

Comparing astrobiologists' confidence in extraterrestrial life claims for K2-18 b and Cheyava Falls

In the format provided by the
authors and unedited

Methodological Information

Preregistration

Our study was pre-registered on 9 July 2024, using the Open Science Framework, with the title ‘The next big announcement of possible extraterrestrial life detection: planned surveys of astrobiology community opinion’; see: <https://osf.io/xusv7/overview>.

Method

A spreadsheet of astrobiologists’ names and email addresses were already available from a prior project (Vickers et al. 2025). Potential participants had previously been selected from (i) lead authors and corresponding authors of articles published in the top three astrobiology-focused journals in the last ten years, and (ii) presenters at the astrobiology conference AbSciCon. In all cases, names and email addresses were found in the public domain. The top three astrobiology-focused journals were deemed to be *Astrobiology*, *International Journal of Astrobiology*, and *Discover Life* (previously *Origins of Life and Evolution of Biospheres*).

One week prior to each survey, we sent a short ‘forewarning email’ to our database of astrobiologists. This served two purposes: (i) to prime participants to expect a survey email, and (ii) to check our database for discontinued email addresses. Due to discontinued email addresses, our stock of potential participants was reduced, but at the same time we added new names from new publications in the top three astrobiology-focused journals. The final result was a list of 1,270 astrobiologists for the K2-18b survey and, by coincidence, a list of 1,270 astrobiologists for the Mars survey.

For an earlier sister project (Vickers et al. 2024), Research Software Engineers (RSEs) at Durham had prepared bespoke surveying software and the survey platform. The completed spreadsheets could be entered into the platform to generate unique voting tokens for every potential participant, ensuring that only those targeted could vote, and each person could only vote once. Whilst the survey was open, it was possible for those with access to the server (the PI and the software engineers) to see which targeted scientists had voted, since they could see which tokens had been used. When the survey closed (after 11 days), all tokens were deleted from the server, and it was then impossible for anyone to ascertain even *if* a given targeted scientist had voted.

The surveys were conducted in April/May and September/October 2025. In both cases a reminder email was sent after 7 days, and the survey was closed after 11 days.

Mappings from Likert scale results to credence values

In the article we used a specific mapping from Likert scale results to credence values (Supplementary table 1).

Response	Description	Credence Weight
Strongly Disagree	Almost certainly not true	0.05
Disagree	Probably not true	0.25

Response	Description	Credence Weight
Neutral	50-50	0.50
Agree	Probably true	0.75
Strongly Agree	Almost certainly true	0.95

Supplementary table 1. Linear mapping from Likert-scale responses to credences.

We then brought this to bear on the K2-18b and Mars data (**Supplementary table 2**).

Response	Credence Weight	K2-18b %	K2-18b Contribution	Mars %	Mars Contribution
Strongly disagree	0.05	35.1 %	1.76	11.1 %	0.55
Disagree	0.25	30.2 %	7.55	33.5 %	8.37
Neutral	0.50	28.0 %	14.00	40.3 %	20.13
Agree	0.75	6.3 %	4.73	13.0 %	9.75
Strongly agree	0.95	0.4 %	0.38	2.1 %	2.00
Total		100 %	28.4 %	100 %	40.8 %

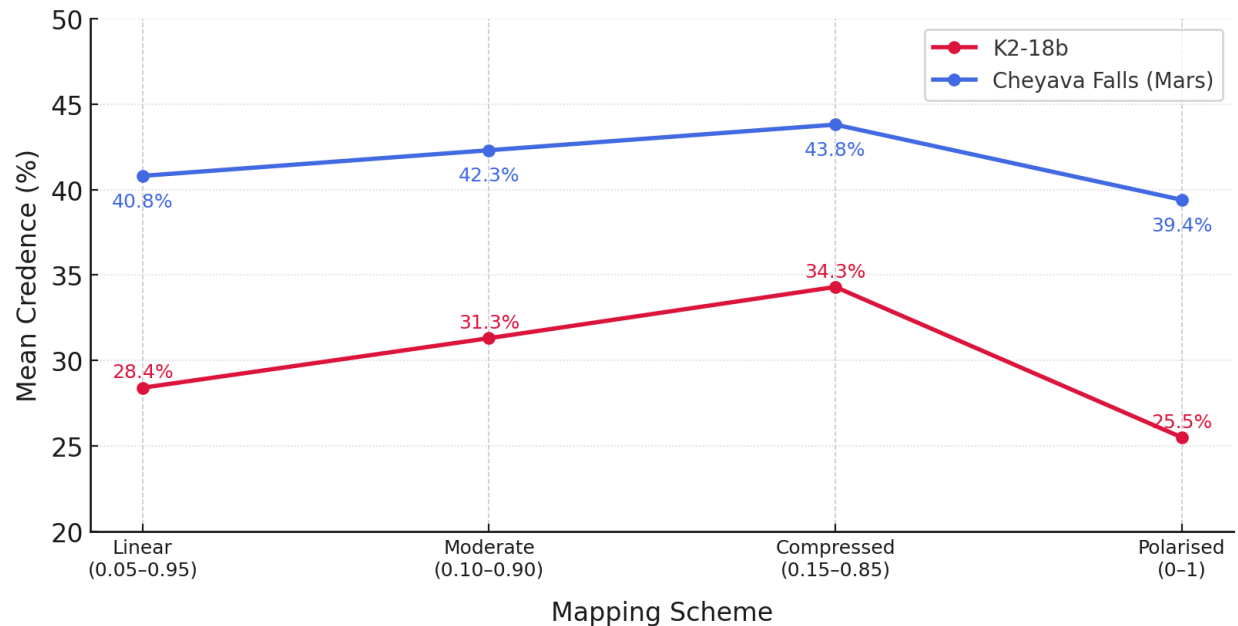
Supplementary table 2. Mean credence calculations for K2-18b and Mars.

