Expert Systems With Applications 296 (2026) 128938

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Contribution of the ratio scale of expert judgments in the analytic hierarchy process

Carlos Sáenz-Royo [a,*], Francisco Chiclana [b]

[a] University of Zaragoza, Department of Business Organization and Management, C/Pedro Cerbuna, 12, 50009 Zaragoza, Spain
[b] Institute of Artificial Intelligence, School of Computer Science and Informatics, De Montfort University, Leicester, UK

ARTICLE INFO

Keywords:
AHP
IBR
Decision support systems
Expertise
Intensity judgment

ABSTRACT

The pairwise comparison ratio scale that the Analytic Hierarchy Process (AHP) acquires a central role because it allows the incorporation of intangible information through expert judgments. Existing literature has been shown that preference results are sensitive to the scale used in ratio judgments. This paper applies the Intentional Bounded Rationality Methodology (IBRM) to study the performance of the most used AHP ratio scales in the literature. By adapting the AHP ratio scales to the IBRM, several scenarios are proposed based on the distribution of the latent performance of the alternatives for two different problems. In problem 1, it is assumed that the AHP comparison judgements are numerical and known to the expert, while in problem 2, the AHP comparison judgements are described linguistically without the expert being aware of their transformation into numerical values. Problem 1 is used to answer the following research question (1) which among a set of seven different ratio scales used in literature favor AHP expected performance?, while Problem 2 is used to answer the following research question (2) how much the expected performance in AHP deteriorates when the expert is only guided by verbal judgments? The expected performance of each of the considered ratio scales is obtained in each scenario for different levels of expertise of a decision-maker. For the first problem, the balanced scale and the power scale show the best and worst expected performances, respectively. Problem 2 compares these two scales and the results show their performance differences are below 8%, which is interpreted as a stability property of AHP with respect to scale changes. Finally, it is also shown that no matter which ratio scale is used, the requirement of the consistency property in AHP contributes positively to the expected performance.

1. Introduction

The Analytic Hierarchy Process (AHP) (Saaty, 1977, 1980) is a multi-criteria decision-making support system for solving complex problems characterized by the existence of multiple actors, scenarios, and criteria (tangible and intangible). When the complexity and uncertainty of the problem prevent an objective decision through quantifiable criteria, AHP provides the possibility of incorporating intangible aspects through expert judgment with great versatility. This AHP quality requires having a ratio scale to map expert's judgments on the level of preference of one criterion or alternative over another, from which a hierarchical structure and the decision solution to the complex problem are developed and obtained, respectively.

Since experts' judgments may be inaccurate, one of the aims of decision support systems is to minimize the impact of experts' erroneous judgments (Liu et al., 2021). In AHP, the analytical resolution integrates all information, tangible and intangible, with judgments required to be quantified numerically. This means that the solution obtained by AHP could be affected by the measurement process' numerical ratio scale used to map the experts' judgment (Hershey et al., 1982; Hershey & Schoemaker, 1985). Indeed, Harker & Vargas, (1987) show that ratio scale of judgments with arbitrary assignation of numerical values affect the results of processing the experts' stated preferences in AHP and, therefore, it cannot be said that any preference revelation method is completely independent of the measurement scale.

When incorporating intangibles, i.e. elements that are difficult to quantify due to their complexity, into decision-making processes, a flexible and realistic rather than merely substantive rationality is required (Saaty 1980; Brans 1996; Moreno-Jiménez et al. 1999). The intentional bounded rationality (IBR) can help locate the scales most sensitive to the different judgments of experts and their characteristics, which may contribute to providing a new systematic approach to this

* Corresponding author at: C/Pedro Cerbuna, 12, 50009 Zaragoza, Spain.
E-mail addresses: csaenz@unizar.es (C. Sáenz-Royo), chiclana@dmu.ac.uk (F. Chiclana).

https://doi.org/10.1016/j.eswa.2025.128938
Received 7 October 2023; Received in revised form 4 July 2025; Accepted 5 July 2025
Available online 11 July 2025

unresolved discussion (Harker & Vargas, 1987; Ishizaka et al., 2011; Ishizaka & Labib, 2011; Liu et al., 2021; Meesariganda & Ishizaka, 2017; Siraj et al., 2015; Zavadskas & Turskis, 2011).

The methodological requirements for expert judgments are determining aspects of their performance. Current trends in multicriteria decision-making are focused on evaluating expert judgments, incorporating activities such as learning, intensity, consistency, negotiation, and the search of consensus among the actors involved in the information process (Aguarón & Moreno-Jiménez, 2000; Han et al., 2016; Lin & Kou, 2021). AHP requires experts to express the intensity of their preferences, thus obtaining a higher level of information than a simpler ordinal comparison. The Intentional Bounded Rationality Methodology (IBRM) (Sáenz-Royo et al., 2023a) provides allows to carry out a sensitivity analysis of the final classification of the priorities of the alternatives and their probabilities of error through the expected performance based on the capabilities of the experts and the distribution of the alternative performances. Furthermore, the evaluation of the consistency requirement to accept the judgments of the experts can be added to this analysis, which results in valuable information for the exploitation of AHP regarding negotiation and monitoring processes. Thus, in the described research context, this paper proposes to use the IBRM to determine among a number of different ratio scales of interest the one with best expected performance in the AHP method. In addition, whether the incorporation of the consistency property can provide value for each of the considered ratio scales is also investigated.

By achieving the above objectives, the proposed theoretical development will broaden the knowledge of decision support systems. In particular, the representation of the value judgments of the experts is improved, which can be proved a very useful tool in the treatment of complex situations characterized by a high level of uncertainty and dynamism. The contributions of this study are summarized as follows:

1. The problems of the ratio scale in AHP are discussed and the main ratio scales used in the literature are compiled, showing them as a transformation of Saaty's 1–9 ratio scale;
2. A mechanism to adapt the AHP ratio scales to the IBR is proposed so that the IBRM is applied to their study;
3. Some scenarios are proposed based on the distribution of the latent performance of the alternatives;
4. The expected performance of each ratio scale is obtained in each scenario for each level of decision-maker expertise and for two different problems; with the Balanced scale presenting the best results;
5. The AHP is shown to be very stable in both problems with respect to scale changes;
6. Finally, the contribution of the AHP consistency requirement to the expected performance is scarce, and in some cases it is negative.

The rest of the paper is structured as follows: Section 2 discusses the problems of the ratio scale in AHP and different ratio scales are compiled. Section 3 describes the concept of IBR and its implications in the considered research framework. Section 4 defines three scenarios with different distributions of alternative performances. Section 5 describes the necessary steps in the IBRM methodology for its application to the problems and scenarios previously discussed, while Section 6 reports on the results obtained. Section 7 provides a discussion of the results and draws conclusions from this investigation.

2. Judgment ratio scales

Science tends to use absolute measurement approach rather than relative measurement approach, although the absolute measurement approach can be seen as type of the relative measurement with respect to the particular auxiliary unit (meter, liter, minute) that is established as reference element of the measurement process. The advantage of the absolute measurement approach is that, for a set of n objects to compare, it saves time and resources by reducing the set of relative pairwise judgments, $n!/(2!(n-2)!)$, required by each auxiliary unit of measurement (Sáenz-Royo et al., 2024). Its drawback though is that it links the measurement values to a reference unit that may not be compared with a different unit, for example absolute measurements in meters may not be comparable with absolute measurements in kilograms, leaving the different dimensions of a problem unconnected.

Absolute measurement is far from providing a method to deal with intangible human judgments of complex problems. Indeed, an absolute judgment refers to the identification of the magnitude of some single stimulus compared to a reference stimulus held in memory while a relative judgment refers to the identification of the quantified relationship between two stimuli, both present to the same observer (Blumenthal, 1977). Thus, although the complexity of a problem can show the intangibles in a diffuse way, the relative judgment helps to quantify them.

Saaty's AHP allows the incorporation of human judgments as a synthesis tool and, for this very reason, it is a highly valued tool in complex decisions. This, however, implies that AHP requires a tools for representing the understanding, tracking, and evaluation of human judgments. The talent of the human brain has been underutilized in science because it has not been learned to measure it. Relative measurement through comparative judgment is intrinsic to thought and should not be carried as an appendage whose actual function is not well understood (Saaty, 1993). The connection established by Sáenz-Royo et al. (2022) between the thinking way and decision-making helps to quantify the way of treating human judgments, opening up new lines of research.

The ratio scale plays a central role in the comparison by pairs because the points of the scale are independent of the measurement units, which allows the integration of different measures in a single judgment. Although pairwise judgments can be provided using an additive scale, the use of a ratio scale is considered more appropriate for measuring relative intensities in AHP (Harker & Vargas, 1987; Krantz, 1972; Stevens, 1957). Relative comparison allows for any statistical test and the order of the scales multiplicatively indicates preference (Blalock, 1968). The limitations of the relative comparison are that zero is absolute and the preference relationships between values must be known (Barzilai, 2005).

The first problem (Problem 1) arises because the cardinal information in a pairwise judgment of intensity, $a_{ij} = V_i/V_j$ where V_i and V_j are the latent performance of each alternative, is perceived diffusely by the expert and it becomes difficult to require a high level of precision. To facilitate this aspect, AHP provides a discrete and bounded ratio scale on which the expert's judgment must be located. In the case of tangible criteria, it is not necessary to establish any scale since its value can be obtained from the information measured directly, such as weights (in kilograms) or distances (in meters). In the case of intangibles, to facilitate the judgment of the experts, a discrete scale that delimited the ratio judgments to a scale of 1 to 9 was proposed by Saaty (1977). This ratio scale was chosen in an apparently arbitrary way, and it is worth asking if setting a different ratio scale may favor AHP performance.

