Epistemology and Anomaly Detection in Astrobiology

David Kinney[1] and Christopher Kempes[2]

[1]Princeton University

[2]Santa Fe Institute

June 23, 2022

Abstract

We examine the epistemological foundations of a leading technique in the search for evidence of life on exosolar planets. Specifically, we consider the "transit method" for spectroscopic analysis of exoplanet atmospheres, and the practice of treating anomalous chemical compositions of the atmospheres of exosolar planets as indicators of the potential presence of life. We propose a methodology for ranking the anomalousness of atmospheres that uses the mathematical apparatus of support vector machines, and which aims to be agnostic with respect to the particular chemical biosignatures of life. We argue that our approach is justified by an appeal to the "hinge" model of epistemic justification first proposed by Wittgenstein (1969). We then compare our approach to previous work due to Walker et al. (2018) and Cleland (2019a, 2019b).

1 Introduction

The search for life beyond planet Earth is fraught with epistemological challenges. Perhaps most saliently, astrobiologists must be open to the possibility that extraterrestrial organisms are very unlike organisms on Earth.[1] Indeed, even the most basic indicators that an object on Earth is an organism (e.g., that it produces carbon dioxide or water as a product of respiration or photosynthesis) may not be typical features of organisms living in environments very different to any found on Earth. Even on Earth, many biosignatures would not have been classified as such until the

[1]There is a rich recent tradition of philosophical commentary on astrobiology, not all of which we are able to comment on here; see for instance Fry (2000), Chela-Flores (2011), and Dick (2012).

recent discovery of exotic metabolic processes (see Vance and Jacobs (2005)). Moreover, when it comes to the detection of signs of life on planets outside of our solar system (hereafter, "exosolar planets"), astrobiologists are severely limited with respect to the kind of data that they can even collect. Using the "transit method" of spectroscopic analysis, to be described in more detail in what follows, scientists are able to make inferences regarding the chemical compositions of the atmospheres of exosolar planets. However, it is far from obvious how we might distinguish between chemical compositions that indicate the presence of life and those that do not, especially since, for the reasons given above, we cannot assume that life on exosolar planets will produce atmospheric biosignatures similar to those produced by living organisms on Earth.

In this paper, we advance the thesis that planets with highly *anomalous* atmospheres relative to a broader sample should be regarded as the most likely settings for extraterrestrial life. We argue that the parameters for any measure of anomalousness should be data-dependent, rather than set *a priori*, so that they do not encode Earth-centric assumptions regarding the chemical signatures of life. We propose a specific measure of anomalousness, the applicability of which is justified by the following assumptions: 1) that a given sample of exosolar planets can be representative (i.e., that it can be large enough and sampled with sufficient randomness so as to allow for meaningful statistical inference about the entire population of exosolar planets), 2) that living organisms tend to leave biosignatures in the atmospheres of the planets that they inhabit, due to metabolism, 3) that life in the set of observable exosolar planets is rare, and 4) that there are not common abiotic processes that mimic the effects of biological metabolism. Each of these assumptions lacks strong empirical justification, and is therefore subject to skepticism. Nevertheless, we argue that these assumptions can still be justified by the role that they play in the cognitive project of astrobiological inquiry. In this respect, our argument for an anomaly-detection-based approach to astrobiological inquiry is in keeping with the "hinge epistemology" developed by Wittgenstein (1969), who held that all cognitive projects depend on assumptions that lack empirical justification. Such assumptions, it is argued, are justified by the essential role that they play in establishing a context of inquiry. This epistemic framework is also in keeping with work by Achinstein (2018) and Currie (2021) defending the value of *speculation* in scientific practice.

Our proposed measure of anomalousness is inspired by work in machine learning that uses support vector machines to detect anomalies in visual images, as developed by Banerjee et al.

(2006), with further inspiration coming from the specific techniques for data-driven parameter estimation proposed by Ghafoori et al. (2018). This measure is presented in full mathematical detail in Appendix A. We argue that our proposed measure can be contrasted with both a Bayesian approach to astrobiological inquiry advocated by Spiegel and Turner (2012) and Walker et al. (2018). As such, our conclusions in this paper have implications both for the philosophical study of the epistemology of astrobiology, and for the actual practice of astrobiology in the near-to-medium-term future.

