scientific reports

OPEN

Integrating regression and multiobjective optimization techniques to analyze scientific perception

Sandra González-Gallardo[1], María Isabel Sánchez-Rodríguez[2]✉, Ana B. Ruiz[1] & Mariano Luque[1]

Science holds high prestige in society and understanding public perception of what is considered scientific is essential. The *scientificity* of a profession is the degree of scientific legitimacy and is determined by the quality of its scientific procedures. Higher levels of scientificity are achieved when scientific results are more objective, impartial, and neutral. In this work, we first estimate the scientificity levels attributed to various professions using a logistic regression model. Then, we explore ways to simultaneously improve their scientific perception by means of multiobjective optimization techniques. To this aim, the statistical results are used to formulate a multiobjective optimization model that maximizes the scientific perception of all the professions considered. The findings provide insights into science policy measures to optimize resource allocation in order to increase the scientific perception of the professions.

Keywords Scientific perception, Science policy, Logistic regression, Multiple criteria decision making

Science is a powerful source of authority in the society. Individuals, organizations, and science politicians frequently make use of scientific evidence and experts to provide information and advice regarding a wide variety of issues. A strong research system needs the support of society, that is, educated citizens who appreciate, are interested and engaged in science, since they are called to play an important role in the decision-making that accompanies scientific research and innovation. In consequence, we need to understand the dynamics underlying who or what is perceived to be "scientific". This is especially important in science policy, whose main concern is the design of policies enabling scientific knowledge to effectively respond to people's needs and the general public to effectively contribute to science.

In this work, we focus on the study of the *scientificity* attributed to various professions. Theoretically, *scientificity* is defined as the degree of scientific legitimacy and can be shown as a measure of the quality of scientific work. Our work is mainly based on the estimation of the scientificity of various professions, i.e. the perception of how scientific the different professions are. The level of acceptance of each discipline is closely related to the extent to which the related professions and practices are seen or perceived as "real" sciences; in other words, it depends on the assumption that it has attributes identified with science. This degree of scientificity may condition many decisions in science policy, such as the justification for public financial support for scientific institutions, the educational and job choices of students and young graduates, and something as important as the transfer of knowledge to society[1].

In the literature, several proxies have been used to analyze the perception of science, such as interest in science[2], knowledge and risk attributed to scientific findings[3], or trust in science[4]. However, measuring the scientificity level of a discipline is not easy, mainly because science is often seen as a homogeneous reality, while there are many differences among the wide variety of scientific disciplines.

Nevertheless, the public already has a certain predetermined image of some disciplines. On the one hand, social sciences are systematically undervalued[5], especially when compared to the so-called STEM disciplines (Science, Technology, Engineering, and Mathematics). Indeed, some social sciences may be considered activities of a political nature (for example, economics). On occasions, this association may lead to transferring the perception of politics to these social sciences, which may have a negative effect in contexts where there is

[1]Department of Applied Economics (Mathematics), Universidad de Málaga, C/ Ejido s/n, 29013 Málaga, Spain. [2]Department of Statistics and Business, Faculty of Law and Business, University of Córdoba, Av/ Puerta Nueva s/n, 14002 Córdoba, Spain. ✉email: td1sarom@uco.es

a lack of confidence in the prestige of politicians and political institutions. On the other hand, the so-called pseudosciences constitute another special case. With an ample social presence, they entail complementary or alternative therapies (homeopathy, acupuncture, and some forms of meditation, among others) that have not been validated by scientific methods but are often described as if they were branches of medicine. They present themselves as scientific in appearance by using institutions and working styles similar to those of science. This means that, in terms of perception, many people also consider them as science[6,7].

In this paper, we study the scientific perception of a set of professions in Spain by exploring the degree of scientificity that Spanish society attributes to them. Our main purpose is to analyze the possible causes that influence the level of scientific perception assigned to the following professions: doctor of medicine, physicist, psychologist, sociologist, economist, physiotherapist, and homeopath. Indeed, we want to determine the extent to which the professions closer to the social sciences (or even pseudosciences) are appreciated as much as the ones that are commonly perceived as being more scientific. In addition, we intend to study whether the relevant (independent) variables in each case coincide or not, and if they do, how relevant they are for each profession.

Research on the perception of science has identified a wide range of cognitive, informational, sociodemographic, and cultural determinants[4,6,8–11]. In our study, the independent (or explanatory) variables used as predictors to measure the perception of professions' scientificity are related to aspects such as scientific culture, politics, education, and sociodemographic characteristics. A first group of variables is related to educational aspects and scientific culture. This includes the degree of support and exposure to scientific practices, which act as mechanisms of socialization and learning. These factors help to discriminate the status of sciences, although they can also introduce higher levels of criticism. Scientific education and active practices of consuming scientific information are important predictors of support for science[9,12]. Greater contact and familiarity with science predispose people to have a greater knowledge and appreciation of the diversity of scientific disciplines and practices. Another set of variables is associated with access to information[10] and trust in political and social institutions[13]. Finally, we have also considered several sociodemographic variables: gender[4], income level[2], political ideology[14] and religion[15]. In short, the social mechanisms that function in attributing social acceptability and support for science can be applied with the same logic to the perception of science. Regression logistic models are obtained to estimate the perception of professions' scientificity on these explanatory variables and to gain an understanding of the relative impact of these predictors and their level of significance. The observational data used was taken from an official survey carried out in 2018 by the Spanish Foundation for Science and Technology (FECYT), whose purpose is to get information about the society's perception in Spain of science and technology (S&T). The questionnaire, the raw data and the reports corresponding to the survey are public and permanently accessible from the FECYT website: https://www.fecyt.es/es/noticia/encuestas-de-percepcion-social-de-la-ciencia-y-la-tecnologia-en-espana.

Besides learning about the aspects that are decisive for each profession, our study also aims to determine if it is possible to improve the scientific perception of the professions considered and how to reach a compromise improvement situation among them in the future. For this, several regression models are built so that it is necessary to optimize several objective functions simultaneously. In such a situation, the application of multiobjective optimization techniques (in conjunction with the previously applied statistical techniques) can help us obtain very valuable information. In general, *multiobjective optimization* deals with problems where several objective functions must be optimized (maximized or minimized). Since it is usually impossible to find a solution that optimizes all the objective functions, we can find several Pareto optimal or efficient solutions reflecting different tradeoffs. At them, no objective function can be improved without worsening, at least, one of the others[16].

Therefore, we propose to build a multiobjective optimization problem to maximize the scientificity's perception of the seven professions considered to determine the extent to which each profession can increase its scientificity. The objective functions (i.e. the perception of professions' scientificity) are formulated according to the regression models obtained in the statistical analysis. To ensure only realistic values for the variables (according to the data used), we include a set of constraints defined based on a correlation analysis performed between the independent variables.

The methodology followed in this paper, which combines statistical and multiobjective optimization methodologies, is based on the theoretical framework proposed by Luque et al.[18]. This framework has been previously applied in different contexts: to study a macroeconomic policy problem in Finland[19]; to analyze the satisfaction level of Spanish workers[20,21]; in the economy of education field[22,23]; or to study the profile of the most profitable insurers[24,25]. Besides, other authors[26] examine and assess the efficiency of several feature selection algorithms and classification models (such as logistic regression) to find out the features that are the most important for determining the quality of the apple. Other applications of multiobjective optimization in social science can be found in other studies[27,28].

To the best of our knowledge, our study is the first one to apply a mix of statistical and multiobjective optimization techniques in a science polity context, to analyze the scientific perception of a set of professions. Indeed, as said, we regress the perception of the scientificity level of the professions using a logistic model. In comparison to the previously mentioned works, this logit statistical regression model is the first time used when using the framework suggested in Luque et al.[18]. In addition, a novelty of the formulated multiobjective optimization problem is that it considers all the information obtained when estimating the scientificity's perception according to the explanatory variables. To be more precise, we build the objective functions representing the perception of the scientificity assigned to each profession using the regression coefficients of all the variables considered regardless of their significance, but we weight these coefficients according to their statistical significance level to obtain a model adjusted as best as possible to the observational data used. It should be noted that these weights are not preferential and they are just included in the formulation to reflect the significance of each variable coefficient. In summary, the contributions of the paper are:

- It is the first time these quantitative tools are used simultaneously, statistical techniques and multiobjective optimization, to analyze the perception of scientificity by citizens.
- It is also the first time a logit regression is used to estimate the multiobjective optimization model, in particular, to formulate the objective functions.
- It is the first time that the regression coefficients are weighted taking into account their statistical significance in the multiobjective optimization model.In what follows, firstly we describe the data used and the regression analysis performed to estimate the perception of scientificity level of the seven professions. Next, the multiobjective optimization study allowing us to obtain a better insight into the real perception improvements that are possible for the different professions and how to achieve them is shown. Finally, we discuss the results obtained and lastly, the main contributions of this paper are summarized.

