APPENDIX: SUPPORTING INFORMATION

Participants

For all the online studies, a participant was defined as a row in the data file that had completed the task, was over 18, and was the first occurrence of that IP and MTurk ID (both within and across studies). The vagaries of participant recruitment meant that our pre-chosen sample sizes were often specified as minima, with some overshooting. Power calculations for mixed-effects models can be difficult, so we primarily based our estimates on simpler tests (e.g., a t-test to contrast two conditions) that are typically less sensitive than our actual analyses. Power calculations used G*Power [1].

Study 1. The sample size was based on previous impression-formation studies, which have found that 20-25 participants per stimulus gives reliable measures (e.g., [3]). One participant was removed for having zero variance in their responses, demonstrating little engagement with the task. The final sample of participants rating the faces on the predictor variables comprised 9 males and 44 females, ages 18-50 (M = 20.0, SD = 4.6); 50.9% reporting English as their first language. The participants who rated the faces on the criterion variables ("Interest" or "Good Scientist") comprised 16 males and 38 females, ages 18-40 (M= 21.4, SD = 4.9), 94% first-language English.

Study 2. Akin to study 1, sample size was based on obtaining 25-30 judgments per each dimension, so we sought to recruit at least 780 participants for the predictor variables and 120 for the criterion variables. The initial ratings of the criterion variables showed rather low reliability, so the criterion sample was boosted by 100 people. Four participants were removed for having zero variance in their responses, another was removed because of a computer error. The final sample who rated faces on the predictor dimensions comprised 450 males and 380 females, ages 18-72 (M = 35.3, SD = 10.8), 98% first-language English; the sample rating the faces on the criterion variables comprised 107 males and 99 females, ages 20-75 (M = 34.3, SD = 10.3), 97% first-language English. The numbers of participants rating each predictor trait/criterion variable are listed in Table S1.

Variable	*N*
Age	68
Capable	56
Competent	66
Effective	58
Fair	53
Friendly	58
Honest	59
Intelligent	62
Likeable	55
Moral	59
Physically Attractive	58
Sociable	62
Trustworthy	60
Warm	56
Interest judgments	103
"Good Scientist" judgments	103

Table S1. Number of participants rating each dimension in Study 2.

Study 3. The sample size was based on 80% power to detect a small effect (w=0.1) in a chi-square test of whether face type influences the article that is chosen, resulting in a desired sample size of at least 785 participants. (This study was based on a pilot study that indicated a small effect size; full details are available from the authors.) No participants were excluded. The final sample comprised 526 males, 323 females, ages 18-73 (*M* = 32.4, *SD* = 10.6), 93% first-language English, with 427 participants assigned to the Text condition and 422 to the Video condition.

Study 4. The sample size was based on 95% power to detect a small-to-medium effect size (d=0.2, based on the modest effect found in Study 3) in a within-subject t-test for an effect of face-type, resulting in a minimum required sample of 330 participants. Two participants were excluded for reporting technical issues (i.e., the photos did not load properly). The final sample comprised 192 males, 216 females, ages 18-74 (*M* = 35.9, *SD* = 11.1), 98% first-language English.

Study 5. The sample size was based on obtaining 95% power to detect a small-to-medium effect (d = 0.15) in a within-subjects t-test comparing the two face types. Seventy participants were excluded for recognizing either of the two articles they read. The final sample comprised 261 males, 297 females, ages 18-81 (*M* = 36.4, *SD* = 12.5), 97% first-language English, with the Male-Biology, Male-Physics, Female-Biology, and Female-Physics conditions having 150, 144, 129, and 135 participants, respectively."

Study 6. The minimum sample size of 800 participants was calculated based on 80% power to detect a small effect (d = 0.1, estimated from the effect size from Study 5). Participants were asked a simple memory/attention check question after reading the science stories; those who failed were redirected away from the survey and counted as "non-completers". Of those who completed the task, 3 were excluded for reporting technical problems, and one was excluded for recognizing all of the articles. This study only excluded people who recognized all articles, whereas Study 5 excluded people who recognized any articles; this discrepancy arose because of an error when we submitted our pre-registration for Study 6, which was intended to have the same policy as Study 5. We decided it was best to keep to the publicly pre-registered plan for this study. The final sample comprised 369 males, 455 females, ages 19-73 (*M* = 37.5, *SD*= 12.0), 98% first-language English.

Stimuli

Study 1. We randomly sampled photos of scientists from the departmental websites of the top-200 US Universities (National University Rankings, 2014). We randomly selected a university; if it had a genetics/human genetics department, then we randomly selected 10 photos of scientists from their departmental web pages; only photos of main faculty were selected. If the university did not have a genetics/human genetics department, then we randomly selected another university and randomly sampled 10 photographs of scientists from their web pages, and so on. Sampling continued until we had acquired at least 250 photos. This procedure was repeated for physics departments. We edited the photos (254 geneticists and 271 physicists) to have a grey background and cropped them to start at the top of the head and finish immediately below the chin, and to be reasonably centred. Images that were below 130 pixels in height were removed, and the remaining images were resized to have a height of 130 pixels. Poor-quality images were excluded, resulting in a final stimuli set of 108 photos of geneticists and 108 photos of physicists.

