CAMBRIDGE
UNIVERSITY PRESS

RESEARCH ARTICLE

Information gain as a tool for assessing biosignature missions

Benjamin Fields[1,2*], Sohom Gupta[1,3*] (iD) and McCullen Sandora[1] (iD)

[1]Blue Marble Space Institute of Science, Seattle, WA 98104, USA
[2]Wheaton College, Wheaton, IL 60187, USA
[3]Indian Institute of Science Education and Research Kolkata, Mohanpur, 741246, West Bengal, India
Corresponding author: McCullen Sandora; Email: mccullen@bmsis.org

Abstract
We propose the mathematical notion of *information gain* as a way of quantitatively assessing the value of biosignature missions. This makes it simple to determine how mission value depends on design parameters, prior knowledge and input assumptions. We demonstrate the utility of this framework by applying it to a plethora of case examples: the minimal number of samples needed to determine a trend in the occurrence rate of a signal as a function of an environmental variable, and how much cost should be allocated to each class of object; the relative impact of false positives and false negatives, with applications to Enceladus data and how best to combine two signals; the optimum tradeoff between resolution and coverage in the search for lurkers or other spatially restricted signals, with application to our current state of knowledge for solar system bodies; the best way to deduce a habitability boundary; the optimal amount of money to spend on different mission aspects; when to include an additional instrument on a mission; the optimal mission lifetime; and when to follow/challenge the predictions of a habitability model. In each case, we generate concrete, quantitative recommendations for optimizing mission design, mission selection and/or target selection.

Contents

---

*Equal contribution

Introduction

The overarching goal of the field of astrobiology is to search for signs of life elsewhere in the universe. As it is a highly interdisciplinary endeavour, the strategies proposed to achieve this are multifaceted, spanning a range of solar system and exoplanet targets and detection techniques. As such, it is sometimes challenging even for seasoned practitioners to assess the value of upcoming missions. Recently, there have been efforts to address this. The Ladder of Life Detection (Neveu *et al.*, 2018), Confidence of Life Detection score (Green *et al.*, 2021), and Standards of Evidence framework (Meadows *et al.*, 2022) aim to provide a language for communicating the evidential strength of any observed biosignature, both to the public and amongst scientists. The nine axes of merit has been proposed as a framework for describing the utility of any technosignature detection mission in Sheikh (2020). Lorenz (2019) provides a risk versus payoff framework for life detection missions, which even factors in the mission success probability.

While many of these frameworks are fantastic at isolating the relevant mission aspects necessary for judging the merits of a mission, they are fundamentally descriptive processes, and leave up to the judgement of the practitioner how to prioritize each mission dimension. As such, they are not suited for any final decision making process, whether it be selecting between missions or optimizing mission design.

In this note we propose a simple method of assessing mission value in terms of *information gain*, and demonstrate the utility of this formalism through considering several idealized mission designs. This framework allows to distil the specifics of a mission down to a single number, which is useful for many purposes. At the outset, we should acknowledge the shortcomings of this procedure: clearly, there is no unique way of ranking missions; our approach is capable of providing one such quantification, but other factors are surely important. While our approach encompasses many mission aspects, it specifically disregards, for instance, any ancillary benefits a mission would provide apart from biosignature detection. Secondly, we wholeheartedly disavow mistaking performance metrics for the indefinite concept of mission value. Any practitioner interested in using a method like this should be well aware of the dangers inherent in using performance metrics for decision making- this is exemplified in Goodhart's law, whereby when a quantitative proxy is used to drive decision making, there is a tendency for people to optimize for the proxy itself, sometimes to the detriment of the original intended goal (Chrystal *et al.*, 2003). As such, we do not endorse a plan to 'maximize information gain', nor do we believe it would be particularly useful to very precisely calculate the information of any particular mission. The intent of our framework is to provide a rough heuristic that is easily calculable in a manner that makes the dependence on mission parameters highly apparent, so that these can be quickly and straightforwardly varied to determine their effect on the overall mission value.

In the remainder of this section we go into more detail on the process of calculating information gain for a given mission setup. In the following sections, we demonstrate the utility and flexibility of this approach by applying it to several disparate scenarios, and deriving concrete, quantitative recommendations for optimizing mission design. In section 'Ratio of 2 signals' we apply this formalism to the challenge of determining whether the occurrence rate of a biosignature depends on some system variable. In section 'The tradeoff between false positives and false negatives' we consider the effects false positives and false negatives have on information. In section 'Lurkers' we consider a tradeoff between resolution and coverage in searching for a biosignature. Section 'Further applications' contains more applications, including: how to best determine the limiting value of some system parameter for life, how to apportion budget to various aspects of a mission, when to include an instrument on a mission, how to determine the optimal mission lifetime and when to follow a habitability model, and when to challenge it.

Bayes estimation and information gain

In this work we propose to operationalize the dictum of many proposed missions, to 'maximize science return', in the following way- by equating science return with *information gain*, in the information theoretic sense, we may write the science return as

$$S = \int df \, p(f) \, \log p(f) \tag{1}$$

This explicitly incorporates a value for how to reduce raw mission data, in the form of a bit stream, down to the distilled representation relevant for our immediate purposes. As such, this framework is more general, and can suitably be applied to a variety of scientific goals across disciplines.

As a simple application of the above framework, let us imagine the simplest setup, with a single biosignature, in the absence of false positives and negatives, and drawing from a uniform population. We wish to estimate the fraction of locations possessing this biosignature f, given a total number of systems surveyed by a mission N_{tot}, which returns N_{det} detections. If the occurrence rate were known, the number of detections would follow a Bernoulli distribution, and so when the number of detections is known instead, the signal occurrence rate f follows the conjugate distribution $f \sim \beta$ (N_{tot}, N_{det}, f), where β refers to the beta distribution[1] (Sandora and Silk, 2020).

$$\beta(N_{tot}, N_{det}, f) = B(N_{tot}, N_{det}) f^{N_{det}} (1 - f)^{N_{tot} - N_{det}}$$
$$B(N_{tot}, N_{det}) = \frac{(N_{tot} + 1)!}{N_{det}!(N_{tot} - N_{det})!} \tag{2}$$

This distribution has the following mean and variance:

$$\mu = \frac{N_{det} + 1}{N_{tot} + 2}, \quad \sigma^2 = \frac{(N_{det} + 1)(N_{tot} - N_{det} + 1)}{(N_{tot} + 2)^2 (N_{tot} + 3)} \tag{3}$$

If the mean is well separated from 0, this has $\mu \to f_o = N_{det}/N_{tot}$ and $\sigma^2 \to f_o(1 - f_o)/N_{tot}$. If the mean is consistent with 0 to observed uncertainty, $\mu \to 1/N_{tot}$ and $\sigma^2 \to 1/N_{tot}^2$. While simple, already this setup indicates how precision depends on survey size, for multiple different regimes. This can be related to science yield (information gain) through equation (1).

For a broad class of distributions, including the setup above, information reduces to $S \approx \log \sigma$, so maximizing information gain becomes equivalent to minimizing variance, which sets the measurement uncertainty. In this limit, 1 bit of information is gained if the error bars for the measured quantity are decreased by a factor of 2.

The above expression assumes initial ignorance. For the case where we are updating knowledge obtained from a previous mission, we can instead use the Kullback–Leibler (KL) divergence as the generalization:

$$\Delta S = KL(p\|q) = \int df \, p(f) \, \log\left(\frac{p(f)}{q(f)}\right) \tag{4}$$

This is a measure of information gain as one moves from an initial distribution p to a final one q. This has the property that $KL(p\|q) \geq 0$, with equality only if $p = q$, guaranteeing that information is always gained with new missions. This reduces to the prior case if we take the initial distribution q to be uniform.

---

[1]This is dependent on the choice of the prior distribution for f, which we take here to be uniform. However, using other priors does not change the behaviour of the population mean and variance by much (Sandora and Silk, 2020).

In the case where we are updating our knowledge of biosignature occurrence rate from a previous mission, both subject to the conditions outlined above, we have $p \sim \beta(N_{\text{tot}}, \ N_{\text{det}}, \ f)$ and $q \sim \beta(N_{\text{tot}}^{(0)}, \ N_{\text{det}}^{(0)}, f)$ then we can find the exact formula for information gain from Rauber *et al.* (2008):

$$
\begin{aligned}
\Delta S = \log \frac{B(N_{\text{tot}}, N_{\text{det}})}{B\left(N_{\text{tot}}^{(0)}, N_{\text{det}}^{(0)}\right)} \\
+ \Delta(N_{\text{tot}} - N_{\text{det}})\psi(N_{\text{tot}} - N_{\text{det}} + 1) \\
+ \Delta N_{\text{det}}\psi(N_{\text{det}} + 1) - \Delta N_{\text{tot}}\psi(N_{\text{tot}} + 2)
\end{aligned}
\tag{5}
$$

where $\psi(x) = d\Gamma(x)/dx/\Gamma(x)$ is the digamma function. Using the approximation $\psi(x) \sim \log x - 1/2x$, we can see that this is a generalization of the formula $\Delta S \approx \log \sigma$ to situations where we start with some information. Additionally, in the limit where the sample size is incremented by 1 in an updated experiment, we find $\Delta S \approx 1/N_{\text{tot}}$.

This framework can be compared to other, more rigorous statistical methods of estimating survey efficacy. If one had data from a set of observations, the usual procedure would be to perform a battery of statistical tests (e.g. Neyman Pearson, binomial, kernel tests) aimed at determining the precise degree to which the tested hypothesis is (dis)favoured compared to some null hypothesis. While rejecting the null hypotheses to some specified confidence level (0.05 is a standard choice) is suitable for large data sets, this procedure must be treated with care when the data samples are small and difficult to obtain (Levine *et al.*, 2008). A well-known way for estimation of parameters is to use the method of moments (Casella and Berger, 2001), which yields potentially biased but consistent estimators. Another method is bootstrapping from existing data to estimate the parameters of the parent distribution via the plug-in principle (Mukherjee and Sen, 2019). However, these are often of little help for determining the amount of data needed to gain a specified level of significance, nor which targets we should favour when collecting this data. Modelling the output of future missions can help to proactively determine the amount needed, and the authors regard this as a necessary step of the mission planning process, but this is laborious, must be redone from scratch for each new mission, and, in our opinions, does not always lead to an enlightening understanding of why a certain amount of data is required. Our framework aims to augment these existing procedures and guide intuition by showing how back of the envelope calculations can lead us to concrete, actionable recommendations for mission design.

We now demonstrate the wider applicability of our framework by applying it to extensions of these simple setups.

Ratio of 2 signals

As a first application of our framework, we turn our attention to the case where we want to determine how the occurrence rate of a biosignature depends on some system parameter. This is a generalization of the analysis done in Sandora and Silk (2020), which made the idealization that all systems being measured were equally probable of hosting life. Here, we are interested in determining the minimum number of systems required to be observed to detect a population trend, and how best to allocate mission time towards systems of varying characteristics to most effectively distinguish whether a trend exists.

This analysis has many applications. Indeed, we may expect the probability of biosignatures to depend on planet mass, stellar mass, incident irradiation, planet temperature, age, metallicity, etc. For illustrative purposes, we focus our attention on a particular case, the system's position within our galaxy, though the analysis will be general enough to easily adapt to any of the other quantities mentioned.