Statistical analysis

Survey: questionnaire, data and variables

Since 2002, FECYT has carried out a survey every two years, to learn more about the relationship between science, technology, and society in Spain and analyze the public's perception of scientific and technological advances and their capacity to improve the quality of life. In particular, our study is based on data corresponding to the *Social perception of Science and Technology Questionnaire 2018* (https://www.fecyt.es/es/noticia/encuestas-de-percepcion-social-de-la-ciencia-y-la-tecnologia-en-espana), as the subsequent ones do not include some variables identified as relevant in this study.

Statistical information was collected through a semi-structured questionnaire, administered through personal and home interviews of about 25 min. The population under study consisted of people who were at least 15 years old and had been residents of Spain for a minimum period of 5 years. A multistage, stratified sampling was used, with proportional random selection of primary (municipality) and secondary (sections) units, and the selection of sampling units (individuals) by using random routes and gender and age quotas. Therefore, the sample size was 5,200, proportional to the population of each one of the 17 Spanish autonomous communities (In Spain, an autonomous community is a territorial entity with its institutions and representatives and certain legislative, executive, and administrative powers.), which takes into account the size of the habitat, weighted concerning its real population. This ensures that all relevant subpopulations are adequately represented, minimizing biases and enhancing representativeness. The resulting sampling error (For a certain sample size n and confidence level $1-\alpha$, the sampling error in the estimation of a proportion p is given by $\epsilon=\sqrt{(pqz^2_{\alpha/2})/n}$, being $q=1-p$ and $z_{\alpha/2}$ the $(\alpha/2)$-quantile of the Normal distribution.) was ±1.36 percentage points for a confidence level of 95.5%, for a two-sided interval, $p=q=0.5$, and considering simple random sampling and non-proportional samples.

The questionnaire structure includes information on the following items: (i) interest in scientific and technological topics; (ii) social image of science; (iii) social image of the scientific profession; (iv) scientific literacy; (v) science and technology and media; (vi) policies to support science and technology; (vii) health and science; (viii) attitudes and positions of citizens; (ix) segmentation of the population regarding its position towards science and technology based on sociodemographic characteristics.

In the questionnaire, participants were asked how scientific they believed a set of different professions were in question Q.8.B: *People may have different opinions about what is scientific and what is not. I'll read out a list of professions. For each one of them, please tell me to what degree you think they are scientific, using a scale from 1 to 5, where 1 means they are 'not scientific at all' and 5 means you think they are 'very scientific'. With the intermediate numbers, you can nuance your answer*. Given that our objective is to study the scientificity level attributed to different professions in the fields of science, social science, and pseudoscience in Spain, we used the answers to this question to obtain a dichotomous variable for each of the following professions: medical doctor, physicist, psychologist, physiotherapist, sociologist, economist, and homeopath. That is, these dichotomous variables were obtained from the answers to Q.8.B (measured on a 1-5 Likert scale) about the degree of scientificity attributed to these professions (Considering a Likert-scaled variable as the dependent variable in an ordinary least squares regression model can lead to a violation of the assumption of normality of errors or make it difficult to interpret the marginal effect of each independent variable on the dependent variable. For this reason, a dichotomous 0-1 variable is considered as the dependent variable, and a binary logistic regression model is fit instead, with maximum likelihood estimation.). To achieve an equal distribution of frequencies, scores 1, 2, and 3 of the Likert scale were re-coded as 0 in the dichotomous variable of each profession, and scores 4 and 5 were re-coded as 1. This implies that, the closer the value of the dichotomous variable is to 1, the more scientific the corresponding profession is perceived to be. In our statistical study, we considered the seven dichotomous variables obtained (for doctor of medicine, physicist, psychologist, physiotherapist, sociologist, economist, and homeopath) as the dependent variables.

On the other hand, the independent variables used were obtained from different items in the questionnaire and were grouped into several categories, namely scientific culture, politics, education, and sociodemographic characteristics. The names of the independent variables, their corresponding question reference in the questionnaire, their mathematical notation, and their descriptions can be seen in Table 1. In all cases, the NR/DK (no response/don't know) values have been filtered. As shown, some variables corresponded to a non-numerical or categorical question of the questionnaire. If the answers to the question could take m different categories, $m-1$ dichotomous 0/1 variables, named as *dummy* variables, were introduced to control the selected category. Thus, each dummy variable represented the selection or not (taking the value 1 or 0, respectively) of each category in the original non-numerical question. One category was always used as a reference (and, therefore, no dummy variables were created for this category).

Variable	Question reference	Notation	Description, reference category and dummy variables
Scientific culture			
Media to be informed about S&T	Q.10		Reference category: Internet.
			Dummy variables
		x_1	Books.
		x_2	Radio.
		x_3	Other means (TV, scientific and technical journals, written press,...).
Opinion about S&T	Q.13		Reference category: Benefits of S&T greater than damages.
			Dummy variables
		x_4	Benefits of S&T equal to damages.
		x_5	Benefits of S&T less than damages.
Successful statements about S&T	Q.24	x_6	Number of correct answers from the six pairs of sentences about S&T knowledge (ranging from 0 to 6)
Willingness to financially support science	Q.28		Reference category: Financial support for science
			Dummy variables:
		x_7	No financial support for science
		x_8	No possibilities to support science
Scores to scientific techniques	Q.17	x_9	Numerical continuous variable obtained from the average of the scientific level scores attributed to vaccines and chemotherapy
Scores to social science techniques	Q.17	x_{10}	Numerical continuous variable obtained from the average of the scientific level scores attributed to public opinion polls and economic growth forecast
Scores to pseudoscience techniques	Q.17	x_{11}	Numerical continuous variable obtained from the average of the scientific level scores attributed to homeopathy and acupuncture
Politics			
Political ideology	D.4	x_{12}	Numerical continuous variable, ranging from 1 to 10, where 1 means far left and 10 means far right.
Education			
Educational level	D.5		Reference category: Until first level studies
			Dummy variables:
		x_{13}	Second level studies
		x_{14}	University studies
Level of S&T knowledge	Q.25		Reference category: very low/low/normal S&T knowledge
			Dummy variable:
		x_{15}	High/very high S&T knowledge
Sociodemographic characteristics			
Religion	D.8		Reference category: catholic
			Dummy variables:
		x_{16}	Non-catholic religion
		x_{17}	Atheist, agnostic or indifferent
Employment situation	D.10.A		Reference category: Inactive, students and unemployed
	D.10.B		Dummy variables:
		x_{18}	Labourers
		x_{19}	Small entrepreneurs, technicians and middle managers
		x_{20}	Directors and professionals
Gender	D.1		Reference category: Man
			Dummy variable
		x_{21}	Woman
Income (st)	D.9.B	x_{22}	Numerical variable obtained from the interval class marks of net family income, with the values standardized (st) to eliminate the effect of measurement units

Table 1. Independent variables considered in the statistical study, obtained from the Social perception of Science and Technology Questionnaire, 2018.

Logistic regression analysis

Background and notation

In general, statistical causal models are used to explain the relationship between a dependent variable, denoted by *y*, from other explanatory or independent variables, referred to as $x_1, \ldots, x_n$, as follows:

$$y = h(x_1, \ldots, x_n) + \epsilon, \tag{1}$$

where the term ϵ, known as *error*, contains the variability of the dependent variable not explained by the independent variables, and $h : \mathbb{R}^n \to \mathbb{R}$ is the function considered to estimate the dependent variable on the independent ones. There are different types of causal models depending on the initial hypotheses assumed about variable distributions (for example, normality or not), the nature of the variables involved (numerical or not), and the type of function *h* considered to regress the dependent variable on the independent ones, which may be linear or not.

In particular, in our study, we have applied a binary logistic regression model. This model considers a dichotomous non-numeric variable as a dependent variable (normally, with "no/yes" categories coded as 0/1 values), and a non-linear function with the following expression is assumed:

$$y = \frac{e^{(\alpha+\beta_1 x_1+\cdots+\beta_n x_n)}}{1+e^{(\alpha+\beta_1 x_1+\cdots+\beta_n x_n)}} + \epsilon = \frac{1}{1+e^{-(\alpha+\beta_1 x_1+\cdots+\beta_n x_n)}} + \epsilon. \tag{2}$$

By definition, this expression takes values in the interval [0, 1] (in fact, the result is a probability). Once estimated by maximum likelihood, a negative estimation for any of the parameter β_i ($i = 1, \ldots, n$) increases the denominator of the last expression, which causes the prediction of the dependent variable to approach the zero value (meaning that the estimated value for the dependent variable is "no"). On the contrary, a positive estimation for a parameter β_i causes the prediction of the dependent variable to approach the one value (i.e. the predicted value for the dependent variable is "yes"). The default cut-off point, 0.5, can be modified depending on the particular characteristics of the study.

This binary logistic model allows us to analyze the effect of non-numerical or categorical independent variables (like the ones in Table 1), on a dependent dichotomous variable (like each of the seven dichotomous variables created to quantify the perception of scientificity of the professions considered). Let us recall that we have created $m - 1$ dichotomous dummy independent variables for the non-numerical questions (where *m* is the number of categories), each one representing the verification (value 1) or not (value 0) of each category, using one of them as reference. In the binary logistic model, the regression coefficients for these dummy variables must be interpreted taking into account the reference category for which no dummy variable has been introduced.

