Divergent Perspectives on Expert Disagreement: Preliminary Evidence from Climate Science, Climate Policy, Astrophysics, and Public Opinion

Supplementary Materials

Peer Disagreement

According to the recent philosophical literature, at a rough approximation, two or more people are epistemic peers in a given domain of enquiry *E* if (1) they have equal knowledge, training, cognitive skills such as powers of reasoning, epistemic virtues such as rationality and impartiality, intelligence, and background information, relevant to *E*, (2) there is no substantial difference in their cognitive abilities and limitations, (3) they are equally well positioned to consider the available evidence regarding *E*, and (4) they have considered the available evidence regarding *E* with equal care and attention. Epistemic peers disagree when they have opposing and incompatible beliefs regarding *E*: two people have a disagreement if one of them believes a certain proposition *P* and the second disbelieves *P* or believes not *P*.

In cases of peer disagreement, most contributors to the debate defend some version of the view that one should move closer to one's peers' opinion, e.g., by suspending judgment or by adopting an intermediate level of confidence between the disagreeing peer and one's former self (e.g., Christensen 2007, Feldman 2009, Elga 2007). This family of views is known as *conciliationism*. In contrast, *steadfastness* holds that one should 'stand one's ground' in the face of peer disagreement, i.e., continue to have the same beliefs and levels of confidence as one did before the disagreement.

Although this is certainly a minority view in the literature, it does have its proponents (e.g., Kelly 2005, 2010).

Question 2.2 asks a general question about the possibility of reasonable and apparently faultless disagreement—i.e., disagreements where neither side seems to be making any obvious errors and disputants are equally well informed and possess the same evidence. It concerns the following principle that figures centrally in philosophical discussions of how one should respond to equally well informed peers with whom one disagrees (White 2005):

> *Uniqueness*. In any given evidential situation, there is only one attitude that it is rational to take toward a proposition in light of the evidence one possesses.

Cf. Christensen (2009) for an overview of how a principle like this figures in contemporary philosophical debates, although Christensen focuses his attention on a slightly stronger version of uniqueness.

Participants

	Total	**Ave. Age**	**% Female**	**Ave. Years Experience**	**% with Doctoral Degree**	**% with Master's Degree only**
Climate scientists	457	43	41	17	66	26
Climate policy experts	200	48	31	18	54	38
Undergraduates	697	23	52	n/a	n/a	n/a
Alumni	1,914	52	44	n/a	28	40%
Astrophysicists	99	49	17	22	91	4

Table S1. Participant demographics.

	UK or Ireland	Rest of Europe	U.S.	Rest of Americas	Other
Climate scientists	6%	28%	45%	8%	13%
Climate policy experts	14%	23%	30%	10%	23%
Astrophysicists	6%	59%	17%	9%	9%

Table S2. Nationalities of expert participants.

Additional Analyses, Figures, and Tables

D'Agostino-Pearson omnibus tests for normality (D'Agostino 1986) were performed on the distributions of each participant group's answers to each question. The D'Agostino-Pearson test works better for large samples than the Kolmogorov-Smirnov or Shapiro-Wilk tests. The D'Agostino-Pearson tests, along with visual inspection of histograms and P-P plots, revealed that a large majority of the distributions of responses were non-normal. The distributions of responses from climate scientists, undergraduates, and alumni were non-normal on every question, and approximately half of the answer distributions of climate policy experts and astrophysicists were non-normal. Shapiro-Wilk tests for normality, which are better suited for smaller sample sizes, were also performed of the distributions of responses from climate policy experts and astrophysicists. These tests indicated that the relevant answer distributions were non-normal. Logarithmic and square root transformations of the data failed to result in normal distributions. Therefore, with only a few exceptions, non-parametric tests were used in the analyses below when comparing the answers of different groups to the same question or the answers of members of the same group to different questions. Parametric one-sample *t*-tests were used above to test whether the mean response of each of the five participant groups to particular items differed from the neutral midpoint because there is not a good non-parametric equivalent and because it is the normality of sampling distributions rather than of sample data that is most central to the validity of the *t*-test.

Kruskal-Wallis tests were used on each question to examine overall between-group differences. Pairwise differences were then examined with post-hoc Mann-Whitney tests, and Benjamini-Hochberg corrections were used in each case to control for multiple comparisons.

