# Utilizing Surrogate Numbers for Probability Elicitation

Mats Danielson, Love Ekenberg and
Andreas Paulsson

Additional information is available at the end of the chapter

## Abstract

Comparatively few of the vast number of suggested decision-analytical methods have been widely spread in actual practice. The majority of those methods call for exact and accurate numbers as input, which could be one of several reasons for this lack of actual use; people frequently seem to be unfamiliar with, or reluctant to express those, in a sense, "true" values required. Many alternative methods to resolve this complication have been suggested over the years, including procedures for dealing with incomplete information. One way, which has proliferated for a while, is to introduce so-called surrogate numbers in the form of ordinal ranking methods for multi-criteria weights. In this chapter, we show how those can be adapted for use in probability elicitation. Furthermore, when decision-makers possess more information regarding the relative strengths of probabilities, that is, some form of cardinality, the input information to ordinal methods is sometimes too restricted. Therefore, we suggest a testing methodology and analyze the relevance of a set of cardinal ordering methods in addition to the ordinal ones.

**Keywords:** decision analysis, probability elicitation, cardinal ranking, rank order, imprecise probability

## 1. Introduction

Elicitation of subjective beliefs has been applied in numerous areas, such as game and decision theory [1, 2], agriculture [3], statistics [4], and various disciplines of economics [5–7], and many methods have been designed for elicitation purposes [8–10]. It is in these contexts often assumed that at least experts in the various areas are capable of acting rationally and can provide reliable information so long as they are eliciting probabilities sensitively. Those assumptions are contrary

to the fact that even expert estimates may differ significantly from the true probabilities and that there are still no universally accepted methods of probability elicitation available. The process of eliciting adequate quantitative information is one of the substantial challenges within decision analysis [11, 12].

In a classical framework (cf., e.g., [13]), numerical probabilities are assigned to the different events in tree representations of decision problems. The assignments are made after the set of parameters, whose values need to be elicited and whom to choose for providing those estimates, have been identified. Domain experts could, for example, be asked to express their beliefs about the likelihood of a particular event in probabilistic form. Such beliefs cannot be measured objectively, and neither should they be judged in such a manner. Besides, the success of an elicitation process depends on how well a representation of the present subjective opinions can be constructed rather than on some set of objectively true values [14–16].

Methods for eliciting utilities and probabilities have been thoroughly investigated, resulting in a large number of recommendations and handbooks on the subject. Procedures range from using direct elicitation, gamble, and lottery techniques, as well as more elaborate methods to reduce biases, aversions, and a multitude of other causes of errors while producing as reliable estimates as possible (c.f., e.g., [17–22]). Here, it is generally assumed that procedures for elicitation should give rise to adequate preference orders, but this assumption is nevertheless often violated in empirical studies (c.f., e.g., [23–25]). A multitude of methods for ordinal rankings or interval approaches have been suggested to provide more realistic models. The goal is to be able to utilize the information the decision-makers can supply without forcing them to express unrealistic, misleading, or meaningless statements.

To elicit probabilities, methods based on ordinal rankings already constructed for obtaining weights for multi-criteria decision analysis (MCDA) (see [26–28]) can be adopted. The fundamental idea is that ordinal information that stems from importance can be converted to a set of normalized surrogate numbers, the values of which are consistent with the elicited ordinal rankings. The actual conversion can be done using a variety of methods, such as rank sum (RS) and rank reciprocal (RR) [29], and centroid-based methods (ROC) [30]; originally used for handling criteria weights in MCDA. However, even in MCDA, the use of only ordinal information is sometimes perceived as being too vague or imprecise, resulting in a lack of confidence in the alternatives' final values or a too large class of non-dominated alternatives.

In this chapter, a set of methods that allow for more expressive information as input when eliciting probabilities, while maintaining the relative correctness and simplicity of ordinal ranking procedures, is proposed. In the following section, we compare and discuss a set of significant features, including correctness and relevance, of various extensions to some existing ranking methods. Following a brief recapitulation and adaptation of some ordinal ranking methods in the following section, we continue with cardinal ranking methods and discuss a set of appealing candidates for probability elicitation. Using simulations, we investigate some properties of the treated methods and conclude with pointing out, according to the results, a particularly attractive method for eliciting probabilities.

## 2. Ordinal ranking methods in MCDA

Different elicitation formalisms by which a decision-maker can express preferences in MCDA decision situations have been proposed. Such formalisms are sometimes based on scoring points, as in point allocation (PA) or direct rating (DR) methods. In PA, the decision-maker is given some number of points, for example, 100, to distribute over a set of criteria or consequences, depending on the type of decision [31]. Hence, for $N$ criteria or consequences, there are $(N-1)$ degrees of freedom (DoF). Direct rating methods, on the other hand, put no limit on the number of points to be allocated.[1] The decision-maker allocates as many points as desired and the points are subsequently normalized. Thus, in DR, there are $N$ degrees of freedom for $N$ criteria. Regardless of elicitation method, the assumption is that all elicitations are made relative to a distribution held by the decision-maker.[2]

Surrogate methods are utilized in [26–28] and many others for handling such problems. Regardless of method, however, a key property must be to retain as much information as possible in the surrogate numbers, yet accommodating for the various constraints required by certain types of values.

