

**The CWI World Cup Competition  
Eliciting Sets of Acceptable Gambles**

Quaeghebeur, Erik; Wesseling, C.; Beauxis-Aussalet, E.; Piovesan, T.; Sterkenburg, T.

**Publication date**

2017

**Document Version**

Final published version

**Published in**

Proceedings of Machine Learning Research

**Citation (APA)**

Quaeghebeur, E., Wesseling, C., Beauxis-Aussalet, E., Piovesan, T., & Sterkenburg, T. (2017). The CWI World Cup Competition: Eliciting Sets of Acceptable Gambles. *Proceedings of Machine Learning Research*, 62, 277-288.

**Important note**

To cite this publication, please use the final published version (if applicable).  
Please check the document version above.

**Copyright**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights.  
We will remove access to the work immediately and investigate your claim.

# The CWI World Cup Competition: Eliciting Sets of Acceptable Gambles

**Erik Quaeghebeur**

*Delft University of Technology (TU Delft)  
Delft (The Netherlands)*

E.R.G.QUAEGHEBEUR@TUDELFT.NL

**Chris Wesseling**

**Emma Beauxis-Aussalet**

**Teresa Piovesan**

**Tom Sterkenburg**

*Centrum Wiskunde & Informatica (CWI)  
Amsterdam (The Netherlands)*

CHRIS.WESSELING@CWI.NL

EMMANUELLE.BEAUXIS-AUSSALET@CWI.NL

T.PIOVESAN@CWI.NL

TOM@CWI.NL

## Abstract

We present an interface for eliciting sets of acceptable gambles on a three-outcome possibility space, discuss an experiment conducted for testing this interface, and present the results of this experiment. Sets of acceptable gambles form a representation for imprecise probabilities that is close to human behavior and eliciting them directly may improve the quality of the resulting uncertainty model. The experiment consisted of a betting competition for the 2014 FIFA World Cup: For each match bets were assigned based on the sets of acceptable gambles elicited from the participants. A new algorithm was designed for generating fair bets for assignment. Participant feedback indicated that improving the usability and transparency of the interface would ease the elicitation procedure. The experiment's results underlined that imprecision is an essential aspect of real-life uncertainty modeling.

**Keywords:** elicitation; gamble; acceptability; desirability; user interface; experiment; fair bet.

## 1. Introduction

In practical applications of uncertainty models, e.g., in expert systems, we need concrete values for their parameters. For example, the conditional probability tables of a Bayesian network need to be filled in. Such values can be obtained by learning them from data or by eliciting them from domain experts, or a combination thereof ([Druzdzel and van der Gaag, 2000](#)). This paper introduces a procedure for eliciting quantities describing the uncertainty about some phenomenon or experiment.

Typically, the uncertainty is modeled in terms of probabilities, such as in a Bayesian network. Elicitation of probabilities is commonplace and well-studied ([Spetzler and Staël von Holstein, 1975](#); [Cooke, 1991](#); [Renooij, 2001](#); [O'Hagan et al., 2006](#)). Apart from other issues such as various biases, there is agreement that eliciting probabilities directly and as precise numbers is often problematic due to a lack of familiarity with probability theory and the absence of a concrete context. Therefore, (i) targeted graphical interfaces such as scales and lottery wheels are designed, (ii) verbal descriptions of probability values are used, or (iii) the elicitation problem is reformulated as a betting problem. Another recurring idea is to use qualitative information such as comparative probability (see, e.g., [Druzdzel and van der Gaag, 1995](#)).

Uncertainty can also be modeled in alternative ways. One approach is to use generalizations of probabilities, such as imprecise-probabilistic models ([Walley, 1991](#); [Augustin et al., 2014](#)). Given that

the theory of imprecise probabilities encompasses probability intervals and comparative probability, imprecise-probabilistic techniques are better suited to deal with the results of an elicitation procedure, as even some of the most ardent ‘precise’ probabilists admit (O’Hagan and Oakley, 2004, Section 3.3).

In this paper, assuming the elicitation problem can be formulated in betting terms, we discuss an interface to elicit coherent sets of acceptable gambles, also called desirable gambles (Walley 1991, Appendix F; Quaeghebeur 2014). Roughly speaking, the gambles (random variables) in such a set are those for which the elicitee’s expectation is at least zero. Our interface can then, e.g., be used for eliciting the parameters of a credal network that is defined in terms of sets of acceptable gambles (De Bock and de Cooman, 2015). Moreover, sets of acceptable gambles can equivalently be transformed into the more classical imprecise-probabilistic models, credal sets (convex sets of probabilities) and lower expectations (previsions); so the procedure can also be used for eliciting, e.g., the parameters of classical—credal set based—credal networks. One can always obtain a single probability measure by selecting it from an elicited credal set in a principled way (see, e.g., Druzdzel and van der Gaag, 1995), opening up the option for also eliciting, e.g., Bayesian networks.

The interface for eliciting sets of acceptable gambles we present is designed for three-outcome possibility spaces, i.e., involving three mutually exclusive and exhaustive events. The design ideas could be adapted for the much simpler case of a two-element possibility space. For larger possibility spaces, the interface can be used in combination with an appropriate decomposition thereof. For example, a marginal extension theorem (Quaeghebeur, 2014, Theorem 1.2) guarantees that we can coherently combine a (marginal) coherent set of desirable gambles on some partition and (conditional) coherent sets of desirable gambles on the partition elements. So working with a hierarchical partitioning of the possibility space with partition elements of cardinality three or less is an option; using two-outcome spaces only would result in reduced expressiveness.

Next to the interface itself, we discuss a real-life experiment conducted as an exploratory test of our elicitation interface. It was organized around the 2014 FIFA World Cup. But first, we start by giving a brief primer on the theory of coherent sets of acceptable gambles.