The mean responses of the five participant groups to each of the items in Questionnaires 1 and 2 are summarized below in Tables S3 and S4. For ease of reference, the rightmost column indicates the midpoint of the response scale for each item. One-sample *t*-tests were run on the set of responses represented by each cell, in order to see whether they differed significantly from the relevant midpoint. All statistical tests reported in this paper are two-tailed. To correct for multiple comparisons, initial p values were adjusted using the Benjamini-Hochberg method. Mean responses whose adjusted *p* values were less than .05 are marked with a '*', those that were less than .01 are marked with a '**', and those below .001 are marked with a '***'.

Question	Climate scientists			Climate policy experts			Undergraduates			Alumni			Astrophysicists			Midpoint
	M	***SD***	***r***	***M***	***SD***	***r***	***M***	***SD***	***r***	***M***	***SD***	***r***	***M***	***SD***	***r***	
Q1.1	1.2***	0.8	.39	1.4	0.8	n/a	1.7***	0.8	.27	1.6***	0.8	.12	2.0***	0.5	.69	1.5
Q1.2	2.3**	1.3	.22	2.4**	1.2	.33	3.0***	1.1	.70	3.0***	1.1	.68	1.8	1.3	n/a	2
Q1.3	2.6***	1.2	.47	2.9***	1.0	.69	2.9***	1.0	.65	3.0***	1.0	.72	2.9***	1.0	.69	2
Q1.4	2.6***	1.2	.43	2.3*	1.2	.26	3.0***	1.0	.72	3.0***	1.0	.71	2.5**	1.0	.42	2
Q1.5	2.0	1.3	n/a	2.0	1.2	n/a	2.4***	1.2	.31	2.5***	1.2	.38	2.0	1.3	n/a	2
Q1.6	2.4***	1.2	.30	2.4***	1.1	.37	2.1*	1.2	.12	2.6***	1.2	.43	2.9***	1.1	.64	2
Q1.7	2.3**	1.3	.20	2.7***	1.1	.52	2.3***	1.2	.24	2.3***	1.3	.25	2.6***	1.0	.53	2
Q1.8	2.1	1.4	n/a	2.3	1.2	n/a	2.3***	1.3	.20	2.2***	1.4	.14	3.0***	0.9	.76	2
Q1.9	2.5***	1.5	.29	2.7***	1.4	.45	3.1***	1.2	.66	3.3***	1.1	.76	0.9***	1.2	.66	2
Q1.10	2.1	1.5	n/a	2.0	1.5	n/a	3.2***	1.1	.73	3.2***	1.2	.71	1.2***	1.4	.52	2
Q1.11	1.9	1.3	n/a	2.2	1.3	n/a	2.6***	1.2	.45	2.6***	1.3	.42	2.3	1.1	n/a	2
Q1.12	1.7***	1.2	.25	1.7	1.3	n/a	1.8**	1.3	.14	2.0	1.3	n/a	1.9	1.2	n/a	2
Q1.13	1.7**	1.3	.23	1.5**	1.3	.35	2.0	1.4	n/a	2.1	1.3	n/a	1.4**	1.2	.48	2

Table S3. Mean responses of the five participant groups to each of the items in Questionnaire 1, along with standard deviations, a measure of effect size, and an additional column indicating the midpoint of the relevant response scale. *R* values represent the size of the mean's difference from the neutral midpoint.