Stillwell et al. [29] compare a set of different methods for eliciting surrogate numbers from ordinal rankings alone, based on the idea of maximizing the power to discriminate between values. Among those are rank sum and rank reciprocal, for which surrogate weights are derived solely from the rank order of the attributes. Take a simplex $S_w$ generated by $w_1 > w_2 > \cdots > w_N$, $\sum w_i = 1$ and $0 \leq w_i$.[3] Assign an ordinal number to each item ranked, starting with the highest ranked item as number 1. Denote the ranking number $i$ among $N$ items to rank. Then the rank sum (RS) surrogates for all $i = 1, \ldots, N$ are defined by

$$w_i^{\text{RS}} = \frac{N+1-i}{\sum_{j=1}^{N} (N+1-j)}. \tag{1}$$

The surrogate numbers produced by the rank reciprocal (RR) method are, as the name implies, based on the reciprocal of the rank of each item. Let the number 1 corresponds to the highest ranked item, the number 2 to the second highest ranked item, and so on. Then for the $i$:th ranked item, the RR surrogate, given a total of $N$ items, are obtained by

$$w_i^{\text{RR}} = \frac{1/i}{\sum_{j=1}^{N} 1/j}. \tag{2}$$

---

[1] It could be that the sum is limited in direct rating methods as well but then as a consequence of a uniform limit to the individual numbers.

[2] See, for example, [32, 33] and some others from the same authors on methodological and cognitive aspects of inexactness in decision-making.

[3] Unless stated otherwise, the component vectors of decision problems will be modeled as simplexes, $S_x$, consisting of $x_1 > x_2 > \cdots > x_N$, where $\sum x_i = 1$ and $0 \leq x_i$.

A decade later, Barron [30] suggested a method based on vertices of the simplex of the feasible space. The rank order centroid (ROC) weights are the centroid vector components of the simplex $S_w$. The weights then become the centroid (mass point) of $S_w$. The ROC weights for the ranking number $i$ among $N$ items to rank are given by

$$w_i^{\text{ROC}} = \frac{1}{N} \sum_{j=i}^{N} 1/j. \tag{3}$$

However, RS, RR, and ROC perform well only for specific assumptions on decision-maker behavior. If we assume that the decision-maker stores its preferences in a way similar to a given point sum, considering normalization, there are $(N-1)$ degrees of freedom (DoF) for $N$ items. On the other hand, if we assume that the decision-maker stores its preferences in a way that puts no limit to the total number of points (or mass) allocated, and the normalization is made afterward, then there are $N$ degrees of freedom for $N$ criteria. It remains an open question as to how a decision-maker perceives the nature of the basis of a particular preference order. Whether the linear dependence in $(N-1)$ DoF models is accounted for, or if preference values are allocated with no particular limits in accordance with $N$ DoF models. Surrogate numbers obtained by RR and ROC models agree with a preference structure based on $(N-1)$ degrees of freedom, while the RS model conforms to a preference structure based on $N$ degrees of freedom. Due to this apparent difference, a model, SR, in which features from both RS and RR were incorporated, was proposed by [34]. The SR method is an additive combination of the Sum and the Reciprocal functions as in

$$w_i^{\text{SR}} = \frac{\frac{1}{i} + \frac{N+1-i}{N}}{\sum_{j=1}^{N} \frac{1/j(N+1-j)}{N}}. \tag{4}$$

To exemplify the above, given any probability simplexes such as $(p_1, p_2, p_3)$ and $(p'_1, \ldots, p'_6)$ that satisfy the previously laid out assumptions, the various methods would assign to them numbers as in **Tables 1** and **2**.

In theory, the choice of elicitation method should preferably be based on the manner in which a decision-maker constructs a preference order; whether it is based on a $(N-1)$ DoF or $N$ DoF model, or even a mixed model. However, since we cannot know the ratio of $(N-1)$ DoF to $N$ DoF employed in the mind of the decision-maker, an elicitation method, to be robust, should work about equally well regardless. The SR method, assuming the principal features of both RS and RR, was found to be the method best accommodating for such a need.

| Method | $p_1$ | $p_2$ | $p_3$ |
|---|---|---|---|
| Rank sum (RS) | 0.50 | 0.33 | 0.17 |
| Rank reciprocal (RR) | 0.55 | 0.27 | 0.18 |
| Rank order centroid (ROC) | 0.61 | 0.28 | 0.11 |
| Sum reciprocal (SR) | 0.52 | 0.30 | 0.17 |

**Table 1.** Probabilities corresponding to the ordinal ranking, $p_1 > p_2 > p_3$, rounded to two decimal places.

| Method | $p_1{}'$ | $p_2{}'$ | $p_3{}'$ | $p_4{}'$ | $p_5{}'$ | $p_6{}'$ |
|---|---|---|---|---|---|---|
| Rank sum (RS) | 0.29 | 0.24 | 0.19 | 0.14 | 0.10 | 0.05 |
| Rank reciprocal (ROC) | 0.41 | 0.20 | 0.14 | 0.10 | 0.08 | 0.07 |
| Rank order centroid (ROC) | 0.41 | 0.24 | 0.16 | 0.10 | 0.06 | 0.03 |
| Sum reciprocal (SR) | 0.34 | 0.22 | 0.17 | 0.13 | 0.09 | 0.06 |

**Table 2.** Probabilities corresponding to the ordinal ranking, $p_1' > p_2' > p_3' > p_4' > p_5' > p_6'$, rounded to two decimal places.

In the following, the methods for weight elicitation within MCDA presented above will be augmented with information denoting the relative difference between adjacent items and modified to meet the requirements of probability elicitation. We adhere to a standard, one-level decision-tree model, in which each of $M$ alternatives has $N$ consequences. Hence, there are $M$ times $N$ consequences in total.