Usually, once a regression model has been fitted, it is interesting to know which independent variables are statistically relevant, that is, the ones that provide significant information to the model. For this purpose, every coefficient β_i is independently tested to determine if it is statistically different from zero or not ($H_0 : \beta_i = 0$, $H_1 : \beta_i \neq 0$). As a result, an independent variable x_i is more relevant as the p_{value} obtained in the test of its coefficient β_i is closer to zero.

Finally, the variance inflation factor (VIF) is a measure used to assess multicollinearity among independent variables in a regression model. The higher the VIF value, the greater the presence of multicollinearity. Typically, a VIF between 1 and 5 indicates a moderate (acceptable) multicollinearity and a VIF greater than 5 indicates strong multicollinearity that can affect the precision of the model.

Results

Table 2 provides the estimated coefficients $\hat{\beta}_{ij}$ and $\hat{\alpha}_j$ of the fitted binary logistic regression models obtained, which estimate the scientificity's perception of the professions considered, denoted by $\hat{y}_j$, with $j = 1$ (medical doctor), 2 (physicist), 3 (psychologist), 4 (sociologist), 5 (economist), 6 (physiotherapist) and 7 (homeopath), as functions of the independent variables x_i ($i = 1, \ldots, 22$) given in Table 1:

$$\hat{y}_j = \frac{e^{(\hat{\alpha}_j+\hat{\beta}_{1j} x_1+\cdots+\hat{\beta}_{nj} x_n)}}{1+e^{(\hat{\alpha}_j+\hat{\beta}_{1j} x_1+\cdots+\hat{\beta}_{nj} x_n)}} = \frac{1}{1+e^{-(\hat{\alpha}_j+\hat{\beta}_{1j} x_1+\cdots+\hat{\beta}_{nj} x_n)}}. \tag{3}$$

In what follows, we discuss some conclusions, mainly obtained by interpreting the sign of the statistically significant coefficients (respectively, the super-index *** means that the estimation is significant at 1%, ** significant at 5%, and * significant at 10%), taking into account the reference category for the dummy variables (when needed).

Firstly, the media used to be informed about S&T has an influence on the perception of some of the professions. We observe that people who use books to find S&T information perceive that economists and physiotherapists are less scientific than those who use the Internet. Besides, people who obtain information through the radio attribute a lower degree of scientificity to the physical profession than those who are informed using the Internet,

Variables		Medical doctor ($\hat{\beta}_{i1}$)	Physicist ($\hat{\beta}_{i2}$)	Psychologist ($\hat{\beta}_{i3}$)	Sociologist ($\hat{\beta}_{i4}$)	Economist ($\hat{\beta}_{i5}$)	Physiotherapist ($\hat{\beta}_{i6}$)	Homeopath ($\hat{\beta}_{i7}$)
Ref.: Internet								
Books	x_1	− 0.934	− 0.452	− 0.253	− 0.286	− 0.744*	− 0.858**	0.100
		(0.158)	(0.425)	(0.481)	(0.455)	(0.066)	(0.018)	(0.826)
Radio	x_2	− 0.521	− 1.062**	− 0.148	− 0.170	− 0.031	− 0.065	− 0.248
		(0.346)	(0.012)	(0.651)	(0.619)	(0.929)	(0.842)	(0.592)
Other means	x_3	− 0.097	− 0.080	0.133	− 0.156	0.403**	0.064	0.170
		(0.761)	(0.763)	(0.397)	(0.339)	(0.014)	(0.683)	(0.408)
Ref.: Benefits of S&T greater than damages								
Benefits of S&T equal to damages	x_4	0.294	− 0.306	0.190	0.317*	0.231	0.305*	0.046
		(0.398)	(0.251)	(0.276)	(0.079)	(0.201)	(0.082)	(0.831)
Benefits of S&T less than damages	x_5	− 0.386	− 1.121***	0.412	− 0.013	− 0.216	0.054	− 0.141
		(0.442)	(0.005)	(0.212)	(0.969)	(0.520)	(0.867)	(0.748)
Successful statements about S&T	x_6	0.138	− 0.044	− 0.070	− 0.143**	0.032	0.021	− 0.013
		(0.232)	(0.661)	(0.279)	(0.033)	(0.644)	(0.741)	(0.875)
Ref.: Financial support for science								
No financial support for science	x_7	− 0.975**	− 0.764**	0.001	0.038	− 0.359**	− 0.128	0.087
		(0.012)	(0.016)	(0.995)	(0.828)	(0.040)	(0.453)	(0.703)
No possibilities to support science	x_8	0.167	− 0.403	0.059	− 0.020	− 0.555***	0.111	0.148
		(0.754)	(0.289)	(0.765)	(0.920)	(0.007)	(0.573)	(0.565)
Scores to scientific techniques	x_9	1.164***	0.846***	0.289**	0.265**	0.148	0.214*	− 0.204
		(0.000)	(0.000)	(0.013)	(0.033)	(0.231)	(0.065)	(0.177)
Scores to social scientific techniques	x_{10}	0.015	0.316***	0.656***	0.834***	0.978***	0.492***	− 0.063
		(0.918)	(0.009)	(0.000)	(0.000)	(0.000)	(0.000)	(0.497)
Scores to pseudoscience	x_{11}	0.062	− 0.046	0.357***	0.313***	0.093	0.410***	1.518***
		(0.660)	(0.699)	(0.000)	(0.000)	(0.178)	(0.000)	(0.000)
Political ideology	x_{12}	− 0.005	0.049	− 0.044	− 0.065	− 0.029	− 0.051	− 0.007
		(0.949)	(0.461)	(0.295)	(0.137)	(0.502)	(0.222)	(0.898)
Ref.: Until first level studies								
Second level studies	x_{13}	0.498	0.552	0.426	0.452	0.043	0.390	0.724*
		(0.266)	(0.137)	(0.149)	(0.150)	(0.889)	(0.184)	(0.060)
University studies	x_{14}	1.190**	0.790*	0.845**	0.868**	0.161	0.829**	0.989**
		(0.042)	(0.089)	(0.011)	(0.013)	(0.642)	(0.013)	(0.023)
Ref.: Very low/low/normal S&T knowledge								
High/very high S&T knowledge	x_{15}	0.674	0.079	0.448**	0.483**	0.375*	0.382*	0.671***
		(0.158)	(0.825)	(0.029)	(0.019)	(0.075)	(0.065)	(0.008)
Ref.: catholic								
Non-catholic religion	x_{16}	1.935*	− 0.951*	0.415	0.914*	0.749	0.420	− 0.113
		(0.087)	(0.088)	(0.436)	(0.084)	(0.161)	(0.429)	(0.841)
Atheist, agnostic or indifferent	x_{17}	0.682**	0.419	0.247	0.362**	0.492***	0.611***	− 0.212
		(0.047)	(0.123)	(0.113)	(0.025)	(0.002)	(0.000)	(0.306)
Ref.: Inactive, students and unemployed								
Labourers	x_{18}	0.178	0.162	− 0.241	0.004	− 0.419**	− 0.372**	− 0.301
		(0.594)	(0.536)	(0.140)	(0.980)	(0.014)	(0.023)	(0.159)
Small entrepreneurs, technicians, middle managers	x_{19}	− 0.532	0.621	− 0.342	− 0.378	0.112	− 0.401	− 0.269
		(0.259)	(0.188)	(0.174)	(0.152)	(0.668)	(0.115)	(0.404)
Directors and professionals	x_{20}	− 0.158	− 0.445	0.245	− 0.391	− 0.687**	0.084	− 0.309
		(0.794)	(0.455)	(0.429)	(0.226)	(0.034)	(0.789)	(0.448)
Ref.: Man								
Woman	x_{21}	− 0.112	− 0.201	− 0.090	0.130	− 0.269*	0.056	0.327*
		(0.696)	(0.397)	(0.530)	(0.379)	(0.070)	(0.698)	(0.082)
Continued								

Variables		Medical doctor ($\hat{\beta}_{i1}$)	Physicist ($\hat{\beta}_{i2}$)	Psychologist ($\hat{\beta}_{i3}$)	Sociologist ($\hat{\beta}_{i4}$)	Economist ($\hat{\beta}_{i5}$)	Physiotherapist ($\hat{\beta}_{i6}$)	Homeopath ($\hat{\beta}_{i7}$)
Income (st)	x_{22}	− 0.063	− 0.110	− 0.009	− 0.033	− 0.195***	0.184***	0.170*
		(0.664)	(0.355)	(0.893)	(0.639)	(0.007)	(0.009)	(0.070)
Constant	$\hat{\alpha}_j$	− 3.401***	− 2.148**	− 3.980***	− 4.514***	− 3.914***	− 3.605***	− 5.382***
		(0.004)	(0.034)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations		1039	1039	1039	1039	1039	1039	1039

Table 2. Estimated coefficients and p_{value} (in parenthesis) of the binary logistic regression models estimating the scientificity's perception of the seven professions (*** means that p_{value} <0.01, ** that p_{value} <0.05, and * that p_{value} <0.1, $i = 1, \ldots, 22$, $j = 1, \ldots, 7$). A positive coefficient indicates that an increase of the independent variable implies an increase of the dependent one; on the contrary, a negative coefficient means that when the independent variable increases, the dependent one decreases. The p_{value} reports each coefficient's significance to estimate the dependent variable according to the independent one. The smaller the p_{value}, the more significant the coefficient. It is commonly considered that a coefficient is significant if its p_{value} is lower than 0.1.

and those who employ other means to be informed about S&T consider economists more scientific than the ones who prefer the Internet for information.

