Reducing inconsistency in AHP by combining Delphi and Nudge theory and network analysis of the judgements: an application to future scenarios

Simone Di Zio

1. Introduction

The Delphi is a widely used method for collecting data from panels of experts (Dalkey and Helmer, 1963) and its key characteristics are: anonymity, interaction, controlled feedback, and statistical aggregation of responses (Rowe and Wright, 1999), while the main goal is reaching a consensus among the panel members on the issue dealt with (Linstone and Turoff, 2011). Another well-known and widely spread method in the context of decision making is the Analytic Hierarchy Process (AHP), a Multi-Criteria Decision-Making (MCDM) method designed to solve problems containing multiple conflicting criteria (Pirdashti et al., 2011). Developed by Thomas Saaty (Saaty, 1980), it has many advisable properties, such as the combination of subjective aspects, the chance of integrating objective and subjective data, and a way to combine individual and group priorities.

As far as we know, no study takes advantage of the Delphi features for reducing the inconsistency in the AHP matrices, a known problem but practically inevitable, given that it is mostly the product of cognitive biases (Bonaccorsi et al., 2020). In case of high inconsistency, generally experts are asked to evaluate again the AHP matrices, but no expert likes to give again judgements because the first ones are inconsistent, which basically means wrong. Furthermore, even if they accept, there are no guarantees that the new judgements are less inconsistent. Our proposal is to exploit the Nudge theory, which proposes suggestions to influence the behaviour of groups involved in a decision-making process (Thaler and Sunstein, 2008). A Nudge is known as a "gentle push" to make better choices which, in our context, means more consistent evaluations. In this paper we propose a new method that exploits a combination of the Delphi method and the Nudge theory to reduce the inconsistency of the AHP matrices. The method has several advantages. In addition to reducing inconsistency, it allows the collection of textual material (expert comments), a valuable data in any decision-making context. A function of the inconsistency is used as stopping criterion of the Delphi rounds. Given the Delphi logic, the participants know from the beginning that they will be reconsulted, therefore they do not feel scrutinized or pressured and they are never told that their judgments are inconsistent. This, at least in principle, ensure freer and more sincere participation and a more willing attitude to evaluate again the judgments. Finally, since at each Delphi round only the matrices with the highest inconsistency values are sent back to the experts, round after round the length of the questionnaire diminishes, and this help in reducing the dropout.

In the next sections we provide an overview of the AHP method, while section 3 shows how Nudge theory can help in reducing the inconsistency of the AHP matrices. Sections 4 presents a case study and finally the paper ends with some concluding remarks.

2. The inconsistency of the AHP matrices

The Analytic Hierarchy Process (AHP) is a general theory of measurement, useful to derive ratio scales for multi-criteria decision problems, suitable when the decision problem is complex and ill-structured. The decision factors are organized in a hierarchical structure where *criteria* and *alternatives* are compared pairwise using the Saaty scale (Saaty 1980). The goal is to find a set of weights ($W_1, W_2, ...$) for each level of the hierarchy (called *local weights*) and, from these, a vector

Simone Di Zio, Gabriele d'Annunzio University, Italy, s.dizio@unich.it, 0000-0002-9139-1451

of *global weights* $(GW_1, GW_2, \dots, GW_N)$ representing a rating of the alternatives in achieving the decision problem (N denotes the number of alternatives). The AHP can be adapted to group decisions (group AHP), and there are two families of methods for the combination of the individual preferences (Ossadnik et al., 2016), known as the Aggregation of Individual Judgements - AIJ - and the Aggregation of Individual Priorities - AIP - (Wu et al, 2008). Anyhow, no technique considers the variability in the distribution of responses, so that AIJ and AIP approaches do not take account of the degree of consensus/dissensus among participants, a fundamental issue in a group decision setting (Pirdashti et al., 2011). A voting procedure can overcome these limitations (Lai et al. 2002), but a majority vote is a "winner-take-all" system, where the opinions of the losers are completely disregarded (Di Zio and Maretti, 2014). This is why some scholars have proposed an integration of the AHP and the Delphi (Tavana et al. 1993; Di Zio and Maretti, 2014) which gives the Delphi the task of structuring a convergence towards a single solution shared by all.