Study 2. We randomly sampled photos of biologists and physicists who had been submitted to the UK's 2014 Research Excellence Framework (REF), a nationwide audit of university research. The power calculation was based on one of the smallest effects of interest in Study 1, that of sociability of "Good scientist" judgments (partial R^2 = .024); 85% power to detect this effect in a simple multiple regression requires at least 368 participants, so we sampled a total of 400 photos (200 from each discipline). We drew up a list of all scientists submitted to the relevant "unit of assessment" (Biological Science or Physics). After excluding 3 universities for which no photos were available, we randomly sampled from this list in proportion to the number of individuals from each university (e.g., the University of Cambridge constituted roughly 8% of the total researchers evaluated within the Biological Sciences unit of assessment, so Cambridge contributed roughly 8% of the biologists within our set of 200 photos). If the scientist selected did not have a suitable photo on the university webpages then we randomly selected another scientist within that university; if were unable to reach the desired number of photos for a given university then we randomly sampled from the whole list of scientists. The photos were cropped around the top of the head and the shoulders, and standardised to 150 pixels in height. Any photos that were too blurry were replaced using the original sampling procedure.

Studies 3 and 4. The titles of 60 science news stories were collected from ScienceDaily.com, 30 from the "Health and Medicine" category, and 30 from the "Physics" category. In the pre-rating task, 105 participants were presented with either the biology or the physics titles, in a random order, and rated them on how interested they would be in reading the full article (0 – not at all interested, 10 – extremely interested). Mean interest ratings were computed for each title, averaging across participants. The titles selected for later studies had average ratings close to the mid-point of the scale, and similar ratings to each other (Table S2).

Biology Article Titles	Mean Rating
Opinions on vaccinations heavily influenced by online comments	5.12
Confidence in government linked to willingness to vaccinate	5.17
Texting may be more suitable than apps in treatment of mental illness*	5.19
Cow immune system inspires potential new therapies*	5.27
Reasons why winter gives flu a leg up could be key to prevention*	5.35
Stress balls, DVDs and conversation ease pain, anxiety during surgery*	5.37
Risk for autism increases for abandoned children placed in institutions	5.38
Elementary teachers' depression symptoms related to students' learning	5.52
Physics Article Titles	**Mean Rating**
Laser pulse turns glass into a metal: New effect could be used for ultra-fast logical switches	5.13
Doing more with less: Steering a quantum path to improved internet security	5.17
A 'Star Wars' laser bullet -- this is what it really looks like	5.23
'Solid' light could compute previously unsolvable problems	5.25
How to make mobile batteries last longer by controlling energy flows at nano-level	5.26
Universe may face a darker future: Is dark matter being swallowed up by dark energy?	5.32
Hunt for Big Bang particles offering clues to the origin of the universe	5.45
Electronics that need very little energy? Nanotechnology used to help cool electrons with no external sources	5.45

Table S2. Mean interest ratings for article titles used in Studies 3 and 4. Titles marked with an asterisk were used in Study 4.

The faces used in Study 4 were selected to score low or high on competence and attractiveness. Table S3 lists the mean ratings that the chosen faces had received in Study 2. As noted in the main text, the Competence manipulation was stronger than the Attractiveness manipulation: the mean Interest rating for the low-attractiveness faces is 4.89; that for the high-attractiveness faces is 5.38, giving a difference of only 0.49, compared with the difference of 0.99 between the low-competence and high-competence faces. Likewise, the low-attractiveness faces received “Good scientist” ratings that were only are only 0.86 above those of the high-attractiveness faces, as compared with a difference of 1.96 between the high- and low-competence faces.

	Low Competence		High Competence	
	Low attractiveness	High attractiveness	Low attractiveness	High attractiveness
Attractiveness	2.65	5.60	2.81	5.12
Competence	4.62	5.02	6.65	6.69
Interest	4.23	5.05	5.55	5.71
"Good Scientist"	4.96	4.34	7.16	6.06
Age	42.38	26.07	52.62	42.02
Sociability	5.80	4.61	5.64	4.91
Morality	5.16	5.23	6.14	5.74
Warmth	5.48	4.92	5.89	5.32

Table S3. Mean ratings for the face stimuli used in Studies 4 and 6.

Studies 5 and 6. Twenty scientific articles (10 biology and 10 physics) were selected from news websites (e.g., newser.com) and re-written in first person, and in an accessible fashion, simulating "scientist profiles" found in magazines. In the pre-rating task, 128 participants saw 5 biology and 5 physics articles (randomly selected and displayed), and rated them on questions related to the quality of research presented in the article, as well as their comprehension and recognition of the work, using 7-point scales (1=Not at all, 7=Extremely). Trials where the participant recognized the research were excluded, and mean quality ratings were computed by averaging across participants and questions. The articles selected for use had ratings close to the mid-point of the scale, good scores on comprehension and less than 10% recognition rate (Table S4).

Biology Articles	Mean Quality	Mean Comprehension	Recognition
Study Suggests Earth Life Began on Mars	3.98	4.76	6.35%
Slime Mould Is Smarter Than You Think	4.39	4.97	4.76%
Beneath Pacific Lies Ancient, Barely Alive Bacteria	4.52	5.18	1.49%
Earth Holds 8.7M Species, and Most of Them are Still Undiscovered	4.52	5.17	4.76%
Physics Articles	Mean Quality	Mean Comprehension	Recognition
Dark Matter Particles Detected Deep in Mine	4.01	4.84	9.52%
Bloodhound Diary: It's rocket science	4.53	5.29	4.84%
World's Next Timekeeper: Quantum Superclock?	4.67	4.63	0%
Final chapter to be published, in decades-long Gravity Probe B project	4.69	3.91	1.56%

Table S4. Titles of articles use in studies 5 and 6; Study 6 only used the Physics stories.