Within the broader context of the philosophy of biology, a noteworthy feature of our proposed approach to astrobiological inquiry is that it does not require the specification of a set of necessary and sufficient conditions for an object to count as being alive. Thus, our arguments in this paper are in keeping with both the anti-definitionist thesis in philosophy of biology, which claims that 'life' cannot and should not be defined (as advocated by, among others, Cleland and Chyba (2002, 2007), Machery (2012) and Cleland (2019a, 2019b)) as well as "diagnostic" definitions of life due to Knuuttila and Loettgers (2017) and Bich and Green (2018).

The remainder of this paper proceeds as follows. In Section 2, we present our anomaly-based approach to exoplanet life detection. In Section 3, we elaborate on the extent to which our proposal depends on a hinge-epistemology approach to epistemic justification, with connections to philosophical defenses of speculative science. In Section 4, we contrast our approach with a Bayesian framework for astrobiological inquiry. We argue that the Bayesian approach is only useful and well-defined if one makes the sort of fine-grained, quantitative, *a priori* assumptions that our approach avoids. In Section 5, we compare and contrast our view with that of Carol Cleland, who has also defended an approach to astrobiological inquiry based on anomaly detection. We conclude in Section 6.

2 An Anomaly-Based Approach to Exoplanet Astrobiology

We propose a statistical procedure for identifying the exosolar planets that are the most likely possible candidates for hosting living organisms. The scientific underpinning for our procedure is as follows. Seager (2014) argues, based on current astronomical capabilities, that in the near future astronomers will be able to use spectroscopic measurements to infer the chemical composition of

the atmospheres of exosolar planets. To date, inferences as to the chemical compositions of the atmospheres of exosolar planets have been made using the "transit method", for which Burrows (2014) provides the following high-level description: when an exosolar planet passes between Earth-based telescopes and its star, the spectrum of light emitted from the star is altered in ways that are indicative of the presence of certain gases. When the same planet then passes behind its star, so that the star is between the planet and Earth, astronomers are able to observe the spectrum of light emitted from the star when the path of said light is unimpeded by the planet in question. Comparing the two spectra allows astronomers to isolate the spectrum of light emitted by the reflection of starlight off of the atmosphere of the planet. The planet-specific spectrum can then be analyzed in order to determine the chemical composition of the atmosphere of the planet.

Due to the limitations of current technology, the transit method has mostly been used to observe large planets that tightly orbit their stars. These giant, hot worlds are unlikely locations for life, such that exploiting the full potential of transit method techniques for identifying potential life-supporting planets will require improvements in our observational technology. Thus, the inference method that we propose for identifying the most likely candidates for life-supporting exoplanets will only be applicable when better technology for detecting and isolating the emission spectra of exosolar planets is available. There is evidence that scientists are progressing towards this capability; both NASA's proposed FINESSE mission and the European Space Agency's proposed Exoplanet Spectroscopy Mission have the potential to improve the state of the art with respect to exoplanet spectroscopy (see Swain 2010a, 2010b).

We posit that most living organisms engage in *metabolism*. This posit is consonant with work in astrobiology by Seager and Bains (2015), who state that "life uses chemistry and metabolism to store energy and outputs metabolic by-product gases" (2015, p. 9).[2] Here, we follow Schrödinger (1944) in defining metabolism very generally as an object's use of free energy in its environment to maintain its structure. This definition is deliberately highly broad, so that abiotic as well as biological processes could be accurately described as instances of metabolism. However, we hold that living objects nearly always engage in metabolism, whereas abiotic metabolism, while conceptually possible, occurs far more rarely if at all. Moreover, the breadth of this definition

[2]In taking this aspect of Seager and Bains' approach as a scientific basis for a metabolism-based epistemology of life detection, we follow Knuuttila and Loettgers (2017).

of metabolism means that metabolism could involve either the consumption or the production of chemicals *not* typically associated with metabolism on Earth, such that our posit that living organisms engage in metabolism is not an Earth-centric biological assumption. The idea that metabolism is a fundamental part of what it means for an object to be alive has a long history. In addition to Schrödinger, Kauffmann (1993) and Dyson (1999) both advance the metabolism-first thesis with regard to the origins of life, arguing that the replication dynamics necessary for evolution were caused by the emergence of metabolism. This thesis has its detractors (see Anet 2004), but has also been bolstered by recent work showing the possibility of abiotic metabolism-like chemical processes that could have given rise to biotic metabolic processes prior to the emergence of RNA and other molecules necessary for replication (see Patel et al. 2015 and Ralser 2018). While we certainly will not settle this issue here, what is clear is that there are plausible theses in origins of life research such that an astrobiological focus on metabolism is in sync with foundational assumptions about the nature of life in evolutionary theory.

