Valuing life detection missions

Edwin S. Kite* (University of Chicago), Eric Gaidos (University of Hawaii), Tullis C. Onstott (Princeton University). * kite@uchicago.edu

Recent discoveries imply that Early Mars was habitable for life-as-we-know-it (Grotzinger et al. 2014); that Enceladus is habitable (Waite et al. 2017); and that many stars have Earth-sized exoplanets whose insolation favors surface liquid water (Dressing & Charbonneau 2013, Gaidos 2013). These exciting discoveries make it more likely that spacecraft now under construction – Mars 2020, ExoMars rover, JWST, Europa Clipper – could find habitable, or formerly habitable, environments. Did these environments see life? Given finite resources ($10bn/decade for the US\textsuperscript{1}), how could we best test the hypothesis of a second tree of life? Here, we first state the case for and against flying life detection missions soon. Next, we assume that life detection missions will happen soon, and propose a framework (Fig. 1) for comparing the value of different life detection missions:

\[
\text{Scientific value} = \frac{(\text{Reach} \times \text{grasp} \times \text{certainty} \times \text{payoff})}{\$} \quad (1)
\]

After discussing each term in this framework, we conclude that scientific value is maximized if life detection missions are set up as hypothesis tests. With hypothesis testing, even a nondetection outcome is scientifically valuable.

Should the US fly more life detection missions?

The case for emphasizing habitable environments, not life detection: Our one attempt to detect life, Viking, is viewed with hindsight as naïve or at best uncertain. In-space astrobiology experiments are expensive. The histories of other expensive experimental disciplines, such as US laser fusion and US particle physics, are cautionary tales about over-promising. Today, astrobiology sustains Congressional and public enthusiasm for planetary science. Astrobiology hypothesis tests that yield nondetections (even if they are false negatives) will dampen this enthusiasm (e.g. Viking). Astrobiology as a discipline remains immature (Heng 2016), and perhaps the real payoff would be something so totally unexpected that it is missed. To the extent that the science questions cannot be precisely defined in advance, a better motivation for planetary missions is pure exploration – to push the boundaries of what humans can do, visit, and know. This argues that the next generation of astrobiology missions should have only modest life detection capabilities, and should instead emphasize detecting and characterizing habitable environments rather than detecting extinct or extant life.

The case for flying more life detection missions: Life appears near the start of Earth’s geologic record and could be widespread in the Universe. A detection of a second tree of life has the potential to transform the science of biology. It would also provide guidance about our own future (Bostrom 2008), including the human role in the Solar System. If we indefinitely defer decisive astrobiology tests, then astrobiology is simply PR for planetary science and astronomy.

\textsuperscript{1} We pick $10bn/decade as a rough estimate of current US spending on astrobiology. We note that international cooperation gathers talent, brings a reduced probability of cancellation, and is valuable in itself.
To optimally spend the $10 bn allotted to us over the next decade, astrobiologists should aim to test as quickly, as decisively, and as often as possible.

As the number of habitable extraterrestrial environments increases, the arguments for flying life detection missions become stronger. Recent developments, such as the publication of the report of the Science Definition Team for the Europa Lander (Hand 2017), show that life detection missions are again being seriously considered. Therefore, we now need a framework for valuing different life detection mission profiles.

Here we emphasize science goals, not the specifics of mission implementation. A rigorous comparison of any two specific missions would have to consider many nuances to the design of the specific missions – for example instrumentation capabilities, trajectory design, and risk versus cost – and we do not attempt to do that here.

**Reach:**
One measure of value is a mission’s reach - how many independent opportunities for finding life are there at the specific target to be investigated? This depends on (1) the size and diversity of the target environment, and (2) the fraction of the target environment that is effectively sampled by the mission.

