicator function that indicates whether an utterance can be used to describe a state of the world (denoted by the Iverson bracket $[[u]]$). In practice, the latter is represented by a truth or meaning table.

Goodman and Frank [24] provide a short, but illustrative example where a speaker uses the utterance "glasses" to describe his friend to the listener (see Figure 4). Although two faces can be described by the utterance "glasses", the listener can exclude the face with the hat and glasses, because the speaker would have chosen the utterance "hat" to describe that face under the assumption of the speaker being a utility-maximizing agent.

The RSA model is based on the assumption that the speaker chooses the most informative utterance. However, this is too idealistic to reflect reality. Thus, several extended RSA models were introduced to react to different aspects like epistemic uncertainty [24], politeness [25], or honesty [26].

Extended RSA models were also applied to formalize verbal probability elicitation and reasoning. For instance, Herbstritt and Franke applied an extended RSA model to analyze the interpretation of simple uncertainty expressions in situations of higher-order uncertainty[27]. In another paper, van Tiel et al. [28] introduced an extended RSA model which was derived from an experiment using 26 VPEs.

Some issues concerning the applicability of RSA models were addressed by Degen [29]. For instance, RSA models require both speaker and listener to know the full set of utterances as the speaker is considered a utility-maximizing agent. As this is difficult to scale in real-world scenarios, most RSA publications report on toy cases with single-shot utterances.

Application in co-learning process   In the context of verbal probability elicitation, several extended RSA models (e.g., [28]) have been already introduced where the models were based on data retrieved from surveys under laboratory settings with limited external validity. Additionally research must be conducted in the field to validate if the RSA model and its extensions are reliable and robust enough to be applied in the context of probability elicitation with experts.

Nonetheless, we can already identify the appropriate use cases in the co-learning process where RSA models can be applied. Even though we cannot reliably ensure that the human user will purely act as utility-maximizing agent, we can do so for the AI agent. In the role of the rational speaker, the AI agent

can deliberately decide on the best VPE to use in situations where the numerical probability is known, but the user prefers to hear the VPE instead. The required meaning table can be either directly acquired from the initial translation table or by eliciting appropriate VPEs from probabilities (production). The production elicitation must not necessarily been done by the human team member, but can be retrieved from a group, as the meaning table is subject to be personalized in the human-AI team over time.

Another interesting possibility is to extend the RSA models of the human team member by adding social and psychological factors to the utility function. For instance, Yoon et al. extended the RSA model to introduce politeness as additional goal [25]. Vignero used the same approach to consider the agent's honesty [26]. In the context of the co-learning process, the human-AI team could better react on biases which are dominant in probability elicitation like *overconfidence* which would be present and therefore relevant in all co-learning phases.

3.5. Large Language Models (LLMs)

Description and related work   Large language models show human-like language capabilities which makes them also interesting for verbal probability elicitation. As the application of LLMs for verbal probability elicitation is still new, we could only identify two relevant papers so far.

Tang et al. [30] recently compared the use of VPEs between modern LLMs and human subjects. The human dataset (N=123) was retrieved from Fagen-Ulmschneider [31]. Only for 5 of the 17 VPEs, the GPT-4 model provided similar estimates to the human dataset. The closest estimates between the human dataset and the GPT-4 model were observed for VPEs with high probability indications like "highly likely" and "almost certain". Only minor differences were observed between English and Chinese prompts for the GPT-4 model. Lastly, Tang et al. argue that advanced methods like Chain-of-Thought could not significantly reduce the gap between human and LLM estimates. Maloney et al. [32] conducted a coordination game where either a human participant (N=50) or the GPT-4 model had to point out the intended meaning of a VPE. In contrast to Tang et al. [30], Maloney et al. concluded that "based on overall performance we cannot distinguish GPT-4 and human". Two explanations for the different claims might be that (1) a different VPE set and (2) a different sentence set was used. While ordinary sentences had to be completed in Tang et al.'s experiment, only the VPEs were given to the participants in the coordination game within an investment or medical context.