We posit further that metabolism tends to leave distinctive chemical traces in the atmospheres of planets whereon metabolic processes are occurring or have occurred. Thus, if life in the galaxy is *rare*, i.e. if most planets do not host living organisms, and if there are not statistically common abiotic processes that lead to atmospheric conditions similar to those caused by metabolic processes, then the atmospheres of planets that do host life will have anomalous chemical compositions. That is, the atmospheres of the relatively few life-supporting planets will bear the chemical signatures of metabolism, whereas the vast majority of planets that do not support life will not bear these signatures. For example, 20% of the atmosphere of Earth is composed of oxygen (O_2), with an additional percentage of the upper atmosphere composed of ozone (O_3). While ozone is abiotically synthesized from oxygen in the upper atmosphere, Léger et al. (1993) argue that the significant presence of both molecules in Earth's atmosphere is attributable to the presence of metabolizing organisms on Earth. Thus, if most exosolar planets are not hosts for life, then we can expect to observe, on average, much lower concentrations of oxygen in the atmospheres of most exosolar planets. Under these assumptions, the anomalous abundance of oxygen in Earth's atmosphere would count as evidence that Earth is life-supporting. However, we do not want to make the assumption that a high presence of oxygen is the only anomaly in a planet's atmosphere that is evidence for life, or even one of a small handful of such anomalies. We also do not wish to assume

that oxygen is strictly a necessary condition for life (see Lv et al. 2017), nor that the presence of life is the only possible cause of a high concentration of oxygen in an atmosphere (see Narita et al. 2015; see also Hörst et al. 2008 for models showing that oxygen on Titan's atmosphere need not be the result of biotic processes on the surface of Titan). In addition, our understanding of the huge space of planetary chemical processes is incomplete. Thus, we design a method for detecting anomalous chemical compositions in the atmospheres of exosolar planets, without making any commitments as to the nature of those anomalies.

The scientific foundation for our approach is found in work by Seager et al. (2016), who use combinatorial techniques and existing chemical knowledge to compile a list of approximately $14,000$ compounds with less than six hydrogen atoms that are unstable at standard temperature and pressure, and therefore likely to appear as gases in the atmospheres of exosolar planets. For a given planet i, let $\vec{w}_i$ be a vector with n entries, where each entry is the abundance of a specific gas from some subset of Seager et al.'s list in i's atmosphere (namely, some subset of the set of gases that can be measured in a given context). Let the set of all such observed vectors be $\mathbf{W}$. The measure of anomalousness should be a function $\mathcal{A}(\vec{w}_i, \mathbf{W})$, such that the anomalousness of the chemical composition of a given planet's atmosphere is a function solely of that chemical composition and of the broader distribution of chemical compositions in the set of observed exoplanet atmospheres. As we observe more exosolar planets, the relative anomalousness of a given planet within that set could change; a planet once thought to have a highly anomalous atmosphere could become more typical of observed exosolar planets as more observations are made, and a planet thought to be typical could be re-classified as highly anomalous as more planets are observed. As such, we do not care about the precise value of the function $\mathcal{A}(\vec{w}_i, \mathbf{W})$ for any exosolar planet i. Rather, we care primarily about the *ranking* of observed exosolar planets within an observed set, with respect to their anomalousness $\mathcal{A}(\vec{w}_i, \mathbf{W})$.