Counterbalancing

Study 3. We constructed three 8x8 Latin-squares that equally allocated articles to faces for each discipline, and four counterbalancing tables that equally allocated face-types to disciplines. Combining these gave 24 versions of the task, with participants randomly assigned to a version.

Study 4. The four article titles were paired with the four cells of the design (low/high attractiveness and low/high competence) using a 4x4 Latin Square; participants were randomly allocated to a version. One of the two photos with the appropriate attractiveness-competence combination was randomly selected on each trial.

Study 5. For each of the four gender-discipline combinations, we constructed a 4x4 Latin Square that ensured that each article was assigned to each face type equally often. Participants were randomly allocated to one of the resulting 16 versions of the task. On each trial, one of the two faces of the relevant type was randomly selected to be displayed alongside the article.

Study 6. The four articles were assigned to the four cells of the design using a 4x4 Latin square, creating four versions of the task with random allocation to version. On each trial, one of the two faces with the relevant competence-attractiveness combination was selected to be displayed alongside the allocated article.

Consistency of measures

Table S5 shows Cronbach's alphas for the measures in Studies 1 and 2, and indicated good consistency. For Study 1, Cronbach's Alpha was calculated separately for each face-set (participants saw one of two face-sets) and the average is reported. For Study 2 there were 3 pairs of face-sets, where one member of each pair comprised 50% of the faces and the other member comprised the complimentary 50%. Cronbach's alpha was calculated for each combination of 3 non-complimentary sets (i.e., with no members of a pair in the analysis) and an average taken.

Type	Measure	Study 1	Study 2
Predictor variables	Capable	-	0.74
	Competent	0.85	0.78
	Effective	-	0.72
	Intelligent	0.88	0.78
	Friendly	-	0.93
	Likeable	0.91	0.84
	Sociable	-	0.91
	Warm	-	0.88
	Kind	0.92	-
	Fair	-	0.75
	Honest	0.89	0.81
	Moral	-	0.79
	Trustworthy	0.88	0.79
	Age	0.99	0.99
	Physically Attractive	0.95	0.91
Composite measures	Competence	0.92	0.91
	Sociability	0.95	0.95
	Morality	0.95	0.92
Criterion variables	"Good Scientist"	0.89	0.89
	Interest	0.72	0.75

Table S5. Cronbach's Alpha values for the individual traits, outcomes and composite measures in Studies 1 and 2.

Confirmatory Factor Analysis

The correlations between the trait ratings are shown in Table S6 and S7, and are consistent with the three-factor structure we were expecting.

	Intelligent	Likeable	Kind	Trustworthy	Honest
Competent	0.860*	0.385*	0.404*	0.543*	0.522*
Intelligent		0.310*	0.376*	0.515*	0.511*
Likeable			0.914*	0.799*	0.850*
Kind				0.850*	0.904*
Trustworthy					0.903*

Table S6. Correlations among the items forming each trait for Study 1 (indicates p <.05).*

	Intel.	Capab.	Effect.	Lik.	Soc.	Friend.	Warm	Trust.	Hon.	Mor.	Fair
Competent	0.678*	0.715*	0.727*	0.295*	0.207*	0.155*	0.244*	0.454*	0.405*	0.459*	0.338*
Intelligent		0.737*	0.675*	0.123*	0.054	0.069	0.073	0.278*	0.319*	0.315*	0.135*
Capable			0.733*	0.210*	0.099*	0.097	0.132*	0.39*	0.404*	0.387*	0.192*
Effective				0.231*	0.144*	0.093	0.132*	0.361*	0.400*	0.364*	0.241*
Likeable					0.819*	0.806*	0.825*	0.742*	0.727*	0.729*	0.799*
Sociable						0.890*	0.867*	0.597*	0.612*	0.576*	0.737*
Friendly							0.912*	0.64*	0.692*	0.624*	0.744*
Warm								0.677*	0.673*	0.645*	0.753*
Trustworthy									0.786*	0.808*	0.709*
Honest										0.803*	0.702*
Moral											0.704*

Table S7. Correlations between the items forming each trait, for Study 2 (indicates p <.05).*

For both studies, we ran a CFA on the three-factor model: competence (comprising Competent and Intelligent in Study 1, and Competent, Intelligent, Capable, and Effective in Study 2), sociability (Study 1: Likeable and Kind; Study 2: Likeable, Sociable, Friendly, and Warm) and morality (Study 1: Trustworthy and Honest; Study 2: Trustworthy, Honest, Moral, and Fair). The models had an acceptable fit for both Study 1 (SRMR = .018, RMSEA = .102, CFI = .991, TLI = .978, BIC = 2056.26) and Study 2 (SRMR = .064, RMSEA = .128, CFI = .933, TLI = .913, BIC = 5579.77), supporting the three-factor model of social judgement.

For both studies, the three-factor model was a better fit than an alternative "competence and warmth" two-factor model in which "warmth" combines morality and sociability. For Study 1, the two-factor model had SRMR = .051, RMSEA = .247, CFI = .931, TLI = .870, BIC = 2139.70, and a chi-square test for the difference in model fit gave χ^2diff(2) = 94.19, p<.001. For Study 2, SRMR = .107, RMSEA = .190, CFI = .845, TLI = .808, BIC = 6004.34, χ^2diff(2) = 436.56, p<.001). The three-factor model also fit better than a single-factor model; Study 1: SRMR = .142, RMSEA = .419, CFI = .776, TLI = .627, BIC = 2370.55; χ^2diff(3) = 330.41, p<.001; Study 2: SRMR = .204, RMSEA = .284, CFI = .650, TLI = .572, BIC = 6969.18, χ^2diff(3) = 1407.4, p<.001).