Given a_{ij}, the pairwise judgement of alternatives i and j, for a perfect consistent matrix we should have $W_i/W_j = a_{ij}$ ($\forall i, j$) and $a_{ij} = a_{ih} \cdot a_{hj}$ ($\forall i, j, h$) but human judgements are never perfectly consistent and in practical applications the equalities do not occur. Inconsistency in expert judgments has been observed in many fields and, for lack of space, we refer the reader to the vast specialized literature. In short, inconsistency is practically inevitable, because it is the product of cognitive biases (Bonaccorsi et al., 2020) and/or problem complexity. Consequently, there is a need to check the consistency through the calculation of a consistency index. The Consistency Ratio (CR), is the most common index used to check for consistency (Brunelli, 2018), calculated as $CR = CI/RI$, where $CI = (\lambda_{max} - n)/(n - 1)$, λ_{max} is the maximum eigenvalue of the matrix and RI (the random index) is the average of the CIs calculated over many random square matrices, reciprocal and positive. As a rule of thumb, introduced by Saaty (1980), if $CR \leq 0.1$ the judgements of a matrix can be considered consistent, otherwise the matrices must be reviewed by the expert (Liao, 2010), as many times until to have $CR \leq 0.1$. The critical point is going back and stress the expert telling him/her that he/she made a wrong evaluation that needs to be revised.

All that being said, the reduction of the inconsistency in the AHP method remains an open issue, and here we propose a new approach which involves asking the experts for new evaluations according to the Delphi logic, in a structured and iterative procedure that, by means of *nudges*, gently push them towards more consistent solutions.

3. Reducing the inconsistency by combining the Delphi and the Nudge theory

Although it still has open issues (Pill, 1971) - such as how to choose the experts, how many experts to include in the panel or how to measure the expertise - the Delphi is a method that offers undoubted advantages in the context of group decisions. In the Delphi-AHP the experts are consulted more than once, and starting from the second round, for each AHP matrix, we propose to give a nudge as feedback. By using a "nudge approach" we obtain both a reduction of the inconsistency of the AHP matrices and the elimination of the problem of choosing an aggregation method. After the first round (time t_1) we get $R + 1$ matrices for each expert. With $\mathbf{A}_m^{1,t_1}$ we denote the 3D array containing the $N \times N$ pairwise comparison matrices according to the first criterion, at time t_1, where $m = m_1, m_2, \dots, m_M$ denotes the expert and M the cardinality of the panel. Since each participant give $s = N(N - 1)/2$ judgements, we have s vectors of size M. For the first criterion, the vector of the generic pairwise comparison (i, j) is $\left[a_{ij,m_1}^{1,t_1}, a_{ij,m_2}^{1,t_1}, \dots, a_{ij,m_M}^{1,t_1} \right]$. To synthetize these judgements, we use the median (other syntheses are possible) and as a result, we obtain a matrix representing the judgments of the whole panel after the first round, say $\mathbf{A}_{med}^{1,t_1}$. On this matrix we calculate the consistency ratio CR_{med}^{1,t_1} and by using the 17 values of the Saaty scale $(1/9, 1/8, \dots, 8, 9)$ we replace the first element of $\mathbf{A}_{med}^{1,t_1}$, namely $a_{12,med}^{1,t_1}$ (as well as its symmetric $a_{21,med}^{1,t_1} = 1/a_{12,med}^{1,t_1}$) obtaining 17 different matrices. On each resulting matrix we calculate the

consistency ratio among which we find the smallest - say $CR_{12,best}^{1,t_1}$. This value is the result of a specific value of the Saaty scale, say $V_{12,best}^{1,t_1}$. This figure represents the theoretical assessment which, for the cell (1,2), gives the best consistency of the matrix, given all the other values. By repeating the same search for the upper triangular of the matrix, we obtain s different "best CRs" among which we find the smallest, that represents the "best of the bests": $CR_*^{1,t_1} = min\{CR_{ij,best}^{1,t_1} : i = 1, ..., n; j > i\}$. We denote the position of this value with (i_{best}, j_{best}) and its corresponding value of the Saaty scale with SA_{nudge} (actually our nudge), that is the judgement that most improves the consistency of the matrix.