There are numerous schools of thought on how habitability may depend on galactic radius (for a review, see Jiménez-Torres *et al.*, 2013). Metallicity (the abundance of elements higher than hydrogen)

is observed to decrease with galactic radius (Daflon and Cunha, 2004), which may either be a boon or bane for life. On the one hand, a minimum amount of material is required for planet formation (Johnson and Li, 2012), and planets are expected to be more abundant around more metal-rich stars (Balbi *et al.*, 2020). On the other hand, the prevalence of hot Jupiters also has been shown to increase with metallicity and, depending on the formation mechanism, could preclude the presence of terrestrial planets in the habitable zone (Fischer and Valenti, 2005). Additionally, galactic centres (usually) host an active galactic nucleus (AGN), which spews high energy radiation that can strip planetary atmospheres (Jiménez-Torres *et al.*, 2013). These two competing factors have lead to the hypothesis that there is a galactic habitable zone (Lineweaver *et al.*, 2004), analogous to the circumstellar habitable zone, comprised of a possibly thin annular region, only within which habitable planets can exist. This hypothesis has drawn some criticism, not least because stars are known to migrate radially throughout the course of their evolution, defying simple attempts at placing their orbit (Prantzos, 2008).

Here, we treat the GHZ as a hypothesis, and ask what sort of data we would need to either verify or falsify it. We may divide the alternatives into four scenarios: (i) null hypothesis- the biosignature occurrence rate does not depend on galactic radius (ii) Z- the occurrence rate is greater towards the centre (iii) HJ- the occurrence rate is greater towards the edge and (iv) AGN- the occurrence rate is 0 inside some critical radius.

A fully statistically rigorous approach would do something along the lines of fitting a curve for occurrence rate versus galactic radius, then determining to what degree of certainty the slope is away from 0. Here, we opt for a much simpler method of bucketing our population into two bins, then comparing the occurrence rates of each bin and determining if they are different from each other to the statistical power afforded by the mission. While this simplified analysis is not a replacement for the full approach, it allows us to easily track the dependence on mission parameters, and so can be useful for planning purposes.

We therefore assume we have two populations, either from the same or different missions. If mission 1 surveys M_{tot} planets and detects M_{det} signals, and mission 2 surveys N_{tot} planets and detects N_{det} signals, the signal occurrence rate in each population will be given by beta distributions $X_M \sim \beta(M_{tot}, M_{det}, f)$ and $X_N \sim \beta(N_{tot}, N_{det}, f)$ respectively. Then the inferred ratio of these two occurrence rates $X = X_M/X_N$, is given by the following distribution (Pham-Gia, 2000):

$$X \sim \begin{cases} \dfrac{B(\alpha, \beta_N)\,_2F_1(\alpha, 1 - \beta_M; \alpha + \beta_N; x)}{B(\alpha_M, \beta_M)B(\alpha_N, \beta_N)} x^{\alpha_M - 1}, & x \leq 1 \\[3mm] \dfrac{B(\alpha, \beta_M)\,_2F_1\left(\alpha, 1 - \beta_N; \alpha + \beta_M; \dfrac{1}{x}\right)}{B(\alpha_M, \beta_M)B(\alpha_N, \beta_N)} x^{-\alpha_N - 1}, & x > 1 \end{cases} \tag{6}$$

where $\alpha_M = M_{det} + 1$, $\beta_M = M_{tot} - M_{det} + 1$, $\alpha_N = N_{det} + 1$, $\beta_N = N_{tot} - N_{det} + 1$, $\alpha = \alpha_M + \alpha_N$. The hypergeometric function is $_2F_1(a, b; c; x) = \sum_{n=0}^{\infty} (a)_n(b)_n/(c)_n x^n/n!$, with $(a)_k = \Gamma(a + k)/\Gamma(a)$ the k-th order rising Pochhammer symbol.

The moments of this distribution can be written as

$$E[X^k] = \frac{(M_{det} + 1)_k}{(M_{tot} + 2)_k} \frac{(N_{tot} + 2 - k)_k}{(N_{det} + 1 - k)_k} \tag{7}$$

From this we can derive the mean

$$\mu = \frac{M_{det} + 1}{M_{tot} + 2} \frac{N_{tot} + 1}{N_{det}} = \frac{f_M}{f_N} + \mathcal{O}(1/N) \tag{8}$$

where $f_M = M_{det}/M_{tot}$ and $f_N = N_{det}/N_{tot}$

Similarly, the variance can be written

$$
\begin{aligned}
\sigma^2 &= \frac{(M_{\mathrm{det}} + 1)(M_{\mathrm{det}} + 2)N_{\mathrm{tot}}(N_{\mathrm{tot}} + 1)}{(M_{\mathrm{tot}} + 2)(M_{\mathrm{tot}} + 3)(N_{\mathrm{det}} - 1)N_{\mathrm{det}}} - \mu^2 \\
&= \frac{f_M^2}{f_N^2}\left(\frac{1 - f_N}{f_N N_{\mathrm{tot}}} + \frac{1 - f_M}{f_M M_{\mathrm{tot}}}\right) + \mathcal{O}(1/N^2)
\end{aligned}
\tag{9}
$$

The variance is not symmetric about the two ratios f_M and f_N since the associated distribution X_M/X_N is not symmetric. However, the quantity σ^2/μ^2 is symmetric, and has a characteristic 1/N behaviour, but in this case dependent on the number of detected signals $N_{\mathrm{det}} = f_N N_{\mathrm{tot}}$ and $M_{\mathrm{det}} = f_M M_{\mathrm{tot}}$, rather than the total number of systems observed. Note also that this variance contains two additive contributions from the uncertainties in both surveys.

If one occurrence rate is smaller than the other, a correspondingly larger number of total systems need to be surveyed to adequately measure any trend with system parameters, rather than the overall occurrence rate. For rare signals, $f_{M,N} \ll 1$, the number of detections in each survey should be roughly equal. This implies that surveys should collect more signals from regions where the signal is expected to be rarer.

To use this to distinguish between two competing hypotheses predicting different values of the ratio f_M/f_N, we would want the measurement error $e = \sqrt{\sigma^2}$ to be smaller than the difference in predictions. So if we are trying to distinguish between the different galactic habitability scenarios outlined above, we would want $\sigma \lesssim \mu$.

Cost analysis

We now use this analysis to determine the optimal way of apportioning funds to the two populations. We denote the fraction of funding given to mission 2 as c, and as an example assume the cost functions for these surveys are given by the same power law relation $M = M_0(1 - c)^q$ and $N = N_0 c^q$, where M_0 and N_0 are the yields that would be obtained if all resources were put into mission M and N, respectively. As we discuss in section 'What is the optimal amount of money to spend on different mission aspects?', we expect $q > 1$ if the two populations come from two different missions, as increasing budget allows multiple mission aspects to be improved simultaneously. For apportioning observing time within the same mission, we instead expect $q < 1$, as increasing observing time yields diminishing returns, discussed further in Sandora and Silk (2020). We wish to minimize the variance in equation (9), corresponding to maximum information gain. The derivative of eqn (9) yields

$$
\begin{aligned}
\frac{\partial}{\partial c}\sigma^2 = \frac{f_M^2}{f_N^2}\Bigg[&\frac{-q}{N_0}\frac{1 - f_N}{f_N}c^{-(q+1)} \\
&+ \frac{q}{M_0}\frac{1 - f_M}{f_M}(1 - c)^{-(q+1)} \Bigg]
\end{aligned}
\tag{10}
$$

This makes use of the approximation that both M_{det} and N_{det} are substantially removed from 0. Setting this to zero and solving for c gives us the optimal cost

$$
c = \frac{1}{1 + s^{1/(q+1)}}, \qquad s = \frac{N_0}{M_0}\frac{f_N}{f_M}\frac{1 - f_M}{1 - f_N}
\tag{11}
$$

Yielding survey sizes

$$
N = N_0 \frac{1}{\left(1 + s^{1/(q+1)}\right)^q}, \qquad M = M_0 \frac{s^{q/(q+1)}}{\left(1 + s^{1/(q+1)}\right)^q}
\tag{12}
$$

From here, it can be seen that if $f_M = f_N$ and $M_{\text{tot}} = N_{\text{tot}}$, then $c = 1/2$, i.e. both surveys should get equal priority. The relative cost is plotted for generic values of observed fraction in Fig. 1.

This allows us to explore how the optimal cost depends on various parameters. The recommendations are sensible - the better the observations in population M, the more cost should be invested on N and vice versa to get maximum information. If q is large, the optimal cost will deviate more strongly from 1/2 compared to the $q = 1$ case, holding other parameters fixed. Similarly if q is small, we have $c \approx 1/2$ over a broad range of parameter values. In the lower right panel we see that more money should be spent on missions that have lower yields.

Lastly, note that from the above expression, we see that the optimal survey sizes depend on biosignature occurrence rates, which are not necessarily known beforehand. If these are unknown, a well designed survey would be able to adapt to information as it arrives, to alter upcoming target selection as current estimates of the occurrence rates dictate.

The tradeoff between false positives and false negatives

In this section, we apply our procedure to solar system planetary biosignature missions. Here, we address ambiguity in the interpretation of mission results. This ambiguity can arise from two significant contributing factors. Firstly, there could be biosignature experiments where a positive result may be a false positive, *e.g.* some geochemical trickery masquerading as life. Alternatively, there may be situations where a negative result could be a false negative, *e.g.* there may be entire communities of microbes present which the instrument is not sensitive enough to detect. Any realistic mission will have to deal with both. In such a scenario, and assuming finite resources/design constraints, one key decision which will have to be made is which of the two sources of ambiguity to prioritize minimizing to arrive at a higher overall certainty of whether or not life is present. As discussed in Foote *et al.* (2022), the need to mitigate these leads to the drive towards the most definitive biosignatures possible, and as exhaustive a determination of abiogenesis rate as possible.

Unlike large scale surveys of many exoplanet systems, solar system based biosignature search strategies are characterized by information depth rather than information breadth. In other words, in the solar system life detection mission, we are going to be successively surveying the same planet over and over again, gaining more and more information and constraining our statistical parameters. For this reason, the ideal framework from which to approach the problem is a Bayesian one, which can successively adjust prior probabilities based on new information (Catling *et al.*, 2018).