## 3. Cardinal ranking methods

Providing ordinal rankings puts fewer demands on decision-makers; they are, in a sense, effort saving. Furthermore, there are techniques such as those mentioned earlier for handling ordinal rankings with some success. For use in probability elicitation, the same ordering is asked of the decision-maker, but this time in reference to how probable events are as outcomes of a chosen alternative of action.[4] Nevertheless, decision-makers might, in many cases, have more knowledge of the decision situation, even if the information still is not precise. For instance, cardinal probability relation information may implicitly exist, entailing that the surrogates may not really reflect what the decision-maker actually means by a particular ranking.

To improve the conformance of an ordinal ranking to the true subjective beliefs of a decision-maker, information such as the relative differences between adjacent items need to be accounted for. Given a ranking of some number of consequences, relative differences in probability comprise such information. Furthermore, methods that allow for such details can also handle probabilities considered equal, something which purely ordinal methods cannot.

The following notations together with the suggested interpretations are used to exemplify the proposed methods. The symbols denote the relative strength of the difference in probability between consequences.[5]

---

[4]For each choice of action, we assume a decision node to be the root of a standard two-level decision tree with each branch representing an alternative, which, in turn, is followed by a set of exhaustive and mutually exclusive events. It is the probabilities of those events that the decision-maker is asked to rank. Such a setup can easily be generalized to decision trees with multiple levels.

[5]Although using verbal interpretations for illustrative purposes, we do not intend to discuss issues related to difference values and their respective meanings in relation to probabilities. In an actual implementation of the method, we consider cardinal input obtained from graphical sliders in a software tool.

$>_0$ "equally probable"

$>_1$ "slightly more probable"

$>_2$ "more probable"

$>_3$ "much more probable"

To derive probabilities for $N$ consequences using cardinal methods, start by ranking the consequences in order of probability, giving the most probable consequence rank one, and the least probable consequence rank $N$. Let $s_j$ be the number of difference steps between probabilities $p_j$ and $p_{j+1}$, denoted by $p_j >_{s_j} p_{j+1}$. Setting $s_j = 1$ is equivalent to $p_j > p_{j+1}$ as in the ordinal case. The sum, $Q$, of all difference steps for $N$ probabilities will then be $\sum_{i=1}^{N-1} s_i$, and for any probability $p_i$, $i \in \{2, \dots, N\}$, the position on this scale of difference steps from 1 to $Q$ is defined by $t(i) = s_{i-1} + t(i-1)$, and $t(1) = 1$.[6]

Modifying the above ordinal methods for probability elicitation can now be done without difficulty. The key is to normalize the values of the cardinal rank positions such that the higher positions result in the lower probabilities. In the cardinal version of RS (CRS), the probabilities should mirror the differences in cardinal rank positions. Hence, the CRS probabilities are given by

$$p_i^{CRS} = \frac{Q + 1 - t(i)}{\sum_{j=1}^{N} (Q + 1 - t(j))}$$

(5)

where $t(i) \in \{1, \dots, Q\}$ are the cardinal rank positions derived from the cardinal differences provided by the decision-maker such that $t(i) \le t(j)$ if and only if $i < j$. The cardinal variation of rank reciprocal (CRR) is defined in a similar fashion, and the CRR probabilities are obtained by

$$p_i^{CRR} = \frac{1/t(i)}{\sum_{j=1}^{N} 1/t(j)}.$$

(6)

with the usual property that a higher probability is assigned to lower ranking numbers. ROC is generalized in the same way, and the corresponding strength rank order centroid (CRC) probabilities are obtained as

$$p_i^{CRC} = \frac{\sum_{j=t(i)}^{Q} 1/j}{\sum_{k=1}^{Q} \left( \sum_{j=t(j)}^{Q} 1/j \right)}.$$

(7)

Finally, generalizing SR is done in the same way and using the above equation, the corresponding cardinal SR (CSR) probabilities are obtained as

---

[6]Here, $p_i = P(c_i)$, that is, the probability that consequence $i$ obtains.

| Cardinal method | $p_1$ | $p_2$ | $p_3$ | $p_4$ | $p_5$ | $p_6$ |
|---|---|---|---|---|---|---|
| Rank sum (CRS) | 0.29 | 0.23 | 0.19 | 0.13 | 0.13 | 0.03 |
| Rank reciprocal (CRR) | 0.49 | 0.16 | 0.12 | 0.08 | 0.08 | 0.05 |
| Rank order centroid (CRC) | 0.45 | 0.21 | 0.16 | 0.09 | 0.09 | 0.02 |
| Sum reciprocal (CSR) | 0.37 | 0.20 | 0.17 | 0.11 | 0.11 | 0.04 |

**Table 3.** Probabilities corresponding to the ranking, $p_1 >_2 p_2 >_1 p_3 >_2 p_4 >_0 p_5 >_3 p_6$, rounded to two decimal places.

$$p_i^{\text{CSR}} = \frac{\frac{1}{t(i)} + \frac{Q+1-t(i)}{Q}}{\sum_{j=1}^{N} \left( \frac{1}{t(j)} + \frac{Q+1-t(j)}{Q} \right)},$$

(8)

which is a generalization similar to the others. Thus, using the idea of cardinal steps, ordinal methods are easily transformed to their respective cardinal counterparts. See **Table 3** for an example of probabilities correlated with a certain cardinal ranking.

### 3.1. Application to a decision problem

To show the merits of cardinal methods for elicitation of both probabilities and values, we present a decision on the choice of programming language, adapted from [35]. Due to resource constraints in a current Prolog implementation, the staff considered two options:

*A*: Rewriting the whole system in C.

*B*: Trying to find an implementation of Prolog that could handle the system.