In Appendix A, we present a mathematically precise definition of an anomalousness function that can be used to rank the relative anomalousness of exosolar planets within an observed sample. This function is based on work in machine learning by Banerjee et al. (2006), where they propose a function for measuring the anomalousness of images in a high-dimensional space. In addition, we propose to use empirical, unsupervised parameter-estimation techniques outlined in Ghafoori et al. (2018). These techniques allow us to tune the parameters of the anomalousness measure solely by

collecting data, rather than by making use of *a priori* assumptions about the chemical signatures of metabolism. Thus, the precise measure of anomalousness that we propose is genuinely *chemically agnostic*; it makes no assumptions about the specific chemistry of metabolic processes that are taken to be generally characteristic of living organisms. This is necessary in order for approach to remain open-minded about the possibility of extra-terrestrial life having a very different chemical instantiation than life on Earth. In Appendix B, we use our specific measure of anomalousness to analyze simulated data, to demonstrate that it succeeds in identifying those elements of a data set that are generated by a process that is more likely to produce anomalies.

Ultimately, the formal details of our measure of anomalousness are of secondary importance within our broader argument. The techniques we propose do possess the virtue of having performed well in empirical tests on a broad range of non-astrobiological datasets (see Ghafoori et al. 2018, pp. 5064-5069). This lends inductive support to the claim that our measure will be useful in an astrobiological context. However, if other methodologies for anomaly detection in high-dimensional data spaces are shown empirically to perform better, then nothing in our broader argument is in tension with astrobiologists adopting those methodologies.

We claim that the more anomalous a planet's atmosphere is within a given set, the more likely it is to host life. This claim follows from a set of specific assumptions. First, we assume that our set of observed exosolar planets is representative. Second, we have assumed that living organisms produce gases as a by-product of metabolism, which can lead to changes in the chemical composition of a planet's atmosphere as it transitions from not hosting life to hosting life. Third, we have assumed that life in the galaxy is rare, so that most exoplanets whose atmospheric chemical compositions we are able to observe via the transit method, provided that our sample is representative, will not host life. Fourth, we assume that there are not common abiotic processes whose atmospheric effects mimic those of biological metabolism. From this it follows that if the chemical atmosphere of a planet is unusual, in the sense quantified by our measure of anomalousness, then a possible cause of this unusual chemical composition is the presence of life on that planet.

It is worth emphasizing that it does *not* follow from these assumptions that planets with chemically anomalous atmospheres *must* be hosts for life. Indeed, there are many reasons why the chemical composition of a planet's atmosphere might be far from the mean, of which the presence of metabolizing organisms is only one. It is also possible that every planet is unique enough in a

high-dimensional space of chemical combinations that the mean is not well-defined. Nevertheless, where an atmosphere's chemical composition is anomalous we have some reason to believe that life may be present, whereas when no chemically anomalous atmosphere is observed, then from the point of view of a life-detection approach to astrobiological inquiry, we have no reason to believe that life may be present. As such, it stands to reason that we ought to prioritize anomalous atmospheres for further investigations that might yield additional insights with respect to the presence or absence of life. If one rejects the assumption that life is rare in favor of the assumption that it is common, then astrobiologists ought to prioritize planets with low anomalousness when searching for life. If it turns out that chemical compositions of exosolar planets follow a unimodal distribution around some mean (or, if the distribution is multi-modal, that data points tend to be clustered around one of the means), then the assumption that exosolar planets tend to be hosts of life will imply that a large number of planets ought to be prioritized in the search for life, which implies in turn that the transit method is of far less probative value in the search for extraterrestrial life.

It might be argued that given the vast, largely unexplored space of *abiotic* phenomena that we might observe on other planets, our insistence on using anomalousness alone as a criterion in the search for extraterrestrial life is far too weak. Some abiotic phenomena may well be just as statistically anomalous as biological phenomena, such that the chemical signatures of these abiotic phenomena will misleadingly be taken as evidence for the possibility of life within our approach. Instead, it might be argued, we must find a way of refining our search so as to specifically identify biologically promising anomalies. In response, we argue that any definition of 'biologically promising' in this context compromises the agnosticism of a search procedure, since what counts as biologically promising will invariably be indexed to a specific perspective on how life might be realized physically and chemically that could close off possibilities beyond the basic assumption that life involves metabolism. This response highlights a crucial trade-off that one encounters when attempting to develop an epistemology of astrobiology; the more that an approach is refined so as to avoid falsely identifying an observation as a potential biosignature, the less one can claim that approach to be genuinely agnostic with respect to how life is physically and chemically realized.