Question	Climate scientists			Climate policy experts			Undergraduates			Alumni			Astrophysicists			Midpoint
	M	***SD***	***r***	***M***	***SD***	***r***	***M***	***SD***	***r***	***M***	***SD***	***r***	***M***	***SD***	***r***	
Q2.1	1.0***	0.8	.52	1.5	0.8	n/a	1.8***	0.8	.30	1.6**	0.8	.11	2.0***	0.6	.63	1.5
Q2.2	3.6***	1.9	.24	4.1	1.9	n/a	4.6***	1.7	.32	4.6***	1.8	.33	4.8***	1.5	.45	4
Q2.3	3.1***	1.8	.46	3.5**	1.9	.27	4.0	1.8	n/a	4.0	1.9	n/a	4.8**	1.7	.40	4
Q2.4	2.3***	1.6	.74	2.7***	1.8	.58	2.9***	1.8	.52	3.2***	2.0	.36	4.3	1.8	n/a	4
Q2.5	3.3***	1.9	.35	3.7	2.0	n/a	4.6***	1.8	.34	4.2**	1.9	.10	4.1	1.9	n/a	4
Q2.6	3.6***	1.6	.27	3.2***	1.7	.43	3.5***	1.7	.27	2.9***	1.6	.57	4.5*	1.5	.33	4
Q2.7	5.0***	1.5	.56	5.0***	1.6	.53	4.7***	1.6	.41	4.8***	1.7	.45	4.1	1.6	n/a	4
Q2.8	3.1***	1.7	.46	3.7	1.6	n/a	4.4***	1.6	.21	3.8**	1.8	.10	4.0	1.6	n/a	4
Q2.9	2.7***	1.7	.63	3.4**	1.9	.30	4.2	1.8	n/a	3.8***	1.8	.12	3.2***	1.6	.44	4
Q2.10	4.5***	1.8	.25	4.5**	1.9	.27	4.4***	1.9	.20	4.8***	1.8	.42	5.5***	1.5	.70	4
Q2.11	5.1***	1.5	.59	5.2***	1.8	.57	4.8***	1.5	.48	4.8***	1.6	.43	n/a	n/a	n/a	4
Q2.12	4.6***	1.9	.31	4.2	1.9	n/a	4.8***	1.8	.39	4.1	1.8	n/a	3.6	1.7	n/a	4
Q2.13	3.6**	2.0	.19	3.1***	1.9	.45	3.4***	1.8	.31	3.6***	2.0	.20	3.7	1.9	n/a	4
Q2.14	6.0***	1.7	.76	5.8***	1.7	.74	5.2***	1.6	.60	4.8***	2.1	.38	n/a	n/a	n/a	4
Q2.15	n/a	n/a	n/a	n/a	n/a	n/a	2.6***	0.6	.88	2.5***	0.7	.81	n/a	n/a	n/a	1.5
Q2.16	n/a	n/a	n/a	n/a	n/a	n/a	2.7***	0.5	.92	2.7***	0.5	.92	n/a	n/a	n/a	1.5

Table S4. Mean responses of the five participant groups to each of the items in Questionnaire 2, along with standard deviations, a measure of effect size, and an additional column indicating the midpoint of the relevant response scale. *R* values represent the size of the mean's difference from the neutral midpoint.

Question	***df***	***H***	***p***
Q1.1	4	83.537	< .000001
Q1.2	4	99.78	< .000001
Q1.3	4	27.14	< .0001
Q1.4	4	61.76	< .000001
Q1.5	4	40.56	< .000001
Q1.6	4	40.09	< .000001
Q1.7	4	8.77	.067
Q1.8	4	16.83	< .01
Q1.9	4	146.64	< .000001
Q1.10	4	207.4	< .000001
Q1.11	4	54.89	< .000001
Q1.12	4	14.64	< .01
Q1.13	4	34.51	< .000001

Table S5. Results of Kruskal-Wallis tests for participants' responses to Questionnaire 1.

Question	***df***	***H***	***p***
Q2.1	4	130.64	< .000001
Q2.2	4	65.71	< .000001
Q2.3	4	72.42	< .000001
Q2.4	4	79.25	< .000001
Q2.5	4	72.81	< .000001
Q2.6	4	92.08	< .000001
Q2.7	4	18.95	< .01
Q2.8	4	72.97	< .000001
Q2.9	4	110.68	< .000001
Q2.10	4	30.64	< .0001
Q2.11	3	19.73	< .001
Q2.12	4	55.14	< .000001
Q2.13	4	9.31	.054
Q2.14	3	112.28	< .000001

Table S6. Results of Kruskal-Wallis tests for participants' responses to Questionnaire 2.

On Questions 1.2 through 1.8 about epistemic factors that contribute to expert disagreement, post-hoc Mann-Whitney comparisons revealed that the two groups of non-experts differed only on Questions 1.3 (r = .07, a statistically significant but theoretically negligible effect size) and 1.6 (r = .16). Pairwise comparisons also indicated that scientists differed significantly from undergraduates on all questions (r's from .11 to .27), except 1.6, and from alumni on all questions (r's from .15 to .20), except 1.6 and 1.8. The answers of climate policy experts differed significantly from undergraduates on all questions (r's from .11 to .23), except 1.3 and 1.8, and from alumni on all questions (r's from .12 to .17), except 1.3, 1.6, and 1.8.