We define the false positive rate as the probability of a detection given nonlife, $FP = P(D|NL)$ and the false negative rate as the probability of a nondetection given life, $FN = P(ND|L)$. The quantities needed for the information can be derived via Bayes' theorem:

$$P(L \mid D, FP, FN, f) = \frac{P(D \mid L, FP, FN, f)P(L \mid FP, FN, f)}{P(D \mid FP, FN, f)} \tag{13}$$

Where

$$P(L \mid FP, FN, f) = P(L) = f \tag{14}$$

is the probability of life existing in the location where the probe searched. The probability that the instrument registers a signal is

$$P(D \mid FP, FN, f) = (1 - FN)f + FP(1 - f) \tag{15}$$

Therefore, the posterior probability of life given a detection is given by

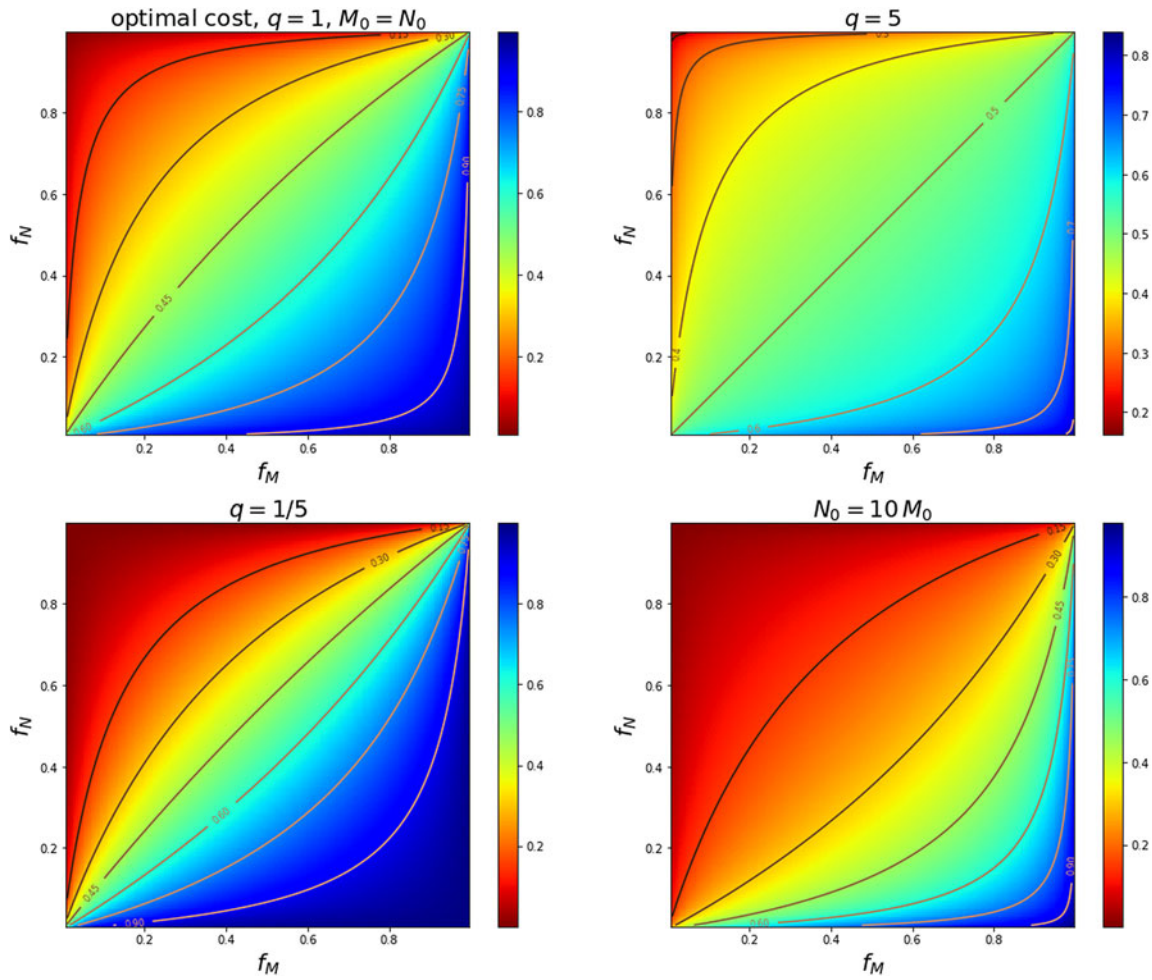$$P(L \mid D, FP, FN, f) = \frac{(1 - FN)f}{(1 - FN)f + FP(1 - f)} \tag{16}$$

Figure 1. *Optimum cost fraction* c *to minimize variance. Unless specified, these plots take* q = 1 *and* M_0 = N_0. *Here, a low value of* c *corresponds to allocating most funds to mission* M.

and

$$P(L|ND, FP, FN, f) = \frac{\text{FN}f}{\text{FN}f + (1 - \text{FP})(1 - f)} \tag{17}$$

To arrive at a decision making framework, we approach the problem in terms of information gain. Here, the information gain measures the uncertainty present in data–the larger the Shannon Entropy, the lower the certainty of the data and the higher the likelihood that it may be random. In this case the probabilities are conditional, so we use the Shannon Entropy for conditional probabilities, given in terms of the event of detection, nondetection and the presence of life below. This is the information of the posterior probability distributions, weighted by their probability of occurrence.

$$\begin{aligned} S(\text{FP}, \text{FN}) = -\int \mathrm{d}f p_f(f) \Big[& P(D)\, s\big(P(L|D)\big) \\ & + P(ND)\, s\big(P(L|ND)\big) \Big] \end{aligned} \tag{18}$$

where $s(p) = p\log p + (1 - p)\log(1 - p)$.

Lastly, as the probability of life f is not known ahead of time, our expression requires that we integrate over all possible values. This requires a prior distribution $p_f(f)$, which incorporates our current estimates on the probability of life existing in a particular locale, given the current uncertainties in the probability of life emerging, as well as any geological history which may have impacted the site's habitability (Westall *et al.*, 2015; Lorenz, 2019). This is an inescapable part of the calculation, and so to investigate the dependence of our results on any assumptions we make about this quantity we explore four different options: (i) the value of f is arbitrarily taken to be 0.001 (ii) the value of f is taken to be 0.5 (iii) the probability of any value of f is uniformly distributed and (iv) the probability is log-uniform, $p_f(f) \sim 1/f$. This last option requires a minimum value for f, which we take to be 10^{-10} in our calculations. The choice of this value is arbitrary, but mainly influences the overall scale of the expected information gain through $S \sim 1/\log(p_{\min})$, rather than the dependence on the variables FN and FP. The expected information gain for these three choices is displayed in Fig. 2.

In these plots, the blue regions throughout the midsection have highest entropy, corresponding to maximum ambiguity in result interpretation. The red regions in the upper right and lower left corners are both places where the information entropy approaches zero, corresponding to certain result interpretation. The rightmost red region is a pathological point in parameter space where both FP or FN approach 1. In our parameterization, this region indicates that the signal (or lack thereof) is almost sure to convey the opposite of what the instrument was designed to do, and the experiment is certain to be wrong. This is what every biosignature mission should deliberately avoid.

In the bottom left corner, we approach our second region where $S \rightarrow 0$, this time where FP and FN both approach zero, representing the ideal scenario. Such unambiguous results will be unobtainable for any realistic mission, but these plots do provide a metric that can be used to evaluate missions with different values of FP and FN. For a preexisting mission design that can be improved upon incrementally, the recommendation given by this analysis is to go in the direction of maximum gradient, which depends on the mission's position in the plot. Through these plots, we have placed a white dashed line demarcating the region where it is more important to prioritize false positives (above the curve) from the region where it is more important to prioritize false negatives (below). This line is exactly equal to the diagonal FP=FN for the uniform and $f = 0.5$ cases, but is substantially different in the others. Notably, in the region where both FP and FN are small, it remains more important to prioritize false positives. This is a consequence of our assumption that $f \ll 1$; the opposite conclusion would hold if we consider $f \approx 1$.

This procedure, is, of course, reliant on a method of estimating false positive and negative rates for a mission, both of which may be difficult in practice. FP can be estimated by characterizing the
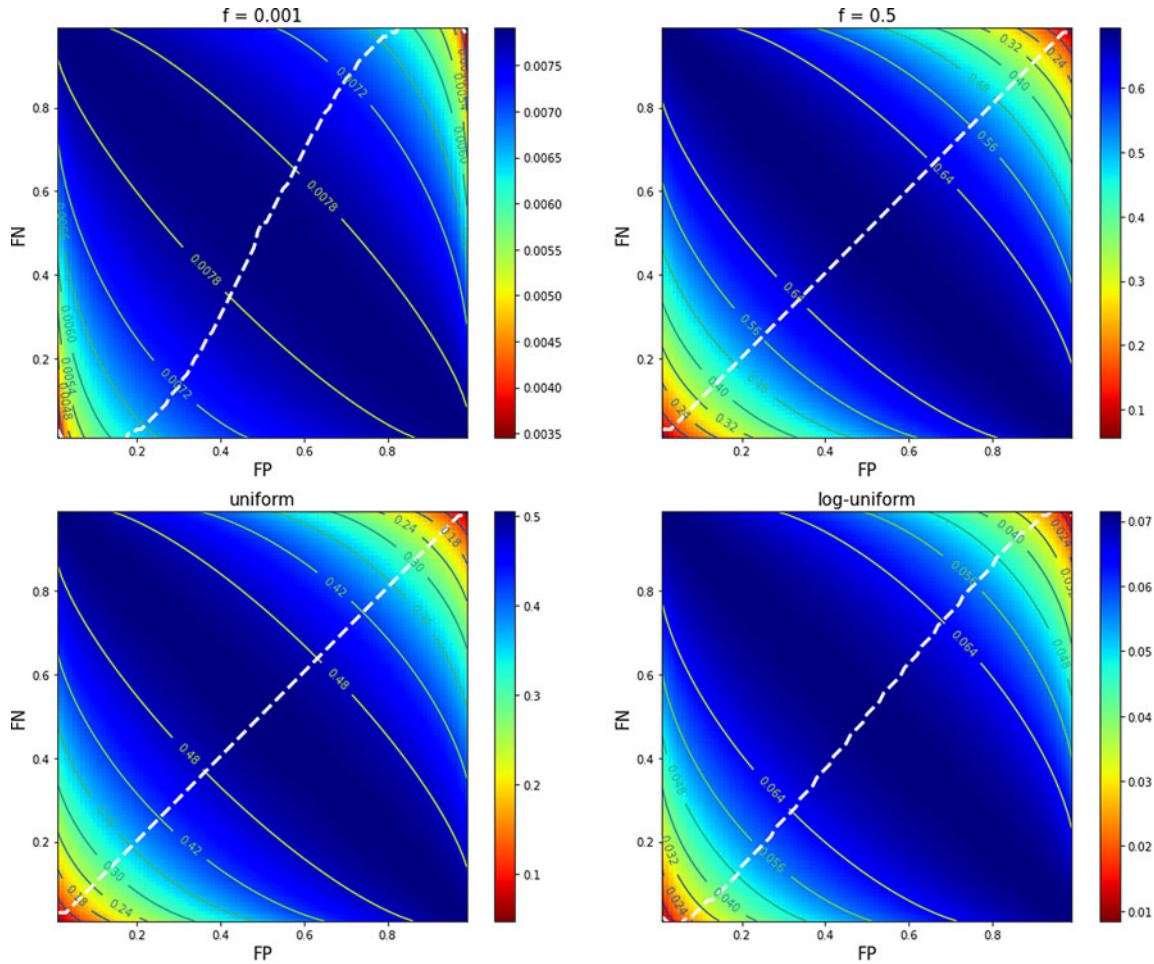
Figure 2. *Heat map of the conditional entropy in terms of FP and FN. The top two plots take particular values for the probability of life existing at the sampling location* f. *The bottom two integrate over a uniform and log-uniform distribution for* f, *respectively. The white dashed line in each separates regions where it is more beneficial to decrease FP (above the line) and FN (below the line).*

instrumental sensitivity, and accurately accounting for all abiotic signal sources with careful geochemical modelling. This is often fraught with controversy (for a recent review see Harman and Domagal-Goldman, 2018), as historical examples can attest (see e.g. Barber and Scott, 2002; Yung *et al.*, 2018; Bains and Petkowski, 2021). FN, on the other hand, may be just as difficult to ascertain, as life may be present in far lower abundances than Earth based analogue environments indicate, yielding uncertainty in overall signal strength (Cockell *et al.*, 2009). Additionally, extraterrestrial life may have entirely different metabolisms, and so may not produce a targeted biosignature (Johnson *et al.*, 2019).