Correlations between the composite traits and the criterion variables in Studies 1 and 2 are shown in Table S8.

	Competence	Sociability	Morality	"Good Scientist"	Interest
Competence		0.39*	0.555*	0.778*	0.505*
Sociability	0.168*		0.893*	0.098	0.632*
Morality	0.424*	0.798*		0.304*	0.624*
"Good Scientist"	0.689*	-0.069	0.163*		0.182*
Interest	0.585*	0.422*	0.534*	0.279*	

*Table S8. Correlations between the composite traits and criterion variables. The top-half of the table (above the diagonal) represents Study 1, the bottom half represents Study 2 (*indicates p <.05).*

Regression coefficients

Tables S9-S13 give the numeric values of the regression coefficients plotted in the main text, along with 95% CIs and *p*-values based on Satterthwaite-adjusted degrees of freedom. In all tables, P = participant, Sci = science engagement. Unless otherwise noted, all predictors were standardized (z-scored), with standardization prior to computing any interaction terms. Standardization was across participants for participant-level variables and across faces for face-level variables.

INTEREST JUDGMENTS								
	Study 1				Study 2			
	B	CI_{low}	CI_{high}	*p*	B	CI_{low}	CI_{high}	*p*
Age	0.047	-0.072	0.166	0.446	0.074	0.012	0.137	0.021
Female	-0.115	-0.242	0.011	0.084	-0.051	-0.141	0.039	0.268
Non-white	-0.009	-0.079	0.060	0.792	0.032	-0.014	0.078	0.176
Physics	0.013	-0.036	0.062	0.613	-0.013	-0.044	0.018	0.406
Attractiveness	0.374	0.233	0.516	<.001	0.213	0.142	0.284	<.001
Competence	0.136	0.022	0.251	0.026	0.200	0.122	0.277	<.001
Sociability	0.059	-0.109	0.226	0.496	0.049	-0.032	0.131	0.236
Morality	0.124	-0.007	0.255	0.068	0.132	0.039	0.225	0.006
P_Age	0.030	-0.332	0.393	0.872	0.020	-0.226	0.265	0.876
P_Female	-0.169	-0.564	0.226	0.409	0.273	0.024	0.523	0.034
P_Sci	0.382	-0.020	0.785	0.073	0.232	-0.017	0.482	0.071
"GOOD SCIENTIST" JUDGMENTS								
	Study 1				Study 2			
	B	CI_{low}	CI_{high}	*p*	B	CI_{low}	CI_{high}	*p*
Age	0.177	0.056	0.298	0.007	0.059	-0.019	0.137	0.140
Female	-0.068	-0.158	0.022	0.143	0.023	-0.072	0.119	0.633
Non-white	0.079	-0.034	0.192	0.179	0.040	-0.014	0.094	0.146
Physics	0.039	-0.015	0.094	0.160	0.024	-0.019	0.067	0.283
Attractiveness	-0.252	-0.382	-0.122	<.001	-0.325	-0.415	-0.235	<.001
Competence	0.698	0.578	0.819	<.001	0.516	0.429	0.604	<.001
Sociability	-0.152	-0.282	-0.022	0.023	-0.123	-0.203	-0.043	0.003
Morality	0.204	0.046	0.362	0.012	0.111	0.003	0.219	0.045
P_Age	-0.247	-0.565	0.070	0.138	-0.054	-0.275	0.167	0.635
P_Female	-0.099	-0.418	0.220	0.548	0.152	-0.072	0.376	0.187
P_Sci	0.026	-0.294	0.345	0.877	0.128	-0.084	0.340	0.239

Table S9. Regression coefficients for Studies 1 and 2.

To obtain a better estimate of the overall effect of each predictor, we pooled the data from Studies 1 and 2. The main text reports the results of simple pooling (with all predictors standardized across the pooled sample); Tables S10 and S11 show the regression coefficients, and those obtained when Study and its interactions are included as fixed-effects (with Study 1 coded -1, Study 2 coded +1; we did not standardize this variable). There is no indication that study modulated the other effects.