After thoroughly discussing the pros and cons of the two options, they arrived at the following possible scenarios. For option *A*, the staff concluded that either ($c_1$) a prototype in C would be ready on time or ($c_2$) a prototype in C would be slightly delayed due to external circumstances. For option *B*, ($c_3$) a prototype in Prolog would be ready on time, ($c_4$) a prototype in Prolog would be slightly delayed due to external circumstances, or ($c_5$) only fractions of a prototype in Prolog would be ready on time.

Having considered the probabilities and the values of the consequences, the staff arrived at the following assumptions, where $p_i$ represents the probability of $c_i$, and $v_i$ represents the value of $c_i$. For option *A*, $p_1$ was thought to be at least 0.67, and consequently, $p_2$ would be at most 0.33. For option *B*, $p_5$ would lie somewhere between 0.40 and 0.90, hence the sum of $p_3$ and $p_4$ could range from 0.10 to 0.60.

With regard to value, $c_3$ was the most desirable with $c_4$ being slightly less so. Both $c_3$ and $c_4$ would be clearly better than $c_1$, which in turn was slightly better than $c_2$. By far, $c_5$ would be the worst consequence.

Applying a cardinal ranking to the above probabilities and values, the staff agreed on the following:

$$p_1 >_3 p_2$$

$$p_5 >_3 p_3 >_0 p_4$$

$$v_3 >_1 v_4 >_2 v_1 >_1 v_2 >_3 v_5$$

Using a purely ordinal ranking, one would end up with the following, less precise ordering:

$$p_1 > p_2$$

$$p_5 > p_3 > p_4 \text{ or } p_5 > p_4 > p_3$$

$$v_3 > v_4 > v_1 > v_2 > v_5$$

Computing the values of the rankings about using SR and CSR yielded the results presented in **Tables 4** and **5**. Let $E_o^m$ be the expected value of option $o$ given elicitation method $m$, we then have $E_A^{SR} = 0.16$ and $E_B^{SR} = 0.19$, resulting in option $B$ being the preferred alternative, while $E_A^{CSR} = 0.16$ and $E_B^{CSR} = 0.14$ resulting in the contrary.

Possible alternatives in the case of ordinal rankings, such as strict uses of either of the rankings $p_5 > p_3 > p_4$ or $p_5 > p_4 > p_3$, do not affect the preference order of the options. However, a change from $p_5 >_3 p_3 >_0 p_4$ to $p_5 >_2 p_3 >_1 p_4$ when applying the cardinal method would alter the outcome, indicating the decision problem's sensitivity to uncertainty, something which is not reflected when using a purely ordinal ranking of the consequences.

When treating the decision problem in the professional software program DecideIT [36], the probability values from the original assumptions can be used, resulting in the constraints:

1.  $p_1 \geq 2/3$ and $p_1 + p_2 = 1$

2.  $0.4 \leq p_5 \leq 0.9$ and $p_3 + p_4 + p_5 = 1$

which in turn can be entered directly into the program. The values are specified as in the cardinal case. Eventually, we end up with the decision tree in **Figure 1**.

| Method | $p_1$ | $p_2$ | $p_3$ | $p_4$ | $p_5$ |
|---|---|---|---|---|---|
| SR | 0.67 | 0.33 | 0.24 | 0.24 | 0.52 |
| CSR | 0.80 | 0.20 | 0.17 | 0.17 | 0.66 |

**Table 4.** The probabilities of the consequences of the options of the decision on a computer programming language.

| Method | $v_1$ | $v_2$ | $v_3$ | $v_4$ | $v_5$ |
|---|---|---|---|---|---|
| SR | 0.18 | 0.12 | 0.38 | 0.25 | 0.07 |
| CSR | 0.17 | 0.13 | 0.38 | 0.26 | 0.05 |

**Table 5.** The values of the consequences of the options of the decision on computer programming language.
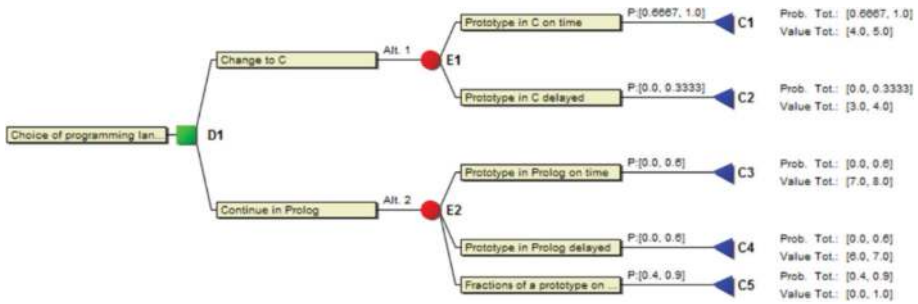
**Figure 1.** Modeling the decision problem of choosing a programming language in DecideIT.
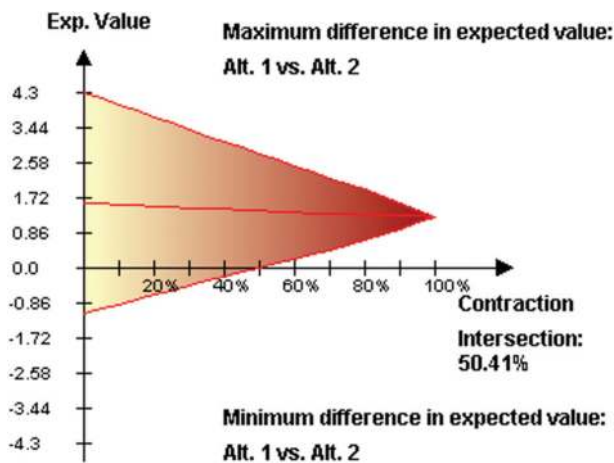


**Figure 2.** The expected value graph generated by DecideIT.

Looking at the expected value graph in **Figure 2**, we note that although about 50% contraction is needed, it is arguable so that either the assumptions need to be revised or option $A$ should be the preferred one. Consequently, a cardinal ranking of the probabilities and the values seem to resonate better with a more detailed examination in which imprecise numbers are taken into account than a purely ordinal one.