For life-as-we-know-it, the odds that life will establish itself and persist get better the greater the area of rock-water interfaces, the greater the dynamical cross-section for panspermia, and the longer-lived the habitable conditions. More diverse environments are more likely to encompass the (unknown) conditions required for life to establish itself. So, looking for microbes on a globally habitable Early Mars may be a better prospect than looking for life in now-frozen impact-generated lakes on Titan (Artemieva & Lunine 2003). On the other hand, big worlds can be so geologically active as to erase ancient fossils (Earth and Venus are examples; Sleep & Bird 2007). The potential reach of exoplanet missions is $\approx 10^8$ habitable-zone worlds, although we do not yet know what fraction of habitable-zone planets are habitable, and so do not yet know how the number of habitable targets depends on telescope specifications (e.g. Postman et al. 2010).
Reach is maximized when geological processes allow a single mission to probe a voluminous environment. For this to happen, life or biosignatures must be conveyed to the probed location by groundwater flow, cryovolcanism, winds, or currents. Winds and currents swiftly mix planet-sized environments. Therefore, the reach of a mission that probes an atmosphere, the surface of a globally habitable planet, or an ocean is large. Exoplanet spectra also probe global environments. By contrast, groundwater flow can be slow and spatially restricted (this is why petroleum can be trapped, and rarely flows more than ~100 km from its source; Hunt 1995). The transport of living cells can be still slower than the transport of biosignatures. Yet subsurface environments may be longer-lived than surface environments (e.g. Mars; Grimm et al. 2017) and can preserve life’s signatures very well (e.g. petroleum again; Peters et al. 2005). Nonetheless, a mission searching for a rock-hosted biosphere has a reach that is a small fraction of the planet’s crustal habitable volume. Proving or falsifying the hypothesis that a rock-hosted biosphere exists deep beneath a hostile-to-life surface in any given planetary crust may simply be too expensive for the current budget constraints.

The scale of an environment is a very indirect measure of its probability for life. For example, Europa has ~100× the seafloor area of Enceladus. Intuitively, Europa might have a higher probability for life (all other things equal), in part because larger environments are more likely to be stable and persist – but not 100× more. This intuition can be captured by using a log prior (or a log log prior; Lacki 2016). These priors say that, if we are very ignorant about the likelihood of the origin of life – and we are – then it is likely that all suitable planet-sized environments are either inhabited (life as self-organizing principle), or all uninhabited (life as chemical accident), and rather unlikely that only some are inhabited. Priors that behave in this way moderate the importance of reach. The more ignorant we are, the more it makes sense to look for life in habitable environments that are small, short-lived, or both. Examples include asteroid parent bodies (Gaidos & Selsis 2006), and rock bodies on Earth that were buried to uninhabitable depths but have since been exhumed (Onstott 2016). For the same reason, a search for life on the hundred closest habitable planets is not much less valuable than a sweep of the entire Galaxy.

**Grasp and Certainty:**

We define grasp as equal to one minus the probability of a life-detection false negative (estimated pre-launch). For example, for a mission to sample cryovolcanic plume materials, both volcanic processes and instrumental effects potentially dilute or damage biosignatures, and so both can contribute to the false negative probability. False-negative probability is defined relative to the best-available pre-launch understanding of the density/distribution of life/biosignatures in the specific materials to be actually investigated (based on Earth’s geologic record, Earth-analog environments, lab work on radiation-degradation / cryovolcanic processes, e.t.c). A mission that asks the wrong question (relative to our pre-launch model of what is there to be found) has low grasp regardless of whether it is ready to successfully answer that question.

We define certainty as equal to one if the post-detection probability of a false positive is low enough to permit scientific consensus, and zero otherwise. Recent examples of scientific certainty include the discoveries of Neanderthal DNA (Krings et al. 1997), gravitational waves (Abbott et al. 2011), and the $^{16}$O enrichment of the Sun relative to the planets (McKeegan et al. 2011). Solid scientific progress demands certainty. Integrating multiple approaches maximizes certainty. For example, by integrating molecular, isotopic, and textural clues, scientific certainty about ancient microbial life is possible (Wacey et al. 2009, Knoll et al. 2016). By contrast,
isotopes-only claims and texture-only claims for Hadean life (Bell et al. 2015, Nutman et al. 2016) have not led to consensus. The histories of claims about Precambrian life and life in Martian meteorites (McKay et al. 1996) illustrate that reaching consensus involves a loop: analyze, interpret, critique, then analyze again. This loop takes time. Time may be in short supply on an in-situ mission (Hand 2008).

Returned-sample life detection, therefore, has better certainty and grasp than in-situ life detection. Sample return allows thorough molecular (e.g. Summons et al. 2008), isotopic (e.g. Stephan et al. 2016) and textural inspection, but only up to the limits set by sample size and by contextual documentation at the sampled site. If too few samples or undersized samples are returned, for example due to an overestimate of the in-space density of interesting samples (Westphal et al. 2014), then a false negative can occur. Nevertheless, sample return to Earth maximizes grasp and certainty (Mustard et al. 2013). This comes at a cost: for sample return from a habitable body, >$1bn. This cost is increased by rules that are set by NASA. Given constraints on spacecraft mass and cost, the coring and sample caching payload on a mission that is intended to be the first in a sample return campaign (e.g. Mars 2020) might seem to maximize hoped-for future certainty (from subsequent sample return) at the expense of instruments that could increase grasp – unless the in-situ analyses are able to complement the package by increasing grasp (Mustard et al. 2013). Certainty for extant life detection comparable to that of sample return might be achieved by using proxies for life, such as motility or perhaps consumption of redox gradients, that can be measured in-situ (Nadeau et al. 2016, Weiss et al. 2000). The grasp of in-situ instruments will improve with further technology investments. However, relative to Earth laboratories, flight instruments have lower certainty for life detection and cannot be swapped out nor upgraded in response to initial results.