In the second round of consultation, the panel is invited to judge each a_{ij}^r (and c_{lm}) inside the proposed interquartile range - $IQR_{i,j} = [Q1_{ij}^{t_1}; Q3_{ij}^{t_1}]$ - where $Q1_{ij}^{t_1}$ and $Q3_{ij}^{t_1}$ are, respectively, the first and third quartile in the distribution of judgements in the cell i, j (Di Zio and Maretti, 2014). The quantity $\Delta = IQR_{i_{best}, j_{best}}/2$ is used to create a symmetric interval around SA_{nudge}. For the judgement in position (i_{best}, j_{best}), in the second round, instead of the IQR, the interval proposed to the panel is $[SA_{nudge} - \Delta; SA_{nudge} + \Delta]$. Therefore, among the s proposed intervals $s - 1$ are $IQRs$ but one is a Nudge which gently pushes the respondents towards a more consistent matrix. The same process applies to all the matrices of the hierarchy and the procedure is repeated iteratively in the following rounds. If the consensus triggers, there will be a progressive reduction of the consistency ratios: $CR_{med}^{r,t_1} \geq CR_{med}^{r,t_2} \geq CR_{med}^{r,t_3} \geq \cdots$.

This method has several advantages. The aggregation of judgements is managed optimally, by considering the degree of consensus, and this reduces, at least in principle, the dropout rate. Simultaneously, we reduce the inconsistency of the matrices in a gentle way, because there are no pressures on the participants. The experts do not perceive any kind of "mistake message" and are softly driven to revise their judgements. A right nudge, in the AHP context, "pushes gently" the participants to more consistent judgements. So, the method stimulates consensus and reduces inconsistency at the same time.

The rule to stop the Delphi iterations is twofold. To make the benefits of the Delphi at least two rounds must be performed, therefore the first stopping criterion is $i \geq 2$ (here i denotes the rounds). During the rounds we have, for each matrix, a sequence of Consistency Ratios $CR^{r,t_1}, CR^{r,t_2}, ..., CR^{r,t_i}$ (here $r = 1, 2, ..., R + 1$ and we removed the subscript med to simplify) and the second stopping criterion is that at least one CR in the sequence is less or equal than 0.1. After the round t_2, for the matrix r, we have four possible cases. 1) $CR^{r,t_1} > 0.1$ and $CR^{r,t_2} \leq 0.1$; the Delphi for the matrix r stops and as result we take the matrix coming from the second round: $\mathbf{A}_{med}^{r,t_2}$. 2) $CR^{r,t_1} \leq 0.1$ and $CR^{r,t_2} > 0.1$; the Delphi stops, but the matrix we take is $\mathbf{A}_{med}^{r,t_1}$. 3) $CR^{r,t_1} \leq 0.1$ and $CR^{r,t_2} \leq 0.1$; the Delphi stops, and we choose between $\mathbf{A}_{med}^{r,t_1}$ and $\mathbf{A}_{med}^{r,t_2}$ the matrix with the lowest inconsistency. 4) $CR^{r,t_1} > 0.1$ and $CR^{r,t_2} > 0.1$; in this last case only the condition $i \geq 2$ holds, therefore the Delphi continues. This double-stopping criterion is appropriate because after a reduction of the inconsistency, there is no guarantee that continuing the rounds the CR decreases monotonically. If after k rounds, for a matrix no index in the sequence is less than 0.1, we suggest the following solution: take the round z such that CR^{r,t_z} is the minimum and hold the matrix used for the calculation of the intervals for the round $z + 1$. This matrix, by definition, has $CR_*^{r,t_z} < CR^{r,t_z}$. Since the above algorithm applies to each matrix of the hierarchy, it may happen that for one matrix only two rounds are necessary while for another matrix we can have three or more rounds. The advantage is that the length of the AHP questionnaires reduces during the rounds. The result of the method is a vector of global weights, with lower levels of inconsistency in the pairwise matrices than the classic AHP.