However, there is no single canonical mapping from Likert scale results to credence values. Three alternative schemes are described in **Supplementary table 3**.

<u>Scheme</u>	<u>Strongly Disagree</u>	<u>Disagree</u>	<u>Neutral</u>	<u>Agree</u>	<u>Strongly Agree</u>	<u>Notes</u>
<u>Linear (our baseline)</u>	<u>0.05</u>	<u>0.25</u>	<u>0.50</u>	<u>0.75</u>	<u>0.95</u>	<u>Equally spaced around 0.5</u>
<u>Moderate-linear</u>	<u>0.10</u>	<u>0.30</u>	<u>0.50</u>	<u>0.70</u>	<u>0.90</u>	<u>Reduces the distance between extremes and middle</u>
<u>Compressed tails</u>	<u>0.15</u>	<u>0.35</u>	<u>0.50</u>	<u>0.65</u>	<u>0.85</u>	<u>Reflects a belief that scientists rarely mean 5% or 95% literally</u>
<u>Strongly polarised</u>	<u>0.00</u>	<u>0.20</u>	<u>0.50</u>	<u>0.80</u>	<u>1.00</u>	<u>Emphasises sharp divisions between agreement and disagreement</u>

Supplementary table 3. Alternative mappings from Likert scale results to credence values.

Across these four mapping schemes, mean credence in the K2-18b claim ranged from 25.5% to 34.3%, and for the Cheyava Falls claim from 39.4% to 43.8%, confirming that results are robust across reasonable variations in mapping scheme; see Supplementary fig 1.



Supplementary fig 1. Shifts in mean credence across four possible mapping schemes, for K2-18b and Cheyava Falls survey results.

The between-survey comparison is robust; within-survey absolute values shift modestly depending on mapping choices, and never enough to change conclusions. K2-18b stays in the low-credence region; Mars in the mid-credence region.

Statistical Results

Neutral responses

In the K2-18b survey, 28% of participants chose ‘neutral’, meaning they neither agreed nor disagreed with the statement that scientists had probably found extraterrestrial life. In the Mars (Cheyava Falls) survey, this proportion rose to around 40%. In other words, roughly twelve more people in every hundred moved to the middle option. A standard two-sample test for equality of proportions (two-proportion z-test) confirms that the twelve-percentage-point increase in neutral responses between surveys is highly statistically significant ($z = 3.94$, $p \approx 8 \times 10^{-5}$).

Strong disagreement

At the same time, the proportion of respondents choosing ‘strongly disagree’ dropped sharply, from 35% in the K2-18b survey to only 11% in the Mars survey. That is a fall of about twenty-

four percentage points, or slightly over two-thirds reduction in the strongest form of rejection. This difference is highly statistically significant ($p < 10^{-16}$).

Overall agreement

Combining the two positive response categories (agree and strongly agree), overall agreement increased from about 7% for K2-18b to 15% for the Mars sample. Although still a minority, this more than doubles the share of respondents willing to endorse the claim. Again, the probability that such a jump occurred by chance is extremely small ($p \approx 0.00003$).

Quantifying spread of opinion

We found a modest but discernible difference in the spread of opinion between the two cases. Responses to Cheyava Falls exhibited slightly higher dispersion (Shannon entropy = 1.91) than those to K2-18b (1.85), indicating a somewhat greater diversity of views and, conversely, a higher degree of consensus in the K2-18b case (see [1], Chapter 2, Equation 2.1; see also ‘Data’). The higher entropy for Cheyava Falls results primarily from a larger proportion of participants selecting the agreement categories, which broadens the overall distribution of responses.

This modest difference is unsurprising. If scientists tend to be more cautious about expressing agreement than disagreement or neutrality, we would expect a natural clustering of responses toward the lower end of the scale unless the evidence is particularly compelling. A greater spread of opinion emerges only when some respondents are sufficiently persuaded to move from cautious disagreement or neutrality toward agreement, as occurred in the Cheyava Falls case.

We therefore hypothesise that, if subsequent evidence for Cheyava Falls strengthens, the distribution of responses will become even more dispersed, and the Shannon entropy correspondingly higher.

Ethics Approval

We have complied with all relevant ethical regulations. The sister project – which used an almost identical survey methodology – went through a comprehensive ethical approval at Durham University, UK, equivalent to IRB. A full Data Protection Impact Assessment (DPIA) was completed, including a Legitimate Interests Assessment (LIA). The assessment result was ‘low risk’. This was first completed on 23rd May 2023, and refreshed on 29th January 2025.

The need for participants’ explicit prior consent was waived by the ethics committee. We stress that if participants did not consent, they could simply ignore or delete the survey invitation email. Clicking the survey link embedded in the email is equivalent to consenting for one’s opinion to be included, anonymously, as part of the survey. It is ethically sound to take a scientist’s decision to click the survey link in the invitation email as signaling their consent to their opinion being included in the survey. Scientists were able to permanently ‘opt out’ in two clicks.

Supplementary References

- [1] Cover TM and Thomas JA (2006). *Elements of Information Theory* (2nd ed.). Wiley.