As further illustration of the application of this formalism, we next turn to two more examples: the combination of multiple instruments, and the recent detection of methane on Enceladus.

Combining signals from multiple instruments

Here we investigate the effect of combining signals to mitigate false positives and negatives, which can provide redundancy and ameliorate the difficulties in interpretation present for a single signal in isolation. Generally, this can be done in a variety of different ways, to either increase the coverage or certainty of a detection. We illustrate the various approaches in a simple setup, and show that our framework yields recommendations for which choice is preferable.

Let's restrict our attention to the combination of two signals. The first is fixed, say the detection of a certain biosignature (say, methane), which carries with it a probability of false positive and false negative. We then have two options for including a second instrument. In option A, we're concerned with false positives, so we include a second instrument to detect a secondary signal which we expect to be present in the case of life (say, homochirality). We then claim detection of life only if both these signals register. While this increases our confidence in a detection, it comes at an increased risk of false negatives, as if either instrument fails to register, we would reject the signal of the other.

Alternatively, we may be more concerned with false negatives, and so may include an additional instrument to measure a third biosignature (say, a spectral red edge), which would be complimentary to the first. This improves our original coverage, and so alleviates concern for false negatives. However, this also increases the chances of false positives, for if either instrument registers a signal, we would claim detection.

The question we wish to address is which of these two options is preferable, given the parameters of the setup. For simplicity, we assume that the false positive and false negative rates of all three instruments are identical, to avoid a proliferation of variables. We also assume that the signals which they measure are uncorrelated with each other (as a high degree of correlation would render additional instruments not worthwhile). In this case, we have

$$
\begin{aligned}
P_A(\text{FP}) &= \text{FP}^2, \quad P_A(\text{FN}) = 1 - (1 - \text{FN})^2 \\
P_B(\text{FP}) &= 1 - (1 - \text{FP})^2, \quad P_B(\text{FN}) = \text{FN}^2
\end{aligned}
\tag{19}
$$

With this, we display the difference in information gained from these two setups in Fig. 3, using the value $f = .001$. We notice that if FP > FN, then scenario *A* is preferred, where both signals must register to count as a detection. Also of note is that the difference in the two setups is greatest when either one of the two probabilities is either very small or very large.

Direct application: the Enceladus data

The Cassini spacecraft, as part of its mission to study Saturn and its moons, passed through a plume erupting from the south pole of Enceladus to collect a sample. Instrumental analysis of the sample revealed a relatively high abundance of methane (CH_4), molecular hydrogen (H_2) and carbon dioxide (CO_2) (Waite *et al.*, 2017). Methane is considered a potential biosignature because on Earth, microbial communities around hydrothermal vents obtain energy through chemical disequilibrium present in H_2
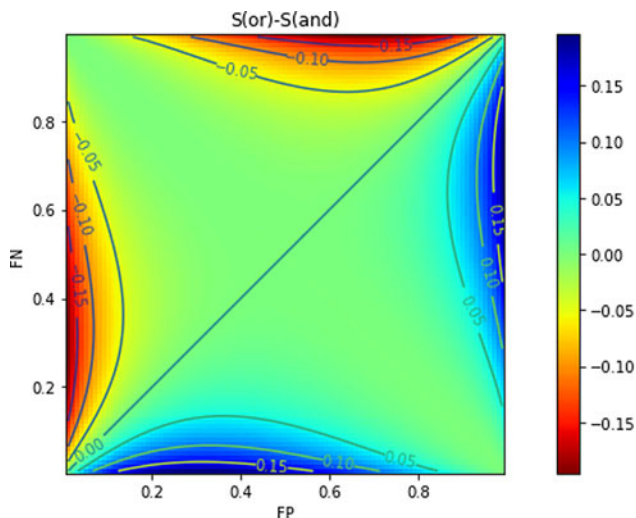
Figure 3. *Difference in two signal entropy for two different signal combination strategies. If this difference is positive, it is more beneficial to use the 'and' strategy, so that both signals must register to count as a detection. If the difference is negative, the 'or' strategy, where either signal registering would count as a detection, is preferable.*

emanating from the vents, resulting in the release of methane as a waste product, a process known as methanogenesis Lyu *et al.* (2018). However, this is not an unambiguous signal, as methane can also be produced abiotically in hydrothermal vent systems via serpentization (Mousis *et al.*, 2016). One method of distinguishing between these two scenarios is then to compare the ratio of fluxes, as biogenic production would yield a higher relative methane rate than would occur abiotically.

This situation represents an ideal test case for our decision making tool–an ambiguous biosignature result, with potential for both false positives and false negatives in the data.

An analysis of the Cassini findings by Affholder *et al.* (2021) compared outgassing rates of simulated populations of archaea and abiotic serpentization, and claimed moderate support for the biotic hypothesis on Bayesian grounds. The model also indicated that at the population sizes and rates of methanogenesis demanded by the data, the archaea would have a negligible effect on the abundance of H_2–therefore, the high concentration of H_2 present in the data did not count against the biotic methanogenesis hypothesis.

It should be noted that this model still carries considerable ambiguity, and a substantial portion of their model runs produce the observed signal abiotically. Regardless, they present quantitative estimates for the probabilities of false positives and false negatives in actual data. Therefore, it is well suited for an application of our decision making tool.

According to the biogeochemical models in Affholder *et al.* (2021), the Cassini plume flybys resulted in FP=.76 and FN=.24. Here FP and FN are determined by the limits of instrumental sensitivity; from Waite *et al.* (2017) the sensitivity threshold is estimated to be about 10x smaller than the observed values. This is illustrated in Fig. 4, where we place a white star to indicate where in the diagram out current knowledge of these biosignatures lies.

Lurkers

As a further application, we apply our approach to the spatial problem, where we are concerned with the tradeoff between coverage and resolution when looking for a biosignature. Much of the language we use in this section is couched in terms of technosignature 'lurkers', but our formalism is general and can equally be applied to microbial colonies, molecular biomarkers, etc.
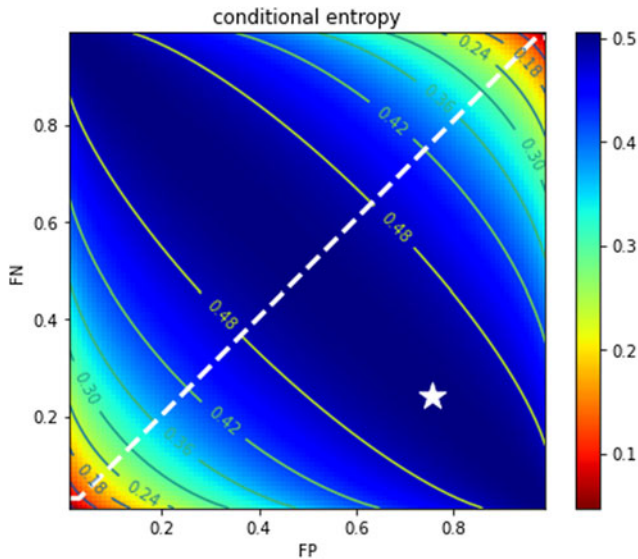
Figure 4. *White star corresponds to the estimated FP and FN of the Enceladus Cassini flythrough mission. The underlying plot is the same as shown in* Fig. 2, *with uniform prior for* f.

Background and context

An ongoing area of technosignature research considers physical artefacts and probes which may be hiding in our solar system, referred to as 'lurkers' (Benford, 2019). The idea is that if extraterrestrial civilizations are present in our galaxy, they may have seeded the galaxy with probes, which, either defunct, dormant or active, could be present in our solar system (Haqq-Misra and Kopparapu, 2012). The plausibility of this was argued by Hart and Tippler, for the time it takes self replicating von Neumann devices to cover the galaxy is modelled to be much shorter than the age of the galaxy itself (Gray, 2015). Therefore, in spite of the considerable energy requirements and engineering challenges inherent in interstellar travel, a reasonable case can be made for past extraterrestrial visitation of our solar system which may have left behind artefacts (Freitas, 1983). Hart and Tippler's argument frames the absence of such lurkers as justification for the Fermi Paradox and pessimism about the prevalence of intelligent life in general (Gray, 2015). However, as deduced by Haqq-Misra and Kopparapu (2012), only a small fraction of the solar system's total volume has been imaged at sufficient resolution to rule out the presence of lurkers. Therefore it remains an open question warranting further investigation and dedicated search programmes.

   This motivates us to apply our formalism to this setup to determine how to effectively allocate resources in order to maximize our chances of finding signs of lurkers. While a survey with unlimited resources may be able to scour every square inch of a planet (or moon, asteroid, etc.), any realistic mission will have budgetary and lifetime constraints. The two extreme strategies would then be to perform a cursory inspection across the entire planet area, but potentially miss signals with insufficient strength, or to perform a high resolution search in a small region of the planet. Here we show that the ideal mission forms a compromise between these two extremes, and deduce the optimal design to maximize chances of success (which in our language corresponds to the mission yielding most information).

Mathematical framework

The setup we consider consists of a survey of resolution R, which is the two dimensional spatial resolution of a survey. Our survey covers a fraction of the planet area f_R. We assume that our survey results in no detection (as otherwise, our concern with interpretive statistics becomes lessened considerably).

We would like to compute the probability that at least one object is present, given a null result, $p(N > 0|\text{null})$, and want to optimize mission parameters so that this quantity is as small as possible given some constraints. Of course, this is equivalent to the formulation where we try to maximize the quantity $p(N = 0|\text{null})$.

To compute this, first note that, in the scenario where there are N objects uniformly spaced on the planet surface, all of size D, we have

$$p(\text{null}|N, D) = (1 - f_R)^N \theta(D - R) + 1\,\theta(R - D) \tag{20}$$

Here $\theta(x)$ is the Heaviside function, which is 1 for positive values and 0 for negative values. From here, we can see that a null result is guaranteed if our mission resolution is above the object size. If the mission does have sufficient resolution, we see that the probability of a null signal diminishes with increasing survey coverage or number of objects.

Inverting this expression requires priors for both the number of objects and their size. If the prior distribution of object sizes is $p_D(D)$, this expression can be integrated to find

$$
\begin{aligned}
p(\text{null}|N) &= \int dD\, p_D(D)\, p(\text{null}|N, D) \\
&= 1 - \left(1 - (1 - f_R)^N\right) P_D(D > R)
\end{aligned}
\tag{21}
$$

And if the prior distribution of number of objects is $p_N(N)$, the total probability of nondetection is

$$p(\text{null}) = \sum p(\text{null}|N) p_N(N). \tag{22}$$

Then by Bayes' theorem,

$$p(N|\text{null}) = \frac{p(\text{null}|N) p(N)}{p(\text{null})}. \tag{23}$$

Note that here, in the limit $R \to \infty$, the survey is completely uninformative, and we have $p(N|\text{null}) \to p(N)$.