Application in co-learning process   As current research is still inconclusive about the qualities that an LLM could provide in the context of verbal probability elicitation, we base the application in the co-learning process on assumptions. A first application could be the retrieval of flexible translation tables which cannot be achieved by the methods before. While we recommend to use an established translation table as initial VPE set, there might be a situation where none of the translation tables contains the desired VPE set. An LLM could be used to provide a situational appropriate VPE set with further definitions and descriptions. Additionally, the limitation of the availability of translation tables in different languages might be solved by the application of LLMs.

In the third phase, we have described the granular adjustments of numerical estimates to be solved by mainly visual means. An human-AI team with an LLM-powered chatbot might use purely verbal means instead which might be preferred by the human team member or due to the environmental constraints (e.g., no access to a display).

4. Conclusion

Although many methods have been proposed since the 1960s, verbal probability elicitation remains a prevailing challenge. The main reason is that the individual use of verbal probability expressions is context-sensitive and difficult to aggregate due to a proven between-subject variability. A conceptual co-learning approach between a human individual and an AI agent has been proposed instead where individual translation tables between verbal probability expressions and numeric probability values

are developed. In this paper, we have provided summaries of relevant methods to communicate verbal probability expressions and descriptions how the methods should be applied in the co-learning process. Lastly, we provide code examples of each method online.

Resource Availability Statement: The code examples are available from GitLab at https://gitlab.com/EAVISE/CFL/synergy25_codeExamples.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools to create this document. The code examples to demonstrate the application of LLMs use the model `gemma3:4b`.

References

[1] European Food Safety Authority, Guidance on expert knowledge elicitation in food and feed safety risk assessment, EFSA Journal 12 (2014) 3734.

[2] J. P. Delgrande, B. Glimm, T. Meyer, M. Truszczynski, F. Wolter, Current and future challenges in knowledge representation and reasoning, 2023. `arXiv:2308.04161`.

[3] B. M. Ayyub, Elicitation of expert opinions for uncertainty and risks, CRC press, 2001.

[4] I. Erev, B. L. Cohen, Verbal versus numerical probabilities: Efficiency, biases, and the preference paradox, Organizational behavior and human decision processes 45 (1990) 1–18.

[5] T. S. Wallsten, D. V. Budescu, R. Zwick, S. M. Kemp, Preferences and reasons for communicating probabilistic information in verbal or numerical terms, Bulletin of the Psychonomic Society 31 (1993) 135–138.

[6] C. Fleiner, J. Vennekens, Towards effective management of verbal probability expressions using a co-learning approach, in: HHAI 2024: Hybrid Human AI Systems for the Social Good, IOS Press, 2024, pp. 124–133.

[7] D. A. Clark, Verbal uncertainty expressions: A critical review of two decades of research, Current Psychology 9 (1990) 203–235.

[8] M. K. Dhami, D. R. Mandel, Communicating uncertainty using words and numbers, Trends in Cognitive Sciences 26 (2022) 514–526.

[9] E. Van Zoelen, T. Mioch, M. Tajaddini, C. Fleiner, S. Tsaneva, P. Camin, T. S. Gouvêa, K. Baraka, M. H. De Boer, M. A. Neerincx, Developing team design patterns for hybrid intelligence systems, in: HHAI 2023: Augmenting Human Intellect, IOS Press, 2023, pp. 3–16.

[10] K. van den Bosch, T. Schoonderwoerd, R. Blankendaal, M. Neerincx, Six challenges for human-ai co-learning, in: Adaptive Instructional Systems: First International Conference, AIS 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26–31, 2019, Proceedings 21, Springer, 2019, pp. 572–589.

[11] E. M. Van Zoelen, K. Van Den Bosch, M. Neerincx, Becoming team members: Identifying interaction patterns of mutual adaptation for human-robot co-learning, Frontiers in Robotics and AI 8 (2021) 692811.