## 2. Sets of Acceptable Gambles

Coherent sets of acceptable gambles are an imprecise-probabilistic model originally introduced by Williams (1976, Section IV). The idea essentially lay dormant until this model was advocated by Walley (1991, Appendix F; 2000, Section 6) using the term ‘desirable gambles’.

### 2.1 Essential Concepts

The *possibility space*  $\Omega$  describes the events about which there is uncertainty. For this paper we may assume it is a finite set. Formally, a *gamble* is a real-valued function on the possibility space. It represents a positive or negative payoff that depends on the unknown actual realization  $\omega \in \Omega$ .

An elicitee finds a gamble  $g$  on  $\Omega$  *acceptable* if she is in some sense committed to the following transaction: Once the realization  $\omega \in \Omega$  is determined, she gets the payoff  $g(\omega)$ . The set of gambles the elicitee assesses to be acceptable is denoted by  $\mathcal{A}$ . We assume it to be finite, which is reasonable in an elicitation context. A consequence is that all nontrivial checks and computations can be done using linear programming (Quaeghebeur, 2013).

We consider accepting a gamble that is everywhere negative to be irrational. Based on the assumption that the gamble payoffs are expressed in a linear utility—e.g., small amounts of money—we also consider positive linear combinations of acceptable gambles to be acceptable. Consequently,

an assessment  $\mathcal{A}$ , even if it does not contain negative gambles, may nevertheless be irrational; to wit, the elicitee can be forced to *incur a sure loss* by combining some of the gambles she accepts. Formally, this happens if there are real coefficients  $\lambda_g \geq 0$  such that  $\sum_{g \in \mathcal{A}} \lambda_g g < 0$ , where the sum and inequality are taken pointwise, i.e., hold for all  $\omega$  in  $\Omega$ .

To the above assumptions, we add that it is irrational to not accept nonnegative gambles. So we arrive at the following set of *coherence* axioms, which describes the essential properties a *deductively closed* set of acceptable gambles  $\mathcal{D}$  should satisfy:

$$\begin{aligned} \text{Avoiding Sure Loss: } & g < 0 \Rightarrow g \notin \mathcal{D}, & \text{Addition: } & g, h \in \mathcal{D} \Rightarrow g + h \in \mathcal{D}, \\ \text{Accepting Partial Gains: } & g \geq 0 \Rightarrow g \in \mathcal{D}, & \text{Positive Homogeneity: } & g \in \mathcal{D}, \lambda_g > 0 \Rightarrow \lambda_g g \in \mathcal{D}. \end{aligned}$$

This set of axioms forces  $\mathcal{D}$  to be a convex cone in the linear space of gambles on  $\Omega$  that includes the positive orthant and does not intersect the negative orthant. Note that this set of axioms allows gambles that are strictly negative on some nontrivial event  $B \subset \Omega$  and zero on its complement to be acceptable: this is interpreted as the elicitee considering the event  $B$  to be (practically) impossible.

Given an elicited assessment  $\mathcal{A}$  that avoids sure loss, we can use the last three axioms in a generative way, to extend the assessment to a coherent set of acceptable gambles

$$\mathcal{D} := \left\{ f + \sum_{g \in \mathcal{A}} \lambda_g g : \text{gamble } f \geq 0, \text{ coefficients } \lambda_g \geq 0 \right\}.$$

This deductive closure is called the *natural extension* of  $\mathcal{A}$ . An illustration of an assessment that avoids sure loss and its natural extension is given in Figure 1a.

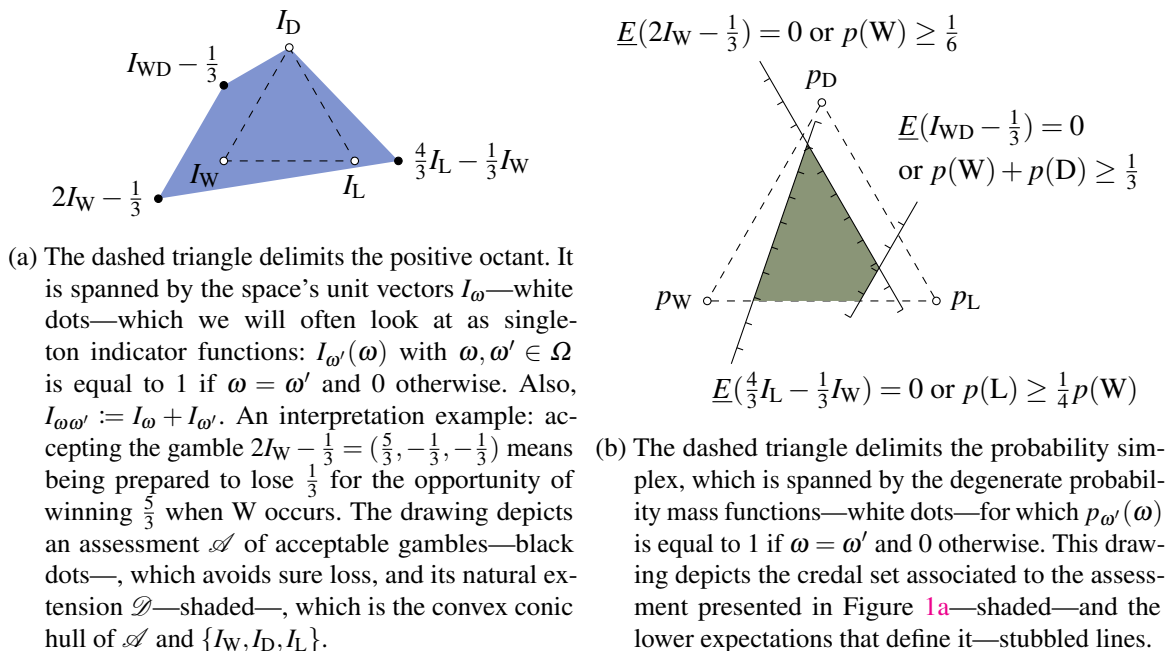


Figure 1: We consider the possibility space  $\{W, D, L\}$  (for ‘Win’, ‘Draw’, and ‘Loss’). Of the resulting 3-dimensional space of gambles, Figure 1a shows the plane with gambles whose payoffs sum to one. Figure 1b shows the plane containing the resulting probability simplex.