Post hoc pairwise comparisons on Questions 1.9 through 1.13, which concerned nonepistemic factors that give rise to expert disagreement, revealed that the responses of climate scientists differed significantly from those of undergraduates on all questions (r's .12 to .38), except 1.12, which was about careerism. Climate scientists differed significantly from alumni on all questions (r's .09 to .30). The responses of climate policy experts differed significantly from those of both groups non-experts on all questions, except Question 1.12 (r's .11 to .34). The responses of undergraduates and alumni to Questions 1.9 through 1.13 differed significantly on the role of political ideology (r = .08), but this theoretically negligible effect size achieved significance only because 1,290 participants were being compared in the analysis. The answers of undergraduates and alumni did not differ significantly on any of the other nonepistemic factors. Astrophysicists' responses did not differ significantly from any group on Questions 1.11 and 1.12. On Question 1.13, astrophysicists did not differ significantly from climate scientists and climate policy experts but did differ from undergraduates (r = .15) and alumni (r = .11) to a small extent. Undergraduates and alumni gave higher ratings to the importance of defending theories that represent one's life's work (Q1.11) than the three groups of experts.

The mean composite epistemic score of each participant group fell significantly above the midpoint (r's .39 to .68). However, the mean composite nonepistemic scores of climate scientists and climate policy experts failed to differ from the midpoint, while those of undergraduates (r = .51) and alumni (r = .58) fell significantly above the midpoint, and those of astrophysicists (r = .45) fell significantly below.

		Mean Composite Epistemic Score		**Mean Composite Nonepistemic Score**	
		American	**Non-American**	**American**	**Non-American**
Climate Expertise	**Expert**	2.3	2.4	2.1	2.0
	Non-Expert	2.7	2.6	2.6	2.5

Table S7. Mean composite scores, grouped by epistemic vs. nonepistemic, expert vs. non-expert, and American vs. non-American.

On Question 2.6, American climate experts, non-American climate experts, and non-American non-experts all had identical mean scores (3.5), whereas American non-experts had a significantly lower mean score (2.9).

The mean response of astrophysicists to Question 2.4 fell near the midpoint. In response to Question 2.5, the mean responses of undergraduates (r = .34) and alumni (r = .10) fell significantly above the midpoint, while that of climate scientists fell significantly above (r = .35).