INTEREST JUDGMENTS

	Studies 1 and 2 Pooled Data				Pooled Data with Effect of Study			
	B	CI_{low}	CI_{high}	*p*	B	CI_{low}	CI_{high}	*p*
Age	0.073	0.017	0.130	0.012	0.061	-0.005	0.128	0.070
Female	-0.063	-0.138	0.012	0.101	-0.084	-0.173	0.006	0.069
Non-white	0.026	-0.017	0.069	0.232	0.015	-0.032	0.062	0.531
Physics	-0.007	-0.034	0.019	0.586	-0.000	-0.031	0.030	0.979
Attractiveness	0.266	0.197	0.336	<.001	0.285	0.204	0.365	<.001
Competence	0.215	0.139	0.292	<.001	0.177	0.090	0.263	<.001
Sociability	0.057	-0.018	0.132	0.140	0.051	-0.041	0.143	0.276
Morality	0.149	0.058	0.240	0.002	0.139	0.044	0.234	0.005
P_Age	-0.024	-0.235	0.186	0.820	0.028	-0.620	0.677	0.932
P_Female	0.185	-0.032	0.402	0.097	0.048	-0.254	0.350	0.756
P_Sci	0.265	0.048	0.481	0.018	0.319	0.018	0.620	0.040
Study					-0.195	-0.876	0.486	0.576
Study*Age					0.015	-0.051	0.081	0.650
Study*Female					0.032	-0.058	0.122	0.485
Study*Non-white					0.023	-0.024	0.069	0.347
Study*Physics					-0.013	-0.044	0.017	0.391
Study*Att					-0.037	-0.118	0.043	0.365
Study*Comp					0.070	-0.016	0.156	0.114
Study*Soc					0.001	-0.091	0.093	0.979
Study*Mor					0.034	-0.061	0.129	0.483
Study*P_Age					0.000	-0.648	0.649	0.999
Study*P_Female					0.204	-0.098	0.506	0.187
Study*P_Sci					-0.095	-0.397	0.206	0.537

Table S10. Regression coefficients for Interest Judgments when data from Studies 1 and 2 are pooled, either with or without including Study and its interactions with other variables.

"GOOD SCIENTIST" JUDGMENTS								
	Studies 1 and 2 Pooled Data				Pooled Data with Effect of Study			
	B	CI_{low}	CI_{high}	*p*	B	CI_{low}	CI_{high}	*p*
Age	0.094	0.027	0.160	0.006	0.115	0.040	0.191	0.003
Female	-0.004	-0.077	0.070	0.921	-0.022	-0.109	0.064	0.613
Non-white	0.056	0.005	0.107	0.034	0.056	0.002	0.110	0.044
Physics	0.029	-0.006	0.063	0.102	0.031	-0.007	0.069	0.109
Attractiveness	-0.331	-0.414	-0.247	<.001	-0.297	-0.392	-0.202	<.001
Competence	0.600	0.518	0.681	<.001	0.592	0.503	0.682	<.001
Sociability	-0.139	-0.206	-0.072	<.001	-0.135	-0.215	-0.056	0.001
Morality	0.167	0.073	0.262	0.001	0.158	0.058	0.258	0.002
P_Age	-0.009	-0.193	0.175	0.927	-0.258	-0.654	0.138	0.204
P_Female	0.072	-0.112	0.257	0.444	0.024	-0.218	0.266	0.846
P_Sci	0.100	-0.083	0.283	0.284	0.087	-0.151	0.325	0.475
Study					0.290	-0.147	0.726	0.195
Study*Age					-0.054	-0.130	0.021	0.160
Study*Female					0.046	-0.041	0.132	0.301
Study*Non-white					-0.009	-0.063	0.046	0.759
Study*Physics					-0.008	-0.046	0.031	0.694
Study*Att					-0.081	-0.176	0.014	0.098
Study*Comp					0.044	-0.045	0.134	0.333
Study*Soc					0.003	-0.077	0.083	0.945
Study*Mor					-0.013	-0.112	0.087	0.803
Study*P_Age					0.206	-0.190	0.602	0.310
Study*P_Female					0.126	-0.116	0.367	0.310
Study*P_Sci					0.029	-0.209	0.267	0.812

Table S11. Regression coefficients for "Good Scientist" judgments when data from Studies 1 and 2 are pooled, either with or without including Study and its interactions with other variables.

Testing for modulation by participant gender

Study 1 had more female participants than males (Study 2 showed a slight preponderance of males). To test whether participant gender modulates our results we re-ran our regression analyses including participant gender and its interaction with all other variables as additional predictors. (The analysis was as before, including the standardization of variables prior to computing interaction terms. We did not include by-face random slopes for any of these interaction terms because of convergence problems.) The regression coefficients are shown in Tables S12 and S13 and indicate that our results are consistent across male and female participants (only one of 40 interaction terms has $p < .05$). Likewise, comparison of the models with and without the interaction terms indicated that the simpler models are to be preferred: Interest judgments for Study 1, χ^2 (10) = 5.50, p = .955, BIC_{no_int} = 22137, BIC_{int} = 22218; for Study 2, χ^2 (10) = 11.90, p = .292, BIC_{no_int} = 82123, BIC_{int} = 82210; "Good Scientist" judgments for Study 1 χ^2 (10) = 11.89, p = .292, BIC_{no_int} = 21846, BIC_{int} = 21921; for Study 2, χ^2 (10) = 8.13, p = .616, BIC_{no_int} = 81706, BIC_{int} = 81797.