Despite great reach, exoplanet life detection using inner-Solar-System telescopes has low certainty. Spectroscopic detection of high levels of O2 and/or chemical disequilibrium have been proposed as exoplanet biomarker candidates (Schwietermann et al. 2017). Both are really bio-hinters, because most detectable gas combinations can be produced without life. Exceptions, such as isoprene or CFCs, are too rare in Earth’s atmosphere for detection at interstellar distances (Seager et al. 2012). Unfortunately, the CH4+O2 combination yielded by Earth’s biosphere has been undetectable in long-range transmission spectroscopy throughout Earth’s history: today because of low CH4 and refraction (Misra et al. 2014), and in the Precambrian due to low CH4 (Reinhard et al. 2017) or negligible O2. If a large fraction of planets have both abundant CH4 and abundant O2, then this would be hard to explain in terms of abiotic transients (Kasting & Catling 2017). However, we have no reason to think that a large fraction of inhabited planets will be so cooperative. Non-gas biosignatures such as the vegetation red edge (Seager et al. 2005) are intriguing, but for these, little effort has yet been spent on modeling to identify false positives.

These problems cannot be evaded by probabilistic approaches, because our prior uncertainty on life’s abundance is so broad (Lacki 2016), and rocky planets are diverse. Thus, a statement such as “planet Z has an 80% probability of life” can only make sense if we have a comparative sample of similar rocky planets. Earth may be too narrow a template for an inhabited planet due to anthropocentric selection effects. We cannot use uninhabitable rocky planets as the control set, because abiotic false positives are correlated with some abiotic processes that promote habitability. Although we can test the hypothesis that biospheres are “uncommon” (they would stand out with respect to other habitable but uninhabited planets), we cannot deal with the “rare” biosphere case (sample size will always be insufficient to detect the rare biosphere itself – as
well as to rule out equally rare, but expected, abiotic false-positive scenarios) nor the “prolific” biosphere case (all or almost all habitable planets have life). Thus, we might detect a true biosignature, but not know with certainty that life is the source. For example, suppose that, 99% of \( O_2 \)-rich atmospheres have \( O_2 \) as the result of life. Solar System telescopes could never approach 99% certainty that even one of those atmospheres roofs a biosphere, because abiotic \( O_2 \) production scenarios cannot be ruled out to this confidence level\(^2\). This has practical consequences for the allocation of JWST time. Should we look for biosignatures around a few planets, or instead for indicators of habitability in a much larger sample of planets (Bean et al. 2017)?

**Payoff:**

To find an independent origin of life would be a scientific breakthrough. The breakthrough would have a payoff that would depend on the nature of the evidence. Ancient-fossil evidence would be studied using the same techniques used to study Precambrian fossils on the Earth. Depending upon the preservation, these techniques constrain metabolism, composition and cell size and structure, but say little about genetics (Knoll et al. 2016). Therefore, ancient-fossil life would have limited direct impact on sciences outside astrobiology unless the fossils preserved their molecular structure. Space-telescope detection of an exoplanet biosphere (Dalcanton et al. 2015) would offer tantalizingly little information about the organisms themselves. This might stimulate interstellar flight if the biosphere orbits a nearby star (Lubin 2016), or the construction of very large single-target space telescopes to study/monitor the biosphere. Detection of extant life (or young fossils that retain DNA or equivalent) would offer the biggest payoff. For example, information about intact life might transform the biological sciences – which, via the health sector, underpin >10% of Gross World Product. Microbial life that shares a common ancestor with life-as-we-know-it might be easiest to exploit economically, but analysis of life that evolved completely independently could solve a wider range of scientific puzzles.

Even a low-payoff detection would supercharge space exploration, and thus potentially speed up the discovery of high-payoff life elsewhere. This moderates the importance of payoff. Similarly, confirmed in-situ detection of a living organism (with no characterization) is almost as good as retrieval, because a retrieval mission would then be launched by one or more countries with minimal delay.