The second problem (Problem 2) arises from the use of a set of verbal judgments that are assigned to the Saaty ratio scale. Verbal judgments are widely used (Liu et al., 2020; Tavana et al., 2023; Vargas, 1990; Yu & Hong, 2022) despite raised criticisms and shortcomings (Belton & Gear, 1983; Dyer, 1990). Fixing different ratio scales to the same verbal judgment affects all methods of preference induction with cardinal intensity. These methods should maintain a minimum ordinal preference, which requires that preference measurements are not affected by the numerical scale used in the measurement process. However, since the quantification procedure of the ratio judgment is a relative value, its modification breaks the assumption of invariance of the transitivity of preferences with respect to scale transformations. It has been shown that the results of the elections in AHP are sensitive to changes in the ratio scale used when the judgments must present a cardinal intensity.

Hershey et al. (1982) and Hershey & Schoemaker (1985) have shown that a linear transformation of the interval scale can drastically alter the results that are obtained. Thus, although it is accepted that verbal judgments are an adequate way to measure the human response to stimuli (Stevens, 1957; Stevens & Galanter, 1957), the numerical values used in the ratio scale affect the comparison operations of criteria and alternatives, altering the final priority vector, which means that AHP is not independent of the ratio scale used in measuring the expert judgments.

In this study, we used the IBR to evaluate which is the ratio scale that presents the best performance in AHP. Experiments reported by Saaty (1980) and the experience of many AHP users tend to support the view that the 1–9 scale captures an individual's preferences fairly well. However, the scale can be modified to suit an individual's needs. For example, in situations where the alternatives are clearly preferred, yet contradictory for different criteria, setting number 9 as the upper bound on the ratio scale may limit the results. On the other hand, the ratio scale must be linked to the human way of thinking, so the upper bound may be reduced with less precision and increased with more precision. These and other arguments have led to different ratio scale proposals.

Harker & Vargas (1987) compared Saaty's *1–9 linear scale*, $x = (x_1 = 1, x_2 = 2, x_3 = 3, x_4 = 4, x_5 = 5, x_6 = 6, x_7 = 7, x_8 = 8, x_1 = 9)$, with other interesting scales, including the reducing and expanding *1–5 and 1–15 linear scales*, $c = (1, 2, 3, 4, 5)$ and c=(1–15), and the *two non-linear scales*, $c = x^2 = (1, 4, 9, 16, 25, 36, 49, 64, 81)$ and $c = x^{1/2} = (1, 1.414, 1.732, 2, 2.236, 2.449, 2.646, 2.828, 3)$. The authors looked for alternatives that better reflect the cognitive biases of the human way of thinking. Despite giving some arguments in favor of Saaty's linear scale, the problem of the choice of ratio scale was left unresolved.

Lootsma (1993) preferred the use of the *geometric scale* $c = 2^{x-1} = (1, 2, 4, 8, 16, 32, 64, 128, 256)$ to the linear one, justifying its choice in psychophysical factors through preference experiments in the heuristics of the human mind to make quick decisions regarding quantifiable information. The proposed geometric scale was a good fit for how the human brain summarizes quantitative information on a ratio scale.

Salo & Hämäläinen (1997) proposed the *balanced scale* $c = \frac{0.45+0.05x}{0.55-0.05x} = (1, 1.222, 1.500, 1.857, 2.333, 3, 4, 5.667, 9)$ stating that its values are more uniformly dispersed and present better properties (transitivity, integrity, independence (Keeney & Raiffa, 1976)) than the values of Saaty's linear scale.

Finan & Hurley (1999) proposed to recalibrate the verbal scale, transforming it into the following *geometric progression scale*, $c = 1.25^{(0.5x-0.5)} = (1, 1.118, 1.25, 1.398, 1.563, 1.747, 1.953, 2.184, 2.441)$, that they claimed to achieve better performance than linear scales.

Using standard consumer theory, which states that the decision-maker, in general, would prefer the compromise alternative to the extreme alternatives, Ishizaka et al. (2011) proposed the *logarithmic scale*, $c = \log_2(x + 2 - 1) = (1, 1.585, 2, 2.322, 2.585, 2.807, 3, 3.170, 3.322)$, that the authors justified with a comparative study with other ratio scales from which it was found that the normalization of the weights of the criteria has little impact on the final result and the "compromise" alternatives have little chance of being selected in AHP.

An interesting analysis of AHP and how elements of its methodology affect performance are available in Ishizaka et al. (2011). Using the fact that ratio scales have received little research, the authors proposed the list of transformations of Saaty's linear scale described above and summarized in see Table 1. However, they concluded that no single scale is the most appropriate for all situations. In relation to this, Franek & Kresta (2014) proposed the selection of the most appropriate scale for a problem rests in the decision-maker must dealing with such problem. Although they provided results that suggest that Saaty's linear scale is a good option, they showed that root-log or square-root scales are to be selected instead when consistency is to be maximized, while power or geometric scales are to be selected when sensitivity to the values of the selection priorities is the main criterion to implement in the resolution problem. In any case, the authors insist on the need to investigate the scales in different decision-making problems and suggested their combining with consistency criteria.

The redundancy of the AHP pairwise judgments may lead to inconsistent judgments to appear, which could be valuable to validate the information acquired from the experts after their mapping to the ratio scales. Inconsistency may happens due to psychological reasons (lack of concentration, cognitive biases, …) or errors in information management, and may leave out important aspects of the problem (Sugden, 1985). An incorrect ratio scale can also lead to inconsistencies inherent to the used scale (Siraj et al., 2015). Our work tries to go further in this analysis by applying the IBRM, combining the ratio scale with the quality of information provided by the consistency property.

3. Intentional bounded rationality (IBR) and ratio scales

When there is no clear measurement of the problem solution, experts are asked to provide judgments that condense a large amount of information. However, experts' judgements may be wrong and, therefore, decision support systems must establish mechanisms to avoid this situation.

One way to evaluate decision support systems is through an ex-post analysis, providing a problem with a known solution and asking experts to evaluate it. The literature has resolved this question by looking for signals (proxy variables) of the quality of the ex-post decision. The level of consensus among experts is perhaps the most used (Chiclana et al., 2013; Herrera-Viedma et al., 2014; Zhang et al., 2019) since it is based on the hypothesis that experts can be wrong but they have a greater probability of being right than of being wrong and, therefore, a higher level of consensus indicates greater certainty about the choice (Cai et al., 2017). Consensus presents positive aspects in the participation of decisions (Sáenz-Royo & Lozano-Rojo, 2023) but it does not guarantee the best solution (Moreno-Jimenez et al., 2013). Another way of evaluating the quality of the ex-post solution is the level of coherence or logical soundness of the choices made, i.e. with the transitivity of the judgments expressed (Ishizaka & Lusti, 2004; Saaty, 1977, 1980). If judgments are error-free, then they are consistent; however, the opposite does not stand, i.e. the fact that judgments are consistent does not mean that they are error-free (Sáenz-Royo et al., 2023a; Sugden, 1985; Temesi, 2011).

3.1. The Intentional bounded rationality (IBR)

Sáenz-Royo et al. (2022) propose a theoretical framework of IBR where experts can be wrong, but errors and successes follow certain

Table 1
Seven AHP ratio scales used in literature.

Linear (Saaty, 1977)	1	2	3	4	5	6	7	8	9
Root (Harker and Vargas, 1987)	1	1.414	1.732	2	2.236	2.449	2.646	2.828	3
Power (Harker and Vargas, 1987)	1	4	9	16	25	36	49	64	81
Geometric L (Lootsma, 1993)	1	2	4	8	16	32	64	128	256
Balanced (Salo & Hämäläinen, 1997)	1	1.222	1.5	1.857	2.333	3	4	5.667	9
Geometric FH (Finan and Hurley, 1999)	1	1.118	1.25	1.398	1.563	1.747	1.953	2.184	2.441
Logarithmic (Ishizaka, Balkenborg and Kaplan, 2011)	1	1.585	2	2.322	2.585	2.807	3	3.17	3.322

regularities (Simon, 1997). The IBR links the mechanisms that govern human cognition and the error probability, making it possible to evaluate decision support systems. This conceptual framework establishes two exogenous variables that determine the error probability: (i) the complexity of the decision understood as the difference in the latent returns of the alternatives; (ii) the decision-maker's expertise understood as the reliability of the person making the judgments. This representation of the error probability is different from the traditional ones that are limited to introducing random noise in the judgments of the decision-maker (Hogarth, 1975; O'Hagan et al., 2006; Ravinder, 1992; Vargas, 1982; Wallsten & Budescu, 1983).

The IBR proposes a logistic probability distribution, in which the probability that a decision-maker chooses alternative A_i with latent value V_i over all other alternatives, $p_{\beta i}$, is obtained as its below relative weight:

$$p_{\beta i} = p(A_i) = \frac{e^{\beta \frac{V_i}{\sum_{k=1}^{n} V_k}}}{\sum_{j=1}^{n} e^{\beta \frac{V_j}{\sum_{k=1}^{n} V_k}}} = \frac{1}{1 + \sum_{j \neq i}^{n} e^{\frac{\beta(V_j - V_i)}{\sum_{k=1}^{n} V_k}}}. \quad (1)$$

The parameter β is a measure of the decision-maker's ability to process information (decision-maker's expertise). The value of parameter β is limited to the problem studied and, therefore, the same decision-maker can show different capacities in different problems. The value $\beta = 0$ is used to represent that the decision-maker knows nothing about the problem and makes all the alternatives have the same probability of being chosen regardless of the relative latent performance of each of them. As the value of β increases so does the expertise level of the decision maker, which implies an increase in the capacity to discern the best alternative and translates into an increase probability of getting it right.

3.2. The ratio scale in IBR

In order to adapt the IBR to the ratio scale and make relative comparisons, it is necessary to transform the absolute latent performances into comparisons by pairwise judgments. When the decision-maker evaluates the relative performance of only two alternatives at a time, $\{A_i, A_j\}$, then the application of Eq. (1) provides two probability values: the value $p_{\beta i}$ representing the probability of choosing A_i over A_j, and the (additive) reciprocal value $p_{\beta j} = 1 - p_{\beta i}$ representing the probability of choosing A_j over A_i. In this case, instead of $p_{\beta i}$ a more appropriate notation would be $p_{\beta ij}$:

$$p_{\beta ij} = \frac{e^{\beta \frac{V_i}{V_j}}}{e^{\beta \frac{V_i}{V_j}} + e^{\beta \frac{V_j}{V_i}}} = \frac{1}{e^{\beta \left(\frac{V_j}{V_i} - \frac{V_i}{V_j} \right)} + 1} = \frac{1}{e^{\beta (a_{ji} - a_{ij})} + 1} \quad (2)$$

Notice that the IBR Eq. (2) can be interpreted as the probability ($p_{\beta ij}$) that the decision-maker, with expertise level β, judges that $a_{ij} = \frac{V_i}{V_j}$ is greater than $a_{ji} = \frac{V_j}{V_i}$. Recall that the ratio scale uses the numerical evaluation of how many times a_{ij} is superior to a_{ji}.