Finally, we wish to clarify that while we have proposed applying our anomaly-detection approach to gas-abundance data obtained via the transit method, the mathematical apparatus described in this section, and presented in greater detail in Appendix A, can be applied to any data set in which

individual data points can be represented as elements of a vector space. Indeed, the techniques we propose were first developed in a machine learning context for identifying anomalies in images, not transit method observations of the atmospheres of exosolar planets. Our proposed technique could just as easily be used to detect chemical anomalies in a set of soil samples from exosolar planets, if such a set were available to us. It is also worth pointing out that, while we have primarily discussed transit-method observations of exosolar planet atmospheres, there is no in-principle reason why one could not include gas abundance vectors for the atmospheres of planets within our solar system in a data set that was analyzed via our method. Indeed, the presence of Earth data in such a set can serve as a kind of sanity check on our method, meant to ensure that data from the one planet that is known to be a host of life is treated as anomalous. We have focused here on transit-method data not because it is the *only* use case for our proposed methodology, but rather because we take it to be the most salient one in an astrobiological context.

3 The Role of Hinge Epistemology

One might question at this point whether the inference method that we have defended above is of any actual scientific value. After all, given the vastness of the galaxy, let alone the universe, we have little way of knowing whether life really is rare, or whether a given sample of exosolar planets is representative. Thus, it could be argued, the argument that we have presented for prioritizing planets with anomalous atmospheres when searching for life is fundamentally vacuous. In response, we acknowledge that these assumptions are indeed unlikely to be verified in the near to medium term, if ever. However, this does not render our model of astrobiological inquiry useless. The assumptions made above should be treated as akin to *hinge propositions*, in the sense of Wittgenstein's final notebooks, which were published as the essay *On Certainty* (1969).

Wittgenstein is interested in the conditions under which the most basic assumptions underlying empirical inquiry (e.g., that there is an external world in which one is situated) can be justified. On Pritchard's (2017) reading, the central conclusion of *On Certainty* is that "all rational evaluation is essentially local, in that it takes place relative to fundamental commitments which are themselves immune to rational evaluation, but which need to be in place for a rational evaluation to occur" (p. 565). For Wittgenstein, basic assumptions of empirical inquiry, like the existence of the external

world, are epistemically justified despite being subject to skepticism precisely because of the crucial role that they play in empirical inquiry. To use Wittgenstein's memorable phrase, "[w]e just can't investigate everything, and for that reason we are forced to rest content with assumption. If I want the door to turn, the hinges must stay put" (1969, p. 44). Empirical inquiry, it is argued, depends crucially on these so-called *hinge propositions*, which are assumptions that cannot be verified, but are essential to the operation of a particular mode of inquiry.[3] As Wright (2004) puts it, sceptics may be able to show that we have no way of knowing any such hinge proposition, but it does not follow from this that we have no reason for accepting or assuming such a proposition. Indeed, strategically it may be in our best interest to accept or assume the proposition for the purpose of inquiry, even while accepting that it cannot be known.

To illustrate this idea further, consider a classic example from Reichenbach (1938). Suppose that a person is trapped on an island, *à la* Robinson Crusoe, and is surrounded by many colorful fruits, and no other possible sources of nutrition. If they do not eat the fruits, then they will starve, but in order to be able to eat the fruits, they must first come to believe that the fruits are not poisonous. Even under the supposition that there is no way for the person, in that moment, to know whether the fruits are poisonous, it does not follow that there is no reason for the person to assume that the fruits are not poisonous, and to base their subsequent actions on this belief, so that they might have some chance of survival. While the belief that the fruits are not poisonous may be vulnerable to skepticism, it does not not mean that the belief is not *warranted* by the pragmatic context of the person stuck on the desert island. Eating the fruits is the dominant strategy for this person (they can do no worse by eating the fruits than by not eating them), and so if believing that the fruits are not poisonous is necessary for the person trapped on the island to choose this Pareto-superior option, then their adopting the belief is thereby warranted.