## 2.2 Relationship with Other Models

Although modeling the uncertainty of an elicitee with the set of gambles she accepts is quite natural and direct, it does differ from the usual, probability-based approach. Let us therefore have a look at how a coherent set of acceptable gambles is related to more common models: expectation operators and sets of probability mass functions.

A coherent set of acceptable gambles  $\mathcal{D}$  determines, for any gamble  $h$  on  $\Omega$ , the supremum acceptable buying price  $\underline{E}(h) := \sup\{\alpha \in \mathbb{R} : h - \alpha \in \mathcal{D}\}$  and infimum acceptable selling price  $\bar{E}(h) := \inf\{\beta \in \mathbb{R} : \beta - h \in \mathcal{D}\} = -\underline{E}(-h)$ . The nonlinear operators  $\underline{E}$  and  $\bar{E}$  are called *lower* and *upper expectations* or *previsions* (Walley, 1991). They satisfy  $\underline{E} \leq \bar{E}$  pointwise, i.e., for all gambles. So the gambles the elicitee finds acceptable are essentially those with lower expected payoff greater than or equal to zero.

With a lower expectation  $\underline{E}$ , we can associate a *credal set*  $\mathcal{M} := \{p : \underline{E} \leq E_p\}$ , consisting of all the probability mass functions  $p$  whose expectation  $E_p$  dominates  $\underline{E}$ . In this definition, the inequalities are again pointwise. Our set of axioms forces  $\mathcal{M}$  to be a convex subset of the probability simplex. We give an illustration in Figure 1b.

## 3. The Elicitation Interface

We first briefly discuss elicitation of probability mass functions and credal sets to provide some context and contrast. Then we move on to acceptable gambles.

### 3.1 Probability Elicitation

Looking at Figure 1b, an interface to elicit a probability mass function on a possibility space of three elements presents itself naturally: Allow the elicitee to indicate a point of the probability simplex.

To elicit a credal set, the above idea should be extended in a way that allows the elicitee to delimit a convex subset of the probability simplex. The three most obvious general approaches are:

- A *direct approach* is to allow multiple points to be selected and take their convex hull. The main advantage is the point-and-click nature, but the elicitee will have difficulty interpreting her actions.
- *Bounding the expectation of gambles* or, more specifically, providing probability intervals is easier to interpret. This can be achieved by ‘placing’ stubbled lines (see Figure 1b), but will result in a more involved interface.
- An *interpretation-agnostic approach* is to partition the simplex into a limited number of points and convex sets which can be selected and combined, e.g., based on comparative probabilities. The main advantage is the point-and-click nature, but there will be non-expressible elicitee attitudes.

The literature focuses mostly on interfaces for eliciting precise probabilities and continuous distributions, nowadays often interactive and on-line (Bastin et al., 2013; Morris et al., 2014). In the imprecise probabilities literature we can find thoughtful consideration of the issue of elicitation (see, e.g., Piatti et al., 2010), but mostly only elicitation interfaces for binary variables—e.g., ‘Win’ vs. ‘No Win’—are considered.

### 3.2 Gamble Space Representation

Walley (1991, Section 4.1) already considered the direct elicitation of acceptable gambles. But, since the credal set representation of imprecise probabilities has received most attention, it seems that these ideas never led to the concrete design of elicitation interfaces until now.

The space of gambles we have to consider is three-dimensional, because—as with the probability simplex interface—we require our interface to be two-dimensional due to practical display technology limitations. But doing this for sets of acceptable gambles is not so straightforward as for credal sets (cf. Section 3.1): The representation of Figure 1a was based on the restriction to the plane of gambles whose values sum to one. However, not all coherent sets of acceptable gambles can be compactly depicted in such a representation; as an extreme example, the above-mentioned plane is strictly contained in the coherent set consisting of those gambles with components that sum to zero or more.

Nevertheless, because of the Positive Homogeneity axiom, we know that we can represent a coherent set of acceptable gambles—a convex cone—on a two-dimensional surface. For example, we could take its intersection with a sphere centered at the origin or some other suitable two-dimensional surface and then do a projection.

The nature of the projection is influenced by the following considerations: (i) because of the Accepting Partial Gains and Avoiding Sure Loss axioms, the positive octant and the negative octant do not need to be represented prominently or faithfully; (ii) the representation should be essentially invariant under a permutation of the elementary events to avoid introducing biases between them; and (iii) to allow for intuitive exploration by the elicitee, the representation should be a continuous deformation of the points in all but the positive and negative octants.

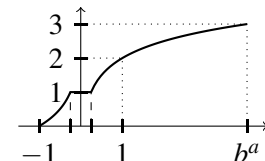
These considerations lead us to a polar projection, where the poles are defined by the line corresponding to constant gambles, i.e., those for which the payoff is equal for all possible outcomes. On the right, we show the example of a spherical such projection found in the United Nations emblem.