A judgment of intensity with respect to a discrete ratio scale provides a collection of distinct values $x = (x_1, \cdots, x_K)$ on which the expert must rank her/his judgment about the latent performance of two alternatives. When comparing $\{A_i, A_j\}$, the expert uses x_k to favor A_i over A_j when her/his judgment shows that V_i is x_k times higher than V_j. In this case, the (multiplicative) reciprocal value $1/x_k$ is assumed to favor A_j over A_i.

The mechanism to relate a discrete ratio scale to the IBR is by associating to the distinct values x_k in the ratio scale distinct intervals, with critical points $Z(x_k)$ representing the values of a_{ij} from which the expert chooses x_k on the ratio scale. Thus, the following rule connects the distinct values x_k in the ratio scale to the critical points $Z(x_k)$: if the expert considers $Z(x_k) - \frac{1}{Z(x_k)} < a_{ij} - a_{ji} < Z(x_{k+1}) - \frac{1}{Z(x_{k+1})}$, then her/his equivalent judgment in the ratio scale will be $a_{ij} = x_k$. The interval associated to value x_k in the ratio scale can be referred to, without causing confusion, as interval x_k.

3.3. Adaptation of ratio scales

The distinct values of the ratio scale are used to obtain the critical points their corresponding distinct intervals. Thus, the values in Table 1 no longer represent the central points of each interval but the values $Z(x_k)$ from which the expert judges the interval as appropriate (Sáenz-Royo, Chiclana, et al., 2022; Sáenz-Royo et al., 2023a). This adaptation is necessary to avoid the indifference and it simplifies the analysis considerably. For example, in Saaty's linear ratio scale, the second interval (weak importance of A_i over A_j) begins when the expert considers that $V_i/V_j > 1$ and, therefore, it would not be centered on 2 as Saaty postulated, but on 1.5. If the expert judges an intensity of preference slightly above $Z(x_1) = 1$, for example 1.1, then the expert will choose the interval x_1 in which the center point is $a_{ij}(x_1) = 1.5$, while if the expert judges the intensity to be slightly above $Z(x_2) = 2$, for example 2.1, then the expert will choose the interval x_2 whose center point is $a_{ij}(x_2) = 2.5$. Thus, $Z(x_1)$ is the value from the Saaty ratio scale that corresponds to the interval x_1, and $Z(x_2)$ for x_2, and so on.

The critical point of interval x_k, $Z(x_k)$, is obtained by comparing the value from which an interval $Z(x_k)$ is chosen with its inverse $1/Z(x_k)$, that is, allowing to compare the minimum required intensity with the mean of the distribution in Eq. (2) and obtaining the probability that the relative performance $a_{ij} - a_{ji}$ is at least $Z(x_k) - 1/Z(x_k)$:

$$p_{\beta,ij}(Z) = \frac{1}{e^{\beta\left(\left(Z - \frac{1}{Z} \right) - \left(\frac{V_i}{V_j} - \frac{V_j}{V_i} \right) \right)} + 1} \quad (3)$$

The probability that an expert judges the interval of the ratio scale to be x_k is bounded by its critical point $Z(x_k)$ and the critical point of the upper interval $Z(x_{k+1})$, and therefore the probability of the interval x_k is equal to the difference between the probability that the intensity is greater than $Z(x_k)$ and the probability that the intensity is greater than $Z(x_{k+1})$: $p_{\beta,ij}(x_k) = p_{\beta,ij}(Z(x_k)) - p_{\beta,ij}(Z(x_{k+1}))$. Using Eq. (3), the probability assigned to the interval x_k is:

$$p_{\beta,ij}(x_k) = \frac{1}{e^{\beta\left(\left(Z(x_k) - \frac{1}{Z(x_k)} \right) - \left(\frac{V_i}{V_j} - \frac{V_j}{V_i} \right) \right)} + 1} - \frac{1}{e^{\beta\left(\left(Z(x_{k+1}) - \frac{1}{Z(x_{k+1})} \right) - \left(\frac{V_i}{V_j} - \frac{V_j}{V_i} \right) \right)} + 1} \quad (4)$$

This equation allows the assignation of a probability value to each intensity interval of a ratio scale. The interval where the real difference in relative performance lies will be the most probable, those next to it, the following ones with the greatest probability, and so on. The IBR indicates that the expert is most likely to hit the interval of the ratio scale with his judgment, but that it is also possible, although it is less likely, to be wrong by choosing a different interval, with the intervals closest to the correct one being more likely than the farthest.

4. Scenarios

In this research framework, scenarios are a tool used to explore different situations that could affect the performance of a decision support system. We carry out a detailed description of a set of exogenous conditions and their consequences on the AHP expected performance for the different ratio scales previously considered. The IBRM commits to study all possible pairwise comparison matrices, the only exogenous aspects being the skill of the decision-maker (β) and the difficulty of the problem, differences between latent performances, $V_j - V_i; i,j = 1, \cdots, n$, (Sáenz-Royo et al., 2023b).

The scenarios are useful to give information about the "resilience" that checks the robustness of the solutions proposed by the decision support system (AHP) (Han et al., 2016). By considering different situations and their possible impacts, it is possible to identify the risks and opportunities that decision-makers will face based on the problem analyzed and to design strategies that guarantee a level of performance based on the organization's objectives (commission errors or omission errors) (Catalani & Clerico, 1996; Sáenz-Royo, Salas-Fumás, et al., 2022; Sah & Stiglitz, 1986).

The IBRM presents a dual effect that is difficult to anticipate when the difference between alternative performances changes. As the difference between the performance of the two alternatives increases, the probability that the decision-maker selects the correct alternative increases, but the error cost (opportunity cost) increases. Conversely, if the difference between the performance of the two alternatives decreases, the decision-maker is more likely to be wrong, but the penalty for doing so is lower.

This paper recreates the below three distribution scenarios of ordered latent performances, $V_1 > V_2 > V_3$, of three alternatives, $\{A_1, A_2, A_3\}$:

- ***Scenario 1:*** The latent performance value of alternative A_2 is closer to the latent performance value of alternative A_3 than to the latent performance value of alternative A_1. In other words, in scenario 1 there is one alternative A_1 with a clearly superior performance value than the other two alternatives, A_2 and A_3, which present similar performance values.
- ***Scenario 2:*** The ratio of the performance value of alternative A_2 with respect to the performances of the other two alternatives is the same, i.e. the latent performance of the middle alternative is the geometric mean of the performances of the other two alternatives.
- ***Scenario 3:*** The latent performance value of alternative A_2 is closer to the latent performance value of alternative A_1 than to the latent performance value of alternative A_3.

For each of the three scenarios, different levels of decision-maker expertise are established within 0.001 to 2, which are selected due to the convergence of the expected performances. As mentioned before, when the expert's judgment capacity converges to 0, all the alternatives tend to have the same probability of being chosen, no matter the performance values the ratio scale used, while when the expert's judgment capacity is very high (high expertise), the easier is to select the best solution no matter which ratio scale is used.

5. Intentional bounded rationality methodology (IBRM)

Decision support systems aim to establish requirements and systematics processes that improve human decisions (Keeney, 1992; Moreno-Jiménez et al., 1999). Sáenz-Royo et al. (2023a) develop the methodology to evaluate decision support systems under the assumption that experts have IBR. This methodology is based on the definition of an automaton that represents a decision-maker with IBR and analyze its expected performance assuming that the alternative performances are known. The computation of the probability of choosing each of the available interval allows to evaluate ex-ante the expected performance of the elements that make up the decision support systems to evaluate. This paper aims to analyze how the choice of a ratio scale of expert judgments improves or degrades the expected performance of the AHP decision support systems.

The steps of the IRBM are described below and they are incrementally illustrated with the particular example of using AHP with linear ratio scale and three alternatives with known performances:

Step 1. The automaton chooses the interval of the ratio scale with the probabilities obtained in Eq. (4).

The IBRM automaton assigns to each pairwise judgment a probability of choosing each interval of the ratio scale. For example, if the automaton has a reliability parameter $\beta = 1$, the comparison of two alternatives A_i and A_j whose performances are $V_i = 25$ and $V_j = 10$, according to the linear ratio scale ($x = [1.5, 2.5, 3.5, 4.5, 5.5, 6.5, 7.5, 8.5, 9.5]$) without the possibility of expressing indifference, would lead to the following automaton probability values for each of the 18 intervals ($p_{\beta,ij}(x)$):

With an intensity greater than 9, A_j is preferred over A_i
$p_{1,ij}(1/9.5) = 0.0000$.
With an intensity greater than 8, A_j is preferred over A_i
$p_{1,ij}(1/8.5) = 0.0000$.
With an intensity greater than 7, A_j is preferred over A_i
$p_{1,ij}(1/7.5) = 0.0001$.
With an intensity greater than 6, A_j is preferred over A_i
$p_{1,ij}(1/6.5) = 0.0002$.
With an intensity greater than 5, A_j is preferred over A_i
$p_{1,ij}(1/5.5) = 0.0006$.
With an intensity greater than 4, A_j is preferred over A_i
$p_{1,ij}(1/4.5) = 0.0019$.
With an intensity greater than 3, A_j is preferred over A_i
$p_{1,ij}(1/3.5) = 0.0056$.
With an intensity greater than 2, A_j is preferred over A_i
$p_{1,ij}(1/2.5) = 0.0182$.
With an intensity greater than 1, A_j is preferred over A_i
$p_{1,ij}(1/1.5) = 0.0825$.
With an intensity greater than 1, A_i is preferred over A_j
$p_{1,ij}(1.5) = 0.2452$.
With an intensity greater than 2, A_i is preferredd over A_j
$p_{1,ij}(2.5) = 0.2836$.
With an intensity greater than 3, A_i is preferred over A_j
$p_{1,ij}(3.5) = 0.2009$.
With an intensity greater than 4, A_i is preferred over A_j
$p_{1,ij}(4.5) = 0.0981$.
With an intensity greater than 5, A_i is preferred over A_j
$p_{1,ij}(5.5) = 0.0396$.
With an intensity greater than 6, A_i is preferred over A_j
$p_{1,ij}(6.5) = 0.0148$.
With an intensity greater than 7, A_i is preferred over A_j
$p_{1,ij}(7.5) = 0.0054$.
With an intensity greater than 8, A_i is preferred over A_j
$p_{1,ij}(8.5) = 0.0020$.
With an intensity greater than 9, A_i is preferred over A_j
$p_{1,ij}(9.5) = 0.0011$.