INTEREST JUDGMENTS

	Study 1				Study 2			
	B	CI_{low}	CI_{high}	*p*	B	CI_{low}	CI_{high}	*p*
Intercept	4.912	4.503	5.320	0.000	4.831	4.584	5.079	0.000
Age	0.047	-0.069	0.162	0.433	0.074	0.012	0.137	0.021
Female	-0.115	-0.241	0.011	0.082	-0.051	-0.141	0.039	0.266
Non-white	-0.009	-0.079	0.060	0.792	0.032	-0.014	0.077	0.176
Physics	0.013	-0.036	0.062	0.613	-0.013	-0.044	0.018	0.408
Attractiveness	0.374	0.235	0.514	0.000	0.213	0.143	0.284	0.000
Competence	0.136	0.022	0.251	0.026	0.200	0.123	0.277	0.000
Sociability	0.059	-0.107	0.225	0.492	0.049	-0.032	0.131	0.235
Morality	0.124	-0.008	0.255	0.069	0.132	0.039	0.224	0.006
P_Age	-0.071	-0.718	0.577	0.832	0.036	-0.214	0.286	0.776
P_Female	-0.218	-0.700	0.264	0.383	0.275	0.027	0.524	0.032
P_Sci	0.388	-0.007	0.782	0.064	0.247	-0.005	0.499	0.057
P_Female*Age	-0.072	-0.185	0.041	0.222	-0.037	-0.096	0.022	0.220
P_Female*Female	-0.021	-0.146	0.104	0.744	0.037	-0.050	0.124	0.401
P_Female*Non-white	0.006	-0.060	0.073	0.852	-0.025	-0.068	0.018	0.251
P_Female*Physics	-0.018	-0.062	0.026	0.416	-0.020	-0.047	0.007	0.147
P_Female*Att	0.027	-0.110	0.164	0.701	0.024	-0.043	0.091	0.491
P_Female*Comp	0.004	-0.109	0.117	0.943	0.060	-0.014	0.133	0.115
P_Female*Soc	0.084	-0.075	0.243	0.306	0.008	-0.068	0.085	0.831
P_Female*Mor	-0.032	-0.149	0.085	0.592	0.036	-0.047	0.120	0.396
P_Female*P_Age	0.235	-0.715	1.184	0.632	0.015	-0.237	0.266	0.910
P_Female*P_Sci	0.237	-0.173	0.647	0.268	-0.131	-0.384	0.121	0.311

Table S12. Regression coefficients for Interest judgments when interactions between Participant Gender and all other predictors are included.

"GOOD SCIENTIST" JUDGMENTS								
	Study 1				Study 2			
	B	CI_{low}	CI_{high}	*p*	B	CI_{low}	CI_{high}	*p*
Intercept	5.664	5.360	5.967	0.000	5.752	5.526	5.977	0.000
Age	0.177	0.057	0.297	0.006	0.059	-0.019	0.136	0.141
Female	-0.068	-0.156	0.019	0.132	0.023	-0.072	0.119	0.635
Non-white	0.079	-0.034	0.192	0.178	0.040	-0.014	0.093	0.147
Physics	0.039	-0.015	0.094	0.160	0.024	-0.019	0.067	0.282
Attractiveness	-0.252	-0.381	-0.123	0.000	-0.325	-0.415	-0.235	0.000
Competence	0.698	0.579	0.818	0.000	0.516	0.429	0.604	0.000
Sociability	-0.152	-0.282	-0.022	0.023	-0.123	-0.203	-0.043	0.003
Morality	0.204	0.046	0.362	0.012	0.112	0.004	0.220	0.043
P_Age	-0.205	-0.578	0.167	0.289	-0.113	-0.354	0.128	0.360
P_Female	-0.137	-0.448	0.175	0.397	0.167	-0.057	0.390	0.148
P_Sci	0.023	-0.298	0.344	0.888	0.133	-0.078	0.344	0.218
P_Female*Age	0.058	-0.054	0.169	0.320	0.019	-0.042	0.080	0.544
P_Female*Female	-0.047	-0.124	0.029	0.236	0.063	-0.018	0.144	0.131
P_Female*Non-	0.024	-0.084	0.133	0.662	0.014	-0.025	0.054	0.469
P_Female*Physics	0.019	-0.023	0.061	0.385	0.002	-0.024	0.027	0.907
P_Female*Att	0.032	-0.085	0.149	0.597	0.019	-0.056	0.095	0.617
P_Female*Comp	0.046	-0.067	0.158	0.434	0.030	-0.042	0.102	0.422
P_Female*Soc	0.113	0.013	0.212	0.028	-0.014	-0.062	0.034	0.569
P_Female*Mor	-0.100	-0.228	0.028	0.128	-0.020	-0.084	0.045	0.550
P_Female*P_Age	-0.151	-0.637	0.335	0.548	0.145	-0.091	0.380	0.232
P_Female*P_Sci	0.181	-0.179	0.540	0.334	0.025	-0.186	0.236	0.818

Table S13. Regression coefficients for "Good Scientist" judgments when interactions between Participant Gender and all other predictors are included.

Results for each gender and ethnicity

For completeness (and as requested by a reviewer), Tables S14 and S15 present the results of analysing the pooled data from Studies 1 and 2 split by face-gender and face-ethnicity. (The sample sizes are small for the female- and non-white groups, resulting in low power.)

INTEREST JUDGMENTS

	Male Faces				Female Faces			
	B	CI_{low}	CI_{high}	*p*	B	CI_{low}	CI_{high}	*p*
Intercept	4.748	4.534	4.963	0.000	5.103	4.851	5.355	0.000
Age	0.099	0.038	0.161	0.002	-0.061	-0.146	0.024	0.160
Non-white	0.034	-0.011	0.080	0.137	-0.007	-0.080	0.065	0.843
Physics	-0.005	-0.034	0.024	0.738	-0.020	-0.072	0.031	0.442
Attractiveness	0.237	0.169	0.305	0.000	0.230	0.119	0.341	0.000
Competence	0.212	0.129	0.296	0.000	0.198	0.107	0.289	0.000
Sociability	0.065	-0.010	0.141	0.091	0.001	-0.109	0.111	0.985
Morality	0.129	0.045	0.214	0.003	0.118	-0.014	0.250	0.083
P_Age	-0.015	-0.230	0.200	0.891	-0.070	-0.323	0.184	0.590
P_Female	0.155	-0.066	0.376	0.172	0.288	0.028	0.548	0.032
P_Sci	0.303	0.082	0.524	0.008	0.107	-0.151	0.366	0.417