To decide on the exact surface to project and the projection center—our ‘North Pole’—we refocus on the interpretation of the projection points as gambles. To provide the elicitee with a reference value when selecting acceptable gambles, we should anchor them by fixing either their maximum or minimum value. We fix the minimum value, as this bounds potential losses and so may—hypothetically—mitigate effects of risk-aversion. We here take  $-1$  as the normalized minimum value. Consequently, the surface we consider is the set of gambles  $\{f : \min f = -1\}$ , namely, the convex cone with apex  $(-1, -1, -1)$  and extreme rays  $(1, 0, 0)$ ,  $(0, 1, 0)$ , and  $(0, 0, 1)$ . The apex is also the projection center. The projection is illustrated in Figure 2.

The illustration of our representation provided by Figure 3a allows us to pinpoint an important deficiency: the limited range due to the linear scale used. For example, the difference between gamble values one and two times the stake is practically speaking much more important than between five and six times the stake. (The same argument can be used for probability values.) In such a context where relative magnitude takes precedence over absolute magnitude, using a logarithmic scale is a better choice. Because nonpositive values are used in our representation and the constant gamble  $-1$  corresponds to the center of our representation, we use a custom scaling that is based on a ‘saturating’ logarithm:

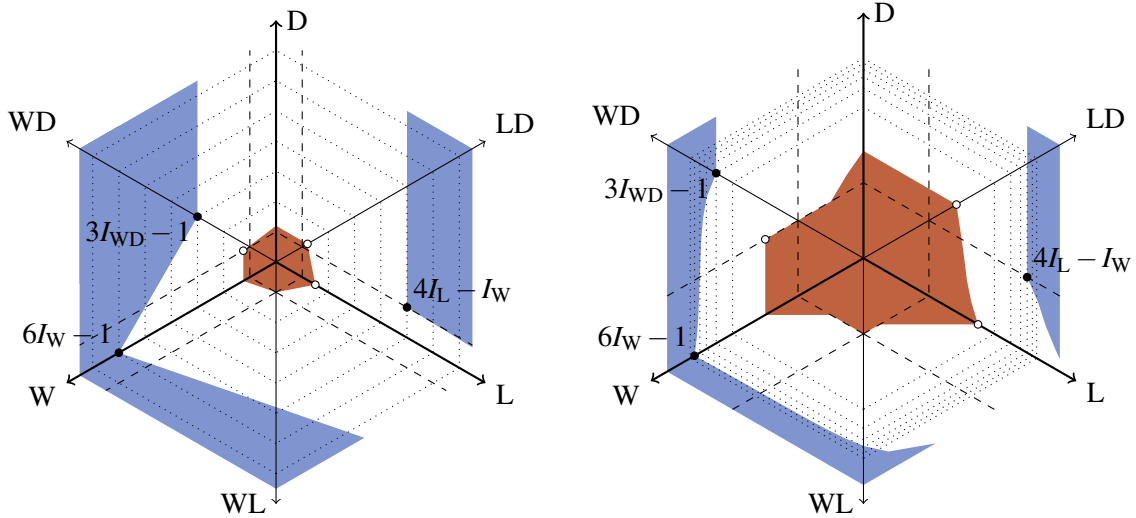
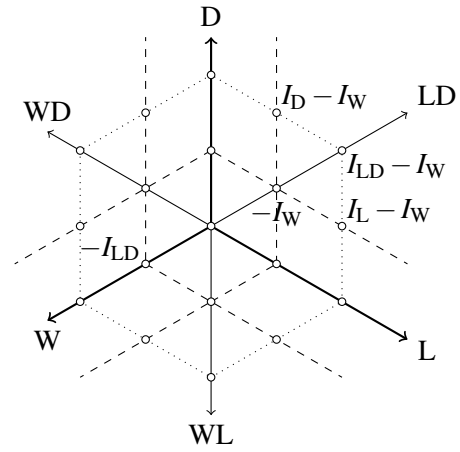
$$f_b^a(x) = \begin{cases} 0 - \frac{1}{a} \log_b(-x), & x \in [-1, -b^{-a}], \quad (\text{values in } [0, 1)) \\ 1, & x \in [-b^{-a}, b^{-a}], \quad (\text{saturating value } 1) \\ 2 + \frac{1}{a} \log_b(x), & x \in (b^{-a}, b^a], \quad (\text{values in } (1, 3]) \end{cases} \quad (1)$$



where  $b > 1$  and  $a > 0$  determine the smallest representable absolute value  $b^{-a}$  and a practical upper bound  $b^a$ . The impact on our representation of this scale is illustrated in Figure 3b.

The considerations underlying our representation and its technical details are quite involved. However, the elicitee need not be aware of these to use our concrete implementation, described next.

Figure 2: This is a polar projection of the gamble-space subset  $\{f : \min f = -1\}$ . The white dots correspond to gambles of interest: The central one represents the apex, the gamble with constant value  $-1$ . The others—some labeled—represent differences of event indicator functions; for example  $-I_{LD} = (0, -1, -1)$  and  $I_L - I_W = (-1, 0, 1)$ . The thick and thin axes point towards gambles with higher payoffs for the corresponding events. The dashed lines form the set of *contingent gambles*, i.e., those that are zero on some event. The dotted line indicates the set of ‘even’ gambles, with a maximum payoff equal to the stake, i.e., one.



(a) This drawing uses our proposed projection of the gamble space. In comparison to Figure 2, the scale we use here is smaller and we now show dotted lines for the loci of gambles with maximum payoff one to six.

(b) This is Figure 3a rescaled using the function of Equation (1) with parameters  $b = 2$  and  $a = 4$ . The effect of the logarithm can, e.g., be seen in the nonlinearity of lines connecting the assessment gambles—black dots. A consequence of saturation is the disappearance of lines between white dots ‘into’ the negative octant (central hexagon).