This process should be repeated for all possible pairwise judgments, that is, A_1 vs. $A_2(a_{12}(x))$; A_1 vs. A_3 $(a_{13}(x))$;...; A_{n-1} vs. A_n $(a_{(n-1)(n)}(x))$, where the alternatives number is n. In AHP, a judgment matrix M requires $n!/(2!(n-2)!)$ number of pairwise judgments, those above the non-necessary main diagonal elements since their reciprocal values are the pairwise judgments below the main diagonal:

$$\mathrm{M} = \begin{pmatrix} - & a_{12}(x_k) & a_{13}(x_k) \\ a_{21}(x_k) & - & a_{23}(x_k) \\ a_{31}(x_k) & a_{32}(x_k) & - \end{pmatrix}$$

Step 2. All possible AHP pairwise comparison matrices are obtained.

The number of possible AHP judgment matrices that can be constructed with a ratio scale with m intervals would be the number of variations with repetition of m taken $n!/(2!(n-2)!)$ by $n!/(2!(n-2)!)$:

$m^{\frac{n!}{2!(n-2)!}}$. For example, with three alternatives ($n = 3$), the number of pairwise judgments required by a judgement matrix is $3 = 3!/(2!(3-2)!)$; and the linear ratio scale of step 1 example, with 18 (= m) intervals, would result in $18^3 = 5832$ possible pairwise comparison matrices:

$$\mathrm{M}(1)=\begin{pmatrix} - & 9.5 & 9.5 \\ \frac{1}{9.5} & - & 9.5 \\ \frac{1}{9.5} & \frac{1}{9.5} & - \end{pmatrix}, \mathrm{M}(2)=\begin{pmatrix} - & 9.5 & 9.5 \\ \frac{1}{9.5} & - & 8.5 \\ \frac{1}{9.5} & \frac{1}{8.5} & - \end{pmatrix}, \mathrm{M}(3)=\begin{pmatrix} - & 9.5 & 9.5 \\ \frac{1}{9.5} & - & 7.5 \\ \frac{1}{9.5} & \frac{1}{7.5} & - \end{pmatrix}, \cdots, \mathrm{M}(5832)=\begin{pmatrix} - & \frac{1}{9.5} & \frac{1}{9.5} \\ 9.5 & - & \frac{1}{9.5} \\ 9.5 & 9.5 & - \end{pmatrix}$$

Step 3. The probability of each pairwise comparison matrix, which represents one combination of judgments of the automaton obtained in the previous step, is calculated.

The automaton probability for a pairwise comparison matrix is calculated as the product of the probabilities of the pairwise judgments above the main diagonal:

$$p_\beta(M(h)) = \prod_{i=1}^{n-1}\prod_{j>i}^{n} p_{\beta,ij}(x_k) \tag{5}$$

Continuing with the example from steps 1 and 2: an automaton with $\beta = 1$; three alternatives A_1, A_2, and A_3 with latent performances $V_1 = 62.5$, $V_2 = 25$, and $V_3 = 10$; and the linear ratio scale ($x = [1.5, 2.5, 3.5, 4.5, 5.5, 6.5, 7.5, 8.5, 9.5]$); the probability of the pairwise comparison matrices would be:

$p_1(M(1)) = p_{1,12}(9.5)p_{1,13}(9.5)p_{1,23}(9.5);$

$p_1(M(2)) = p_{1,12}(9.5)p_{1,13}(9.5)p_{1,23}(8.5);$

$p_1(M(3)) = p_{1,12}(9.5)p_{1,13}(9.5)p_{1,23}(7.5);$

…

$p_1(M(5831)) = p_{1,12}(1/9.5)p_{1,13}(1/9.5)p_{1,23}(1/8.5)$

$p_1(M(5832)) = p_{1,12}(1/9.5)p_{1,13}(1/9.5)p_{1,23}(1/9.5)$

Step 4. The priority vector, $w(h)$, and the consistency ratio, $CR(h)$, of the pairwise comparison matrices obtained in step 2 are calculated using the corresponding AHP processes.

The priority vector in AHP is the normalized evaluation of the alternatives that satisfy $Mw = dw$. The equation $(M - dI)w = 0$ is solvable solution if and only if d is an eigenvalue (λ_i) of M. Recall that elements of M are $a_{ij} = \frac{V_i}{V_j}$, which means that M has rank 1, so all its eigenvalues are zero ($\lambda_i = 0$) except one, denoted λ_{max} which is equal to n, since the sum of all eigenvalues of a matrix is equal to its trace (or sum of main diagonal elements, in this case $a_{ii} = 1\forall i = 1,\cdots,n$). Thus, each column of M is a solution to the above equation. Since the AHP solution is unique, the normalized solution or priority vector $w = (w_1, \ldots, w_n)$, where w_i is the normalized score of alternative A_i, is proposed and obtained by normalizing any one of the columns of M. By choosing an interval of the ratio scale, the solution offered by AHP must match the evaluation ratio to its latent performances ($w_i/w_j = V_i/V_j$).

Consistency attempts to ensure that the judgments expressed by an expert have a logical relationship. Consistency in AHP requires that the pairwise judgment matrix M satisfies the following cardinally consistency property: $a_{ik} = a_{ij}a_{jk}|\forall i \neq j \neq k$. When performances of alternatives are known, the matrix M has elements defined as $a_{ij} = \frac{V_i}{V_j}$, which obviously verifies $\frac{V_i}{V_k} = \frac{V_i}{V_j}\frac{V_j}{V_k}$ and therefore is consistent and the above computation process of the priority vector applies. Indeed, algebraically, a matrix M is fully consistent if and only if $\lambda_{max} = n$.

Saaty (1980) defines the value $CI_M = (\lambda_{max} - d)/(d-1)$ as the consistency index of a matrix M, where $\lambda_{max} = \sum_{i=1}^{n} a_{ij}\frac{w_j}{w_i}$ is the sum of the product between the real relative value between two alternatives and the inverse of the normalized score obtained by AHP. If matrix M is consistent, then $CI_M = 0$. Otherwise, it is not zero, with values closer to zero meaning greater degree of consistency. An issue with CI_M is its dependency of the dimension of M. To solve this problem, Saaty (1980) introduces the Consistency Ratio (CR) criterion as the quotient between CI_M and the Random Consistency Index (RI), which represents the mean of CI values of a set of randomly generated matrices of dimension n. In this way, $CR_M = CI_M/RI$ is no longer dependent on the dimension of M, which allows to set the universal consistency acceptance requirement for a matrix M to be of acceptable consistency if its CR_M value is below 0.1 (Vaidya & Kumar, 2006). Following with the example of previous steps, the priority vector and the consistency ratio for each of the 5832 possible pairwise comparison matrices are computed:

$M(1)w(1) = (w_1(1), w_2(1), w_3(1))CR(1)$

$M(2)w(2) = (w_1(2), w_2(2), w_3(2))CR(2)$

… … …

$M(5832)w(5832) = (w_1(5832), w_2(5832), w_3(5832))CR(5832)$

Step 5. Set constraint.

This step classes pairwise comparison matrices into those that are of acceptable consistency and those that are not. If no consistency constraint is imposed, then all matrices with their probabilities calculated in step 3 are considered in next step 6. Otherwise, i.e. if consistency constraint is imposed, then only the set of matrices with acceptable consistency are considered in next step 6 and their probabilities are normalized: $p'_\beta(M(h)) = p_\beta(M(h))/\sum_{CR(k)<0.1} p_\beta(M(k))$. Thus, we have the following rule to apply to the pairwise comparison matrices.

If constraint set, then

$$p'_\beta(M(h)) = \begin{cases} \dfrac{p_\beta(M(h))}{\sum_{CR(k)<0.1} p_\beta(M(k))} & if CR(h) < 0.1 \\ 0 & otherwise. \end{cases};$$

otherwise, $p'_\beta(M(h)) = p_\beta(M(h))$.

The information is grouped according to the goals of the analysis. Sometimes it is interesting to know the ranking of all alternatives. In our case, we assume an ϵ-type problem where the alternatives are mutually exclusive, only the one alternative ranked first is chosen and errors in the order of the rest of the alternatives whose position are not relevant (Aguarón & Moreno-Jiménez, 2000; Sáenz-Royo et al., 2023a). In this case, the matrices leading to the selection of the same alternative are grouped together and the addition of their probabilities is defined as the probability of choosing such alternative by the AHP:

$$p_\beta(A_i) = \sum_{w_i(j)=\max_c w_c(j)} p_{\beta'}(M(j))$$

In the considered example with only three alternatives, the probability that the AHP automaton chooses A_1 is obtained by adding the probability values $p_{\beta'}$ of those matrices with priority vector with w_1 the largest component element. The same is done for alternatives A_2 and A_3, respectively. Thus, at the end of this step, the probability of each alternative to be chosen by the AHP are obtained: $p_\beta(A_1)$, $p_\beta(A_2)$, $\cdots$, $p_\beta(A_n)$.

Step 6. Computing the expected performance of the automaton with reliability β for the considered decision support system and constraints.