Then, the probability of at least one object being present given a null result is

$$p(N > 0|\text{null}) = 1 - \frac{p_N(N = 0)}{p(\text{null})}. \tag{24}$$

In the limit of complete survey $f_R \to 1$, we find

$$p(N > 0|\text{null}) \to \frac{P_D(D < R)}{\Theta + P_D(D < R)}, \qquad \Theta = \frac{P_N(N = 0)}{P_N(N > 0)} \tag{25}$$

reproducing the results of Haqq-Misra and Kopparapu (2012).

To simplify this general analysis, we now specialize to the case where the number of objects is constrained to be either 0 or N^*, $p_N(N) = (1 - p^*)\delta_{N,0} + p^*\delta_{N,N^*}$, where $\delta_{i,j}$ is the Kronecker delta. Then we have

$$p(N > 0|\text{null}) = \frac{p^* \left(P_D(D < R) + (1 - f_R)^{N^*} P_D(D > R)\right)}{1 - p^* \left(1 - (1 - f_R)^{N^*}\right) P_D(D > R)} \tag{26}$$

We display this expression for various parameter values in Fig. 5, which plots the ratio of mean lurker size to resolution D/R versus the coverage fraction f_R. In the first three panes, we assume the distribution of lurker diameters follows a normal distribution (with mean 1 and standard deviation 0.5). The aim of a mission should be to get as close to the 'red region' of these plots as possible
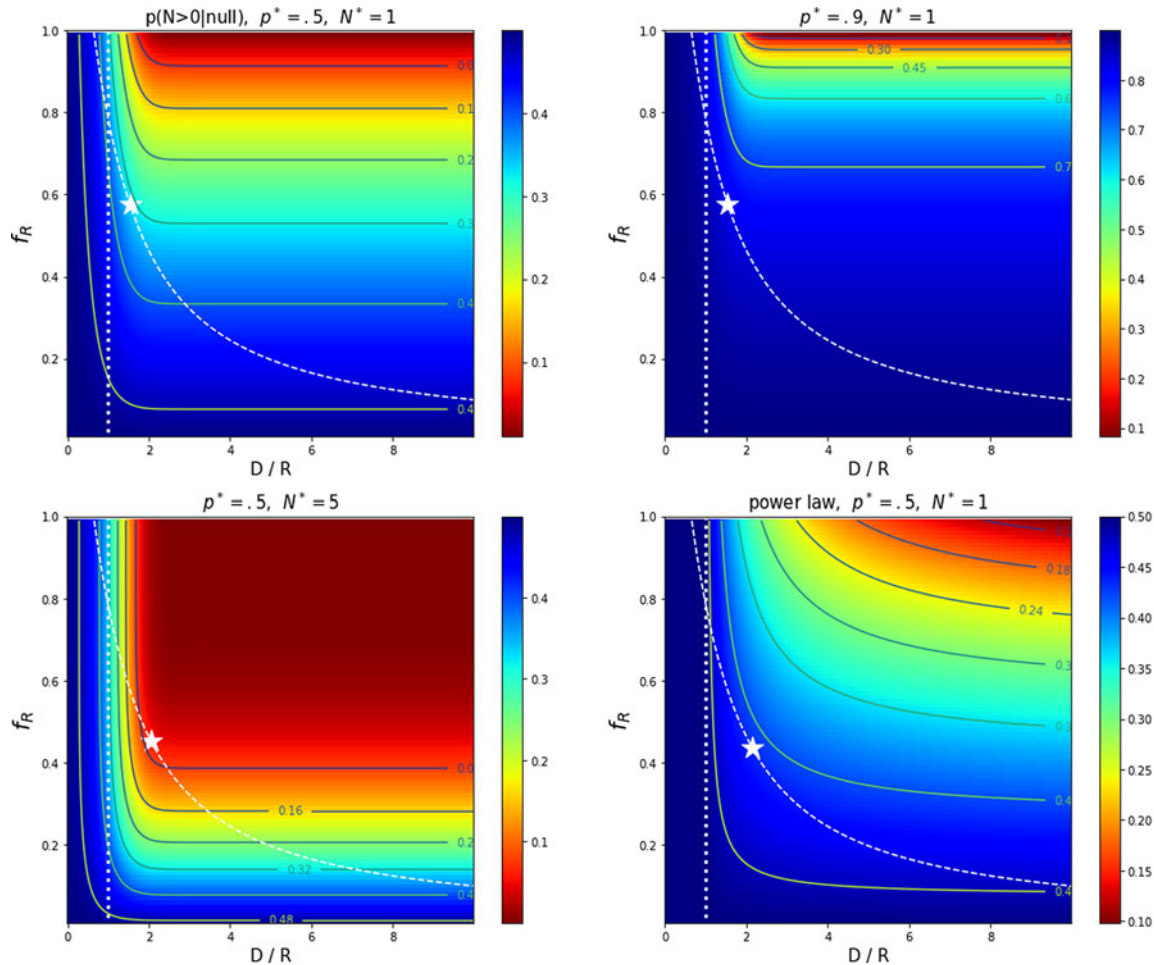
Figure 5. *The probability that lurkers exist on a surface as a function of mean diameter over resolution* D/R *and fraction of surface covered* f_R. *The vertical dashed line corresponds to mean lurker diameter, and the dashed curve corresponds to a constraint on mission design. The white star is the point of lowest* p(N > 0|null) *restricting to the constraint. The top two and bottom left plots take a normal distribution for lurker diameter, and the lower right plot takes a power law.*

given mission constraints. We overlay a constraint curve $A \sim \tan^{-1}(1/R)$ to encapsulate a typical area/resolution tradeoff, but this form specifies a family of curves with differing base values. For a given constraint curve, this will single out a unique mission design that minimizes this quantity (thus maximizing science return). This point is denoted by the white star in each of these plots.

Comparing the top two plots, we can see that the optimum point does not depend on the lurker probability $p*$ at all, a fact that can be derived from the form of equation (26). Comparing to the lower left plot, we can see that the optimum point is shifted towards lower f_R and higher D/R for larger $N*$, which is reasonable, as there would be a greater probability of spotting a lurker in a given area if more are present on the planet surface. The bottom right plot assumes a power law distribution (with $p(D) \propto D^{-2}$) of lurker sizes to highlight the effect this has on our recommendation; again, this shifts the optimum point towards lower f_R and higher D/R.

Direct application to solar system bodies

Here, we will demonstrate a direct application of this statistical framework, by examining the recommendations we find for a series of different bodies in the solar system. We use the spatial resolutions of planetary missions and surveys to establish upper limits on the probability of lurkers which may have gone undetected up until now, allowing us to make a series of recommendations for each solar system body/region of space. For this, we display the resolution and fraction of area covered for multiple missions to bodies around our solar system in Table 1.

The estimates quoted are for illustrative purposes only, and there are some nuances which the table does not account for. Our model makes a simplifying assumption by assuming that lurkers will be unambiguously distinguishable from background geological features. This may not necessarily be true, especially if deliberately camouflaged. For the same reason, even excluding the possibility of deliberate camouflage or stealth, using diameter as our only proxy for detectability has its limitations because the lurkers' albedo relative to that of its surroundings is also a factor. A low albedo lurker on a dark planetary surface may blend in, appearing to be a shadow of another feature or a dark rock. Conversely, a lurker with an albedo that contrasts more with surrounding geology may be easier to detect. Additionally, these are remote sensing missions, and so any subsurface/sub-ocean area would be inaccessible and are not accounted for in this application.

The values in Table 1 are displayed alongside our estimate for $p(N > 0|\text{null})$ in Fig. 6. As fiducial values, we use $p* = 0.5$, $N* = 10$, and $\langle D \rangle = 10$ m, which is a reasonable upper estimate for the size of an interstellar probe (Freitas, 1983). From this, we can draw a number of conclusions:

- Of the bodies considered, we are most certain that no lurkers exist on the moon, followed by Earth. This is because although the Earth has been mapped to higher resolution on the continents, the oceans prevent coverage for the majority of the surface.
- The gradient at each point dictates which aspect of a particular mission should be improved. In a nutshell, whichever is worse between the covered fraction and resolution is what should be improved, but this formalism allows for that comparison to be made quantitatively.
- Several bodies have near-complete low resolution maps, with small patches of high resolution. This allows us to compare which are more valuable for confirming the absence of lurkers. For Mars, for example, the high resolution images give us most confidence that lurkers are absent.
- Of the bodies considered, we are least certain about the absence of lurkers on the outer solar system moons. Areas even further out, such as the Kuiper belt and Oort cloud, are even less certain.
- Despite mapping efforts, many of these bodies still have $p(N > 0|\text{null}) \approx 0.5$, which is our reference probability that lurkers are present. Therefore we are nowhere close to being able to claim that no lurkers are present in the solar system.

We may concern ourselves with the overall probability that lurkers are present in the total solar system, which is given by the compound probability of each body separately,

Table 1. *Upper limit diameters for lurkers on the surfaces of solar system bodies and regions of surveyed space*

Location	R	f_R	Mission	Source
Mercury	250 m	1	MESSENGER	Hawkins *et al.* (2007)
Venus	10 m	10^{-9}	Venera	Freitas (1985)
	250 m	1	Magellan	Saunders *et al.* (1992)
Earth	1 cm	1/3	various	Freitas (1985)
Moon	0.5 m	1	LRO	Robinson *et al.* (2010)
Mars	150 m	.86	MRO	Zurek and Smrekar (2007)
	1 m	.01	HiRISE	Kirk *et al.* (2008)
Ceres	35 m	.97	Dawn	Park *et al.* (2019)
Europa (Jupiter)	230 m	.01	Galileo	Figueredo and Greeley (2000)
Enceladus (Saturn)	110 m	.96	Cassini	Roatsch *et al.* (2008)
Ganymede (Uranus)	3.5 km	.94	Galileo	Patterson *et al.* (2010)
Triton (Neptune)	3 km	1/3	Voyager 2	Abelson (1989)
Pluto	660 m	.45	New Horizons	Olkin *et al.* (2017)
	77 m	.003	New Horizons	Keeter (2017)

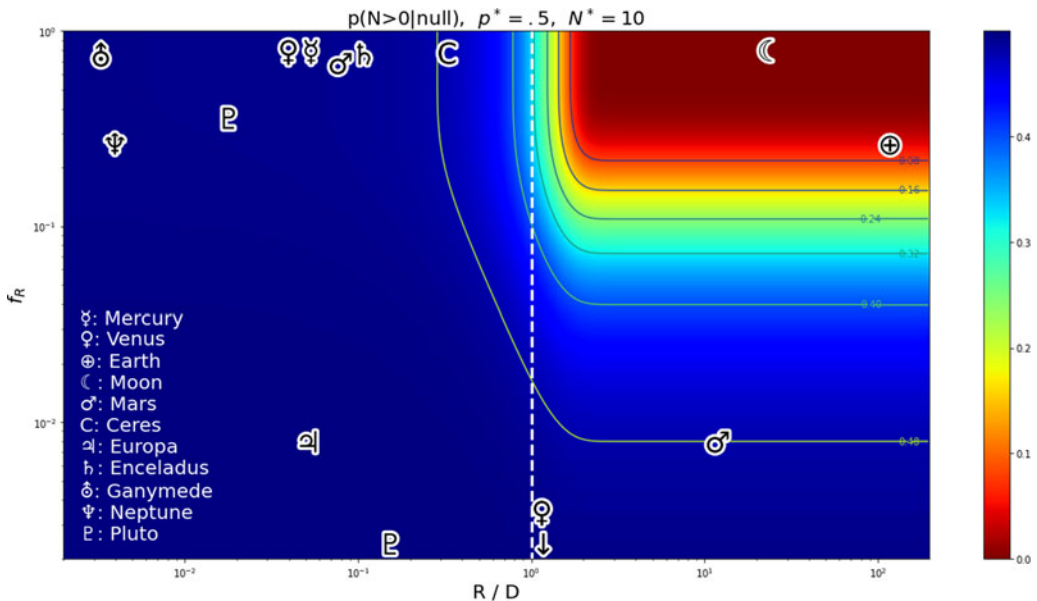Original concept for this table and all numbers adapted from (Lazio, 2022).