Figure 3: Translations of Figure 1a to alternate gamble space representations (renormalization to satisfy  $\min f = -1$ ). The gambles in the set  $\mathcal{A}$  assessed to be acceptable are represented by black dots. Their natural extension  $\mathcal{D}$ , represented by a closed convex polytope before, is now (partially) represented by the disconnected shaded area on the outside. We have added the pointwise additive inverse of the gambles assessed to be acceptable—white-filled dots. These determine the open convex polytope of ‘rejected’ gambles—shaded area in the center—that would cause a sure loss if one or more of them were to be assessed acceptable.

### 3.3 Eliciting Acceptable Gambles

With deployment convenience in mind, our representation was turned into a concrete elicitation interface by implementing it in SVG and Javascript (ECMAScript) so that it can be used in current web browsers. We relied on the library d3js (Bostock et al., 2011) for visualization and the library NumericJS (Loisel, 2012) for linear programming functionality.

The biggest and only substantive change we had to make was a discretization of the gamble space. There are two reasons for this:

- We have not found a way to calculate and represent the sets of acceptable and rejected gambles (cf. Figure 3a) fast enough to obtain a responsive interface. This is mainly due to the nonlinear character of their borders, which is a result of the logarithmic rescaling. We can work around this issue by discretizing the representation.
- We wish to show the values of the gamble over which the elicitee is hovering with her pointer. We do not want to show a large number of significant digits of these values, because it is unrealistic to expect the elicitee’s uncertainty attitudes to be so fine-grained; this would therefore be distracting. But now, if we show only a few significant digits and wish to make sure that the numbers shown correspond to the gamble under the pointer, we must discretize the representation.

The result is shown in Figure 4a; it can be used without detailed knowledge of the representation.

The biggest computational challenge we faced when implementing the interface was finding the natural extension  $\mathcal{D}$  efficiently enough to make it responsive. To tackle this, we split up the problem into different subroutines. The ones that provide the most important efficiency gain are the *propagation* routines:

- In our interface, for each dot—i.e., gamble  $f$  in the discretization—, there is a unique dot in its negation’s neighborhood whose negation—up to scaling—strictly dominates the gamble  $f$ ; e.g.,  $(-\frac{1}{2}, 1, -1)$  negation-dominates  $(\frac{1}{4}, -1, \frac{1}{2})$ . We pre-calculate this *negation structure*. So when a dot is marked as accepted, we can mark the dot it is negation-dominated by as rejected due to the Avoiding Sure Loss and Addition axioms.
- In our interface, for each dot, the neighboring dots either pointwise dominate it or are dominated by it; e.g.,  $(\frac{1}{4}, -1, \frac{1}{2})$  dominates  $(\frac{1}{4}, -1, \frac{1}{4})$ . We pre-calculate this *dominance structure*. So when a dot is marked as accepted (or rejected), we can recursively propagate this status to all its dominating (dominated) dots due to the Accepting Partial Gains and Addition axioms. All unmarked dots that neighbor an accepted (rejected) dot are gathered in a list of accept (reject) *candidates*, which is kept up to date while propagating.

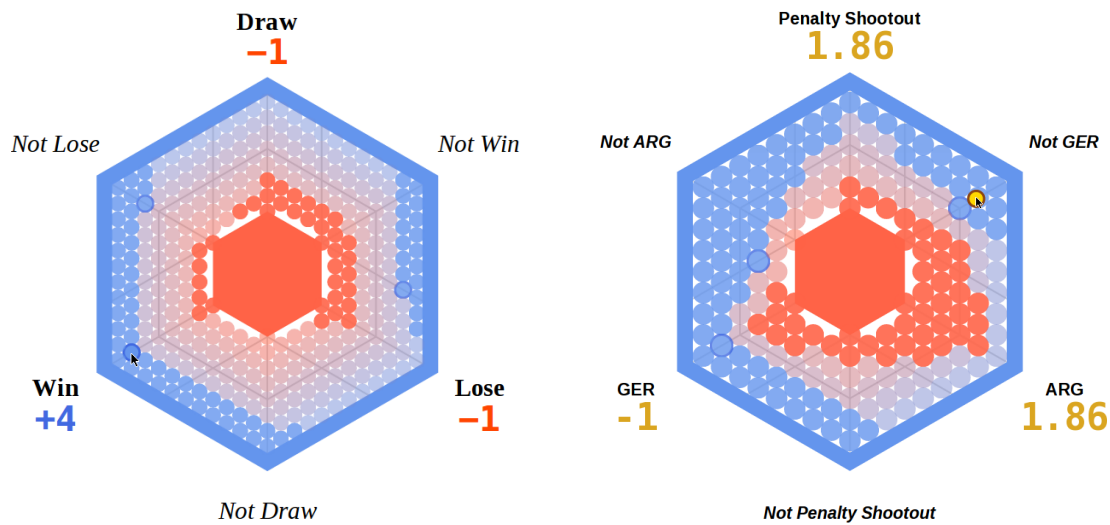
In our procedure, whenever a dot is marked as accepted or rejected, this change is fully propagated before continuing with the outer *search* routine:

- We iterate over the accept (or reject) candidates and check whether they should be marked acceptable (rejected). The iteration order is determined by the heuristic that dots ‘low’ (‘high’) in the dominance structure should come first, as they are most promising from the propagation perspective. The candidate lists are pre-populated by applying propagation to the assessment.

The subroutine that checks the status of a dot effectively calculates the natural extension by determining whether the dot’s lower (or upper) expectation is nonnegative (strictly negative). Calculating a lower (upper) prevision requires linear programming, a relatively computationally demanding task in a web browser. Propagation sufficiently reduces the number of prevision calculations in practice.