When there is no tie between alternatives, the expected performance will be computed as:

$$E(DSS\_constrains) = \sum_i^n p_\beta(A_i) V_i$$

When there is ties between alternatives, then these must be resolved by multiplying them by the mean of the performances of the tied alternatives. For the considered AHP example with three alternatives and linear order, if the consistency criteria of the automaton's judgments is not required, then the expected performance with ties between the three alternatives would be:

$$\begin{aligned} E(AHPwoC) &= p_1(A_1)V_1 + p_1(A_2)V_2 + p_1(A_3)V_3 \\ &\quad + p_1(A_1 = A_2 = A_3)\left(\frac{V_1 + V_2 + V_3}{3}\right) \end{aligned}$$

Note that two alternatives cannot tie for first place because indifference is not allowed. The tie can only result from a non-transitive circular ordering of the three alternatives.

6. Ratio scales analysis

The case presented in Sáenz-Royo et al. (2023a) is used as the basis to evaluate the expected performance of each ratio scales in AHP previously described. Thus, it is assumed that there are three alternatives $\{A_1, A_2, A_3\}$ with latent performance values $V_1 = 62.5$, $V_2 = 25$ and $V_3 = 10$, respectively. The three scenarios described in Section 4 are created by varying the value of the latent performance of alternative A_2. In Scenario 1, $V_2 = 15$. In Scenario 2, $V_2 = \sqrt{625} = 25$. In Scenario 3, $V_2 = 55$.

6.1. Analysis of the AHP ratio scales with the expert being aware of the relative performance values (Problem 1)

As mentioned before, different reliability values of the automaton (β) ranging from 0.01 (an automaton with a very low ability to discriminate between alternatives) to 2 (an automaton with high reliability) are simulated for each of the three scenarios. In addition, as described in Section 5, the automaton can manifest different preference intensities, V_i/V_j, with different probabilities, $p_{\beta ij}$, for each pairwise comparison. This analysis allows to evaluate the results of each of the considered ratio scales at different levels of expertise of the automaton.

Tables 2, 4, and 6 report expected performances for each considered AHP ratio scale, at different levels of expertise, with and without consistency constraint, for Scenarios 1, 2 and 3, respectively. At each level of expertise, background graded colors were added to indicate the proximity to the corresponding maximum expected performance (blue color) or the minimum expected performance (yellow color). Also, at each level of expertise, the difference between the maximum and minimum expected performances was divided into four equal ranges represented with highlighted bars of different heights to the left of the expected performances to indicate where in expected performance range they lie. Expected performance values closer to the minimum expected performance value than to the maximum expected performance value are indicated with one or two blue bars; otherwise, three or four blue bars are used instead. In addition, bold blue and blue italics numbers are used to highlight the maximum and minimum expected performances of all ratio scales without and with consistency, respectively. The lower value labeled "Total" shows the corresponding cumulative expected performances for the considered AHP ratio scale, which is proposed to use as an indicator of the ratio scale favoring the AHP expected performance if the expertise level the decision-maker involved in the analysis were unknown.

Scenario 1 analysis. Since alternative A_1 performs significantly better than the other two alternatives, it is more likely to be selected as the best alternative (but errors have a high opportunity cost). Table 2 analysis shows that the scales with greater precision in the real performance area (greater number of intervals in values close to 2.5) sometimes present worse performance than the more dispersed scales (geometric FH and logarithmic scales perform worse than geometric L scale). The consistency constraint does not provide the same gain for all compared scales. Without consistency constraint, the power scale is most striking because it achieves the minimum expected performances for all levels of expertise but 0.2 and 0.4. With consistency constraint, the expected performances are closer to the corresponding maximum expected performances. Another striking result is that the geometric FH scale with consistency constraint achieves the best expected performance values when the expertise level is low (0.001–0.2) and the worst expected performance values when the expertise level is high (1–2), while its worst expected performance value is achieved consistency constraint and expertise level of 0.2. Regarding total expected performance, the balance scale is the best of all compared scales with or without consistency constraint. In the second case, without consistency constraint, the balance scale stands out because it presents the highest expected

Table 2
Scenario 1 expected performances for each considered AHP ratio scale, at different levels of expertise, with and without consistency constraint.

V_2=15	Linear		Root		Power		Geometric L		Balanced		Geometric FH		Logarithmic	
β	AHPwoC	AHPwC	AHPwoC	AHPwC	AHPwoC	AHPwC	AHPwoC	AHPwC	AHPwoC	AHPwC	AHPwoC	AHPwC	AHPwoC	AHPwC
0.001	29.2307	29.2307	29.2306	29.2306	**24.8797**	*17.3519*	**29.2339**	29.2296	29.2307	29.2307	29.2305	*29.2518*	29.2306	29.2306
0.2	44.0379	45.8048	43.3580	*44.3487*	43.9355	45.9357	44.1175	46.3872	**44.1650**	46.3534	**43.1894**	*46.5938*	43.4404	44.4433
0.4	54.7076	56.8082	53.9183	*55.4240*	54.0009	56.6910	54.5575	57.0927	**54.8928**	*57.1278*	**53.6981**	56.4275	54.0270	55.5474
0.6	59.6303	60.7602	59.0125	*59.9325*	**58.8235**	60.7468	59.4037	60.9442	**59.7767**	*60.9522*	58.8508	60.1793	59.1030	60.0262
0.8	61.4832	61.9767	61.1039	*61.5422*	**60.8894**	61.9960	61.3085	62.0828	**61.5751**	*62.0883*	61.0093	61.5687	61.1647	61.5996
1	62.1382	62.3409	61.9344	62.1281	**61.7759**	62.3582	62.0357	62.3922	**62.1885**	*62.3929*	61.8848	*62.1100*	61.9704	62.1587
1.2	62.3693	62.4509	62.2679	62.3511	**62.1680**	62.4606	62.3170	*62.4727*	**62.3944**	62.4718	62.2431	*62.3324*	62.2875	62.3661
1.4	62.4522	62.4847	62.4040	62.4394	**62.3461**	62.4891	62.4274	*62.4931*	**62.4639**	62.4924	62.3919	*62.4272*	62.4140	62.4463
1.6	62.4824	62.4952	62.4602	62.4751	**62.4282**	62.4970	62.4711	*62.4983*	**62.4876**	62.4979	62.4543	*62.4683*	62.4651	62.4782
1.8	62.4934	62.4985	62.4835	62.4897	**62.4665**	62.4992	62.4885	*62.4996*	**62.4957**	62.4994	62.4807	*62.4862*	62.4859	62.4911
2	62.4976	62.4995	62.4932	62.4958	**62.4844**	62.4998	62.4954	*62.4999*	**62.4985**	62.4998	62.4918	*62.4940*	62.4943	62.4963
Total	623.5227	629.3504	620.6664	624.8572	**616.1982**	*617.5252*	622.8563	630.5921	**624.1689**	*630.6065*	619.9246	628.3392	621.0828	625.2838

Table 3
Scenario 1 comparison of maximum and minimum expected performance, at different levels of expertise, with and without consistency constraint.

$V_2 = 15$	AHPwoC			AHPwC			Global		
β	Max	Min	%	Max	Min	%	Max	Min	%
0.001	29.2339	24.8797	17.50 %	29.2518	17.3519	68.58 %	29.2518	17.3519	68.58 %
0.2	44.1650	43.1894	2.26 %	46.5938	44.3487	5.06 %	46.5938	43.1894	7.88 %
0.4	54.8928	53.6981	2.22 %	57.1278	55.4240	3.07 %	57.1278	53.6981	6.39 %
0.6	59.7767	58.8235	1.62 %	60.9522	59.9325	1.70 %	60.9522	58.8235	3.62 %
0.8	61.5751	60.8894	1.13 %	62.0883	61.5422	0.89 %	62.0883	60.8894	1.97 %
1	62.1885	61.7759	0.67 %	62.3929	62.1100	0.46 %	62.3929	61.7759	1.00 %
1.2	62.3944	62.1680	0.36 %	62.4727	62.3324	0.22 %	62.4727	62.1680	0.49 %
1.4	62.4639	62.3461	0.19 %	62.4931	62.4272	0.11 %	62.4931	62.3461	0.24 %
1.6	62.4876	62.4282	0.10 %	62.4983	62.4683	0.05 %	62.4983	62.4282	0.11 %
1.8	62.4957	62.4665	0.05 %	62.4996	62.4862	0.02 %	62.4996	62.4665	0.05 %
2	62.4985	62.4844	0.02 %	62.4999	62.4940	0.01 %	62.4999	62.4844	0.02 %
Total	624.1689	616.1982	1.29 %	630.6065	617.5252	2.12 %	630.6065	616.1982	2.34 %

performances at all levels of expertise but 0.001; in the first case, with consistency constraint, the expected performances of the balance scale in comparison to the values of the geometric L scale are very close when the first are lower but higher when the level of expertise range from 0.4 to 1. Another result to highlight in this scenario refers to the total expected performance values achieved with consistency constraint being higher than without consistency constraint for all compared scales, although when the level of expertise is very low (0.001 and/or 0.2) some scales (power and geometric L scale) have higher expected performance values without consistency constraint than with consistency constraint.

Table 3 shows the maximum and minimum expected performance, at different levels of expertise, with and without consistency constraint, as well as their corresponding totals. The percentage differences between maximum and minimum expected performance show that: (i) they decrease monotonically with the level of expertise with and without consistency constraint; (ii) for expertise levels lower than (higher than or equal to) 1 the percentage differences with consistency constraint are higher (lower) than without consistency constraint; (iii) above the expertise level $\beta = 1$, the overall percentage difference between the best and the worst performance is less than 1 %; (iv) for the lowest level of expertise (0.001) the percentage differences are very high with and without consistency constraint; (vi) for level of expertise different to 0.001, the biggest global percentage difference does not exceed 8 %, of which 5 % is due to the scale and 3 % to the consistency constraint or lack of it.