## 4. Assessing models for cardinal relations

Given that we have a set of cardinal methods as in the previous section, how can they be validated? For ordinal relations in MCDA, simulation studies similar to [27, 37–39] and others have become a kind of de facto standard. The simulations herein are based on the fundamental idea that a set of genuine, or "true," probabilities, in fact, exists in the mind of

the decision-maker. No elicitation method is capable of completely mirroring these probability values, but in the simulations, the potency of the probability ranking approaches is judged by comparing the "true" values to those elicited by the methods mentioned earlier.

The modeling caters to the two extremes of decision-makers' mindsets outlined earlier in the way the decision problem vectors are randomly generated. Following an $N$ DoF model, a vector is generated, where the components are kept within [0%, 100%], and subsequently normalized, that is, a process with $N$ degrees of freedom. Details on this kind of simulation can be found, for example, in [40]. For an $(N-1)$ DoF model, the components are generated such that they sum to 100% already from the outset; that is, using a process of $(N-1)$ degrees of freedom. This simulation is based on a homogeneous $N$-variate Dirichlet distribution generator. Details on this kind of simulation can be found, for example, in [41].

The "true" probabilities in the minds of decision-makers might, of course, follow a distribution different from the ones used in this study, and there might eventually be models available to elicit values following those. Nonetheless, the crucial observation is that the validity and reliability of the results of the simulations are highly dependent on how the minds of decision-makers are modeled. Although the difference in the number of degrees of freedom is only one of several parameters related to cognitive behavior, it still offers a meaningful way of distinguishing between cognitive models.

### 4.1. Biases of simulation studies

The results of the simulations are highly dependent on the type of generator used. A generator corresponding to an $N$ DoF model is referred to as an N-generator. In the same manner, a generator corresponding to an $(N-1)$ DoF model is referred to as an $(N-1)$-generator. In applying the $N$-generator, probabilities elicited by the RS method outperforms those elicited by other methods. Not because the RS method is superior per se, but because it produces numbers at regular intervals. Likewise, ROC outperforms the other methods when an $(N-1)$-generator is used for the reason that the values elicited by the ROC method are similarly skewed toward the lower end of the interval.

In actual fact, it is impossible to determine whether decision-makers, in general (or even some), elicit values in particular accordance with $(N-1)$ or $N$ DoF representations of their knowledge. As a group, or as individuals, it is possible that they completely adhere to either one or that they follow an arbitrary mix of the two. Due to this uncertainty pertaining the cognitive processes of decision-makers, a rank ordering mechanism, to be robust, must elicit values that conform to both types of representations reasonably well. Therefore, to find the method that yields the most robust and efficient assignments, the evaluations in this study employ both types of generators and combinations of them.

### 4.2. Comparing the methods

To evaluate the validity of the RS, RR, and ROC methods for multi-criteria weight elicitation, Barron and Barrett [27] simulated a large set of "true" weights, using an $(N-1)$-generator.

Based on those, they then produced a corresponding set of surrogate weights for each of the elicitation methods. As shown earlier, the generation procedure does have significant effects with regard to such a comparison. The set of "true" weights is dependent on how we model the minds of decision-makers. Barron and Barrett [27] presented a computer simulation consisting of four main steps, which for probability elicitation is modified as follows:[7]

### 4.2.1. Generation procedure

1.  For a decision problem with $M$ alternatives, where each alternative can result in one of $N$ consequences, generate $M$ probability vectors, $\mathbf{p}_1, \ldots, \mathbf{p}_M$, in $N$ dimensions. These vectors contain the so-called true probabilities. Then, for each elicitation method $\mu'$, produce vectors $\mathbf{p}_1^{\mu'}, \ldots, \mathbf{p}_M^{\mu'}$, according to the order of the "true" probabilities.

2.  After that, generate an $M \times N$ matrix of random numbers $v_{ij}$ corresponding to consequence $j$, of the $i$th alternative. These are the values of each consequence.

3.  Let $p_{ij}^{\mu}$ be the probability obtained by method $\mu$ for consequence $j$ of alternative $i$ (where $\mu$ is either $\mu'$ or "true"). For each method $\mu$ and each alternative $i$, calculate the expected value $E_i^{\mu} = \sum_{j=1}^{N} p_{ij}^{\mu} v_{ij}$. Note the rank order of $E_i^{\mu}$ for each $\mu$, that is, the preference order of the alternatives for each method.

4.  Lastly, determine if method $\mu$ resulted in the same most preferred alternative as "true." If that is the case, then record a hit.

The abovementioned procedure (a simulation round) is repeated a large number of times, with the ratio of the number of hits to the total number of simulation rounds used as a measure of efficacy. In some MCDA studies, two additional measures of efficacy have been reported, namely the average value loss and the average proportion of the achieved maximum value range. These measures do not, however, add anything in particular in terms of value due to their strong correlation with hit ratio.

Using an $(N–1)$-generator MCDM simulation model over the simplex $S_x$, ROC outperforms the other two methods. But a study by Roberts and Goodwin [40] came up with a different result where RS performed better than ROC with RR in third place, by employing a different distribution generating function where a fixed number, say 100, is given to the most important criterion and the others are uniformly generated as U[0,100]. This $N$-generator is, of course, different from $(N–1)$-generators based on a Dirichlet distribution, and thus, their simulation study instead yields the result that RS outperforms ROC with RR in third place. As we see later, using surrogate weights for probability elicitation yields similar results.