Once the interface was ready for action, we moved to test it in a practical experiment with the goal of getting usage data and general feedback. This is the topic of the next section.





(a) This shows the interface we implemented with  $\{-1, -\frac{1}{2}, -\frac{1}{4}, -\frac{1}{8}, 0, \frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1, 2, 4, 8\}$  as the discretization values. We have chosen the assessment gambles—e.g.,  $(4, -1, -1)$ —to mimic Figure 3b. (b) This shows the experiment’s interface variant, with  $\{-1, -\frac{1}{2}, -\frac{1}{4}, 0, \frac{1}{4}, \frac{1}{2}, 1, 2, 4\}$  as the set of discretization values. The assigned gamble appears as the light (yellow) dot hovered over by the pointer.

Figure 4: Elicitation interface screenshots. Each dot represents a gamble; when hovering over it, the payoff for each outcome is shown; assessment dots are a bit darker and have a border.

## 4. The Experiment

In 1982, [Walley \(1991, Appendix I\)](#) ran an experiment for eliciting lower and upper probabilities concerning the outcome of matches in that year’s FIFA World Cup. There were 17 academic participants. Their assessments were evaluated using the possible pairwise bets between them (cf. Section 4.2). This experiment has later been repeated in the imprecise probabilities community, but more as a diversion than in search of data. Others, such as [Winkler \(1971\)](#), ran earlier, precise probability elicitation experiments organized around sports competitions.

We organized our experiment around the 2014 FIFA World Cup. Whereas in 1982 pen and paper was used, we can now deliver a point-and-click interface accessible over the internet. Moreover, we can calculate the natural extension on-line and thus make sure the elicitee avoids sure loss.

### 4.1 A Betting Competition Website

We set up the competition as a betting website. We provided the following instructions:

The [...] Competition is a game in which you gamble against other participants. Each of the possibly many matches for which you enter the ‘gambling pool’, you stake €1 (or [...], e.g., \$1), so you can only lose this amount or less per match. The exact amounts you can win or lose depend on the other players’ choices.

You play by expressing your opinion about the outcome of the match in terms of gambles that are acceptable to you. [...] An algorithm will look for a fair bet between you and the other participants. [A bet is a set of gambles assigned to the participants.] If a bet is found, and you are included in it, a single acceptable gamble will be assigned to you and appear in the gamble selection interface for that match. This assigned gamble determines your potential winnings [...]

and losses. You may change your set of acceptable gambles up until an hour before the start of the match; [...]. The match's outcome determines your actual winnings or losses.

Actually, any winnings will not go to you, but, by participating, you commit yourself to pay your losses to the Red Cross/Crescent [...]

The website and its backend were developed using the [Django framework](#). Figure 4b shows a match screen with an assigned gamble.

Participants were recruited by word of mouth initially, then through academic mailing lists, and via social networks near the end of the competition. Participation was anonymous; only a hash of the sign-up email address was stored. A total of 80 people participated, providing assessments for 32 of the 64 matches (due to time constraints, the website was launched and tweaked while the World Cup was already ongoing), for a total of 488 gambles assessed to be acceptable. For 20 of those 32 matches bets were possible, for a total of 100 assigned gambles (cf. Section 4.2). The sum of the lower common expected winnings for those gambles was 37.86 currency units, and in the end the total amount won—and lost—was 47.19 currency units. Even though we could not enforce participants paying to the Red Cross/Crescent, adding the element of having real money at stake, even if not much, was important because it incentivizes them to take the elicitation task more seriously.

This experiment is not just meant as a one-off setup to test our gamble selection interface. Many of its elements can be used as inspiration for practical elicitation experiments. The competitive aspect can be used when eliciting from multiple experts (cf. [Lichtendahl and Winkler, 2007](#)). The repetition aspect is relevant when considering time series, e.g., in a context of weather forecasting.

## 4.2 Generating Fair Bets

For his experiment, [Walley \(1991, Appendices I and H6\)](#) scored the participants by arranging specific pairwise bets between them for each match, whenever possible. These bets assigned opposite gambles to each that were acceptable given their assessments. The gambles were moreover fair in the sense that their lower expectations—supremum acceptable buying prices (see Section 2.2)—coincided.

We used the same acceptability requirement but moved away from the pairwise approach to a global one, in which for each match a single bet was generated, i.e., the assigned gambles sum to zero. One reason is that we wished the stake per match to be at most one. An advantage of such a setup is that the set of potential bets is in general much larger—e.g., it includes all convex combinations of pairwise bets. Therefore we needed a criterion to choose a single one: we decided on a utilitarian one by maximizing the sum of *identical* acceptable buying prices for the assigned gambles. Our notion of fairness is this maximal common price instead of [Walley's](#) common maximal price.

A bet satisfying the constraints described above can be computed using mixed-integer linear programming. Its formulation is independent of the size of the possibility space  $\Omega$ . Participant  $j$  in  $J$  has specified an assessment  $\mathcal{A}_j$ . He may be included in the bet or not, encoded by the binary variable  $b_j$ . If included, he will be assigned a nontrivial gamble  $h_j$ . His acceptable buying price for this gamble is  $\alpha_j$  and must—by fairness—be identical to the common buying price  $\alpha_*$ . So we have the following program (gamble constraints must be read pointwise):

$$\text{maximize } \sum_{j \in J} \alpha_j = \alpha_* \sum_{j \in J} b_j \tag{2}$$

$$\text{subject to } \sum_{j \in J} h_j = 0, \tag{3}$$

$$\text{and for all } j \text{ in } J: h_j - \alpha_j \geq \sum_{g_j \in \mathcal{A}_j} \lambda_{j,g_j} g_j \text{ with } \lambda_{j,g_j} \geq 0, \tag{4}$$

$$0 \leq \alpha_j = b_j \alpha_* \text{ and } -1 \leq h_j = b_j h_j \text{ with } b_j \in \{0, 1\}. \tag{5}$$

Constraint (4) expresses that  $h_j - \alpha_j$  must lie in the cone  $\mathcal{D}_j$  spanned by  $\mathcal{A}_j$  and the first orthant (cf. Section 2.1). Objective (2) then forces  $\alpha_j$  to be an acceptable buying price of  $h_j$  (cf. Section 2.2). Constraint (3) guarantees that the  $h_j$  gambles form a bet between the participants. Constraints (5) give us the freedom to exclude participants from the bet, force the acceptable prices for those included to coincide, and force the stakes to be one or less.