Scenario 2 analysis. In this case, the probability of getting it right (alternative A_1 is chosen as best alternative) is reduced and so is the opportunity cost, making the overall balance uncertain. Table 4 shows many similarities with Table 2: the balanced scale is again the scale with highest "total" expected performances with and without consistency constraint, while the power scale has the worst expected performances without consistency constraint. The differences with respect to Scenario 1 are that the geometric FH scale is no longer the scale with the worst expected performance with consistency constraint when the level of expertise is between 1 and 2; the root scale shows the worst "total" expected performance with consistency constraint, which is also true at all the levels of expertise but 0.001.

As with Table 3, Table 5 shows that both the maximum and minimum expected performances monotonically increase with the level of expertise. However, the percentage difference between the best and worst ratio scales' expected performances monotonicity varies with the level of expertise range: without consistency constraint the monotonicity is increasing (decreasing) between 0.2 and 0.8 (1 and 2); with consistency constraint the same happens but with monotonicity turning point being the level of expertise 0.6. In general, the choice of scale is more important in Scenario 2 than in Scenario 1, except without consistency constraint at the level of expertise 0.2, and with consistency constraint and levels of expertise 0.001 and 0.2. The same is true globally, from a level of expertise 0.2, the differences between the best and worst expected performance are greater in Scenario 2 than in Scenario 1.

Scenario 3 analysis. Since alternatives A_1 and A_2 have close performances (62.5 and 55), the probability of choosing alternative A_2 instead of A_1 is high, while the opportunity cost of being wrong is relatively low. In this scenario, the balanced scale is still the scale with best expected performance with and without consistency constraint while the power scale shows the worst "total" expected performance with and without consistency constraint (Table 6). A notable difference with respect to the two previous scenarios is that the performance of the linear scale for high level of expertise in [1–2], which had the best performance range in Scenarios 1 and 2, has now the worst performance range in Scenario 3. The power scale and the geometric L scale show the same behavior: the first one without consistency constraint and level of expertise between 0.2 and 2, the second one with consistency constraint and level of

Table 4
Scenario 2 expected performances for each considered AHP ratio scale, at different levels of expertise, with and without consistency constraint.

V_2=25	Linear		Root		Power		Geometric L		Balanced		Geometric FH		Logarithmic	
β	AHPwoC	AHPwC	AHPwoC	AHPwC	AHPwoC	AHPwC	AHPwoC	AHPwC	AHPwoC	AHPwC	AHPwoC	AHPwC	AHPwoC	AHPwC
0.001	32.5539	32.5539	32.5538	32.5538	**27.4483**	*21.6968*	**32.5566**	32.5530	32.5539	32.5539	32.5538	*32.5723*	32.5538	32.5538
0.2	44.0018	45.5283	43.4390	*44.4630*	43.9244	45.5738	44.0609	*45.9571*	**44.0950**	45.9271	**43.2953**	45.9564	43.5087	44.5301
0.4	51.8193	54.0709	51.0230	*52.3838*	51.1173	53.9607	51.6895	*54.4225*	**52.0237**	54.4008	**50.8132**	53.4160	51.1270	52.5757
0.6	56.1728	58.0784	55.1811	*56.1057*	**54.9249**	58.1142	55.8534	58.4034	**56.4663**	*58.4189*	54.9850	56.4427	55.3009	56.2223
0.8	58.6183	60.0545	57.6144	*58.0767*	**57.1927**	60.2657	58.1990	60.3812	**58.9922**	*60.4414*	57.4539	58.3042	57.7337	58.2304
1	60.0635	61.0791	59.1907	*59.3469*	**58.7533**	61.3976	59.6302	61.4304	**60.4785**	*61.4684*	59.0692	59.5511	59.2917	59.5439
1.2	60.9526	61.7033	60.2587	*60.2795*	**59.8854**	61.9706	60.5540	61.9775	**61.3514**	*61.9830*	60.1727	60.4360	60.3337	60.4527
1.4	61.5125	62.0404	60.9879	*60.9413*	**60.7029**	62.2388	61.1709	*62.2510*	**61.8548**	62.2391	60.9308	61.0691	61.0386	61.0886
1.6	61.8693	62.2366	61.4836	*61.4335*	**61.2809**	62.3755	61.5908	*62.3852*	**62.1403**	62.3669	61.4482	61.5184	61.5157	61.5327
1.8	62.0975	62.3503	61.8187	*61.7997*	**61.6812**	*62.4457*	61.8790	62.4455	**62.3003**	62.4312	61.7982	61.8329	61.8381	61.8409
2	62.2435	62.4153	62.0442	*62.0139*	**61.9539**	62.4739	62.0772	*62.4744*	**62.3893**	62.4626	62.0333	62.0501	62.0556	62.0532
Total	611.9049	622.1109	605.5950	*609.3978*	**598.8651**	612.5132	609.2616	624.6813	**614.6456**	*624.6933*	604.5537	613.1491	606.2975	610.6244

Table 5
Scenario 2 comparison of maximum and minimum expected performance, at different levels of expertise, with and without consistency constraint.

$V_2 = 25$	AHPwoC			AHPwC			Global		
β	Max	Min	%	Max	Min	%	Max	Min	%
0.001	32.5566	27.4483	18.61 %	32.5723	21.6968	50.12 %	32.5723	21.6968	50.12 %
0.2	44.0950	43.2953	1.85 %	45.9571	44.4630	3.36 %	45.9571	43.2953	6.15 %
0.4	52.0237	50.8132	2.38 %	54.4225	52.3838	3.89 %	54.4225	50.8132	7.10 %
0.6	56.4663	54.9249	2.81 %	58.4189	56.1057	4.12 %	58.4189	54.9249	6.36 %
0.8	58.9922	57.1927	3.15 %	60.4414	58.0767	4.07 %	60.4414	57.1927	5.68 %
1	60.4785	58.7533	2.94 %	61.4684	59.3469	3.57 %	61.4684	58.7533	4.62 %
1.2	61.3514	59.8854	2.45 %	61.9830	60.2795	2.83 %	61.9830	59.8854	3.50 %
1.4	61.8548	60.7029	1.90 %	62.2510	60.9413	2.15 %	62.2510	60.7029	2.55 %
1.6	62.1403	61.2809	1.40 %	62.3852	61.4335	1.55 %	62.3852	61.2809	1.80 %
1.8	62.3003	61.6812	1.00 %	62.4457	61.7997	1.05 %	62.4457	61.6812	1.24 %
2	62.3893	61.9539	0.70 %	62.4744	62.0139	0.74 %	62.4744	61.9539	0.84 %
Total	614.6456	598.8651	2.64 %	624.6933	609.3978	2.51 %	624.6933	598.8651	4.31 %

Table 6
Scenario 3 expected performances for each considered AHP ratio scale, at different levels of expertise, with and without consistency constraint.

V_2=55	Linear		Root		Power		Geometric L		Balanced		Geometric FH		Logarithmic	
β	AHPwoC	AHPwC	AHPwoC	AHPwC	AHPwoC	AHPwC	AHPwoC	AHPwC	AHPwoC	AHPwC	AHPwoC	AHPwC	AHPwoC	AHPwC
0.001	42.5702	42.5703	42.5701	42.5702	**35.2013**	*34.7749*	**42.5738**	42.5691	42.5702	42.5703	42.5701	*42.5935*	42.5701	42.5702
0.2	54.3893	55.1878	53.9177	*54.7188*	54.3232	55.0950	54.3794	55.3276	**54.3989**	55.3672	**53.7687**	*55.7967*	53.9874	54.7512
0.4	58.2458	58.6600	58.0994	*58.4974*	58.1081	58.6169	58.2227	*58.7169*	**58.2595**	58.7148	**58.0006**	58.6716	58.1369	58.4984
0.6	59.0290	59.2455	58.9532	*59.0057*	**58.9057**	59.2318	59.0111	*59.2999*	**59.0692**	59.2937	58.9212	59.0611	58.9650	59.0096
0.8	59.2156	59.3998	59.1431	*59.1080*	**59.1155**	59.3732	59.1930	59.4375	**59.2938**	*59.4766*	59.1333	59.1664	59.1481	59.1142
1	59.3055	59.4809	59.2414	*59.1915*	**59.2274**	59.4422	59.2809	59.4843	**59.4366**	*59.6044*	59.2376	59.2482	59.2441	59.1962
1.2	59.3803	59.5427	59.3307	*59.2854*	**59.3233**	59.4990	59.3579	59.5103	**59.5714**	*59.7230*	59.3289	59.3330	59.3322	59.2879
1.4	59.4553	59.5993	59.4205	*59.3846*	**59.4164**	59.5561	59.4374	59.5411	**59.7076**	*59.8405*	59.4196	59.4213	59.4213	59.3856
1.6	59.5338	59.6566	59.5110	*59.4845*	**59.5088**	59.6172	59.5207	59.5851	**59.8461**	*59.9596*	59.5105	59.5112	59.5114	59.4846
1.8	59.6159	59.7171	59.6015	59.5827	**59.6003**	59.6830	59.6069	59.6425	**59.9864**	*60.0814*	59.6013	59.6015	59.6017	*59.5825*
2	59.7003	59.7817	59.6915	59.6786	**59.6909**	59.7531	59.6944	59.7105	**60.1274**	*60.2056*	59.6914	59.6915	59.6916	*59.6782*
Total	630.4409	632.8416	629.4800	630.5074	**622.4208**	*624.6425*	630.2782	632.8249	**632.2672**	*634.8369*	629.1832	632.0960	629.6095	630.5588

Table 7
Scenario 3 comparison of maximum and minimum expected performance, at different levels of expertise, with and without consistency constraint.