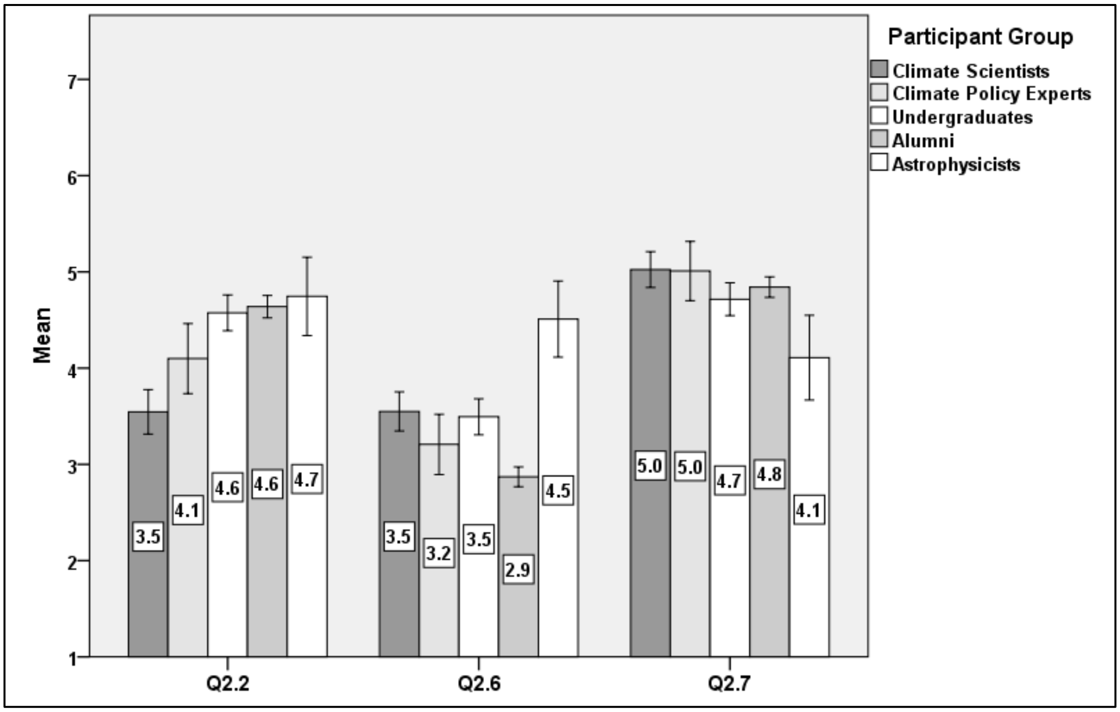

Figure S1. Mean responses to Questions 2.2, 2.6, and 2.7, organized by participant group.

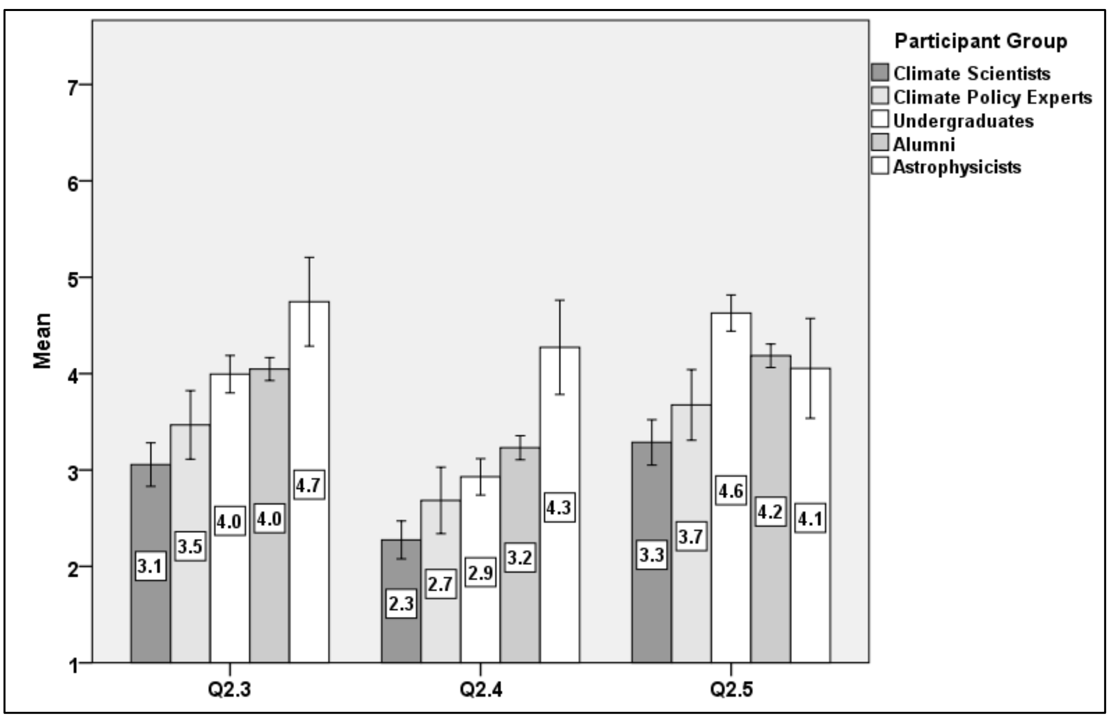

Figure S2. Mean responses to Questions 2.3, 2.4, and 2.5, organized by participant group.