According to our results, confidence in science also plays a determining role in the degree of scientificity perceived by some professions. Citizens who consider that science includes both benefits and drawbacks attribute more scientificity to sociologists and physiotherapists than those who believe that the benefits are greater than the drawbacks. However, physicists are considered less scientific by those who see that science's benefits are fewer than its drawbacks, compared to those who think that there are more benefits than drawbacks.

Furthermore, we can observe that the higher the number of successful statements about S&T, the lower the degree of scientificity attributed to sociologists. In addition, people who do not financially support science perceive that the professions of medical doctors, physicists, and economists are less scientific than those who are in favor of financial support for science. In the case of economists, this difference is also significant for those who do not have the possibility of financially supporting science.

Clear statistically significant differences are detected in the degree of scientificity's perception based on the scores assigned by citizens to scientific, social science, and pseudoscientific techniques. Our results reveal that the higher the scores attributed to scientific techniques, the more scientific medical doctors, physicists, psychologists, sociologists, and physiotherapists are considered. Similarly, the degree of scientificity assigned to the professions of physicists, psychologists, sociologists, economists, and physiotherapists is directly proportional to the scores given to social science techniques. Finally, the scientific perception of psychologists, sociologists, physiotherapists, and homeopaths is higher for people who attribute higher scores to pseudoscience techniques.

The level of education also plays an important role in the degree of scientificity attributed to the different professions. Thus, citizens with secondary-level studies perceive the profession of homeopath as more scientific than citizens who only have primary studies, and people with university studies consider that all professions (except economists) have a higher perception of scientificity than people who have first-level studies at most. Concerning their level of S&T knowledge, people stating they have a high or very high level of scientific knowledge attribute a higher perception of scientificity to psychologists, sociologists, economists, physiotherapists, and homeopaths than people who consider that they have a normal, low, or very low level of scientific knowledge.

Religious practice also represents a key factor in attributing a scientificity degree to the professions. In comparison with Catholics, we see that citizens who profess another religion consider medical doctors and sociologists to be more scientific, and physicists are less scientific. However, the scientific perception of medical doctors, sociologists, economists, and physiotherapists is higher among those who define themselves as atheists, agnostics, or indifferent than among Catholics.

Concerning employment status, statistically significant differences were only detected for economists and physiotherapists. Laborers perceive economists and physiotherapists as less scientific than people who are inactive, students, and unemployed people. Besides, the latter also consider economists to be more scientific than directors and professionals. Moreover, compared to men, women attribute a lower degree of scientificity to economists and have a better scientific perception of homeopaths. Furthermore, the (standardized) income level is shown to be relevant for the professions of economist, physiotherapist, and homeopath: the higher the income level, the less scientific economists are perceived to be and the more scientific physiotherapists and homeopaths are considered to be.

Finally, it should be noted that, although the number of independent variables considered is high, the majority of them correspond to dummy variables defined for categorical ones, which makes the presence of multicollinearity unlikely. Nevertheless, the VIF was calculated, yielding an average value of 1.5 and a maximum value of 3, indicating the absence of a problem of multicollinearity.

Multiobjective optimization analysis

According to our statistical analysis, we observe that the regression coefficients of the variables have different signs. Mathematically, this justifies the formulation of a multiobjective optimization problem to simultaneously optimize the scientificity's perception of the seven professions. The application of multiobjective optimization approaches will let us gain insights into the real possibilities of enhancing the scientific perception of all professions at the same time, identifying the key factors to attain such improvements, and being able to anticipate the impacts of the possible improvement policies on all professions.

Background and notations

Many real-world problems involve dealing with several conflicting criteria, which must be optimized simultaneously. These problems, called *multiobjective optimization problems*, are defined by objective functions that represent the criteria and by constraints and bounds for variables that define the feasible set of alternative solutions. Formally, a *multiobjective optimization problem* can be formulated as follows:

$$\begin{array}{ll} \text{maximize} & \{f_1(\mathrm{x}), f_2(\mathrm{x}), \ldots, f_k(\mathrm{x})\} \\ \text{subject to} & \mathrm{x} \in S, \end{array} \tag{4}$$

where $k \geq 2$ is the number of *objective functions*, denoted by $f_j : S \rightarrow \mathbb{R}$ ($j = 1, \ldots, k$), and $S \subset \mathbb{R}^n$ is the *feasible set* formed by vectors of *decision variables* $\mathrm{x} = (x_1, x_2, \ldots, x_n)^T$ satisfying the constraints and bounds. The *feasible objective region* is the image of *S* in the objective space $\mathbb{R}^k$ ($\mathrm{f}(S)$), which is composed of *objective vectors* as $\mathrm{f}(\mathrm{x}) = (f_1(\mathrm{x}), f_2(\mathrm{x}), \ldots, f_k(\mathrm{x}))^T$, for any $\mathrm{x} \in S$. In problem (4), without loss of generality, all objectives are considered to be maximized. In case any of them is to be minimized, it can be transformed into the maximization form by multiplying by -1.

Because of the degree of conflict among the objective functions, it is very unlikely to find a single solution where all of them can reach their individual optima. Therefore, we consider so-called *Pareto optimal solutions* with different tradeoffs, in which no objective function can be improved without worsening, at least, one of the others. Specifically, a solution $\mathrm{x} \in S$ is said to be a *Pareto optimal* of problem (4) if and only if there is no other $\bar{\mathrm{x}} \in S$ such that $f_j(\bar{\mathrm{x}}) \geq f_j(\mathrm{x})$ for all $j = 1, \ldots, k$, and $f_r(\bar{\mathrm{x}}) > f_r(\mathrm{x})$ for, at least, one index *r*. The corresponding objective vector $\mathrm{f}(\mathrm{x})$ is referred to as a *Pareto optimal objective vector*. The set of all Pareto optimal solutions is called the *Pareto optimal set*, denoted by *E*, and the set of all Pareto optimal objective vectors is called the *Pareto optimal front*, denoted by $\mathrm{f}(E)$.

Lower and upper bounds of the objective function values are contained in the so-called *ideal* and *nadir points*, denoted by $\mathrm{z}^\star = (z_1^\star, \ldots, z_k^\star)^T$ and $\mathrm{z}^{\mathrm{nad}} = (z_1^{\mathrm{nad}}, \ldots, z_k^{\mathrm{nad}})^T$, respectively. Formally, for every $j = 1, \ldots, k$, we define $z_j^\star = \max_{\mathrm{x} \in E} f_j(\mathrm{x}) = \max_{\mathrm{x} \in S} f_j(\mathrm{x})$ and $z_j^{\mathrm{nad}} = \min_{\mathrm{x} \in E} f_j(\mathrm{x})$. While the ideal point can be easily obtained, the nadir point is, in general, more difficult to calculate and it is normally estimated[29]. As a result, the range of values for each f_j is defined by the interval $[z_j^{\mathrm{nad}}, z_j^\star]$.

As multiple Pareto optimal solutions are available and cannot be distinguished mathematically, an expert *decision maker* (DM) in the problem domain is needed to finally find the most preferred solution as the final solution. The DM's preferences are decisive in this task, and many different ways to express preferences have been proposed, such as marginal rates of substitution, surrogate values for tradeoffs, classification of objective functions, or a reference point defined by desirable objective function values, among others (see Miettinen (1999)[16] and references therein).

Typically, the purpose of multiobjective optimization methods is to support a DM in finding the most preferred solution. Depending on the moment of the solution process when the DM's preferences are considered, these methods can be classified as a priori, a posteriori, or interactive methods. In a priori methods, the DM specifies their preferences before the optimization process begins. The optimization algorithm then uses this information to find a solution that best matches these preferences. A posteriori methods generate a set of Pareto optimal solutions without requiring the DM to specify their preferences beforehand. Then, after the optimization process, the DM reviews the solutions found and selects the most preferred one. Finally, interactive methods involve the DM throughout the whole solution process. The DM provides feedback and preferences iteratively, and the algorithm uses this information to refine the search for the best solution. This approach allows the DM to adjust their preferences based on the solutions generated and the tradeoffs observed.