$V_2 = 55$	AHPwoC			AHPwC			Global		
β	Max	Min	%	Max	Min	%	Max	Min	%
0.001	42.5738	35.2013	20.94 %	42.5935	34.7749	22.48 %	42.5935	34.7749	22.48 %
0.2	54.3989	53.7687	1.17 %	55.7967	54.7188	1.97 %	55.7967	53.7687	3.77 %
0.4	58.2595	58.0006	0.45 %	58.7169	58.4974	0.38 %	58.7169	58.0006	1.23 %
0.6	59.0692	58.9057	0.28 %	59.2999	59.0057	0.50 %	59.2999	58.9057	0.67 %
0.8	59.2938	59.1155	0.30 %	59.4766	59.1080	0.62 %	59.4766	59.1080	0.62 %
1	59.4366	59.2274	0.35 %	59.6044	59.1915	0.70 %	59.6044	59.1915	0.70 %
1.2	59.5714	59.3233	0.42 %	59.7230	59.2854	0.74 %	59.7230	59.2854	0.74 %
1.4	59.7076	59.4164	0.49 %	59.8405	59.3846	0.77 %	59.8405	59.3846	0.77 %
1.6	59.8461	59.5088	0.57 %	59.9596	59.4845	0.80 %	59.9596	59.4845	0.80 %
1.8	59.9864	59.6003	0.65 %	60.0814	59.5825	0.84 %	60.0814	59.5825	0.84 %
2	60.1274	59.6909	0.73 %	60.2056	59.6782	0.88 %	60.2056	59.6782	0.88 %
Total	632.2672	622.4208	1.58 %	634.8369	624.6425	1.63 %	634.8369	622.4208	1.99 %

expertise between 1.4 and 2. Therefore, in general, a deterioration of the performance ranges is observed (Table 7).

Scenario 3 presents the lowest percentage difference between maximum and minimum expected performance with the only exception of level if expertise of 0.001 without consistency constraint. In this scenario, the choice of scale is unimportant for all levels of expertise above 0.001, even at the global level.

6.2. Analysis of the AHP ratio scales when the expert is not aware of the relative performance values (Problem 2)

Finally, an analysis is carried out under the assumption that the AHP verbal judgments (Weak, Moderate importance, Moderate plus, Strong importance, Strong plus, Very strong or demonstrated importance, Very very strong, Extreme importance) (Saaty, 1977) are the only guidelines in the pairwise comparisons of alternatives, i.e. there is no requirement for the expert to make a quantitative comparison. The use of verbal responses is intuitively attractive, easy to use, and more common in our daily lives, but it implies accepting an ambiguity in the comparisons that can have important consequences. In any case, the AHP needs numerical ratio scales in pairwise comparisons to derive priorities. In the following, we evaluate what can happen when verbal statements are transformed into numbers without the expert being aware of the transformation. The expert provides a verbal preference; its quantification is unknown to AHP and could be assigned to any of the numerical ratio scales analyzed in this study. With this approach, there is theoretically no reason to restrict the numbers assigned to verbal gradations. Although some authors try to justify this assignment, we will show, through two cases, the distortions generated depending on the assignment of unconscious values by the expert versus the assignment of

Table 8
Differences in expected performance between problem 1 and problem 2.

	Balanced						Power					
$V_2 = 25$	AHPwoC			AHPwC			AHPwoC			AHPwC		
β	Problem 1	Problem 2	%	Problem 1	Problem 2	%	Problem 1	Problem 2	%	Problem 1	Problem 2	%
0.001	32.5539	32.5539	0.00 %	32.5539	32.3149	0.74 %	27.4483	32.5539	−15.68 %	21.6968	32.5539	−33.35 %
0.2	44.0950	44.3107	−0.49 %	45.9271	45.4848	0.97 %	43.9244	43.9717	−0.11 %	45.5738	45.5695	0.01 %
0.4	52.0237	52.7229	−1.33 %	54.4008	52.5956	3.43 %	51.1173	51.6539	−1.04 %	53.9607	54.4536	−0.91 %
0.6	56.4663	57.5133	−1.82 %	58.4189	57.4529	1.68 %	54.9249	55.8653	−1.68 %	58.1142	58.9132	−1.36 %
0.8	58.9922	60.0338	−1.74 %	60.4414	59.9314	0.85 %	57.1927	58.2622	−1.84 %	60.2657	60.9707	−1.16 %
1	60.4785	61.3236	−1.38 %	61.4684	61.2826	0.30 %	58.7533	59.7256	−1.63 %	61.3976	61.8727	−0.77 %
1.2	61.3514	61.9593	−0.98 %	61.9830	61.9419	0.07 %	59.8854	60.6570	−1.27 %	61.9706	62.2507	−0.45 %
1.4	61.8548	62.2593	−0.65 %	62.2391	62.2474	−0.01 %	60.7029	61.2638	−0.92 %	62.2388	62.4033	−0.26 %
1.6	62.1403	62.3955	−0.41 %	62.3669	62.3974	−0.05 %	61.2809	61.6655	−0.62 %	62.3755	62.4632	−0.14 %
1.8	62.3003	62.4555	−0.25 %	62.4312	62.4560	−0.04 %	61.6812	61.9345	−0.41 %	62.4457	62.4862	−0.06 %
2	62.3893	62.4813	−0.15 %	62.4626	62.4815	−0.03 %	61.9539	62.1160	−0.26 %	62.4739	62.4949	−0.03 %
Total	614.6456	620.0091	−0.87 %	624.6933	620.5864	0.66 %	598.8651	609.6694	−1.77 %	612.5132	626.4319	−2.22 %

another scale in the method (AHP). To analyze this Problem 2, we assume that the expert unconsciously assigns the linear scale to the verbal judgments, and that AHP uses the balanced or the power scale (which the previous Problem 1 showed to have the best and worst performance, respectively) under the conditions of Scenario 2 to compute the vector of priorities. In Problem 2, the probability dispersion from the true value obtained from Eq. (4) uses the values of the linear scale, and the balanced and power scales are used to model the intensity value for each verbal judgment in the matrix (*M*) (step 2 of IBRM).

Table 8 shows the differences in expected performance between Problem 1, where the expert is aware of the cardinality of the scale and her/his errors are due to her/his IBR, and Problem 2, where the expert focuses on interpreting the verbal judgments and she/he can make a double error, the interpretation of the judgment and the errors due to her/his IBR.

The differences between problems in the balanced and power scales are small with and without consistency constraint. A counterintuitive result is that Problem 2 gives better results than Problem 1: the error caused by the change in scale biases the expert's decisions towards the best option, this phenomenon occurs with the balanced scale without consistency constraint and almost all level of expertise, with consistency constraint at some of the levels of expertise cases, and in the power scale in all but one level of expertise. The balanced scale continues to produce better results than the power scale except with consistency constraint where the favorable bias of the scale change makes the overall expected performance of the power scale superior to the balanced scale, something that never happened in Problem 1. Another noteworthy result is that the expected performance with consistency constraint is lower than without consistency constraint for the three compared scales, which was also verified in Problem 1, although their differences in values is very small.

7. Discussion and conclusions

These results show that despite the fact that Hershey et al. (1982) and Hershey & Schoemaker (1985) demonstrated that a linear transformation of the interval scale can drastically alter the generated results, their expected performance is not very different because the probability of these changes according to IBR is low. Remember that the equations proposed by IBR are used in situations where the experts have no cognitive biases or prejudices, so the AHP decision support system is evaluated in a strictly logical sense. It has been shown that, except for situations where the experts show almost total ignorance ($\beta = 0.001$), the maximum difference between the best and the worst expected performances never reaches 8 %, which demonstrates the stability of AHP.

Given that verbal judgments are an adequate way to measure human response to stimuli (Stevens, 1957; Stevens & Galanter, 1957), Problem 2 was used to evaluate whether the numerical values used in the ratio scales drastically affect the comparison operations of criteria and alternatives, altering the final choice of the expert. It was shown that AHP is stable regardless of the ratio scale used to measure expert judgments.

According to our simulations, the balanced scale is very stable at all levels of expertise and in all scenarios of Problem 1 and also in Problem 2. Although the differences between the scales are not very high, if the analysis requires great precision, our evaluation recommends the use of the balanced scale as a guarantee of performance with the AHP decision support system.

Finally, the doubts raised by Sáenz-Royo et al. (2023a) that the consistency requirement has a significant contribution to the expected performance are confirmed. Perhaps this contribution is stronger when the experts have cognitive biases, which is an interesting possible line of future research.

Funding

This work was supported by the Spanish Ministerio de Economía y Competitividad, [ECO2017-86305-C4-3-R], the Diputación General de Aragón (DGA) and the European Social Fund [CREVALOR], and by Project PID2020-113338RB-I00.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

Aguarón, J., & Moreno-Jiménez, J. M. (2000). Local stability intervals in the analytic hierarchy process. *European Journal of Operational Research, 125*(1), 113–132. https://doi.org/10.1016/S0377-2217(99)00204-0

Barzilai, J. (2005). Measurement and preference function modelling. *International Transactions in Operational Research, 12*(2), 173–183. https://doi.org/10.1111/j.1475-3995.2005.00496.x

Belton, V., & Gear, T. (1983). On a short-coming of Saaty's method of analytic hierarchies. *Omega, 11*(3), 228–230. https://doi.org/10.1016/0305-0483(83)90047-6

Blalock, H. M. (1968). *Methodology in social research*. McGraw-Hill.

Blumenthal, A. L. (1977). *The process of cognition*. Prentice-Hall.

Brans, J.-P. (1996). The space of freedom of the decision maker modelling the human brain. *European Journal of Operational Research, 92*(3), 593–602. https://doi.org/10.1016/0377-2217(96)00012-4

Cai, M., Lin, Y., Han, B., Liu, C., & Zhang, W. (2017). On a simple and efficient approach to probability distribution function aggregation. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 47*(9), 2444–2453. https://doi.org/10.1109/TSMC.2016.2531647

Catalani, M. S., & Clerico, G. F. (1996). How and When Unanimity is a Superior Decision Rule. In W. A. Müller & P. Schuster (Eds.), *Decision Making Structures* (pp. 15–29). Physica-Verlag HD. https://doi.org/10.1007/978-3-642-50138-8_2.