4. Application on four future scenarios with network analysis

We applied the proposed method in the evaluation of four future scenarios on the genetic modification experiments. It is called CRISPR the new technology that allows splicing of DNA molecules, and in the future, it could allow human selection of characteristics of children, including escaping of many diseases. The ethics of this technology is obviously questionable. Starting from these considerations, Theodore J. Gordon (one of the fathers of the Delphi method) sketched four brief future scenarios on CRISPR technology. For lack of space, we do not report the scenarios in full but only their titles: Scenario A. *Genetic tech self-regulation*; Scenario B. *Genetic tech external control*; Scenario C. *Genetic tech uncontrolled*; Scenario D. *Genetic tech downside*. Each scenario represents an alternative of the AHP hierarchy.

Following Gordon and Glenn (2018), the main factors measuring the usefulness of a scenario, and which here constitutes the criteria of the AHP, are the following: *Plausibility*: the paths to the futures must be seen as feasible and may not be viewed as impossible. *Consistency*: the paths to the futures and the resulting images must not be mutually contradictory. *Simplicity*: a good scenario describes paths to the future scenario in a way that is easily understood. Therefore, we had 4 alternatives (the scenarios) and 3 criteria, and we wanted to find a ranking of importance of the scenarios according to these criteria. The survey was performed on Alchemer (https://app.alchemer.com) where each pairwise comparison was built on a radio button that reproduces the whole Saaty scale. This avoided that the respondents fill in the matrices, in general a complicated task for non AHP-experts.

The panel consisted of 26 experts, recruited around the world, diversified according to age, gender, expertise and employment, and having skills both in the field of futures studies and genetics. For each round they gave 21 pairwise judgements and voluntary comments. For each round and for each matrix (A^1, A^2, A^3, C) we obtained the consistency ratios reported in Table 1.

Table 1. Consistency Ratios along three Delphi rounds

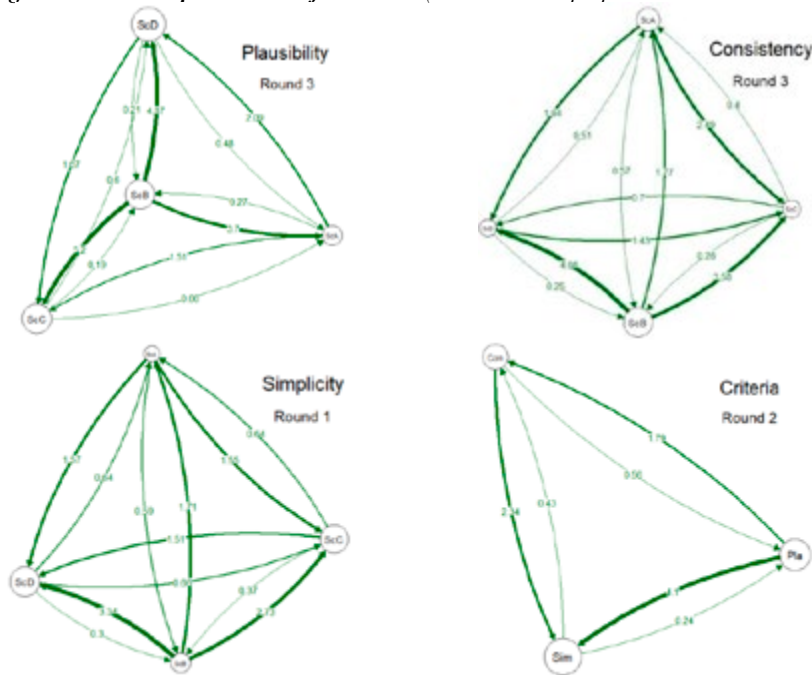| Round	Plausibility	Consistency	Simplicity	Criteria
t_1	0.0936	0.0035	**0.0195**	0.1591
t_2	0.0514	0.0087	0.0225	**0.0056**
t_3	**0.0366**	**0.0018**	0.0295	0.1850