Figure 6. *Probability that lurkers exist on the various solar system bodies listed in Table 1. This assumes an a priori probability of 1/2, and considers the scenario where 10 lurkers would be present, with a mean diameter of 10 m.*

$p(N > 0|\text{null})_{\text{tot}} = 1 - \prod_{b \in \text{bodies}} (1 - p(N > 0|\text{null})_b)$. This total probability is most sensitive to the body with lowest probability, suggesting that if our goal is to minimize this quantity, we ought to prioritize the bodies we are least certain about first.

Further applications

What is the best way to determine a habitability boundary?

Supposing we do begin to detect biosignatures, we will become interested in the problem of determining the range of conditions for which life can persist. Though there are a number of variables that certainly play a role (temperature, planet size, star size, atmospheric mass, water content, ellipticity, obliquity, etc.), in this section we focus on a single abstract quantity t. We wish to outline a procedure for determining the range of t which can sustain life. For simplicity we focus on determining the lower bound, denote the current range as $(0, \ t_1)$, and treat the variable as uniformly distributed[2].

In the scenario where all systems with habitable characteristics are guaranteed to possess biosignatures, this problem reduces to the standard method of bisecting the remaining interval until the desired precision is reached. The subtlety here is that if a planet does not exhibit the biosignature, we would not be able to infer that life cannot exist within that range, because presumably life does not inhabit every potentially habitable environment. Let us discuss the certain case first as a warm up, to outline how information gain can be used to arrive at this classic result.

As stated, before our measurement we know that the lower bound for t lies somewhere in the range $(0, \ t_1)$. We will perform a measurement of a system with t_0. A priori, the probability of a detection (the probability that the lower bound is less than t_0) is $P(\text{det}) = t_0/t_1 \equiv c_0$, and the probability of no detection is $P(\text{miss}) = 1 - c_0$. If a signal is detected, then the updated distribution of possible lower bounds is $Pdet = \mathcal{U}(0, t_0)$, and if no signal is detected, the distribution becomes $P(tlower) = \mathcal{U}(t_0, t_1)$. The expected information gain from this measurement is

$$\begin{aligned}\langle S \rangle &= S(P_{\text{det}})P_{\text{det}} + S(P_{\text{miss}})P_{\text{miss}} \\ &= -c_0 \log c_0 - (1 - c_0) \log(1 - c_0)\end{aligned} \tag{27}$$

This quantity is maximized at $t_0 = t_1/2$, recovering the classic bisection method.

Now, let's generalize this to the case where life is only present on a fraction of all habitable planets, denoted f (independent of t in this section). In this case, the probability of detection is given by $P_{\text{det}} = f c_0$, which is the product of the probability of the lower bound being smaller than the observed system, and that life is present. The absence of a detection could be either because the lower bound is greater than the system's value, or because no life happens to be present. So, the probability of no detection is given by the sum of two terms, $P_{\text{miss}} = P(t_0 < t_l) + P(t_l < t_0)(1 - f) = 1 - f c_0$. The expected information gain is given by

$$\begin{aligned}\langle S \rangle &= fc_0 \log c_0 + (1 - c_0) \log (1 - fc_0) \\ &\quad + (1 - f)c_0 \log \left(\frac{1 - fc_0}{1 - f} \right)\end{aligned} \tag{28}$$

After some manipulation, it is found that this quantity is maximized at

$$c_0 = \frac{(1 - f)^{1/f - 1}}{1 + f(1 - f)^{1/f - 1}} \tag{29}$$

When $f \to 1$, this expression goes to $1/2$, reproducing the certain case. When $f \to 0$ it goes to $1/e$, reminiscent of the classic secretary problem, which aims to determine an optimal value in complete absence of information.

The expected information gained by an optimal measurement is

$$\langle S \rangle |_{\text{opt}} = \log \left(1 + f(1 - f)^{1/f - 1} \right) \tag{30}$$

---

[2]if a quantity is not uniformly distributed, its cumulative distribution function can be used as t everywhere in this section.

This goes to $1/e$ for the certain case, representing 1 nat of information gain, and f/e for small f, which is 1 nat multiplied by the detection probability.

We can also determine the convergence rate towards the true lower bound by looking at the expected maximum possible value for a given time step. Recall that in the certain case, this is given by $1/2t_0 + 1/2t_1 = 3/4t_1$, so that this converges as $(3/4)^N$ after N time steps. In the uncertain case, the formula instead yields $t_0 f c_0 + t1(1 - f c_0)$. We do not display the full expression after substituting the optimal value of c_0 here, but instead report its limit for small f, which becomes $(1 - (e - 1)/e^2 f)t_1$. This series converges as approximately $\exp(-.232fN)$, which is considerably slower than the certain case.

What is the optimal amount of money to spend on different mission aspects?

Another application is in determining how much money to spend on each aspect of a mission. In this section we will abstractly define a number of mission aspects $\{a_i\}$, with the total number of observed planets scaling as $N_{tot} = N_0 \prod_i a_i^{q_i}$. Here, we assume that the total number scales with each aspect to some power. For example, the aspects could correspond to telescope diameter D_t, mission lifetime T, spectral resolution $\Delta\lambda$, etc., as found in Sandora and Silk (2020) where $N_{tot} \propto D_t^{12/7} T^{3/7} \Delta\lambda^{3/7}$. In the next subsection we will treat a step-wise increase as well. Now, we assume that a mission has a fixed total budget c_{tot} apportioned amongst each aspect as $c_{tot} = \sum_i c_i$. We would like to determine the optimal way of assigning each c_i so as to maximize the number of observed planets, as well as the resultant scaling of mission yield with cost.

This is a straightforward convex optimization problem, that can be solved inductively by finding the location where the gradient of the total yield vanishes for every component budget. The solution is

$$c_i = \frac{q_i}{q_{tot}} c_{tot}, \quad N_{tot} = N_0 \frac{\prod_i q_i^{q_i}}{q_{tot}^{q_{tot}}} c_{tot}^{q_{tot}} \tag{31}$$

Here, we have defined $q_{tot} = \sum_i q_i$, and assume nothing other than that each exponent is positive.

This yields a concrete optimal fraction of budget to yield to each mission aspect. Particularly noteworthy is the fact that the total yield scales superlinearly with budget for all practical applications of this formalism. This implies, among other things, that the cost per planet decreases dramatically with increased scale. The cost per information gain, however, does not, but is instead minimized at the value

$$cb = \frac{e}{\left(\left(\prod_i q_i^{q_i}/q_{tot}^{q_{tot}}\right)N_0\right)^{1/q_{tot}}} \tag{32}$$

This minimum occurs because the information return depends only logarithmically on total number, and so these diminishing returns must be balanced against initial startup costs. For the reference scenario above, taking spectral resolution, telescope diameter and mission length (to each scale linearly with cost) gives $q_{tot} = 18/7$, and that the optimal cost is $\$10^9$ when $N_0 = 121.8$.

Several things to note about this expression: optimal pricing decreases slowly for missions with higher yield potential, and is smaller for missions that scale faster. If followed, this gives the optimal number of planets for a mission to observe as $N_{opt} = e^{q_{tot}}$, which results in q_{tot} nats of information. However, minimizing cost per information gain in this manner neglects the cost associated with mission development, which can be significantly higher than the costs for the duration of the mission. This is included in section 'The marginal value theorem and mission lifetimes', where we find that the recommended number of systems to observe is substantially larger.

When should an additional instrument be included?

In the previous subsection, we focussed on mission aspects that affected the mission yield continuously and monotonically. Some aspects do not scale this way; a particular case is which an aspect

is either included or not, and its effectiveness cannot be improved other than it simply being present. An example would be whether to include a starshade for a space mission, or whether to include an instrument capable of making a specific measurement that would lead to higher yields.

First, we will consider the case that the type of biosignatures being measured are not affected by the inclusion of a binary aspect, but that only the total yield is. In this case, we can say that the total yield is $N_{tot}(b, c) = N_0(b)c_{tot}^q$, where b is the binary aspect, and $N_0(b) = N_0$ if the aspect is absent and $N_0(b) = N_1$ if it is present. If the cost of the binary aspect is c_b, then this instrument should be included when

$$N_1(c_{tot} - c_b)^{q_{tot}} > N_0\, c_{tot}^{q_{tot}} \tag{33}$$

It may also happen that a binary aspect does not affect the overall number of planets observed, but allows the measurement of a different class of biosignatures. This is the case, for example, when deciding whether to include an instrument that can detect oxygen. In this scenario, the yield without this instrument is given by N_{tot}, whereas with the instrument it is given by $f_{bio}N_{tot}^2$. In this case, the instrument should be included when

$$f_{bio}\, N_0\, (c_{tot} - c_b)^{2q_{tot}} > c_{tot}^{q_{tot}} \tag{34}$$

From this, we can see that the number of planets observed should be at least as large as $1/f_{bio}$, plus some to counteract the fact that fewer systems will be observed with the instrument. If the occurrence rate is too low, the mission will not be expected to observe a signal anyway, and so the measurement of the second class of biosignatures would be a waste.

The marginal value theorem and mission lifetimes

In this subsection we address the question of how long a mission should be run. The yield scales sub-linearly with mission lifetime, as the brightest, most promising stars are targeted first, leaving the stars which require longer integration times for later in the mission. This gives diminishing returns as the mission goes on.

The marginal value theorem from optimal foraging theory (Charnov, 1976) can be applied to determine the point at which a mission should be abandoned in favour of a new mission with greater yield potential. Indeed, collecting exoplanet data closely resembles the process of sparse patch foraging, wherein samples are collected from a finite resource in order of desirability until the yield of a patch is poor enough to warrant the time investment to travel to another. The analogy is completed when we consider that it takes considerable time to design and construct missions. Though different missions are always being run in parallel, they consume a fixed portion of the budget which could otherwise have gone to some other mission.