As previously indicated, one of the ways used most to elicit preferences is to give a reference point. Mathematically, a reference point is defined as $\mathrm{q} = (q_1, \ldots, q_k)^T$, where each q_j represents a desirable value for the objective function f_j $(j = 1, \ldots, k)$[17]. To find the Pareto optimal solution that best suits the DM's expectations according to q, the multiobjective optimization problem is often scalarized into a single objective optimization problem, which is then solved using an appropriate mathematical programming technique. One of the most frequently used *achievement scalarizing function* is the one proposed by Wierzbicki[17], which can be minimized over the feasible set *S* to obtain the Pareto optimal solution that best satisfies the elicited preferences. This function is formulated as follows:

$$\begin{aligned} & s(\mathrm{f}(\mathrm{x}), \boldsymbol{\mu}, \mathrm{q}) \\ & \quad = \max_{j=1,\ldots,k} \{\, \mu_j(q_j - f_j(\mathrm{x}))\, \} + \rho \sum_{j=1}^{k} \mu_j(q_j - f_j(\mathrm{x})), \end{aligned} \tag{5}$$

where $\mu = (\mu_1, \ldots, \mu_k)^T$ is a vector of positive weights (i.e. $\mu_j > 0$ for $j = 1, \ldots, k$), and the parameter $\rho > 0$ is a small real value that assures the Pareto optimality (for the original problem (4)) of the solution obtained when minimizing (5) over *S*. Any Pareto optimal solution of (4) can be obtained by minimizing (5) over *S* and varying the reference point and/or the weight vector (see Miettinen (1999)[16] for further details). It is important to note that the choice of the reference point q and the weight vector μ can significantly influence the resulting

Pareto optimal solution. Therefore, the DM must carefully consider their preferences and the specific context of the problem when selecting these parameters. Furthermore, the achievement scalarizing function can be adapted to incorporate additional criteria or constraints, making it a flexible and powerful tool for multiobjective optimization.

Multiobjective optimization problem to study the perception of scientificity

In this paper, multiobjective optimization is used to investigate how society's scientific perception of all the professions considered could be improved at the same time. That is, we want to know which situation (regarding the population's scientific culture, politics, education, and sociodemographic characteristics) would allow attaining the best possible levels of scientificity's perception simultaneously for all the disciplines. To do this, we propose building a MOP to maximize seven objective functions representing the perceived scientificity of these professions. The decision variables of the proposed MOP are the 22 independent variables described in Table 1.

As introduced previously, one of the contributions of this paper in comparison to previous works combining regression and multiobjective optimization techniques is related to the formulation of the objective functions. As we have done in previous work, we formulate the objective functions of the MOP using the information provided by the regression model (3) (Table 2). Let us recall that, for each profession, this model estimates its scientificity's perception according to the 22 variables related to scientific culture, politics, education, and sociodemographic characteristics. On the one hand, to the best of our knowledge, this is the first study applying a binary logistic regression model to formulate the resulting MOP. On the other hand, in previous works applying the framework proposed in Luque et al.[18], the objective functions were obtained just using the regression coefficients of the most significant variables (at 1%, 5%, and 10%). Now, we suggest that the objective function of each profession includes the regression coefficients of all the variables considered, regardless of their level of significance. Nonetheless, to have the most realistic estimation, every regression coefficient is weighted using its significance level measured by the p_{value}, as follows. If we denote the seven objective functions by $F_j : \mathbb{R}^{22} \to \mathbb{R}$, where $j = 1$ (medical doctor), 2 (physicist), 3 (psychologist), 4 (sociologist), 5 (economist), 6 (physiotherapist), 7 (homeopath), then they are defined by:

$$F_j(\mathrm{x}) = \frac{1}{1 + e^{-(\hat{\alpha}_j' + \hat{\beta}_{1j}' \cdot x_1 + \cdots + \hat{\beta}_{22j}' \cdot x_{22})}}, \quad \text{with } j = 1, \ldots, 7, \tag{6}$$

where $\mathrm{x} = (x_1, x_2, \ldots, x_{22})^T$ represents the vector of decision variables. The coefficient $\hat{\beta}_{ij}'$ of each variable x_i in each objective F_j $(i = 1, \ldots, 22,\ j = 1, \ldots, 7)$ is defined as $\hat{\beta}_{ij}' = (1 - p_{ij}) \cdot \hat{\beta}_{ij}ij$, where $\hat{\beta}_{ij}ij$ is the regression coefficient of x_i in the binary logistic model (3) obtained for profession *j* and p_{ij} is its p_{value} (all of this can be found in Table 2). In this way, the more significant an independent variable is in one of the regression models, the higher its weight in the corresponding objective function. This enables introducing into the model information obtained according to the data that would be missed if we just include the regression coefficients with less than 10% significant level, and therefore we enrich the formulation of the model to have the best-suited estimations. We would like to clarify that the weights $\hat{\beta}_{ij}'$ do not represent preferential weights that give different importance to improving the seven professions. These weights are just introduced in the formulation of the objective functions to consider not only the most significant variables but all the variables according to their p_{value}. In comparison to the previous works (where only the most significant regression coefficients were used to build the objective functions), we enrich the formulation by introducing information obtained according to the data that would be missed if we do not include the variables with significant levels higher 10%. Similarly, the constant $\hat{\alpha}_j'$ $(j = 1, \ldots, 7)$ is calculated as $\hat{\alpha}_j' = (1 - p_j) \cdot \hat{\alpha}_j$, where $\hat{\alpha}_j$ is the constant in the binary logistic model (3) of profession *j* and p_j represents its p_{value}. All the coefficients of the objective functions are calculated using the information in Table 2 (for example, $\hat{\alpha}_{11} = -0.620 \cdot (1 - 0.180) = -0.508$). Therefore, the objective functions are formulated as follows:

$$\begin{aligned}
F_1(\mathrm{x}) &= \tfrac{1}{1+e^{-(-3.3874-0.7864\cdot x_1-0.3407\cdot x_2+\cdots+0.0212\cdot x_{22})}},\\
F_2(\mathrm{x}) &= \tfrac{1}{1+e^{-(2.0750-0.2599\cdot x_1-1.0493\cdot x_2+\cdots+0.0710\cdot x_{22})}},\\
F_3(\mathrm{x}) &= \tfrac{1}{1+e^{-(-3.9800-0.1313\cdot x_1-0.0517\cdot x_2+\cdots-0.0010\cdot x_{22})}},\\
F_4(\mathrm{x}) &= \tfrac{1}{1+e^{-(-4.5140-0.1559\cdot x_1-0.0648\cdot x_2+\cdots-0.0119\cdot x_{22})}},\\
F_5(\mathrm{x}) &= \tfrac{1}{1+e^{-(-3.9140-0.6949\cdot x_1-0.0022\cdot x_2+\cdots-0.1936\cdot x_{22})}},\\
F_6(\mathrm{x}) &= \tfrac{1}{1+e^{-(3.6050-0.8426\cdot x_1-0.0103\cdot x_2+\cdots+0.1823\cdot x_{22})}},\\
F_7(\mathrm{x}) &= \tfrac{1}{1+e^{-(-5.3820+0.0174\cdot x_1-0.1012\cdot x_2+\cdots+0.1581\cdot x_{22})}}.
\end{aligned} \tag{7}$$

To have a MOP that fits the data as best as possible, some constraints are needed to control the values allowed for the decision variables. On the one hand, we have introduced some technical constraints into the MOP owing to the categorical nature of some of the independent variables used. All the different categories of a categorical question in the original questionnaire cannot all be selected at the same time, meaning that all the dummy variables created for the categories cannot be equal to 1 simultaneously; indeed, only one dummy variable can take the value one. Thus, we have formulated a constraint assuring that the sum of all dummy variables defined for each categorical item is less than or equal to 1 (for more details, see Luque et al. (2022)[18]). This way, six technical constraints are defined for the six categorical items considered, which are:

$$\begin{array}{lc} (C1) & x_1 + x_2 + x_3 \leq 1, \\ (C2) & x_4 + x_5 \leq 1, \\ (C3) & x_7 + x_8 \leq 1, \\ (C4) & x_{13} + x_{14} \leq 1, \\ (C5) & x_{16} + x_{17} \leq 1, \\ (C6) & x_{18} + x_{19} + x_{20} \leq 1. \end{array} \tag{8}$$

On the other hand, other constraints are built to only allow meaningful and sufficiently realistic values of the decision variables according to the data considered. Based on a linear correlation analysis carried out among the continuous decision variables, we have detected dependencies between some of these variables; therefore, their values cannot be set independently. In consequence, we have formulated some bound constraints using the confidence intervals of the correlation coefficients in the pairwise linear relation for the variables with a highly significant dependency (at 95%).