For the calculation of the local and global weights we take the matrix resulting from the last round for *plausibility* ($CR^{1,t_3} = 0.0366$) and *consistency* ($CR^{2,t_3} = 0.0018$). For the criterion *simplicity* the best value derives from the first round ($CR^{3,t_1} = 0.0195$) and for the comparison of the three criteria we take the matrix coming from the second round ($CR^{C,t_2} = 0.0056$). In all cases the values are very good, being all well below the 0.1 threshold. The result consists of a vector of global weights, which quantifies the relative importance of each future scenario. The best scenario, according to the panel of experts, is scenario B, Genetic tech external control ($GW_B = 0.52$). It follows scenario A, Genetic tech self-regulation ($GW_A = 0.28$), and scenario D, Genetic tech downside ($GW_D = 0.10$). The last is scenario C, Genetic tech uncontrolled ($GW_C = 0.09$). About the local weights of the criteria, the experts considered *plausibility* as the most important criterion ($W_{Pla} = 0.47$). Following we have *consistency* ($W_{Con} = 0.43$) and *simplicity* ($W_{Sim} = 0.10$).

After that, we explored the network structure of the scenarios and criteria. A network refers to a structure representing a group of objects and relationships between them, and its mathematical representation is a graph, which consists of nodes and edges. Since each scenario/criterion is linked to the others through a preference ratio, it is useful to represent the results of the AHP through weighted direct graphs, in which the nodes are the scenarios/criteria and the edges are proportional to the geometric mean (or median) of the judgments provided by the experts. By considering the matrices with the lowest CRs (Table 1, bold digits) we obtained the four digraphs of Figure 1.

From each graph emerges, with a single glance, the whole structure of the preferences expressed by the panel in comparing the future scenarios under each criterion and the criteria, as well as the

structure of relationships between the scenarios/criteria, with evident advantages over the representation through matrices. Also, we can build a network for each expert and, even more interesting, we can consider each expert as a layer in a multiplex network, that is a network in which the same set of nodes are connected via more than one type of links (Kyu-Min et. al, 2015). Besides, we can consider each criterion as a layer, to study the interactions between scenarios and criteria, or even each Delphi round as a layer, to explore, within each criterion, the interactions between scenarios and rounds. In short, there are many possibilities to represent and analyse the outputs of a Delphi-AHP through the Network Analysis. So, we can study whether scenarios behave similarly across experts, across Delphi rounds or across criteria. Hence, this is not only a way of visualizing the results but a statistical tool for modelling the Delphi-AHP data in a way that to highlight the structure of relationships between experts, scenarios, criteria and Delphi rounds.

Figure 1. Network representation of the results (nodes sizes are proportional to the closeness)

To give a taste of the measures that can be computed, we calculated the closeness (CL) of the networks of the Figure 1 (where nodes sizes are proportional to CL), which gives information on how close a scenario (or a criterion) to all the others is. Plausibility network: $CL_A = 0.549$, $CL_B = 0.800$, $CL_C = 0.799$, $CL_D = 0.925$. Consistency network: $CL_A = 0.627$, $CL_B = 0.779$, $CL_C = 0.474$, $CL_D = 0.457$. Simplicity network: $CL_A = 0.335$, $CL_B = 0.422$, $CL_C = 0.597$, $CL_D = 0.626$. Criteria network: $CL_{PLA} = 1.246$, $CL_{CON} = 1.014$, $CL_{SIM} = 1.488$. Scenario D is the "closest" to the others under plausibility and simplicity criteria, while under consistency the scenario with major closeness is B, and simplicity is the criterion with the higher closeness.

5. Concluding remarks

We have introduced a new method to use the Delphi method to nudge responses of participants toward better consistency in the AHP pairwise comparison matrices. The network analysis helps to depict the structure of interactions between alternatives and criteria of the AHP hierarchy. We applied the method for the evaluation of four future scenarios, dealing the management of genetic modification technologies. The study confirmed quite well the research hypothesis, since the inconsistency in all the AHP matrices remained under, or dropped below, the threshold of 0.1.