We will consider a mission at fixed cost that takes a development period t_{dev} before it begins collecting data. Then the yield as a function of time is $N(t) = N_0(t - t_{dev})^{q_t}$, where it should be understood that when this quantity is not positive, it is 0. It is convenient to nondimensionalize the variables in this expression by measuring time in years, so that N_0 represents the number of planets returned in the first year of operation. To determine the optimal stopping time we use the marginal value theorem, which maximizes the average information gain per time, this sets $\dot{S}(t_0) = S(t_0)/t_0$. This can then be solved for t_0, yielding

$$t_{operation} = \frac{1}{W_0\left(N_0^{1/q_t} t_{dev}/e\right)} t_{dev} \tag{35}$$

Here, $W_0(x)$ is the Lambert W function, which is the solution to the equation $W_0 e^{W_0} = x$, and asymptotes to a logarithm for large arguments. So, it can be seen that the operation time scales not quite linearly with development time, and decreases with increasing first year yield. The total science return

over the mission lifetime is given by

$$S(t_0) = q_t \left[W_0 \left(N_0^{1/q_t} t_{\text{dev}}/e \right) + 1 \right] = q_t \left[\frac{t_{dev}}{t_{op}} + 1 \right] \tag{36}$$

When is it best to follow a prediction, and when is it best to challenge it?

We now turn our attention to the scenario where, according to some leading model of habitability, a particular locale should be devoid of life. According to this model, any attempt at measuring a biosignature around such a location would be fruitless, and would have been better spent collecting data from a safer bet. However, we are unlikely to ever be 100% confident in any of our habitability models, and so it is possible that testing the model will result in it being overturned. Here, we show that our framework allows us to determine a criterion for when this habitability model should be obeyed, and when it should be challenged, based off of the confidence one assigns to its truthfulness. A concrete example would be whether to check hot Jupiters for biosignatures, as by most reasonable accounts they do not fulfil the requirements for life (Fortney *et al.*, 2021).

As a toy example, let us consider the case where there are two possibilities: the leading theory, which dictates that life is impossible on some type of system, and a second theory, which assigns a nonzero chance that life may exist. The confidence we assign to the standard theory is denoted p. The prior distribution for the fraction of systems harbouring a signal is then

$$p_b(f) = p \, \delta(f) + (1 - p) \, \delta(f - f_2) \tag{37}$$

Here δ is the delta function, which equals 0 when its argument is nonzero and integrates to 1. More thoroughly, we could consider a range of values for f_2, but this significantly complicates the analysis, and leads to little new insight.

Now we determine the expected amount of information gained from a measurement of a system. According to the above prior, the probability of measuring no signal is $P_0 = p + (1 - p)(1 - f_2)$, and the probability of measuring a signal is $P_1 = (1 - p)f_2$. Then the expected information gain $\langle \Delta S \rangle = P_0 S(p(f|\text{no detection})) + P_1 S(p(f|\text{detection})) - S(p_b(f))$ is:

$$\langle \Delta S \rangle = (1 - p)(1 - f_2) \log \left(\frac{(1 - p)(1 - f_2)}{p + (1 - p)(1 - f_2)} \right)$$
$$- p \log \left(p + (1 - p)(1 - f_2) \right) - (1 - p) \log (1 - p) \tag{38}$$

If we further take the limits $p \to 1$, $f_2 \to 0$, this expression simplifies somewhat:

$$\langle \Delta S \rangle \to (1 - p) \left[f_2(1 - \log(1 - p)) + (1 - f_2) \log(1 - f_2) \right] \tag{39}$$

From this, it can be seen that the expected amount of information gain is proportional to $(1 - p)$, the probability that the theory's recommendation is incorrect.

Equation (38) is displayed in Fig. 7, from where it can be seen that expected information is maximized when f_2 is large and p is small. As example values, take $\langle \Delta S(p = .9, f_2 = .1) \rangle = .02$, $\langle \Delta S(p = .99, f_2 = .1) \rangle = .005$.

To determine whether this measurement should be made, we need to compare it to an alternative measurement yielding a different amount of expected information gain, and see which is larger. In the standard case of measuring a biosignature of frequency f well localized from 0, the information is $S = 1/2 \log (f(1 - f)/N)$ after N measurements, and so the information gain is $\Delta S = 1/(2N)$. So, for the case $p = .99$, $f_2 = .1$, the expected value for the risky measurement exceeds the standard after the standard measurement has been made 107 times. This suggests a strategy of collecting
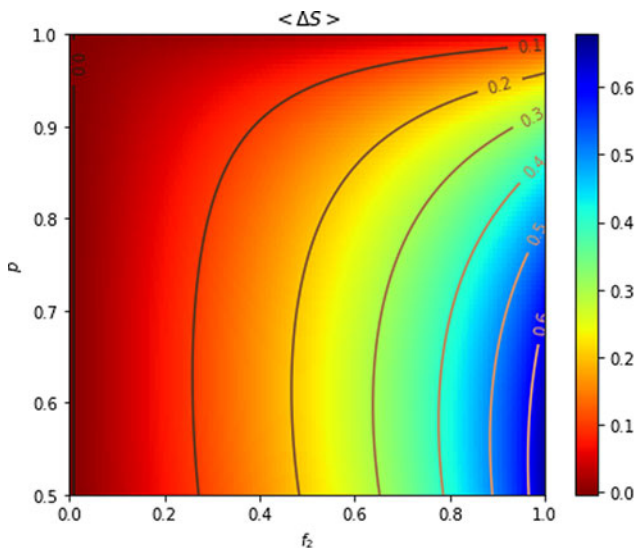
Figure 7. *Information gain for measurement of an unlikely signal as a function of possible signal occurrence rate* f_2 *and confidence in the theory proclaiming its impossibility* p.

measurements around likely places first, until the gains become small enough that measuring the unlikely signal becomes more interesting, on the off chance that it could invalidate the current theory of habitability.

The above assumed that no measurement of the unlikely signal had been attempted previously. We may wish to determine how many times to challenge a theory before accepting it. For this scenario, if N previous measurements have been made, the prior will be

$$p_b(f) \propto p\,\delta(f) + (1-p)(1-f_2)^N \delta(f - f_2) \tag{40}$$

With constant of proportionality enforcing that this is normalized to 1. The expected information gain is, in the limit of $N \rightarrow \infty$,

$$\langle \Delta S \rangle \rightarrow \frac{1-p}{p}\, N f_2\, (1-f_2)^N\, \log\!\left(\frac{1}{1-f_2}\right) \tag{41}$$

The full expression is plotted in Fig. 8. From here, we can clearly see a steep dropoff at $N > 1/f_2$, beyond which empirical evidence strongly indicates the absence of a signal.

Discussion

In this work, we have utilized the concept of *information gain* as a method to assess the value of biosignature missions, and demonstrated that this is capable of generating nontrivial recommendations for mission design and target prioritization, across various domains. Throughout, this has resembled less of a first-principles approach, whereby any mission setup can automatically be loaded into a master equation to straightforwardly generate recommendations, but rather a style of thinking about missions in terms of the quality of the knowledge that they generate as a function of mission parameters. The concept of information explicitly depends on the quantities that are abstracted out of the raw data any mission would return, and so at the outset any analysis of this sort relies on value judgements on the part of the practitioner. We have illustrated that this process does not always yield a unique encapsulation of the data, as for instance when we chose to define how we detect population trends through the distribution of the
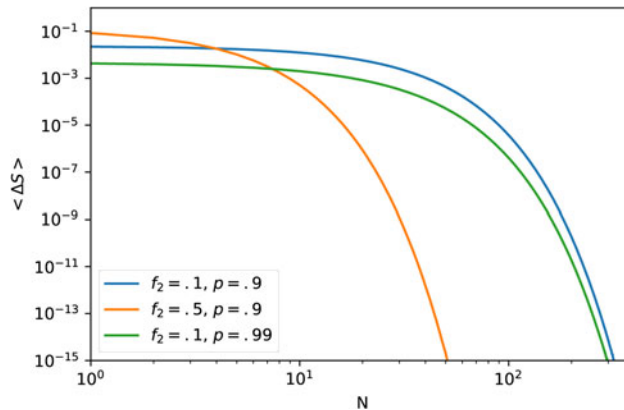
Figure 8. *Information gain of an unlikely signal after a series of* N *unsuccessful measurements. Above* $N \gtrsim 1/f_2$ *the likelihood of measuring a signal becomes exponentially suppressed.*

occurrence rate ratio, rather than some other measure. While there is no guarantee that the recommendations will be independent of the way we set up the problem, the advantage this framework has is in making the mission goals and assumptions explicit when deriving recommendations.

At this point we'd like to reiterate that we are ardently against using information gain as a quantity to maximize for designing missions, since attempts to substitute quantitative proxies for vague overarching goals often results in pathological maximization procedures, often to the detriment of the original intent (Chrystal *et al.*, 2003). Rather, we advocate the use of this framework for generating rules of thumb and an intuitive understanding for when certain mission aspects ought to be improved upon. Additionally, we note that this framework explicitly depends upon an honest and accurate assessment of mission capabilities. Even with rigorous design practices, often there ends up being uncertainties in instrumental sensitivities, background noise and plausibility of modelling scenarios that prevent perfect assessment of the relevant parameters. On top of this, human biases preclude even perfect experts from being able to make impartial assessments for any new venture (Pohorille and Sokolowska, 2020). Nevertheless, our formalism is useful for determining the sensitivity of our recommendations to these unknowns.

We hope that by demonstrating the practicality of this framework to a variety of cases, the reader will find it productive to apply to other, new problems. To this end, our approach has often been the following: first, we seek out a continuous quantity that can be measured with a particular mission setup. Next, we compute the distribution of values compatible with observations, given the model parameters and any other additional assumptions that must be used. Then, lastly, we vary these mission parameters, subject to any relevant mission constraints such as finite budget/lifetime/sample size, and determine the combination of mission parameters that equivalently maximizes information, minimizes uncertainty or maximizes probability of the desired outcome. Our hope is that this may be useful as a tool to aid in determining the design parameters of future missions, providing transparency and impartiality in selecting among upcoming missions and ultimately maximizing our chances of successfully detecting biosignatures.

References

Abelson PH (1989) Voyager 2 at Neptune and triton. *Science* **246**, 1369–1369.

Affholder A, Guyot F, Sauterey B, Ferrière R and Mazevet S (2021) Bayesian analysis of Enceladus's plume data to assess methanogenesis. *Nature Astronomy* **5**, 805–814.

Bains W and Petkowski JJ (2021) Astrobiologists are rational but not Bayesian. *International Journal of Astrobiology* **20**, 312–318.

Balbi A, Hami M and Kovačević A (2020) The habitability of the galactic bulge. *Life* **10**.

Barber DJ and Scott ER (2002) Origin of supposedly biogenic magnetite in the Martian meteorite Allan hills 84001. *Proceedings of the National Academy of Sciences* **99**, 6556–6561.

Benford J (2019) Looking for lurkers: co-orbiters as seti observables. *The Astronomical Journal* **158**, 150.

Casella G and Berger R (2001) *Statistical Inference*. Cambridge: Duxbury Resource Center.

Catling DC, Krissansen-Totton J, Kiang NY, Crisp D, Robinson TD, DasSarma S, Rushby AJ, Del Genio A, Bains W and Domagal-Goldman S (2018) Exoplanet biosignatures: a framework for their assessment. *Astrobiology* **18**, 709–738.

Charnov EL (1976) Optimal foraging, the marginal value theorem. *Theoretical Population Biology* **9**, 129–136.