These are the variables indicating the scores given to scientific techniques (x_9), social science techniques (x_{10}), and pseudoscience techniques (x_{11}), each of them concerning the following: successful statements about S&T (x_6), politic ideology (x_{12}) and standardized family income (x_{22}). For the sake of clarity, next, we describe an example based on the significant dependence observed for x_9 (scores for scientific techniques) concerning x_6 (successful statements about S&T). In our analysis, a significant linear relationship was identified between them as $x_9 = a \cdot x_6 + b$, with the following confidence intervals for the coefficients (at 95%): $a \in [a^l, a^u] = [-0.002, 0.048]$, and $b \in [b^l, b^u] = [4.353, 4.687]$. This linear relation implies that x_9 can only take values between these two bounds:

$$a^l \cdot x_6 + b^l \leq x_9 \leq a^u \cdot x_6 + b^u.$$

Therefore, the following bound constraints are considered in the MOP to control the possible values for x_6 and x_9:

$$\begin{array}{c} (a^l \cdot x_6 + b^l - x_9) \leq 0, \\ x_9 - (a^u \cdot x_6 + b^u) \leq 0 \end{array} \Rightarrow \begin{array}{c} (-0.002 \cdot x_6 + 4.353) - x_9 \leq 0, \\ x_9 - (0.048 \cdot x_6 + 4.687) \leq 0. \end{array}$$

Following the same procedure, the eighteen bound constraints indicated in (9) (including the ones given in the example), which were built taking into account the significant dependencies between the previously indicated variables, have been introduced in the MOP:

$$\begin{array}{lc} (C7) & (-0.008 \cdot x_6 + 4.353) - x_9 \leq 0, \\ (C8) & x_9 - (0.048 \cdot x_6 + 4.687) \leq 0, \\ (C9) & (-0.007 \cdot x_{12} + 4.489) - x_9 \leq 0, \\ (C10) & x_9 - (0.016 \cdot x_{12} + 4.704) \leq 0, \\ (C11) & (0.007 \cdot x_{22} + 4.586) - x_9 \leq 0, \\ (C12) & x_9 - (0.076 \cdot x_{22} + 4.663) \leq 0, \\ (C13) & (-0.007 \cdot x_6 + 2.606) - x_{10} \leq 0, \\ (C14) & x_{10} - (0.006 \cdot x_6 + 3.236) \leq 0, \\ (C15) & (-0.008 \cdot x_{12} + 2.720) - x_{10} \leq 0, \\ (C16) & x_{10} - (0.006 \cdot x_{12} + 3.126) \leq 0, \\ (C17) & (0.076 \cdot x_{22} + 2.793) - x_{10} \leq 0, \\ (C18) & x_{10} - (0.210 \cdot x_{22} + 2.936) \leq 0, \\ (C19) & (-0.171 \cdot x_6 + 2.786) - x_{11} \leq 0, \\ (C20) & x_{11} - (-0.043 \cdot x_6 + 3.426) \leq 0, \\ (C21) & (-0.031 \cdot x_{12} + 2.457) - x_{11} \leq 0, \\ (C22) & x_{11} - (0.014 \cdot x_{12} + 2.870) \leq 0, \\ (C23) & (-0.004 \cdot x_{22} + 2.513) - x_{11} \leq 0, \\ (C24) & x_{11} - (0.005 \cdot x_{22} + 2.661) \leq 0. \end{array} \tag{9}$$

Additionally, the minimum and maximum values achieved by the explanatory variables in the dataset are used as lower and upper bounds for the variables in the MOP, respectively, as shown in Table 3.

With all of this, the MOP to be solved to determine if the perceived scientificity of the seven professions can all be improved simultaneously, and which situation of society would enable this improvement, is the following:

$$\begin{array}{ll} \text{maximize} & \{F_1(\mathrm{x}), F_2(\mathrm{x}), F_3(\mathrm{x}), F_4(\mathrm{x}), F_5(\mathrm{x}), F_6(\mathrm{x}), F_7(\mathrm{x})\} \\ \text{subject to} & (C1) - (C24), \\ & l_i \leq x_i \leq u_i, \text{ for } i = 1, \ldots, 22. \end{array} \tag{10}$$

Next, we will demonstrate which type of information can be extracted from a Pareto optimal solution of problem (10) by solving it. As previously explained, MOPs have multiple Pareto optimal solutions, all of

Description	Decision variable	Type	Lower bound (l_i)	Upper bound (u_i)
Books	x_1	Binary	0	1
Radio	x_2	Binary	0	1
Other means	x_3	Binary	0	1
Benefits of S&T equal to damages	x_4	Binary	0	1
Benefits of S&T less than damages	x_5	Binary	0	1
Successful statements about S&T	x_6	Continuous	0	6
No financial support for science	x_7	Binary	0	1
No possibilities to support science	x_8	Binary	0	1
Scores to scientific techniques	x_9	Continuous	0	5
Scores to social science techniques	x_{10}	Continuous	0	5
Scores to pseudoscience techniques	x_{11}	Continuous	0	5
Political ideology	x_{12}	Continuous	1	10
Second level studies	x_{13}	Binary	0	1
University studies	x_{14}	Binary	0	1
High/very high S&T knowledge	x_{15}	Binary	0	1
Non-catholic religion	x_{16}	Binary	0	1
Atheist, agnostic or indifferent	x_{17}	Binary	0	1
Labourers	x_{18}	Binary	0	1
Small entrepreneurs, technicians and middle managers	x_{19}	Binary	0	1
Directors and professionals	x_{20}	Binary	0	1
Woman	x_{21}	Binary	0	1
Income (st)	x_{22}	Continuous	− 1.112	5.167

Table 3. Lower and upper bounds of the decision variables in the multiobjective optimization problem. They determine the range of values allowed for each of them.

them equal from a mathematical point of view. Preference information about the problem in question must be considered in the solution process to select a final solution. In our case, we will apply a reference point-based methodology. Since this is just an example of use and no real DM is involved in the solution process, we will just consider the ideal point as the reference point, given that it represents the best individual situations that the professions can reach. Note that other solution techniques can be applied, or other preferences can be used to find different solutions and reach different conclusions (for example, to study what happens if higher priorities are assigned to the improvement of some of the professions). The components of the ideal point are calculated by individually maximizing each objective function F_j $(j = 1, \ldots, 7)$ over the feasible set defined by the constraints (C1)-(C24) and the ones with the lower and upper variable bounds. As a result, the reference point obtained is:

$$q = (0.9987, 0.9908, 0.9149, 0.9258, 0.8048, 0.9502, 0.6751).$$

These reference values represent the desirable levels of perceived scientificity that we would like citizens to attribute to each of the seven professions (medical doctor, physicist, psychologist, sociologist, economist, physiotherapist, and homeopath), which are the best that each profession can be assigned individually, according to our results.

Once we have set the reference point, as explained previously, we will find the Pareto optimal solution that fits best by minimizing the achievement scalarizing function given in (5) over the feasible set. In this function, all objective functions have been equally weighted (i.e. $\mu_j = 1$, for every $j = 1, \ldots, 7$), which means, in practice, that reaching the (individual) ideal degree of the perceived scientificity of each profession is equally important. The resulting problem that has been solved is:

$$\begin{array}{ll} \text{minimize} & \max_{j=1,\ldots,7}(q_j - F_j(x)) + \rho\sum_{j=1}^{7}(q_j - F_j(\mathrm{x})) \\ \text{subject to} & (C1) - (C24), \\ & l_i \leq x_i \leq u_i, \quad \text{for } i = 1, \ldots, 22. \end{array} \tag{11}$$

The Pareto optimal solution obtained by solving problem (11) is described in Tables 4 (decision variable values) and 5 (objective function values).

Description	Decision variable	Value
Books	x_1	0
Radio	x_2	0
Other means	x_3	1
Benefits of S&T equal to damages	x_4	1
Benefits of S&T less than damages	x_5	0
Successful statements about S&T	x_6	1.270
No financial support for science	x_7	0
No possibilities to support science	x_8	0
Scores to scientific techniques	x_9	4.700
Scores to social science techniques	x_{10}	3.120
Scores to pseudoscience techniques	x_{11}	2.670
Political ideology	x_{12}	1.000
Second level studies	x_{13}	0
University studies	x_{14}	1
High/very high S&T knowledge	x_{15}	1
Non-catholic religion	x_{16}	1
Atheist, agnostic or indifferent	x_{17}	0
Labourers	x_{18}	0
Small entrepreneurs, technicians and middle managers	x_{19}	0
Directors and professionals	x_{20}	0
Woman	x_{21}	1
Income (st)	x_{22}	2.378

Table 4. Decision variable values at the Pareto optimal solution obtained by solving the scalarizing optimization problem (11).

Description	Objective function	References values	Objective values	Sample mean
Medical doctor	F_1	0.999	0.997	0.937
Physicist	F_2	0.991	0.898	0.907
Psychologist	F_3	0.915	0.875	0.548
Sociologist	F_4	0.926	0.909	0.432
Economist	F_5	0.805	0.686	0.440
Physiotherapist	F_6	0.950	0.891	0.579
Homeopath	F_7	0.675	0.545	0.244

Table 5. Objective function values at the Pareto optimal solution obtained by solving the scalarizing optimization problem (11).

Analysis of the results and discussion

In this section, we will interpret the Pareto optimal solution and analyze its practical implications. Regarding the perception of scientificity achieved by the different professions (Table 5), it can be observed that none of them has been able to reach its most desirable (best individual) scientificity perception level. This highlights the impossibility of enhancing the perception of all the professions at the same time; indeed, the improvement policies that may be applied must be carefully defined, bearing in mind that improving the scientific perception of one profession may have an impact on the level of perceived scientificity of some of the others.