Chiclana, F., Tapia García, J. M., del Moral, M. J., & Herrera-Viedma, E. (2013). A statistical comparative study of different similarity measures of consensus in group decision making. *Information Sciences, 221*, 110–123. https://doi.org/10.1016/j.ins.2012.09.014

Dyer, J. S. (1990). Remarks on the analytic hierarchy process. *Management Science, 36*(3), 249–258. https://doi.org/10.1287/mnsc.36.3.249

Finan, J. S., & Hurley, W. J. (1999). Transitive calibration of the AHP verbal scale. *European Journal of Operational Research, 112*(2), 367–372. https://doi.org/10.1016/S0377-2217(97)00411-6

Franek, J., & Kresta, A. (2014). Judgment scales and consistency measure in AHP. *Procedia Economics and Finance, 12*, 164–173. https://doi.org/10.1016/S2212-5671(14)00332-3

Han, B., Liu, C. L., & Zhang, W. J. (2016). A method to measure the resilience of algorithm for operation management. *IFAC-PapersOnLine, 49*(12), 1442–1447.

Harker, P. T., & Vargas, L. G. (1987). The theory of ratio scale estimation: Saaty's analytic hierarchy process. *Management Science, 33*(11), 1383–1403. https://doi.org/10.1287/mnsc.33.11.1383

Herrera-Viedma, E., Cabrerizo, F. J., Kacprzyk, J., & Pedrycz, W. (2014). A review of soft consensus models in a fuzzy environment. *Information Fusion, 17*, 4–13. https://doi.org/10.1016/j.inffus.2013.04.002

Hershey, J. C., Kunreuther, H. C., & Schoemaker, P. J. H. (1982). Sources of bias in assessment procedures for utility functions. *Management Science, 28*(8), 936–954. https://doi.org/10.1287/mnsc.28.8.936

Hershey, J. C., & Schoemaker, P. J. H. (1985). Probability versus certainty equivalence methods in utility measurement: Are they equivalent? *Management Science, 31*(10), 1213–1231. https://doi.org/10.1287/mnsc.31.10.1213

Hogarth, R. M. (1975). Cognitive processes and the assessment of subjective probability distributions. *Journal of the American Statistical Association, 70*(350), 271–289. https://doi.org/10.1080/01621459.1975.10479858

Ishizaka, A., Balkenborg, D., & Kaplan, T. (2011). Influence of aggregation and measurement scale on ranking a compromise alternative in AHP. *Journal of the Operational Research Society, 62*(4), 700–710. https://doi.org/10.1057/jors.2010.23

Ishizaka, A., & Labib, A. (2011). Review of the main developments in the analytic hierarchy process. *Expert Systems with Applications*. https://doi.org/10.1016/j.eswa.2011.04.143

Ishizaka, A., & Lusti, M. (2004). An expert module to improve the consistency of AHP matrices. *International Transactions in Operational Research, 11*(1), 97–105. https://doi.org/10.1111/j.1475-3995.2004.00443.x

Keeney, R. L. (1992). *Value-focused thinking: A path to creative decisionmaking*. Harvard Univ. Press.

Keeney, R. L., & Raiffa, H. (1976). Decisions with multiple objectives: Preferences and value tradeoffs. *Wiley*.

Krantz, D. H. (1972). Measurement structures and psychological laws. *Science, 175*(4029), 1427–1435. https://doi.org/10.1126/science.175.4029.1427

Lin, C., & Kou, G. (2021). A heuristic method to rank the alternatives in the AHP synthesis. *Applied Soft Computing, 100*, Article 106916.

Liu, F., Qiu, M.-Y., & Zhang, W.-G. (2021). An uncertainty-induced axiomatic foundation of the analytic hierarchy process and its implication. *Expert Systems with Applications, 183*, Article 115427. https://doi.org/10.1016/j.eswa.2021.115427

Liu, Y., Eckert, C. M., & Earl, C. (2020). A review of fuzzy AHP methods for decision-making with subjective judgements. *Expert Systems with Applications, 161*, Article 113738. https://doi.org/10.1016/j.eswa.2020.113738

Lootsma, F. A. (1993). Scale sensitivity in the multiplicative AHP and SMART. *Journal of Multi-Criteria Decision Analysis, 2*(2), 87–110. https://doi.org/10.1002/mcda.4020020205

Meesariganda, B. R., & Ishizaka, A. (2017). Mapping verbal AHP scale to numerical scale for cloud computing strategy selection. *Applied Soft Computing, 53*, 111–118.

Moreno-Jiménez, J. M., Aguarón-Joven, J., Escobar-Urmeneta, M. T., & Turón-Lanuza, A. (1999). Multicriteria procedural rationality on SISDEMA. *European Journal of Operational Research, 119*(2), 388–403. https://doi.org/10.1016/S0377-2217(99)00141-1

Moreno-Jimenez, J. M., Pérez-Espés, C., & Wimmer, M. (2013). The Effectiveness of e-governance experiences in the knowledge society1. *European Conference on E-Government*.

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E., & Rakow, T. (2006). *Uncertain judgements: Eliciting experts' probabilities*. John Wiley & Sons.

Ravinder, H. V. (1992). Random error in holistic evaluations and additive decompositions of multiattribute utility—An empirical comparison. *Journal of Behavioral Decision Making, 5*(3), 155–167. https://doi.org/10.1002/bdm.3960050302

Saaty, T. L. (1977). A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology, 15*(3), 234–281. https://doi.org/10.1016/0022-2496(77)90033-5

Saaty, T. L. (1980). *The analytic hierarchy process: Planning, priority setting, resource allocation*. McGraw-Hill International Book Co.

Saaty, T. L. (1993). What is relative measurement? The ratio scale phantom. *Mathematical and Computer Modelling, 17*(4), 1–12. https://doi.org/10.1016/0895-7177(93)90170-4

Sáenz-Royo, C., Chiclana, F., & Herrera-Viedma, E. (2022). Functional representation of the intentional bounded rationality of decision-makers: A laboratory to study the decisions a priori. *Mathematics, 10*(5), 739. https://doi.org/10.3390/math10050739

Sáenz-Royo, C., Chiclana, F., & Herrera-Viedma, E. (2023a). Intentional bounded rationality methodology to assess the quality of decision-making approaches with latent alternative performances. *Information Fusion, 89*, 254–266. https://doi.org/10.1016/j.inffus.2022.08.019

Sáenz-Royo, C., Chiclana, F., & Herrera-Viedma, E. (2023b). Steering committee management. Expertise, diversity, and decision-making structures. *Information Fusion, 99*, 101888. https://doi.org/10.1016/j.inffus.2023.101888

Sáenz-Royo, C., Chiclana, F., & Herrera-Viedma, E. (2023). Ordering vs. AHP. Does the intensity used in the decision support techniques compensate? *Expert Systems with Applications*. , Article 121922. https://doi.org/10.1016/j.eswa.2023.121922

Sáenz-Royo, C., & Lozano-Rojo, Á. (2023). Authoritarianism versus participation in innovation decisions. *Technovation, 124*, Article 102741. https://doi.org/10.1016/j.technovation.2023.102741

Sáenz-Royo, C., Salas-Fumás, V., & Lozano-Rojo, Á. (2022). Authority and consensus in group decision making with fallible individuals. *Decision Support Systems, 153*, Article 113670. https://doi.org/10.1016/j.dss.2021.113670

Sah, R. K., & Stiglitz, J. E. (1986). The architecture of economic systems: Hierarchies and polyarchies. *The American Economic Review*, 716–727.

Salo, A. A., & Hämäläinen, R. P. (1997). On the measurement of preferences in the analytic hierarchy process. *Journal of Multi-Criteria Decision Analysis, 6*(6), 309–319. https://doi.org/10.1002/(SICI)1099-1360(199711)6:6<309::AID-MCDA163>3.0.CO;2-2

Simon, H. A. (1997). *Administrative behavior: A study of decision-making processes in administrative organizations (4th ed)*. Free Press.

Siraj, S., Mikhailov, L., & Keane, J. A. (2015). Contribution of individual judgments toward inconsistency in pairwise comparisons. *European Journal of Operational Research, 242*(2), 557–567. https://doi.org/10.1016/j.ejor.2014.10.024

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review, 64*, 153–181. https://doi.org/10.1037/h0046162

Stevens, S. S., & Galanter, E. H. (1957). Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology, 54*, 377–411. https://doi.org/10.1037/h0043680

Sugden, R. (1985). Why be consistent? A critical analysis of consistency requirements in choice theory. *Economica, 52*(206), 167–183. https://doi.org/10.2307/2554418

Tavana, M., Soltanifar, M., Santos-Arteaga, F. J., & Sharafi, H. (2023). Analytic hierarchy process and data envelopment analysis: A match made in heaven. *Expert Systems with Applications, 223*, Article 119902.

Temesi, J. (2011). Pairwise comparison matrices and the error-free property of the decision maker. *Central European Journal of Operations Research, 19*(2), 239–249. https://doi.org/10.1007/s10100-010-0145-8

Vaidya, O. S., & Kumar, S. (2006). Analytic hierarchy process: An overview of applications. *European Journal of Operational Research, 169*(1), 1–29. https://doi.org/10.1016/j.ejor.2004.04.028

Vargas, L. G. (1982). Reciprocal matrices with random coefficients. *Mathematical Modelling, 3*(1), 69–81. https://doi.org/10.1016/0270-0255(82)90013-6

Vargas, L. G. (1990). An overview of the analytic hierarchy process and its applications. *European Journal of Operational Research, 48*(1), 2–8. https://doi.org/10.1016/0377-2217(90)90056-H

Wallsten, T. S., & Budescu, D. V. (1983). State of the art—encoding subjective probabilities: A psychological and psychometric review. *Management Science, 29*(2), 151–173. https://doi.org/10.1287/mnsc.29.2.151

Yu, D., & Hong, X. (2022). A theme evolution and knowledge trajectory study in AHP using science mapping and main path analysis. *Expert Systems with Applications, 205*, Article 117675.

Zavadskas, E. K., & Turskis, Z. (2011). Multiple criteria decision making (MCDM) methods in economics: An overview. *Technological and Economic Development of Economy, 17*(2), 397–427. https://doi.org/10.3846/20294913.2011.593291

Zhang, H., Dong, Y., Chiclana, F., & Yu, S. (2019). Consensus efficiency in group decision making: A comprehensive comparative study and its optimal design. *European Journal of Operational Research, 275*(2), 580–598. https://doi.org/10.1016/j.ejor.2018.11.052