Although the rounds of the Delphi must stop after the second round, in the application we performed three rounds for all the matrices, to explore all the potentialities of the method. The method removes the problem of choosing an aggregation method of the individual judgements, because the Delphi produces a convergence toward a synthesis of the evaluations which includes all points of view, even the extremes or the minority ones. By using a multiplex network approach, the structure of relationships between experts, scenarios, criteria and Delphi rounds can be studied.

As future developments we can think of the graph representation as a tool to be included in the Delphi questionnaires, which help to visualize in real time the answers that each participant gives. Also, when considering each expert as a layer of a multiplex network, the similarity measures between layers could be exploited to explore new measures of consensus in the Delphi method and new ways of aggregating the individual judgements could be also studied.

References

Bonaccorsi, A., Apreda R., Fantoni, G. (2020). Expert biases in technology foresight. Why they are a problem and how to mitigate them. *Technological Forecasting & Social Change*, **151**, pp. 1-17.

Brunelli, N. (2018). A survey of inconsistency indices for pairwise comparisons. *International Journal of General Systems*, **47**(8), pp. 751-771.

Dalkey, N.C., Helmer, O. (1963). An Experimental Application of Delphi Method to the Use of Experts. *Management Science*, **9**, pp. 458-467.

Di Zio, S., Maretti, M. (2014). Acceptability of energy sources using an integration of the Delphi method and the analytic hierarchy process. *Quality & Quantity*, **47**(6), pp. 2973–2991.

Gordon, T.J., Glenn, J.C. (2018). Interactive Scenarios, in *Innovative Research Methodologies in Management, Volume II: Futures, Biometrics and Neuroscience Research*, eds. L. Moutinho and M. Sokele, Palgrave Macmillan, London (UK).

Kyu-Min, L., Byungjoon, M., Kwang-Il, G. (2015). Towards real-world complexity: an introduction to multiplex networks. European Physical Journal B, **88**(2), pp. 1-20.

Lai, V.S., Wong, B.K., Cheung, W. (2002). Group decision making in a multiple criteria environment: A case using the AHP in software selection. *European Journal Operational Res.*, **137**, pp. 134-144.

Liao, C.N., (2010). Supplier selection project using an integrated Delphi, AHP and Taguchi loss function. *ProbStat Forum*, **3**, pp. 118-134.

Linstone, H.A., Turoff M. (2011). Delphi: A brief look backward and forward. *Technological Forecasting & Social Change*, **78**(9), pp. 1712-1719.

Ossadnik, W., Schinke, S., Kaspar, R.H. (2016). Group Aggregation Techniques for Analytic Hierarchy Process and Analytic Network Process: A Comparative Analysis. *Group Dec. and Negotiation*, **25**, pp. 421-457.

Pill, J. (1971). The Delphi method: Substance, context, a critique and an annotated bibliography. *Socio-Economic Planning Sciences*, **5**(1), pp. 57-71.

Pirdashti, M., Omidi M., Pirdashti H., Hassim M.H. (2011). An AHP-Delphi Multi-Criteria Decision Making Model with Application to Environmental Decision-Making. *Iranian Journal of Chemical Engineering*, **8**(2), pp. 3-17.

Rowe, G. and Wright, G. (1999). The Delphi technique as a forecasting tool: issues and analysis, *International Journal of Forecasting*, **15**(4), pp. 353-375.

Saaty, T.L. (1980). *The Analytic Hierarchy Process*. McGraw-Hill, New York (NY).

Tavana, M., Kennedy, D.T., Rappaport, J., Ugras, Y.J. (1993). An AHP-Delphi Group Decision Support System applied to conflict resolution in hiring decisions. *Journal of Management Systems,* **5**(1), pp. 49-74.

Thaler, R. and Sunstein, C.R. (2008). *Nudge. Improving decisions about health, wealth and happiness*. Yale University Press, New Haven (CT).

Wu, W.H., Chiang, C.T., Lin, C.T. (2008). Comparing the aggregation methods in the analytic hierarchy process when uniform distribution. *Transactions on Business and Economics*, **3**(5), pp. 82-87.