	White Faces				Non-white Faces			
	B	CI_{low}	CI_{high}	*p*	B	CI_{low}	CI_{high}	*p*
Intercept	4.816	4.606	5.026	0.000	4.873	4.615	5.131	0.000
Age	0.074	0.014	0.133	0.016	0.017	-0.076	0.110	0.722
Female	-0.057	-0.132	0.019	0.145	-0.095	-0.224	0.033	0.153
Physics	-0.013	-0.041	0.015	0.375	0.058	-0.021	0.137	0.155
Attractiveness	0.259	0.187	0.330	0.000	0.299	0.184	0.415	0.000
Competence	0.217	0.140	0.295	0.000	0.125	0.014	0.237	0.034
Sociability	0.063	-0.015	0.141	0.115	0.035	-0.129	0.199	0.680
Morality	0.143	0.049	0.237	0.003	0.149	-0.061	0.358	0.171
P_Age	-0.025	-0.236	0.186	0.816	-0.046	-0.301	0.210	0.727
P_Female	0.193	-0.024	0.410	0.084	0.101	-0.166	0.369	0.458
P_Sci	0.276	0.060	0.492	0.014	0.147	-0.114	0.408	0.273

Table S14. Regression coefficients for Interest Judgments when data from Studies 1 and 2 are pooled, with separate analyses for each gender and ethnicity.

"GOOD SCIENTIST" JUDGMENTS								
	Male Faces				Female Faces			
	B	CI_{low}	CI_{high}	*p*	B	CI_{low}	CI_{high}	*p*
Intercept	5.808	5.625	5.992	0.000	5.564	5.335	5.792	0.000
Age	0.141	0.070	0.212	0.000	-0.155	-0.261	-0.049	0.005
Non-white	0.063	0.008	0.117	0.024	0.035	-0.058	0.128	0.462
Physics	0.026	-0.011	0.064	0.174	0.026	-0.044	0.096	0.462
Attractiveness	-0.271	-0.350	-0.193	0.000	-0.450	-0.579	-0.321	0.000
Competence	0.596	0.511	0.681	0.000	0.511	0.402	0.620	0.000
Sociability	-0.134	-0.204	-0.065	0.000	-0.161	-0.296	-0.026	0.021
Morality	0.161	0.071	0.252	0.001	0.132	-0.027	0.291	0.107
P_Age	-0.014	-0.199	0.170	0.879	-0.013	-0.242	0.215	0.911
P_Female	0.049	-0.137	0.234	0.608	0.186	-0.044	0.415	0.115
P_Sci	0.028	-0.156	0.211	0.767	0.409	0.184	0.634	0.001
	White Faces				Non-white Faces			
	B	CI_{low}	CI_{high}	*p*	B	CI_{low}	CI_{high}	*p*
Intercept	5.701	5.517	5.885	0.000	6.260	6.030	6.490	0.000
Age	0.089	0.019	0.159	0.014	0.027	-0.102	0.157	0.682
Female	0.001	-0.074	0.076	0.979	-0.058	-0.202	0.086	0.434
Physics	0.033	-0.004	0.069	0.082	0.022	-0.077	0.121	0.663
Attractiveness	-0.343	-0.429	-0.257	0.000	-0.260	-0.399	-0.121	0.001
Competence	0.606	0.523	0.688	0.000	0.507	0.370	0.643	0.000
Sociability	-0.150	-0.222	-0.079	0.000	-0.027	-0.231	0.178	0.800
Morality	0.172	0.073	0.271	0.001	0.136	-0.113	0.384	0.289
P_Age	-0.012	-0.198	0.173	0.896	-0.006	-0.222	0.209	0.955
P_Female	0.069	-0.118	0.255	0.471	0.101	-0.114	0.315	0.360
P_Sci	0.100	-0.085	0.284	0.292	0.076	-0.135	0.288	0.480

Table S15. Regression coefficients for "Good Scientist" Judgments when data from Studies 1 and 2 are pooled, with separate analyses for each gender and ethnicity.

	Predictor	B	CI_{low}	CI_{high}	*p*
Study 3	Video	0.104	0.030	0.178	.006
	Physics	-0.096	-0.178	-0.013	.023
	Gender	-0.017	-0.098	0.064	.682
	Video*Physics	-0.024	-0.106	0.059	.574
	Video*Female	0.056	-0.025	0.136	.178
	Physics*Female	0.118	0.025	0.212	.013
	Video*Physics*Female	0.075	-0.018	0.168	.116
	P_Age	-0.089	-0.164	-0.014	.020
	P_Female	-0.134	-0.210	-0.058	<.001
	P_Sci	-0.028	-0.104	-0.048	.467
Study 4	Competence	0.083	0.007	0.158	.032
	Attractiveness	0.052	-0.027	0.131	.196
	Comp*Att	-0.059	-0.129	0.010	.093
	P_Age	0.104	-0.014	0.223	.084
	P_Female	0.124	0.007	0.240	.039
	P_Sci	0.380	0.261	0.499	<.001
	Comp*P_Female	-0.051	-0.129	0.026	.196
	Comp*P_Age	-0.023	-0.099	0.054	.564
	Comp*P_Sci	0.029	-0.049	0.107	.471
	Att*P_Female	0.009	-0.072	0.091	.821
	Att*P_Age	0.060	-0.021	0.140	.148
	Att*P_Sci	-0.038	-0.120	0.043	.357

Table S16. Regression coefficients for Studies 3 and 4.