However, the products  $b_j\alpha_*$  and  $b_jh_j$  make the Constraints (5) nonlinear. Luckily, we can replace them by an equivalent set of linear constraints: Notice first that  $h_j \geq -1$  together with Constraint (3) implies the bound  $h_j \leq |J| - 1$ . Then the Constraint (4) with all  $\lambda_{j,g_j} = 0$  further implies that also  $\alpha_j \leq |J| - 1$ . The existence of these bounds allows us—given  $b_j \in \{0, 1\}$ —to replace (5) by

$$\alpha_* - (|J| - 1)(1 - b_j) \leq \alpha_j \leq \alpha_*, \quad 0 \leq \alpha_j \leq (|J| - 1)b_j, \quad -b_j \leq h_j \leq (|J| - 1)b_j. \quad (6)$$

We used the linear programming library GLPK (Makhorin, 2014) without practical efficiency issues. An instance of an assigned gamble calculated using this program can be seen in Figure 4b.

### 4.3 Experimental Results

The aim of our exploratory experiment was to obtain feedback about the gamble selection interface, get a view of the types of assessments people provide, test the fair bet generation procedure on real-life data, and have some fun doing it.

We provided a form where participants could optionally enter feedback and information about themselves, such as gender, age, and experience relevant to the competition. However, almost no one made use of it. We did get quite a bit of feedback through personal communications with participants we knew, both laymen and people experienced in uncertainty modeling. As was anticipated by the human-computer interaction (HCI) expert in our team, the interface was found to be too complex: it needs to be simplified, explained more extensively, or a combination thereof.

There were 194 match assessments in total, of which a good 20% was complete in the sense that all dots were marked—after natural extension—, so corresponding to some probability mass function. (A nice anecdote: The few participants who used complete models almost exclusively all had greater losses than winnings.) For the others, with strictly imprecise-probabilistic assessments, the degree of completeness varied over the whole range between just a few and all but a few marked dots.

Something generally orthogonal to completeness is the number of selected dots per assessment:

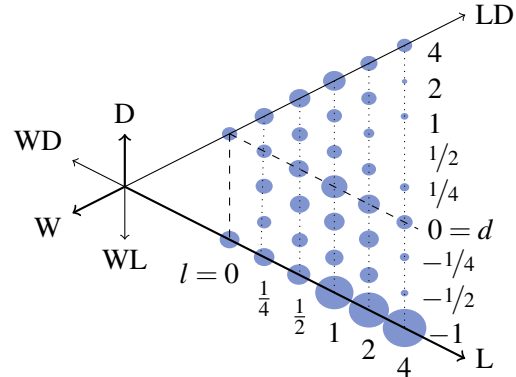
|               |    |    |    |    |   |   |   |   |
|---------------|----|----|----|----|---|---|---|---|
| #dots:        | 1  | 2  | 3  | 4  | 5 | 6 | 7 | 8 |
| #assessments: | 54 | 52 | 47 | 26 | 8 | 5 | 1 | 1 |

We see that the number of selected dots is mostly concentrated in relatively small numbers. In fact, whenever four or more dots were registered, there usually were some that were actually redundant after natural extension, i.e., were implied by other selected dots. So participants kept things simple.

Regarding the distribution of the dot selection, Figure 5 shows that mainly dots on the axes and corresponding to contingent gambles were chosen, but not overwhelmingly so. It nevertheless indicates that restricting attention to these loci may be a way to simplify the interface.

The bet generation procedure worked as intended, but we noticed that on average a third and sometimes up to more than 60% of the participants that gave an assessment for a match—typically relatively more imprecise (incomplete)—were not included in the bet. Once the match outcome is determined, the assigned gamble results in a payoff; such feedback gives the participant an idea of the quality of his assessments (cf. scoring rules; see, e.g., Winkler, 1971). Therefore it would be useful to modify the bet generation procedure to include more participants.

Figure 5: This drawing depicts the observed relative dot selection frequency. Because of symmetry—outcome identity irrelevance—all dots were mapped to the subregion of dots  $(-1, d, l)$  (cfr. Figure 3b). The possible values for  $d$  and  $l$  are respectively shown on the right and at the bottom of the drawing. The *area* of the circles is proportional to the relative number of selections of that dot; the largest circle, at  $(-1, -1, 4)$ , corresponds to 12.5% of selections.



## 5. Conclusions

We designed a gamble selection interface that is based on a representation of the space of gambles tailored to elicitation. We coded an efficient—responsive—implementation and used it in an experiment. In support of this experiment, we developed a novel procedure for generating fair bets.

From the experiment, we learned that the interface can be effectively used, but also that it needs to be made more usable and transparent. Furthermore, given that the majority of assessments made by the participants were imprecise, a more generally important conclusion is that *imprecision is a non-negligible aspect of uncertainty*: models that do not allow for it to be expressed may lead to gambles—i.e., any decision under uncertainty—that its users are actually not willing to commit to.