According to our results, being a medical doctor is the profession that can reach the highest level of scientificity perception, followed by (from the highest to the lowest one) sociologist, physicist, physiotherapist, psychologist, economist, and, finally, homeopath. Indeed, the levels of perceived scientificity closest to their desired values are attained by medical doctors and sociologists, while economists and homeopaths are those with the furthest levels of perceived scientificity from their reference values.

Table 5 also shows the sample mean of the scientificity levels attained by the professions in our dataset, which gives us an average idea of the current perception of each profession by Spanish society (These sample means are obtained using the re-coded values of the answers to the question Q.8.B, as explained in Statistical analysis Section: scores 1, 2, and 3 of the Likert scale are equivalent to 0 in the dichotomous variable generated for each profession, and scores 4 and 5 to 1.). Next, we compare these mean values with the corresponding objective values attained by the Pareto optimal solution found to gain insight into the extent to which the scientific perception of the seven professions can be enhanced. As shown, medical doctors are currently the best considered by Spanish society (its sample mean is the highest one, above 0.9) and, as previously concluded, they can reach the highest level of perceived scientificity among all professions. Nevertheless, when we compare its best possible scientificity's perception level (i.e. its optimal objective value) with this average value, we see that this profession is the one that can improve the least. On the other hand, physicists are the next ones with the highest perception nowadays (the second highest mean), and this profession attains the third-best level of perceived scientificity in the obtained solution (i.e. the third-highest objective value). However, both this objective value and its average value are, in practice, very similar, which indicates that the scientific perception of physicists will remain similar to their current situation.

On the contrary, our results indicate that it is possible to improve the scientificity's perception of the other professions to a quite higher level. Observe that the mean values of the perceived scientificity of psychologists and physiotherapists are only above 0.5, approximately, but their perception can experience a very interesting increase since they can reach levels of 0.875 for psychologists and of 0.891 for physiotherapists. In addition, the scientific images of economists and sociologists can also be enhanced to a large extent, since their scientificity's perception means are both around 0.4. However, the level of perceived scientificity of economists can be improved to values around 0.7 and sociologist has the highest improvement potential, being able to reach a scientificity's perception level of 0.909, which is the second-best optimal value. Finally, although homeopath is the profession with the lowest possible perceived scientificity level, its perception by Spanish citizens can also be significantly improved, moving from a 0.244 mean value to 0.545.

With all this analysis, we can conclude that the professions related to health sciences and STEM are the ones that can reach the highest values of perception as scientific disciplines, although the room for improvement is limited due to the fact that these professions are already well, including sociologists, who obtain the second-best scientificity perception level. Furthermore, according to the comparisons of the current and possible best situations of the profession, it is evident that the professions that can experience the highest possible improvement are sociologist, economist, and homeopath. They are the ones that belong to the social sciences and pseudosciences, which have traditionally had the worst perception by society.

The values achieved by the decision variables in the Pareto optimal solution provide us with very valuable information. These values inform about the social situation that must be promoted with specific policies if we want to simultaneously attain, in the future, the "optimal" scientificity levels attributed to the seven professions achieved at the solution found. In consequence, we analyze next the potential citizen profile obtained, that is, the combination of scientific, political, educational, and sociodemographic features that would allow achieving the desirable scientific perception for the professions, and thus, which should be politically promoted.

Regarding the scientific culture variables, the Pareto optimal solution obtained indicates that the potential person that would perceive the professions with the "optimal" scientificity levels gets informed about S%T using other means (different to the Internet, books, and radio). In addition, (s)he thinks that the benefits and drawbacks of S&T are balanced and is willing to support science financially. The variable with the largest number of successful statements about S&T obtains the value of 1.270. Given that the sample mean of this variable is 4.852, and that it ranges from 0 to 6, it seems that the number of correct answers to the S&T statements does not seem to be very relevant. Finally, this person's scores for scientific, social science, and pseudoscience techniques are 4.700, 3.120, and 2.6704, respectively, which are very close to the sample means of these variables (4.635, 2.901, and 2.589, respectively). This citizen's political ideology is far left. Concerning education, university studies are desirable, and the level of S&T knowledge must be high or very high. This implies that education is a key variable for increasing the scientific perception of the considered professions.

Regarding sociodemographic characteristics, the solution indicates that this person is an inactive woman, a student or unemployed, is not Catholic, with a family income (standardized value) of about 2.378, which is above the sample income mean (around 0.25). It is important to highlight that although not all independent variables are controllable (such as being catholic or non-catholic, gender, etc.) we have considered all of them as decision variables in the MOP formulated in (10) because, with our model, we want to extract information about the potential citizen profile that would perceive the professions with the best possible perceived scientificity levels according to the preferences used (i.e. the scientificity's perception levels at the solution obtained). Therefore, politicians should concentrate their efforts on people not corresponding with the obtained profile, since those people would not have the best possible perception of the professions but could perceive them better if some policies are directed to them. Thus, our analysis is important to know in which directions politicians should implement some specific actions (i.e. what should be promoted) and the public to whom these measures should be directed.

Thus, we can conclude that policy decicion makers must promote the levels of education of Spaniards, especially those related to S&T, as previously said. In addition, measures should be implemented to increase the scientific culture of society. Actions must be taken to explain to the public the benefits and drawbacks of S&T and to promote the importance of a financial system for science involving all of society, highlighting the importance of each individual's support. Although the Internet is one of the most used sources of information, the solution obtained indicates that obtaining information by other means (such as TV, scientific or technical journals, and the written press, among others) plays a very important role in increasing the scientific perception of the professions. This suggests that these media are a more reliable source of information for society. In this sense, policies are needed to guarantee the credibility of the S&T information available in these means and the accessibility of all citizens to them (e.g. promoting the Open Science), even measures to encourage citizens to check the veracity and scientific rigour of the information they find would be very beneficial to this objective.

To address these concerns, specific actions as the ones indicated hereafter could be implemented to promote S&T education, enhance scientific culture, and ensure the credibility and accessibility of scientific information. For example, school curricula could be revise to integrate more comprehensive S&T education at all levels, from primary school through to university, with the emphasis placed on the promotion of critical thinking, problem solving, and understanding of the real-world applications of S&T. Also, national campaigns could be launched to explain the importance of S&T to society, using easily understandable language and examples that resonate with the public. Introducing incentives such as scholarships, career development programs, or public-private partnerships to encourage students to pursue STEM careers could be another interesting action. Furthermore, more programs should be developed where scientists and experts engage directly with the public (e.g. public lectures, science fairs, or social media Q&A sessions), to demystify science, making it more relatable and accessible.

In relation to sources of information, collaborations between scientists and journalists could be very positive to improve the quality of science communication, ensuring that the information is accurate, reliable and accessible. Obviously, the implementation of policies to support open access to science would be essential to ensure that scientific publications, datasets, and research findings are freely available to all citizens. This could be facilitated through government-funded platforms or incentives for private institutions to adopt open science practices. In addition, politicians should develop educational programs that teach citizens how to assess the credibility of scientific information they encounter in various media, as well as introduce more regulations to ensure that science communicated through traditional media channels adheres to high standards of accuracy, evidence-based reporting and peer-reviewed sources.

To promote the public support of S&T, campaigns that explain how the financial system for science works and why public investment in S&T is crucial for social progress would be desirable, highlighting how it can contribute to innovation, job creation, and better quality of life. This should be coupled with the promotion and support of platforms that allow citizens to directly contribute to specific scientific research projects, thus involving them in the funding process and raising awareness of the importance of collective financial support for science.

Conclusion

Public perception of science may condition many aspects of society, such as the financial support system for science and technology or the educational and job options selected by young people. Indeed, a strong research system is only possible when society has a basic knowledge of the different science and technology disciplines and is fully convinced of supporting them. Therefore, science policymakers must encourage studies aimed at understanding the social perception (scientificity) of scientific professions belonging to different fields, with an emphasis on investigating in which direction should be directed their policies to promote the level of perceived scientificity of certain disciplines.

In this paper, we have studied the scientific perception of a set of professions in Spanish society using several quantitative techniques (statistical analysis and multiobjective optimization). Our purpose has been to analyze the possible key factors that influence the scientific perception in Spain of medical doctors, physicists, psychologists, sociologists, economists, physiotherapists, and homeopaths as scientific disciplines. In addition, we have researched possible ways to improve the degree of scientificity attributed to these professions in the future.

First, we have estimated the level of perceived scientificity of these professions through binary logistic regression models, allowing us to reach interesting findings about the key variables that affect them. Secondly, we have gone a step further and we have built a multiobjective optimization problem where all variables are used to formulate the objective functions taking into account their level of significance. This has enabled us to gain an insight into the scope of possibilities for enhancing, at the same time, the situation of the seven professions. We concluded that social science and pseudo-science professions could see the greatest improvements if certain promotional policies were implemented, although the professions that could still achieve the highest levels of scientific perception are those belonging to the health sciences and STEM fields.

Indeed, our methodology has allowed us to identify the directions in which the required promotion policies. Attention should be paid to improving the educational system, in particular, to promote actions to increase the interest of young people in science, technology, and research. For example, an example of this type of action is *Science is Wonderful!*[30], an annual science fair celebrated in Europe since 2015 that allows primary and secondary school students to interact with researchers, learn more about their work and ask questions about scientific careers.