---

[7]For simplicity in generation procedures but without loss of generality, assume that all alternatives have the same number of consequences, $N$.

### 4.3. Simulations of the cardinal surrogate numbers

This chapter focuses on the performance of cardinal surrogate weights in probability elicitation. The simulations included 20 different scenarios made up of 3, 6, 9, 12, and 15 alternatives, and for each of those cases, the number of consequences was set to 3, 6, 9, and 12, respectively. Each combination was simulated 10 times, and this was then repeated 10,000 times, totaling 2,000,000 decision problems. The probability vectors for the $N$ DoF model was generated using a standard round-robin random number generator with subsequently normalized numbers, and the probability vectors for the $N - 1$ DoF model was generated using an $N$-variate Dirichlet distribution. The value vectors were generated on a uniform interval but left unscaled, analogous to MCDA studies such as [27]. Compared to alternative value distributions, there was no significant difference in the results.

A subset of the results, using a 50% combination of $N$ DoF and $N - 1$ DoF models, is presented in the tables mentioned later. As described earlier, the numbers denote the ratio of the number of times the most preferred alternative according to method $\mu'$ coincides with the most preferred alternative obtained from the "true" probabilities to the total number of simulation runs.[8]

The marginal utility of various levels of cardinal expressibility was evaluated by varying the levels of maximum cardinal differences in the simulations. The numbers, say $j$, in the "Symbols" column denotes the maximum $>$-index, $\max(i)$, of the set of $>_i$ symbols, such that $\max(i) = j$, for $j > 0$. The results are obtained from the ordinal counterparts when $j = 0$. Hence, $j = 1$ implies $\{>_0, >_1\}$, $j = 2$ means $\{>_0, >_1, >_2\}$, and so on. The actual results of the methods, the hit ratios, are given in percentages.[9]

Albeit there is a clear advantage of introducing the possibility of equivalence ($>_0$) between probabilities, the results, in general, indicate diminishing returns with an increase of the maximum $>$-index, in particular, when the number of possible consequences for each alternative is high. For example, at nine or more consequences, cardinal probability elicitation with only $>_0$ and $>_1$ is to prefer for CRC. Also, the cognitive burden on decision-makers tends to increase with the granularity of the scale. Hence, a maximum $>$-index of 2 seems to provide enough expressional power for probability elicitation (**Tables 6–12**).

| Combined DoF | Symbols | ROC/CRC | RS/CRS | RR/CRR | SR/CSR |
|---|---|---|---|---|---|
| Three consequences and three alternatives | 0 | 86.9 | 86.8 | 86.7 | 87.0 |
| | 1 | 88.8 | 87.7 | 87.9 | 88.0 |
| | 2 | 89.3 | 90.1 | 89.4 | 90.3 |
| | 3 | 89.5 | 91.2 | 90.2 | 91.3 |

**Table 6.** Comparing the methods using three consequences to each alternative and three alternatives.

---

[8] Two alternative sets of measurements, not shown in this chapter due to the strong correlation with the hit ratio, exist. One is the number of times to three most preferred alternatives obtained using $\mu'$ as elicitation method agrees with the three most preferred alternatives according to the "true" probabilities (i.e., the "podium"). A second is the number of times the overall rank of the alternative-using method $\mu'$ agrees with the overall rank based on the "true" probabilities.
[9] The results of the sets of 10 runs yielded a standard deviation of around 0.2–0.3%.

| Combined DoF | Symbols | ROC/CRC | RS/CRS | RR/CRR | SR/CSR |
|---|---|---|---|---|---|
| Three consequences and 15 alternatives | 0 | 70.9 | 69.2 | 69.1 | 69.5 |
| | 1 | 73.9 | 71.2 | 70.7 | 71.0 |
| | 2 | 74.6 | 76.1 | 74.4 | 76.2 |
| | 3 | 74.6 | 77.7 | 76.5 | 78.6 |

**Table 7.** Three consequences to each alternative and 15 alternatives.

| Combined DoF | Symbols | ROC/CRC | RS/CRS | RR/CRR | SR/CSR |
|---|---|---|---|---|---|
| Six consequences and six alternatives | 0 | 79.4 | 79.8 | 78.0 | 80.8 |
| | 1 | 83.5 | 80.8 | 81.0 | 82.4 |
| | 2 | 84.0 | 83.9 | 80.5 | 85.4 |
| | 3 | 83.8 | 85.3 | 79.3 | 86.4 |

**Table 8.** Six consequences to each alternative and six alternatives.

| Combined DoF | Symbols | ROC/CRC | RS/CRS | RR/CRR | SR/CSR |
|---|---|---|---|---|---|
| Six consequences and 12 alternatives | 0 | 75.0 | 73.8 | 72.3 | 75.1 |
| | 1 | 78.6 | 75.6 | 75.3 | 77.2 |
| | 2 | 78.3 | 78.3 | 74.1 | 78.4 |
| | 3 | 77.6 | 79.6 | 72.9 | 80.9 |

**Table 9.** Six consequences to each alternative and 12 alternatives.

| Combined DoF | Symbols | ROC/CRC | RS/CRS | RR/CRR | SR/CSR |
|---|---|---|---|---|---|
| Nine consequences and nine alternatives | 0 | 76.7 | 76.5 | 72.2 | 78.3 |
| | 1 | 81.0 | 77.2 | 76.8 | 79.8 |
| | 2 | 80.1 | 78.7 | 73.3 | 81.3 |
| | 3 | 78.5 | 79.7 | 70.2 | 81.6 |