Follow-up work should focus on improvements to the interface, user guidelines, and bet generator, and experimental comparison to alternative interfaces (see, e.g., Section 3.1).

## Acknowledgments

This work is part of the *Safe Statistics* project at the CWI financed by the Netherlands Organisation for Scientific Research (NWO). Erik Quaeghebeur was an ERCIM “Alain Bensoussan” Fellow, receiving funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 246016. Tom Sterkenburg is also affiliated with the Faculty of Philosophy of the University of Groningen. Teresa Piovesan is partially funded by the European Project SIQS.

## References

- T. Augustin, F. P. A. Coolen, G. de Cooman, and M. C. M. Troffaes, editors. *Introduction to Imprecise Probabilities*. Wiley, 2014. doi:[10.1002/9781118763117](https://doi.org/10.1002/9781118763117).
- L. Bastin et al. Managing uncertainty in integrated environmental modelling: The UncertWeb framework. *Environ. Modell. Software*, 39:116–134, 2013. doi:[10.1016/j.envsoft.2012.02.008](https://doi.org/10.1016/j.envsoft.2012.02.008).
- M. Bostock, V. Ogievetsky, and J. Heer. D<sup>3</sup>: Data-driven documents. *IEEE Trans. Visual Comput. Graphics*, 17(12):2301–2309, 2011. doi:[10.1109/TVCG.2011.185](https://doi.org/10.1109/TVCG.2011.185). URL <http://d3js.org/>.
- R. M. Cooke. *Experts in Uncertainty*. Oxford University Press, 1991.

- J. De Bock and G. de Cooman. Credal networks under epistemic irrelevance: The sets of desirable gambles approach. *Int. J. Approx. Reason.*, 56:178–207, 2015. doi:[10.1016/j.ijar.2014.07.002](https://doi.org/10.1016/j.ijar.2014.07.002).
- M. J. Druzdzel and L. C. van der Gaag. Elicitation of probabilities for belief networks: Combining qualitative and quantitative information. In P. Besnard and S. Hanks, editors, *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence*, page 141–148. Morgan Kaufmann, 1995.
- M. J. Druzdzel and L. C. van der Gaag. Building probabilistic networks: Where do the numbers come from? *IEEE Trans. Knowl. Data Eng.*, 12(4):481–486, 2000. doi:[10.1109/TKDE.2000.868901](https://doi.org/10.1109/TKDE.2000.868901).
- K. C. Lichtendahl, Jr. and R. L. Winkler. Probability elicitation, scoring rules, and competition among forecasters. *Manage. Sci.*, 53(11):1745–1755, 2007. doi:[10.1287/mnsc.1070.0729](https://doi.org/10.1287/mnsc.1070.0729).
- S. Loisel. Numeric javascript, 2012. URL <http://numericjs.com/>.
- A. Makhorin. GNU Linear Programming Kit, 2014. URL <http://www.gnu.org/software/glpk/>.
- D. E. Morris, J. E. Oakley, and J. A. Crowe. A web-based tool for eliciting probability distributions from experts. *Environ. Modell. Software*, 52:1–4, 2014. doi:[10.1016/j.envsoft.2013.10.010](https://doi.org/10.1016/j.envsoft.2013.10.010).
- A. O’Hagan and J. E. Oakley. Probability is perfect, but we can’t elicit it perfectly. *Reliab. Eng. Syst. Saf.*, 85(1–3):239–248, 2004. doi:[10.1016/j.res.2004.03.014](https://doi.org/10.1016/j.res.2004.03.014).
- A. O’Hagan et al. *Uncertain Judgements: Eliciting Experts’ Probabilities*. Wiley, 2006.
- A. Piatti, A. Antonucci, and M. Zaffalon. Building knowledge-based systems by credal networks: a tutorial. In A. R. Baswell, editor, *Advances in Mathematics Research*, volume 11, page 227–279. Nova Science Publishers, 2010.
- E. Quaeghebeur. The CONEstrip algorithm. In R. Kruse et al., editors, *Synergies of Soft Computing and Statistics for Intelligent Data Analysis*, page 45–54. Springer, 2013. doi:[10.1007/978-3-642-33042-1\\_6](https://doi.org/10.1007/978-3-642-33042-1_6).
- E. Quaeghebeur. Desirability. In [Augustin et al. \(2014\)](#), chapter 1, page 1–27.
- S. Renooij. Probability elicitation for belief networks: Issues to consider. *The Knowledge Engineering Review*, 16(3):255–269, 2001. doi:[10.1017/S026988901000145](https://doi.org/10.1017/S026988901000145).
- C. S. Spetzler and C.-A. S. Staël von Holstein. Probability encoding in decision analysis. *Manage. Sci.*, 22(3):340–358, 1975. doi:[10.1287/mnsc.22.3.340](https://doi.org/10.1287/mnsc.22.3.340).
- P. Walley. *Statistical reasoning with imprecise probabilities*. Chapman & Hall, 1991.
- P. Walley. Towards a unified theory of imprecise probability. *Int. J. Approx. Reason.*, 24(2–3):125–148, 2000. doi:[10.1016/S0888-613X\(00\)00031-1](https://doi.org/10.1016/S0888-613X(00)00031-1).
- P. M. Williams. Indeterminate probabilities. In M. Przełęczki et al., editors, *Formal Methods in the Methodology of Empirical Sciences*, page 229–246. D. Reidel Publishing Company, 1976. doi:[10.1007/978-94-010-1135-8\\_16](https://doi.org/10.1007/978-94-010-1135-8_16).
- R. L. Winkler. Probabilistic prediction: Some experimental results. *J. Am. Stat. Assoc.*, 66(336):675–685, 1971.