	Predictor	B	CI_{low}	CI_{high}	*p*
Study 5	Facetype	0.161	0.083	0.238	<.001
	Physics	0.119	0.028	0.210	.011
	Female	0.017	-0.074	0.108	.716
	Facetype*Physics	0.001	-0.076	0.078	.978
	Facetype*Female	0.024	-0.054	0.101	.550
	Physics*Female	0.009	-0.082	0.101	.844
	Facetype*Physics*Female	-0.051	-0.129	0.026	.193
	P_Female	0.081	-0.014	0.175	.094
	P_Age	-0.068	-0.160	0.024	.150
	P_Sci	0.150	0.055	0.245	.002
Study 6	Competence	0.142	0.104	0.179	<.001
	Attractiveness	-0.017	-0.053	0.020	.368
	Comp. * Att.	-0.016	-0.052	0.021	.402
	P_Female	0.102	0.041	0.163	.001
	P_Age	-0.080	-0.140	-0.020	.009
	P_Sci	0.094	0.033	0.155	.003
	Comp*P_Female	0.001	-0.037	0.040	.950
	Comp*P_Age	0.013	-0.024	0.050	.497
	Comp*P_Sci	0.037	-0.002	0.075	.060
	Att*P_Female	-0.010	-0.047	0.028	.610
	Att*P_Age	0.006	-0.031	0.042	.758
	Att*P_Sci	-0.022	-0.059	0.015	.252

Table S17. Regression coefficients for Studies 5 and 6

Additional Study

We conducted an additional face-rating study using 200 neutral-expression photos from the Park Aging Mind Face Database [2], comprising 25 men and 25 women in each of 4 age-bands (18-29; 30-49; 50-69; 70-94), with 10-15% non-white faces per group. Using an MTurk sample, we had 80 participants rate the faces on the same traits as in Study 1; the faces were randomly divided into 2 sets and 40 participants rated all the faces in each set on all dimensions. A further 30 participants gave Interest judgments for each face and 30 gave "Good scientist" judgments, like those in Studies 1 and 2. In addition, 31 participants were asked to indicate how far each person looked like "a scientist" (Scientist judgments; this differs from the other judgments, which are predicated upon the person being a scientist, and was an exploratory variable). Other aspects of the procedure and analysis were the same as for Studies 1 and 2.

The regression coefficients are shown in Table S18. The pattern is similar to our main studies: the confidence intervals are often somewhat wider, most likely because of the smaller samples and more heterogeneous stimuli, but interest was again greater for attractive and competent-looking faces and for older individuals (although there was little effect of perceived morality), and good-scientist judgments were positively related to apparent competence and morality but negatively associated with attractiveness and perceived sociability. "Good Scientist" ratings were also lower for non-white than for white faces, possibly because there were more African-American faces in this face-set. Also in contrast to our mains studies is the finding that females received lower "Good Scientist" ratings than did males.

	INTEREST JUDGMENTS			
	B	CI_{low}	CI_{high}	*p*
Age	0.198	-0.047	0.442	0.122
Female	-0.048	-0.185	0.088	0.492
Non-white	0.051	-0.028	0.131	0.213
Attractiveness	0.199	0.016	0.382	0.040
Competence	0.197	0.035	0.358	0.022
Sociability	0.079	-0.059	0.217	0.263
Morality	0.003	-0.161	0.167	0.973
P_Age	0.472	-0.016	0.960	0.068
P_Female	0.279	-0.229	0.787	0.290
P_Sci	0.408	-0.082	0.899	0.113
	"GOOD SCIENTIST" JUDGMENTS			
	B	CI_{low}	CI_{high}	*p*
Age	-0.163	-0.402	0.076	0.185
Female	-0.249	-0.428	-0.069	0.009
Non-white	-0.116	-0.221	-0.011	0.034
Attractiveness	-0.161	-0.359	0.036	0.112
Competence	0.914	0.724	1.104	0.001
Sociability	-0.204	-0.434	0.027	0.085
Morality	0.233	-0.058	0.524	0.119
P_Age	0.055	-0.362	0.471	0.799
P_Female	-0.243	-0.659	0.173	0.262
P_Sci	0.417	0.000	0.833	0.06
	SCIENTIST JUDGMENTS			
	B	CI_{low}	CI_{high}	*p*
Age	-0.268	-0.483	-0.053	0.016
Female	-0.182	-0.346	-0.018	0.034
Non-white	-0.028	-0.122	0.066	0.557
Attractiveness	-0.355	-0.577	-0.132	0.002
Competence	1.074	0.866	1.282	0.000
Sociability	-0.141	-0.390	0.109	0.270
Morality	0.134	-0.188	0.455	0.416
P_Age	0.073	-0.258	0.405	0.667
P_Female	-0.297	-0.651	0.057	0.110
P_Sci	0.038	-0.326	0.401	0.841

Table S18. Regression coefficients from Supplementary Study.

References

[1] Faul, F., Erdfelder, E., Lang, A-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175-191.

[2] Minear, M., & Park, D.C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers*, *36*(4), 630-633.

[3] Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, *17*(7), 592-598.