[12] E. S. Committee, D. Benford, T. Halldorsson, M. J. Jeger, H. K. Knutsen, S. More, H. Naegeli, H. Noteborn, C. Ockleford, A. Ricci, et al., The principles and methods behind efsa's guidance on uncertainty analysis in scientific assessment, EFSA Journal 16 (2018) e05122.

[13] D. R. Mandel, D. Irwin, Facilitating sender-receiver agreement in communicated probabilities: Is it best to use words, numbers or both?, Judgment and Decision Making 16 (2021) 363–393.

[14] Professional Head of Intelligence Assesment, Professional Development Framework for all-source intelligence assessment, Technical Report, 2019. URL: https://assets.publishing.service.gov.uk/media/6421b6a43d885d000fdadb70/2019-01_PHIA_PDF_First_Edition_Electronic_Distribution_v1.1__1_.pdf.

[15] D. Irwin, D. R. Mandel, Communicating uncertainty in national security intelligence: Expert and nonexpert interpretations of and preferences for verbal and numeric formats, Risk Analysis 43 (2023) 943–957.

[16] D. V. Budescu, H.-H. Por, S. B. Broomell, M. Smithson, The interpretation of ipcc probabilistic statements around the world, Nature Climate Change 4 (2014) 508–512.

[17] K. H. Teigen, Dimensions of uncertainty communication: What is conveyed by verbal terms and numeric ranges, Current Psychology 42 (2023) 29122–29137.

[18] P. P. Bonissone, S. S. Gans, K. Decker, Rum: A layered architecture for reasoning with uncertainty., in: IJCAI, volume 87, 1987, pp. 891–898.

[19] C. Fleiner, J. Vennekens, Dataset sefsa - interpretation and production (dutch, french, german), 2025. URL: osf.io/eumxn.

[20] J. P. Gosling, Shelf: the sheffield elicitation framework, Elicitation: The science and art of structuring judgement (2018) 61–93.

[21] M. C. Frank, N. D. Goodman, Predicting pragmatic reasoning in language games, Science 336 (2012) 998–998.

[22] N. D. Goodman, A. Stuhlmüller, Knowledge and implicature: Modeling language understanding as social cognition, Topics in cognitive science 5 (2013) 173–184.

[23] H. P. Grice, Logic and conversation, Syntax and semantics 3 (1975) 43–58.

[24] N. D. Goodman, M. C. Frank, Pragmatic language interpretation as probabilistic inference, Trends in cognitive sciences 20 (2016) 818–829.

[25] E. J. Yoon, M. H. Tessler, N. D. Goodman, M. C. Frank, Talking with tact: Polite language as a balance between kindness and informativity, in: Proceedings of the 38th annual conference of the cognitive science society, Cognitive Science Society, 2016, pp. 2771–2776.

[26] L. Vignero, Updating on biased probabilistic testimony: Dealing with weasels through computational pragmatics, Erkenntnis 89 (2024) 567–590.

[27] M. Herbstritt, M. Franke, Complex probability expressions & higher-order uncertainty: Compositional semantics, probabilistic pragmatics & experimental data, Cognition 186 (2019) 50–71.

[28] B. van Tiel, U. Sauerland, M. Franke, Meaning and use in the expression of estimative probability, Open Mind 6 (2022) 250–263.

[29] J. Degen, The rational speech act framework, Annual Review of Linguistics 9 (2023) 519–540.

[30] Z. Tang, K. Shen, M. Kejriwal, An evaluation of estimative uncertainty in large language models, arXiv preprint arXiv:2405.15185 (2024).

[31] W. Fagen-Ulmschneider, Perception-of-probability-words, 2019. URL: https://github.com/wadefagen/datasets/tree/master/Perception-of-Probability-Words.

[32] L. T. Maloney, M. F. Dal Martello, V. Fei, V. Ma, A comparison of human and gpt-4 use of probabilistic phrases in a coordination game, Scientific reports 14 (2024) 6835.