Chrystal KA, Mizen PD and Mizen P (2003) Goodhart's law: its origins, meaning and implications for monetary policy. *Central Banking, Monetary Theory and Practice: Essays in Honour of Charles Goodhart* **1**, 221–243.

Cockell CS, Kaltenegger L and Raven JA (2009) Cryptic photosynthesis–extrasolar planetary oxygen without a surface biological signature. *Astrobiology* **9**, 623–636.

Daflon S and Cunha K (2004) Galactic metallicity gradients derived from a sample of OB stars. *The Astrophysical Journal* **617**, 1115.

Figueredo PH and Greeley R (2000) Geologic mapping of the northern leading hemisphere of Europa from Galileo solid-state imaging data. *Journal of Geophysical Research: Planets* **105**, 22629–22646.

Fischer DA and Valenti J (2005) The planet-metallicity correlation. *The Astrophysical Journal* **622**, 1102.

Foote S, Sinhadc P, Mathis C and Walker SI (2022) False positives and the challenge of testing the alien hypothesis. preprint arXiv:2207.00634.

Fortney JJ, Dawson RI and Komacek TD (2021) Hot Jupiters: origins, structure, atmospheres. *Journal of Geophysical Research: Planets* **126**, e2020JE006629.

Freitas RA (1983) The search for extraterrestrial artifacts(seta). *British Interplanetary Society, Journal(Interstellar Studies)* **36**, 501–506.

Freitas RA Jr (1985) There is no fermi paradox. *Icarus* **62**, 518–520.

Gray RH (2015) The fermi paradox is neither fermi's nor a paradox. *Astrobiology* **15**, 195–199.

Green J, Hoehler T, Neveu M, Domagal-Goldman S, Scalice D and Voytek M (2021) Call for a framework for reporting evidence for life beyond earth. *Nature* **598**, 575–579.

Haqq-Misra J and Kopparapu RK (2012) On the likelihood of non-terrestrial artifacts in the solar system. *Acta Astronautica* **72**, 15–20.

Harman CE and Domagal-Goldman S (2018) Biosignature false positives. Technical report, Springer International Publishing.

Hawkins S, Boldt J, Darlington E, Espiritu R, Gold R, Gotwols B, Grey M, Hash C, Hayes J, Jaskulek S, Kardian C, Keller M, Malaret E, Murchie S, Murphy P, Peacock K, Prockter L, Reiter R, Robinson M, Schaefer E, Shelton R, Sterner R, Taylor H, Watters T and Williams B (2007) The mercury dual imaging system on the messenger spacecraft. *Space Science Reviews* **131**, 247–338.

Jiménez-Torres JJ, Pichardo B, Lake G and Segura A (2013) Habitability in different milky way environments: A stellar interaction dynamical approach. *Astrobiology* **13**, 491–509.

Johnson JL and Li H (2012) The first planets: the critical metallicity for planet formation. *The Astrophysical Journal* **751**, 81.

Johnson S, Graham H, Anslyn E, Conrad P, Cronin L, Ellington A, Elsila J, Girguis P, House C and Kempes C (2019) Agnostic approaches to extant life detection. *Mars Extant Life: What's Next?* **2108**, 5026.

Keeter B (2017) Pluto's icy plains in highest-resolution views from new horizons. https://www.nasa.gov/image-feature/pluto-s-icy-plains-in-highest-resolution-views-from-new-horizons (Accessed January 31, 2023).

Kirk RL, Howington-Kraus E, Rosiek MR, Anderson JA, Archinal BA, Becker KJ, Cook DA, Galuszka DM, Geissler PE, Hare TM, Holmberg IM, Keszthelyi LP, Redding BL, Delamere WA, Gallagher D, Chapel JD, Eliason EM, King R and McEwen AS (2008) Ultrahigh resolution topographic mapping of mars with mro hirise stereo images: meter-scale slopes of candidate phoenix landing sites. *Journal of Geophysical Research: Planets* **113**.

Lazio J (2022) Technosignatures in the solar system. In *The First Penn State SETI Symposium*. Zenodo.

Levine TR, Weber R, Hullett C, Park HS and Lindsey LLM (2008) A critical assessment of null hypothesis significance testing in quantitative communication research. *Human Communication Research* **34**, 171–187.

Lineweaver CH, Fenner Y and Gibson BK (2004) The galactic habitable zone and the age distribution of complex life in the milky way. *Science* **303**, 59–62.

Lorenz RD (2019) Calculating risk and payoff in planetary exploration and life detection missions. *Advances in Space Research* **64**, 944–956.

Lyu Z, Shao N, Akinyemi T and Whitman WB (2018) Methanogenesis. *Current Biology* **28**, R727–R732.

Meadows V, Graham H, Abrahamsson V, Adam Z, Amador-French E, Arney G, Barge L, Barlow E, Berea A, Bose M, Bower D, Chan M, Cleaves J, Corpolongo A, Currie M, Domagal-Goldman S, Dong C, Eigenbrode J, Enright A, Fauchez TJ, Fisk M, Fricke M, Fujii Y, Gangidine A, Gezer E, Glavin D, Grenfell JL, Harman S, Hatzenpichler R, Hausrath L, Henderson B, Johnson SS, Jones A, Hamilton T, Hickman-Lewis K, Jahnke L, Kacar B, Kopparapu R, Kempes C, Kish A, Krissansen-Totton J, Leavitt W, Komatsu Y, Lichtenberg T, Lindsay M, Maggiori C, Marais D, Mathis C, Morono Y, Neveu M, Ni G, Nixon C, Olson S, Parenteau N, Perl S, Quinn R, Raj C, Rodriguez L, Rutter L, Sandora M, Schmidt B,

Schwieterman E, Segura A, Şekerci F, Seyler L, Smith H, Soares G, Som S, Suzuki S, Teece B, Weber J, Simon FW, Wong M and Yano H (2022) Community report from the biosignatures standards of evidence workshop. preprint arXiv:2210.14293.

Mousis O, Lunine J, Waite J, Magee B, Lewis W, Mandt K, Marquer D and Cordier D (2016) Formation conditions of enceladus and origin of its methane reservoir. *The Astrophysical Journal* **701**, 39–42.

Mukherjee R and Sen B (2019) On efficiency of the plug-in principle for estimating smooth integrated functionals of a nonincreasing density. *Electronic Journal of Statistics* **13**, 4416–4448.

Neveu M, Hays LE, Voytek MA, New MH and Schulte MD (2018) The ladder of life detection. *Astrobiology* **18**, 1375–1402.

Olkin CB, Spencer JR, Grundy WM, Parker AH, Beyer RA, Schenk PM, Howett CJ, Stern SA, Reuter DC, Weaver HA, Young LA, Ennico K, Binzel RP, Buie MW, Cook JC, Cruikshank DP, Ore CMD, Earle AM, Jennings DE, Singer KN, Linscott IE, Lunsford AW, Protopapa S, Schmitt B, Weigle E and the New Horizons Science Team (2017) The global color of Pluto from new horizons. *The Astronomical Journal* **154**, 258.

Park RS, Vaughan AT, Konopliv AS, Ermakov AI, Mastrodemos N, Castillo-Rogez JC, Joy SP, Nathues A, Polanskey CA, Rayman MD, Riedel JE, Raymond CA, Russell CT and Zuber MT (2019) High-resolution shape model of Ceres from stereophotoclinometry using Dawn imaging data. *Icarus* **319**, 812–827.

Patterson GW, Collins GC, Head JW, Pappalardo RT, Prockter LM, Lucchitta BK and Kay JP (2010) Global geological mapping of Ganymede. *Icarus* **207**, 845–867.

Pham-Gia T (2000) Distributions of the ratios of independent beta variables and applications. *Communications in Statistics - Theory and Methods* **29**, 2693–2715.

Pohorille A and Sokolowska J (2020) Evaluating biosignatures for life detection. *Astrobiology* **20**, 1236–1250.

Prantzos N (2008) On the 'galactic habitable zone'. *Strategies of Life Detection*, pp. 313–322.

Rauber T, Braun T and Berns K (2008) Probabilistic distance measures of the Dirichlet and beta distributions. *Pattern Recognition* **41**, 637–645.

Roatsch Th, Waehlisch M, Giese B, Hoffmeister A, Matz K-D, Scholten F, Kuhn A, Wagner R, Neukum G, Helfenstein P and Porco C (2008) High-resolution Enceladus atlas derived from Cassini-ISS images. *Planetary and Space Science* **56**, 109–116.

Robinson MS, Brylow SM, Tschimmel M, Humm D, Lawrence SJ, Thomas PC, Denevi BW, Bowman-Cisneros E, Zerr J, Ravine MA, Caplinger MA, Ghaemi FT, Schaffner JA, Malin MC, Mahanti P, Bartels A, Anderson J, Tran TN, Eliason EM, McEwen AS, Turtle E, Jolliff BL and Hiesinger H (2010) Lunar Reconnaissance Orbiter Camera (lROC) instrument overview. *Space Science Reviews* **150**, 81–124.

Sandora M and Silk J (2020) Biosignature surveys to exoplanet yields and beyond. *Monthly Notices of the Royal Astronomical Society* **495**, 1000–1015.

Saunders RS, Spear AJ, Allin PC, Austin RS, Berman AL, Chandlee RC, Clark J, Decharon AV, De Jong EM, Griffith DG, Gunn JM, Hensley S, Johnson WTK, Kirby CE, Leung KS, Lyons DT, Michaels GA, Miller J, Morris RB, Morrison AD, Piereson RG, Scott JF, Shaffer SJ, Slonski JP, Stofan ER, Thompson TW and Wall SD (1992) Magellan mission summary. *Journal of Geophysical Research: Planets* **97**, 13067–13090.

Sheikh SZ (2020) Nine axes of merit for technosignature searches. *International Journal of Astrobiology* **19**, 237–243.

Waite JH, Glein CR, Perryman RS, Teolis BD, Magee BA, Miller G, Grimes J, Perry ME, Miller KE, Bouquet A, Lunine JI, Brockwell T and Bolton SJ (2017) Cassini finds molecular hydrogen in the Enceladus plume: evidence for hydrothermal processes. *Science* **356**, 155–159.

Westall F, Foucher F, Bost N, Bertrand M, Loizeau D, Vago JL, Kminek G, Gaboyer F, Campbell KA, Bréhéret J-G, Gautret P and Cockell CS (2015) Biosignatures on mars: what, where, and how? implications for the search for Martian life. *Astrobiology* **15**, 998–1029.

Yung YL, Chen P, Nealson K, Atreya S, Beckett P, Blank JG, Ehlmann B, Eiler J, Etiope G, Ferry JG, Forget F, Gao P, Hu R, Kleinböhl A, Klusman R, Lefèvre F, Miller C, Mischna M, Mumma M, Newman S, Oehler D, Okumura M, Oremland R, Orphan V, Popa R, Russell M, Shen L, Lollar BS, Staehle R, Stamenković V, Stolper D, Templeton A, Vandaele AC, Viscardy S, Webster CR, Wennberg PO, Wong ML and Worden J (2018) Methane on mars and habitability: challenges and responses. *Astrobiology* **18**, 1221–1242.

Zurek RW and Smrekar SE (2007) An overview of the Mars Reconnaissance Orbiter (MRO) science mission. *Journal of Geophysical Research: Planets* **112**.