**Table 10.** Nine consequences to each alternative and nine alternatives.

| Combined DoF | Symbols | ROC/CRC | RS/CRS | RR/CRR | SR/CSR |
|---|---|---|---|---|---|
| Twelve consequences and six alternatives | 0 | 80.1 | 80.5 | 73.2 | 82.2 |
| | 1 | 83.6 | 79.9 | 78.3 | 82.5 |
| | 2 | 81.9 | 81.1 | 73.3 | 83.5 |
| | 3 | 80.2 | 81.7 | 70.0 | 83.3 |

**Table 11.** Twelve consequences to each alternative and six alternatives.

| Combined DoF | Symbols | ROC/CRC | RS/CRS | RR/CRR | SR/CSR |
|---|---|---|---|---|---|
| Twelve consequences and 12 alternatives | 0 | 74.8 | 74.4 | 67.9 | 76.3 |
| | 1 | 78.5 | 73.8 | 72.3 | 76.8 |
| | 2 | 77.0 | 75.7 | 67.3 | 78.2 |
| | 3 | 74.8 | 77.2 | 62.0 | 78.7 |

**Table 12.** Twelve consequences to each alternative and 12 alternatives.

The results show that cardinal methods markedly outperform the ordinal methods. Among the cardinal methods, it is CRC and CSR that provide the best outcomes, but the difference between these lies within this investigation's margin of error.

## 5. Summary and conclusion

Elicitation methods available today are often either too cognitively demanding and require too much time and effort or unable to use the available information. The aim of this study was to offer decision-makers a set of methods for probability elicitation with a reasonable balance between simplicity and usability on the one hand and correctness and accuracy on the other hand. In particular, we strived to reduce the issues of applicability by loosening the requirements of precise information, while allowing for more details than what ordinal methods can handle. Also, the methods should be relatively robust and applicable to a wide range of decision problems.

By augmenting a set of ordinal elicitation methods, originally developed for weight elicitation within MCDA, with a notion of a difference between adjacent probabilities, we arrived at a collection of cardinal probability elicitation methods. The robustness of the methods was evaluated using a large number of simulated decision problems, where we also accommodated for two different cognitive models, based on the degree of freedom used during the ranking, including a mix of them. The results of the simulations point toward a significant improvement of cardinal methods over purely ordinal ones. In particular, due to the introduction of equality between probabilities. Among the cardinal methods, in particular, CSR and CRC seem to provide the most robust results. CSR generalizes SR from [42] by also taking the cardinal differences of the probabilities into account in a more straightforward way than, for example, [34].

More fine-grained expressions seem to produce diminishing returns when the number of consequences of each alternative becomes high. For alternatives with 12 consequences or less, a cardinal method with the set $\{>_0, >_1, >_2\}$ seems to supply the decision-maker with adequate options for producing quite reliable probability elicitations. For alternatives with more than 12 consequences, the reduced set $\{>_0, >_1\}$ seems to provide a sufficient granularity.

In conclusion, cardinal methods rather than ordinal ones should be preferred for eliciting probabilities when applicable. More specifically, CSR and CRC have been shown to produce surrogates, which outperform those of their competitors. They keep decision-makers from having to provide too much detail, something which has turned out to be difficult for decision-makers in general, while at the same time reducing the risk of neglecting available information.

## Acknowledgements

## Conflict of interest

All authors report no conflicts of interest relevant to this chapter.

## Author details

Mats Danielson[1,2], Love Ekenberg[1,2] and Andreas Paulsson[1*]

*Address all correspondence to: apaulsson@dsv.su.se

1  Department of Computer and Systems Sciences, Stockholm University, Kista, Sweden

2  International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria

## References

[1] Andersen S, Fountain J, Harrison GW, Rutström EE. Estimating subjective probabilities. Journal of Risk and Uncertainty. 2014 Jun 1;**48**(3):207-229

[2] Rey-Biel P. Equilibrium play and best response to (stated) beliefs in normal form games. Games and Economic Behavior. 2009 Mar;**65**(2):572-585

[3] Menapace L, Colson G, Raffaelli R. Risk aversion, subjective beliefs, and farmer risk management strategies. American Journal of Agricultural Economics. 2013 Jan 1;**95**(2): 384-389

[4]   Veen D, Stoel D, Zondervan-wijnenburg M, van de Schoot R. Proposal for a five-step method to elicit expert judgment. Frontiers in Psychology. 2017;**8**(DEC):2110. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2017.02110. https://www.frontiersin.org/article/10.3389/fpsyg.2017.02110

[5]   Delavande A, Giné X, McKenzie D. Measuring subjective expectations in developing countries: A critical review and new evidence. Journal of Development Economics. 2011 Mar;**94**(2):151-163

[6]   Schotter A, Trevino I. Belief elicitation in the laboratory. Annual Review of Economics. 2014 Aug;**6**(1, 1):103-128

[7]   Harrison GW, Martínez-Correa J, Swarthout JT, Ulm ER. Scoring rules for subjective probability distributions. Journal of Economic Behavior and Organization. 2017 Feb 1; **134**:430-448

[8]   Schlag KH, Tremewan J, van der Weele JJ. A penny for your thoughts: A survey of methods for eliciting beliefs. Experimental Economics. 2015 Sep;**18**(3):457-490

[9]   Smithson M. Introduction to imprecise probabilities. In: Augustin T, Coolen FPA, Coolen FP, de Cooman G, Troffaes MCM, editors. West Sussex. United Kingdom: John Wiley & Sons Inc.; 2014. pp. 318-328