**Valuing life detection missions as hypothesis tests**

Inevitably, our notional attempt to apply the criteria of reach, grasp, certainty and payoff (Table 1) is mottled by our blind spots and prejudices as authors. Our intent is to encourage a broader discussion that draws on the community’s collective expertise. Moreover, Table 1 could be repainted by a scientific wildcard, such as liquid water at <1 km depth on Europa, or by a technology development, such as fission reactors for deep-space missions (McNutt et al. 2015). Nevertheless, two low-cost opportunities appear to have potential out of proportion to current funding. One is SETI. The other is study of natural origin of life experiments in Earth’s subsurface – isolated water pockets that were first sterilized, then exhumed to habitable depths (Holland et al. 2013). These terrestrial environments are dwarfed by the crustal volume of Mars, but using a log prior this should not count against them too strongly. Intraterrestrial origin-of-life

\[^2\] At least as long as high-resolution data to constrain abiotic-planet models are confined to the solar system.
experiments can be investigated by sterile drilling, which is in any case a needed technology for ocean-worlds exploration. This argues for NSF-NASA or DoE-NASA cooperation.

Using origin-of-life research to drive target selection is risky. Because the geologic setting(s) of abiogenesis is (are) unknown (e.g. McCollom & Seewald 2013), geologically diverse targets – and targets with the highest production rates of free energy – are the best bets. However, prioritizing a mission because of any one origin-of-life hypothesis is questionable. For example, the environment targeted for life detection can be distant (physically and chemically) from the environment of abiogenesis: fragmentation during impacts early in Solar System history enables re-inoculation after giant impacts (Wells et al. 2003). Nevertheless, prebiotic systems where life did not arise might inform origin-of-life research. Life might yet be created in the laboratory – perhaps tomorrow. Would this inform the search for life on other worlds? Probably not: there may be many mechanisms for abiogenesis – many roads to life – and because of the timescale and chemical limitations of laboratory work, we should not expect the one that first works in the lab to be the same as the one that happened at planetary scale.

Once a habitable environment has been identified, refined constraints on energy-and-nutrient-fluxes offer (limited) guidance for target selection. Biomarker production/concentration could scale with energy and nutrient fluxes and so these places offer better sensitivity for life detection. However, life endures in nutrient-poor environments (Priscu et al. 1999), many energy conservation strategies are possible (Schulze-Makuch & Irwin 2002), and – if given an initial minimal nutrient budget and an energy source – a biosphere may self-sustain via heterotrophy, recycling and adaptation.

Current reconnaissance missions, such as MRO and Europa Clipper, have a strong science return regardless of astrobiology outcome. However, life detection requires instruments that differ from those used to study habitable environments. Therefore, future Solar System astrobiology planners will have to weigh continued characterization of habitable environments against life detection.

A life detection mission is a hypothesis test if the probability of life is greatly reduced by a nondetection. Missions that are not hypothesis tests – usually due to low grasp – have low value within the framework we propose here. Although it has been said that “exploration often cannot be hypothesis testing” (Chyba & Phillips 2001, Hand 2017), hypothesis-testing has served us well in the past (e.g., Mars Science Program Synthesis Group 2004). Hypothesis-testing also offers a clear basis for reallocating resources in response to negative results (Smolin 2006).

Recent successful missions have uncovered apparently-habitable environments. Each target offers unique tradeoffs. As the number of known habitable environments increases, it will be tempting to rebalance the US astrobiology portfolio away from continued exploration of habitable environments, and towards testing the hypothesis of life. Proposed life detection missions may be valued by sizing up their reach, grasp, certainty, and payoff. Missions that emphasize life detection should test astrobiology hypotheses. Life detection missions have low scientific value unless a negative result can guide future decisions and future missions – for example, whether or not to move on to more promising targets.