On the other hand, our results have also highlighted the importance of designing campaigns to bring research and researchers of different fields closer to the public, showing their impact on citizens' daily lives and emphasizing the importance of a strong financial support system. This is in line with current actions worldwide, such as the *European Researchers' Night*[31], which is an event organized the same night in many European countries to promote research projects across Europe and beyond, display the diversity of science and show the benefits of researchers' work on ordinary lives in ways that are fun and inspiring.

A possible limitation of this study may be the fact that it is supported by survey data corresponding to a specific wage. To extract conclusions on how to increase the scientific perception of the professions in other time periods, or in a different country, new survey data would be needed. However, the methodology applied can be replicated easily just by updating the dataset. Actually, we are also interested in designing a framework able to replicate the study using real-time data so that conclusions can be extracted easily in other contexts. Furthermore, analyzing similar data from other countries could be interesting to compare the results obtained and identify potential similarities or differences, and it represents a future line of research. This comparative approach would enhance the understanding of the findings and provide a broader context for the study's conclusions.

As other future research lines, we would like to interact with real policymakers to consider real preferences regarding the improvement of the perceived scientificity levels of the professions. In addition, we plan to investigate if a difference in men's and women's scientific perceptions of the professions can be identified. We are interested in researching which factors have different impacts for both of them and if specific policies may be needed to reach better scientificity levels separately.

Data availability

Our study is based on data corresponding to the 'Social perception of Science and Technology Questionnaire 2018' (https://www.fecyt.es/es/noticia/encuestas-de-percepcion-social-de-la-ciencia-y-la-tecnologia-en-espana). Dataset supporting this study is available from this website or upon request from the corresponding author.

Received: 28 October 2024; Accepted: 3 February 2025
Published online: 09 February 2025

References

1. Fernández-Esquinas, M., Cárdenas, J. & Sánchez-Rodríguez, M.I. El estatus científico de profesiones y prácticas: Una comparación entre ciencias experimentales, ciencias sociales y pseudociencias. In FECYT (ed.), Percepción Social de la Ciencia y la Tecnología 2018, Fundación Española para la Ciencia y la Tecnología, Madrid, pp. 107-139 (2019).
2. Nisbet, M. C. et al. Knowledge, reservations, or promise?: A media effects model for public perceptions of science and technology. *Commun. Res.* **29**(5), 584–608 (2002).
3. Bauer, M. W., Petkova, K. & Boyadjieva, P. Public knowledge of and attitudes to science: Alternative measures that may end the "science war". *Sci. Technol. Human Values* **25**(1), 30–51 (2000).
4. Myers, T. A. et al. Predictors of trust in the general science and climate science research of US federal agencies. *Public Underst. Sci.* **26**(7), 843–860 (2017).
5. Lilienfeld, S.O. Public skepticism of psychology. Why many people perceive the study of human behavior as unscientific. American Psychologist 67, 2, 119-129. (2012)
6. Rogero, J. & Lobera, J. Márgenes difusos: La confianza en las pseudociencias en España. In FECYT (ed.), Percepción Social de la Ciencia y la Tecnología en España 2016, Fundación Española para la Ciencia y la Tecnología, Madrid, pp. 207-233 (2017).
7. Zaboski, B. A. & Therriault, D. J. Faking science: scientificness, credibility, and belief in pseudoscience. *Educ. Psychol.* **40**(7), 820–837 (2020).
8. Allum, N., Sturgis, P., Tabourazi, D. & Brunton-Smith, I. Science knowledge and attitudes across cultures: A meta-analysis. *Public Underst. Sci.* **17**(1), 35–54 (2008).
9. Muñoz, A. La imagen de la ciencia en España a través de la lente del modelo pica. In FECYT (ed.), Percepción Social de la Ciencia y la Tecnología en España 2016, Fundacion Espanñola para la Ciencia y la Tecnologıa, Madrid, pp. 149-177 (2017).
10. Takahashi, B. & Tandoc, E. C. Media sources, credibility, and perceptions of science: Learning about how people learn about science. *Public Underst. Sci.* **25**(6), 674–690 (2016).
11. Scheitle, C. P. Politics and the perceived boundaries of science: Activism, sociology, and scientific legitimacy. *Socius* **4**, 1–13 (2018).
12. Santos, L., Escobar Mercado, M. & y Quintanilla Fisac, M. á. Dimensiones y modelos de cultura científica: implicaciones prácticas para la financiación y la demarcación de la ciencia. En: FECYT (ed.), Percepción Social de la Ciencia y la Tecnología en España 2016. Madrid: FECYT (2017).
13. Huber, B., Barnidge, M., Gil de Zúñiga, H. & y Liu, J. Fostering public trust in science: The role of social media. *Public Underst. Sci.* **28**(7), 759–777 (2019).
14. Scheitle, C. P. & y Guthrie, S. K. Public responses to scientific research: Does disciplinary attribution matter?. *Public Underst. Sci.* **28**(2), 234–245 (2019).
15. Liu, H. & y Priest, S. Understanding public support for stem cell research: Media communication, interpersonal communication and trust in key actors. *Public Underst. Sci.* **18**(6), 704–718 (2009).
16. Miettinen, K. *Nonlinear Multiobjective Optimization* (Kluwer Academic Publishers, Boston, 1999).
17. Wierzbicki, A. P. The use of reference objectives in multiobjective optimization. In *Multiple Criteria Decis. Mak.* (eds Fandel, G. & Gal, T.) 468–486 (Springer, Theory and Applications, 1980).
18. Luque, M., Marcenaro-Gutiérrez, O. D., González-Gallardo, S. & Ruiz, A. B. Towards a framework to combine multiob- jective optimization and econometrics and an application in economics of education. *RAIRO-Oper. Res.* **56**(3), 2015–2035 (2022).
19. Wallenius, H., Wallenius, J. & Vartia, P. An approach to solving multiple criteria macroeconomic policy problems and an application. *Manage. Sci.* **24**(10), 1021–1030 (1978).
20. Henriques, C. O., Luque, M., Marcenaro-Gutiérrez, O. D. & López-Agudo, L. A. A multiobjective interval programming model to explore the trade-offs among different aspects of job satisfaction under different scenarios. *Socioecon. Plann. Sci.* **66**, 35–46 (2019).
21. Marcenaro-Gutiérrez, O. D., Luque, M. & Ruiz, F. An application of multiobjective programming to the study of workers' satisfaction in the Spanish labour market. *Eur. J. Oper. Res.* **203**(2), 430–443 (2010).
22. González-Gallardo, S., Henriques, C. O., Marcenaro-Gutiérrez, O. D. & Luque, M. A novel approach for exploring the trade-offs between several features of students' well-being. *Int. Trans. Oper. Res.* **29**(3), 1723–1748 (2022).
23. González-Gallardo, S., Ruiz, A. B. & Luque, M. Analysis of the well-being levels of students in Spain and Finland through interval multiobjective linear programming. *Mathematics* **9**(14), 1628 (2021).
24. González-Fernández, A. I., Rubio-Misas, M. & Ruiz, F. Goal programming to evaluate the profile of the most profitable insurers: An application to the Spanish insurance industry. *Int. Trans. Oper. Res.* **27**(6), 2976–3006 (2020).
25. González-Fernández, A. I., Rubio-Misas, M. & Ruiz, F. Multi-objective reference point techniques to optimize profitability, growth, and risk in the non-life insurance industry: International analysis. *Int. Trans. Oper. Res.* **30**(5), 2546–2570 (2023).
26. Ibrahim, A. et al. Apple Perfection: Assessing Apple Quality with Waterwheel Plant Algorithm for Feature Selection and Logistic Regression for Classification. *J. Artif. Intell. Eng. Practi.* **1**(1), 34–48 (2024).
27. El-Kenawy, E.M. & Ibrahim, A.R.A.F. Football Optimization Algorithm (FbOA): A Novel Metaheuristic Inspired by Team Strategy Dynamics. *J. Artif. Intell. Metaheur.* (2024).
28. El-Sayed, M. E. et al. Greylag Goose Optimization: Nature-inspired optimization algorithm. *Expert Syst. Appl.* **238**, 122147 (2024).
29. Deb, K., Miettinen, K. & Chaudhuri, S. Towards an estimation of nadir objective vector using a hybrid of evolutionary and local search approaches. *IEEE Trans. Evol. Comput.* **14**(6), 821–841 (2010).
30. European Commission Science is wonderful! https://marie-sklodowska-curie-actions.ec.europa.eu/science-is-wonderful (2023).
31. European Commission, European researchers' night. https://marie-sklodowska-curie-actions.ec.europa.eu/event/2022-european-researchers-night (2022).

Author contributions

All the authors participated in the analysis and writing of the manuscript.

Funding

This work was funded by the Spanish Ministry of Science and Innovation (project PID2020-115429GB-I00) and by the Andalusian Regional Ministry of Economy, Knowledge, Business and University (PAI group SEJ-532).

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to M.I.S.-R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.