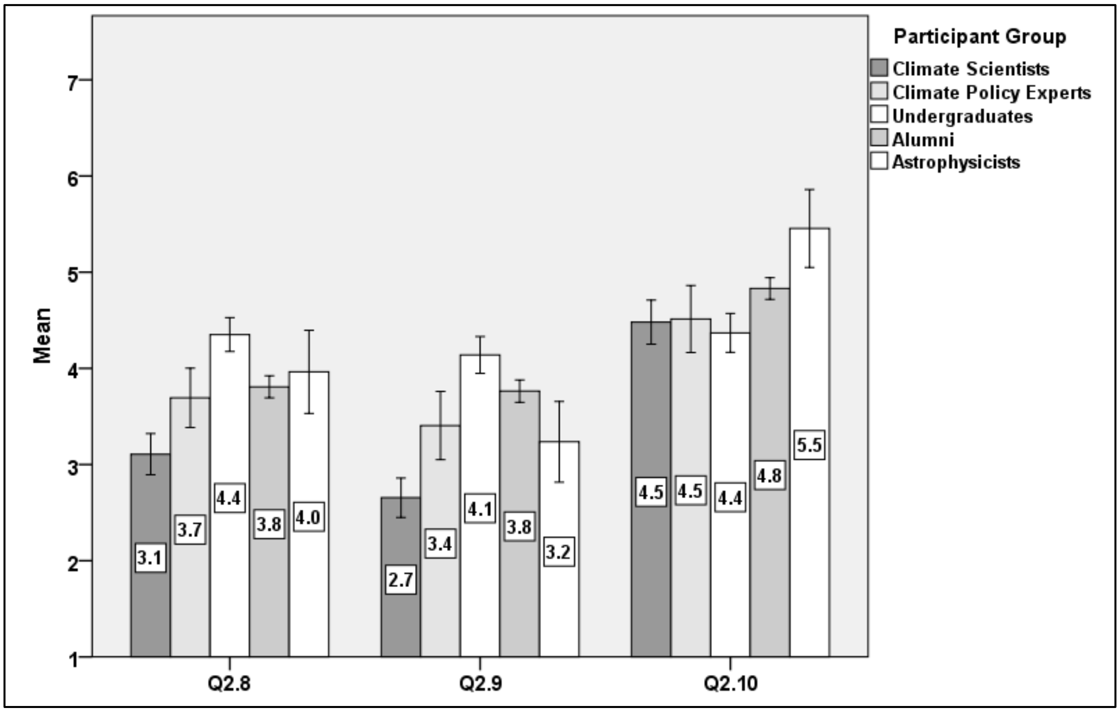

Figure S3. Mean responses to Questions 2.8, 2.9, and 2.10, organized by participant group.

We averaged participants' responses to Questions 2.8 and 2.9 to obtain an institutional bias score. Among climate experts, Americans were less likely to judge there was institutional bias against controversial hypotheses or minority views than non-Americans (t (333) = 2.17, $p < .05$).