**Acknowledgements.** We thank Bethany Ehlmann and Chris House for feedback, without implying agreement with the points made here.
<table>
<thead>
<tr>
<th>Mission profile</th>
<th>Reach</th>
<th>Grasp</th>
<th>Certainty</th>
<th>Payoff</th>
<th>New technologies needed</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space-based</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Return sample of ancient Mars surface environments (e.g., Mars 2020 to Jezero)</td>
<td>++</td>
<td>+/++</td>
<td>++</td>
<td>+</td>
<td></td>
<td>$$$</td>
</tr>
<tr>
<td>Return sample of ancient Mars subsurface environments</td>
<td>n.a./+</td>
<td>+</td>
<td>++</td>
<td>+</td>
<td></td>
<td>$$$</td>
</tr>
<tr>
<td>Mars deep drill, in-situ measurements only (assuming present-day aquifers exist)</td>
<td>n.a./+</td>
<td>+</td>
<td>+</td>
<td>++</td>
<td>Compact high-output power source</td>
<td>$$$$</td>
</tr>
<tr>
<td>Seek refugia on present Mars surface, in-situ measurements only</td>
<td>n.a.</td>
<td>+</td>
<td>+</td>
<td>++</td>
<td></td>
<td>$/$$</td>
</tr>
<tr>
<td>Mars in situ paleontology (e.g., ExoMars lander)</td>
<td>+++</td>
<td>n.a.</td>
<td>n.a./+</td>
<td>+</td>
<td>Enhanced by improved in-situ instruments</td>
<td>$$</td>
</tr>
<tr>
<td>Ocean world drill to ≲1m (e.g. Europa lander)</td>
<td>++</td>
<td>n.a/+</td>
<td>(+)</td>
<td>+</td>
<td></td>
<td>$$(?$$)</td>
</tr>
<tr>
<td>Ocean world, probe liquid water ocean in-situ</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>Compact high-output power source</td>
<td>$$$</td>
</tr>
<tr>
<td>Ocean-sourced plume in-situ (e.g., Enceladus Life Finder)</td>
<td>++</td>
<td>+</td>
<td>+</td>
<td>++</td>
<td>Enhanced by improved in-situ instruments</td>
<td>$$</td>
</tr>
<tr>
<td>Ocean-sourced plume sample return (e.g., Life Investigation For Enceladus)</td>
<td>++</td>
<td>+</td>
<td>++</td>
<td>++</td>
<td></td>
<td>$$</td>
</tr>
<tr>
<td>Retrieve earliest Earth materials (from “Earth’s attic,” the Moon)</td>
<td>+</td>
<td>n.a.</td>
<td>++</td>
<td>n.a./+</td>
<td></td>
<td>$/$$</td>
</tr>
<tr>
<td>Exoplanet survey transit or direct imaging</td>
<td>+++</td>
<td>+</td>
<td>n.a.</td>
<td>+</td>
<td></td>
<td>$$/$$$$$</td>
</tr>
<tr>
<td>Investigate material from interstellar interloper (e.g. ’Oumuamua)</td>
<td>n.a.</td>
<td>n.a.</td>
<td>+</td>
<td>+/++</td>
<td>High ΔV to land, or return samples</td>
<td>$$</td>
</tr>
<tr>
<td>Interstellar probe</td>
<td>++</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>Interstellar propulsion &amp; communication</td>
<td>$10^{11}$-$10^{12}$</td>
</tr>
<tr>
<td>Earth-based</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laboratory origin of life experiments</td>
<td>n.a.</td>
<td>+</td>
<td>++</td>
<td>+</td>
<td></td>
<td>&lt;$</td>
</tr>
<tr>
<td>Probe natural origin of life experiments on Earth</td>
<td>n.a.</td>
<td>+</td>
<td>++</td>
<td>+</td>
<td></td>
<td>$</td>
</tr>
<tr>
<td>SETI</td>
<td>+++</td>
<td>n.a.</td>
<td>++</td>
<td>+++</td>
<td></td>
<td>$</td>
</tr>
</tbody>
</table>

Table 1. A matrix for assessing the life detection case for selected potential mission profiles. We were unable to resist the temptation to make a notional attempt ourselves. This notional attempt is shown in the bottom right corner of each cell as a small box. The small size is to emphasize that this notional attempt is mottled by our blind spots and prejudices as authors, and to allow readers to ‘fill in the blanks’ themselves. Our intent is to encourage a broader discussion that draws on the community’s collective expertise. (n.a. = does not strengthen the life detection case for a mission, + = ambivalent implications for the life detection case for a mission, ++ = Bolsters the life detection case for a mission, +++ = Offers strong support for the life detection case for a mission, * = Depends on geologic history of landing site). Notional costs: $ = <1 bn. $$ = 1-3 bn. $$$/$$ = 3-10 bn. $$$$/$$$/$$ = >10 bn
References.
Gaidos, E., 2013, Candidate Planets in the Habitable Zones of Kepler Stars, Astrophys. J. 770, article id.90
Heng, K., 2016, The Imprecise Search for Extraterrestrial Habitability, American Scientist, 104(3), 146-, doi:10.1511/2016.120.146.


Wacey, D., 2009, Early life on Earth, a practical guide, Springer.