[10]  Hollard G, Massoni S, Vergnaud J-C. In search of good probability assessors: An experimental comparison of elicitation rules for confidence judgments. Theory and Decision. 2016 Mar 1;**80**(3):363-387

[11]  Fox J. Probability, logic and the cognitive foundations of rational belief. Journal of Applied Logic. 2003 Jun;**1**(3–4):197-224

[12]  Biedermann A, Bozza S, Taroni F, Aitken C. The consequences of understanding expert probability reporting as a decision. Science & Justice. 2017 Jan;**57**(1, 1):80-85

[13]  Von Winterfeldt D, Edwards W. Decision Analysis and Behavioral Research. Cambridge [Cambridgeshire]; New York: Cambridge University Press; 1986. 604 p

[14]  Savage LJ. The Foundations of Statistics. 2nd rev. ed. New York: Dover Publications; 1972. 310 p

[15]  De Finetti B. Probability: Interpretations. In: Sills DL, Bos HC, Tinbergen J, editors. International Encyclopedia of the Social Sciences. New York, NY: Macmillan; 1968. pp. 496-504

[16]  Wright G, Ayton P. Subjective Probability. Chichester, New York: Wiley; 1994. 574 p

[17]  Clemen RT, Reilly T. Making Hard Decisions with Decision Tools. 3rd ed. South-Western: Mason, OH; 2014

[18]  Corner JL, Corner PD. Characteristics of decisions in decision analysis practice. Journal of the Operational Research Society. 1995 Mar;**46**(3):304-314

[19]  Hogarth RM. Cognitive processes and the assessment of subjective probability distributions. Journal of the American Statistical Association. 1975 Jun;**70**(350):271-289

[20]  Morgan MG, Henrion M, Small MJ. Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis. Cambridge, New York: Cambridge University Press; 1990. 332 p

[21]  Wallsten TS, Budescu DV. State of the art—Encoding subjective probabilities: A psychological and psychometric review. Management Science. 1983 Feb;**29**(2):151-173

[22]  Johnson EM, Huber GP. The technology utility assessment. IEEE Transactions on Systems, Man, and Cybernetics. 1977;**7**(5):311-325

[23]  Tversky A, Kahneman D. The framing of decisions and the psychology of choice. Science. 1981;**211**(4481):453-458

[24]  Lenert LA, Treadwell JR. Effects on preferences of violations of procedural invariance. Medical Decision Making. 1999 Oct;**19**(4):473-481

[25]  Lichtenstein S, Slovic P. The Construction of Preference. Cambridge, New York: Cambridge University Press; 2006. 790 p

[26]  Barron FH, Barrett BE. The efficacy of SMARTER—Simple multi-attribute rating technique extended to ranking. Acta Psychologica. 1996 Sep;**93**(1–3):23-36

[27]  Barron FH, Barrett BE. Decision quality using ranked attribute weights. Management Science. 1996 Nov;**42**(11):1515-1523

[28]  Katsikopoulos KV, Fasolo B. New tools for decision analysts. IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans. 2006 Sep;**36**(5):960-967

[29]  Stillwell WG, Seaver DA, Edwards W. A comparison of weight approximation techniques in multiattribute utility decision making. Organizational Behavior and Human Performance. 1981 Aug;**28**(1):62-77

[30]  Hutton Barron F. Selecting a best multiattribute alternative with partial information about attribute weights. Acta Psychologica. 1992 Aug;**80**(1–3, 103):91

[31]  Metfessel M. A proposal for quantitative reporting of comparative judgments. The Journal of Psychology. 1947 Oct;**24**(2):229-235

[32]  Danielson M, Ekenberg L, Larsson A. Distribution of expected utility in decision trees. International Journal of Approximate Reasoning. 2007 Oct;**46**(2):387-407

[33]  Danielson M, Ekenberg L, Larsson A, Riabacke M. Weighting under ambiguous preferences and imprecise differences in a cardinal rank ordering process. International Journal of Computational Intelligence Systems. 2014 Jan;**7**(Suppl. 1):105-112

[34]  Danielson M, Ekenberg L, He Y. Augmenting ordinal methods of attribute weight approximation. Decision Analysis. 2014 Mar;**11**(1):21-26

[35]  Malmnäs P-E, editor. Hypermjuk beslutsteori och ekonomisk optimering av det industriella brandskyddet. Stockholm: Thales; 2002

[36] Preference AB. DecideIT [Internet]. Stockholm, Sweden: Preference AB; 2017. Available from: http://www.preference.nu

[37] Stewart TJ. Use of piecewise linear value functions in interactive multicriteria decision support: A Monte Carlo study. Management Science. 1993 Nov;**39**(11):1369-1381

[38] Arbel A, Vargas LG. Preference simulation and preference programming: Robustness issues in priority derivation. European Journal of Operational Research. 1993 Sep;**69**(2): 200-209

[39] Ahn BS, Park KS. Comparing methods for multiattribute decision making with ordinal weights. Computers & Operations Research. 2008 May;**35**(5):1660-1670

[40] Roberts R, Goodwin P. Weight approximations in multi-attribute decision models. Journal of Multi-Criteria Decision Analysis. 2002 Nov;**11**(6):291-303

[41] Rao JS, Sobel M. Incomplete Dirichlet integrals with applications to ordered uniform spacings. Journal of Multivariate Analysis. 1980 Dec;**10**(4):603-610

[42] Danielson M, Ekenberg L. Rank ordering methods for multi-criteria decisions. In: Zaraté P, Kersten GE, Hernández JE, editors. Group Decision and Negotiation A Process-Oriented View. Cham: Springer International Publishing; 2014. pp. 128-135