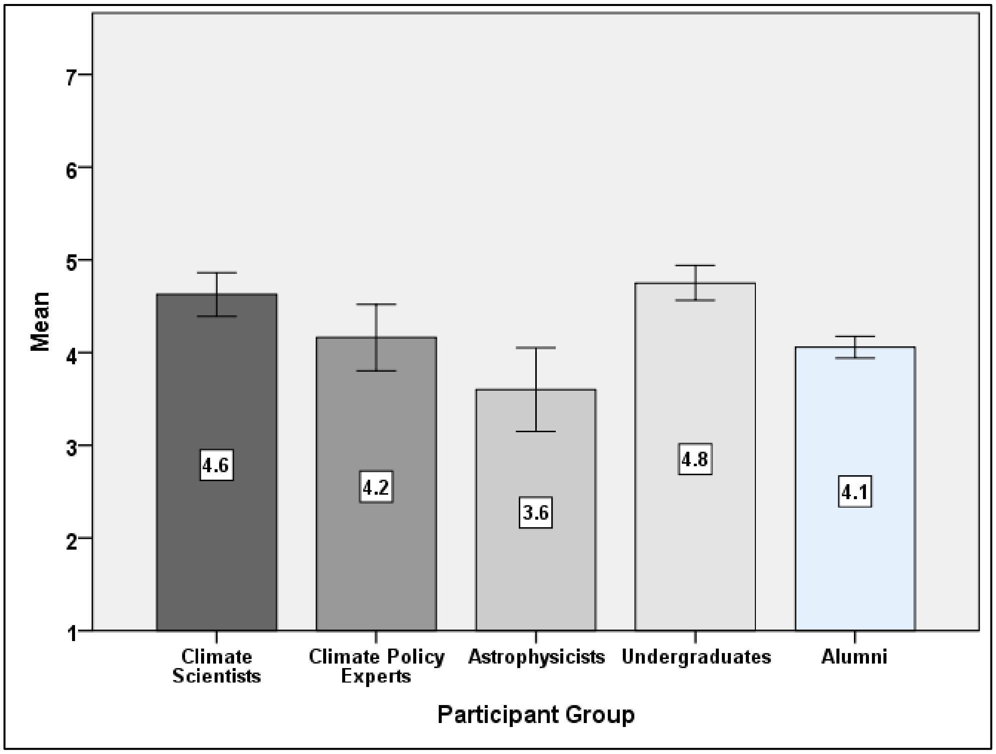

Figure S4. Mean participant responses to Question 2.12, organized by participant group.

On Question 2.14, the differences between the answers of the two groups of climate experts differed significantly from the answers of each of our groups of non-experts (*r*'s .17 to .30).

For Questions 2.15 and 2.16, Wilcoxon signed-rank tests were used to test for between-question differences, and Mann-Whitney tests were used to test for between-group differences. Both undergraduates (r = -.22) and alumni (r = -.29) showed significant between-question differences (Undergraduates: z = -4.04, p < .0001. Alumni: z = -9.01, p < .00001). The Benjamini-Hochberg method was used to correct for multiple comparisons.

Disagreement in Cosmic Ray Physics

The questions we asked about cosmic ray physics concerned the science as a whole and did not introduce finer distinctions between specific topics or areas of disagreement. Because this area of research is likely to be unfamiliar to many readers it may be useful to give a brief overview. The surprising existence of a highly penetrating ionising radiation reaching the Earth from outer space was discovered over one hundred years ago, but many features of these "cosmic rays" are still poorly understood.

Observationally the main issue in cosmic ray physics is that the flux of particles falls off extremely rapidly as one goes to higher particle energies so that no one experiment works over more than a small part of the energy range that needs to be covered. The flux of mildly relativistic particles above the atmosphere is about one per square centimetre per second so that quite small satellite experiments can be used, whereas at the very highest energies we see only about one particle per square kilometre per century requiring large experiments such as the Pierre Auger Observatory, which covers an area in Argentina roughly the size of Luxembourg (of course with sparse sampling).

Furthermore, while at low energies we can relatively easily identify the particles (they turn put to be mainly atomic nuclei), at the higher energies we have to use very indirect techniques. Combining the results from these many different techniques and experiments is complicated, and this is one major source of disagreement in the field. There is room for disagreement about the fundamental observational facts themselves, even before theoretical interpretation. And once one moves away from the basic observations into interpretation, then further uncertainties are introduced. Relating the observed fluxes of particles to their putative production spectra in sources requires a model for their propagation through the intervening magnetic fields, radiation fields and matter distributions, many of which are poorly known.

Of course some facts are indisputable. That the particles exist is certain. The broad features of the all-particle energy spectrum (where one simply lumps all the particles together and just measures their energy) are by now well established except at the very highest energies. The total power requirements are fairly clear. But the composition at the higher energies is quite uncertain, the origin of some fine structure that has recently been seem in precision measurements at lower energies is unclear, there is a long-standing tension between the source spectra favoured by propagation theory and acceleration theory, we do not have a satisfactory propagation model that fits all the data (especially the anisotropy in arrival directions) etc. As in most areas of science one has a mixture of some well-established facts, but a lot of uncertainty in the detail and especially in the interpretation. The challenge is to synthesise a coherent scheme that is also consistent with the rest of astrophysics and physics. Depending on the relative weight one attaches to different pieces of evidence and theoretical preferences, it is possible to have legitimate peer disagreement on many aspects while still agreeing on certain basics.

References

D’Agostino, Ralph B. 1986. “Tests for the Normal Distribution.” In Ralph B. D’Agostino & Michael A. Stephens (eds.), *Goodness-of-Fit Techniques*. Marcel Dekker, Inc.: New York, pp. 367-419.

Elga, Adam. 2007. “Reflection and Disagreement”, *Nous* 41: 478-502.