Analogously, astrobiology is a science that proceeds under conditions of deep uncertainty, and without certain assumptions, such as those that we have made in justifying our methodology, inquiry may simply be impossible. So while the sceptic is of course correct to point out that it may be very difficult to ever verify some of these core assumptions regarding the importance of metabolism to life and the relative frequency with which life appears in the universe, it may still

[3] Having said this, some of our assumptions could, with more data on the chemical compositions of exosolar atmospheres, undergo a "sanity check" by ensuring that Earth is classified as having an anomalous chemical composition in its atmosphere, thereby identifying it as a potential location for life.

be the case that such assumptions are needed in order for us to have any chance of conducting successful inquiry in astrobiology. For now, such assumptions may be the hinges that have to stay put so that the door of astrobiological inquiry can turn properly. As the science develops, it may be possible to refine the assumptions upon which our inquiry is based, whittling down their non-verifiable content. However, the fact that current astrobiology proceeds on the basis of presently unverifiable assumptions does not imply that we ought to wait until such a future time to try and make any progress in the project of detecting potential evidence of life on exosolar planets.

Further, the history of science contains instances in which false assumptions have nevertheless led to fruitful research programs. In what is now a classic case in philosophy of science, theories of light which assumed the existence of an optical aether were used to successfully predict a number of optical phenomena, including "reflection, refraction, interference, double refraction, diffraction and polarization" (Laudan 1981, p. 27). This improved understanding of the behavior of light led to enormous advances in nineteenth century astronomy, even as they were underwritten by the false assumption that all light passes through a substantive medium. In much the same way, the approach to astrobiological inquiry described above could yield significant empirical insights in the search for extraterrestrial life, even if some of the assumptions that underwrite it are eventually demonstrated to be false. Indeed, hinge propositions that are integral to the early success of a scientific research program can eventually be falsified by those working within the very same research program; this was the case with the geocentric model of the solar system.

It might be questioned at this stage whether this notion of a hinge proposition is highly detached from Wittgenstein's original usage of the term. Paradigmatic hinge propositions like 'there is an external world' are not testable in principle. By contrast, a proposition like 'life is rare in the set of observable exosolar planets' may be testable in principle. Ultimately, we are not concerned with the historical question of whether Wittgenstein only meant for hinge propositions to be propositions that are not testable in principle, or if he meant to include as hinge propositions those claims that could, in theory, be tested, but in practice are assumed and not tested in order to set up a context of inquiry.[4] For our part, we wish only to highlight that there is a useful class of propositions, including

[4]That said, we do take the following quote from *On Certainty* to indicate that the hinge epistemological method is meant to be relevant to the assumptions used in science: "All testing, all confirmation and disconfirmation of a hypothesis takes place already within a system. And this system is not a more or less arbitrary and doubtful point of departure for all our arguments: no, it belongs to the essence of what we call an argument. The system is not so much the point of departure, as the element in which arguments have their life" (p. 107).

both those that may not be testable in principle and those which are testable in principle, which: i) are not tested, ii) are assumed in order to set up a context of inquiry, and iii) cannot be tested *by that system of inquiry.* We call such propositions "hinge propositions." To illustrate, consider the proposition that life in the set of observable exosolar planets is rare. It is assumed, but not tested, in order to justify a particular methodology of astrobiological inquiry via anomaly detection. Even if it is not empirically justified, it is *warranted* in virtue of its role in setting up a potentially fruitful research program. Moreover, that research program itself is incapable of testing the proposition; performing anomaly detection on samples of exosolar planets will never, in itself, provide a test of whether or not life is common in the population from which one samples. To summarize, for our purposes here we take a hinge proposition to be one that is warranted primarily by its role in setting up a system for inquiry, and which, though perhaps testable in principle, are not testable via the system of inquiry that they serve to set up.

This discussion serves to clarify the conceptual difference between a hinge proposition (as we use the term) on the one hand and an *a priori* assumption on the other. Science, it might be argued, makes essential use of untested, *a priori* assumptions for essentially pragmatic reasons. To illustrate, when a bench scientist performs an experiment, they tacitly assume that the number of pipettes in their drawer is not relevant to the outcome of the experiment that they are conducting. What we call "hinge propositions," it could be argued, are essentially more speculative versions of these kinds of tacit, *a priori* assumptions, such that speculative inquiry does not really depart in any epistemically relevant way from the practices used in *any* scientific practice.

In response to this argument, we clarify that the *a priori* assumptions described above are not hinge propositions because it is possible, in principle, to test them using the same methodological framework that defines the context of inquiry in which they are assumed. For instance, the assumption that the number of pipettes in the drawer does not affect the outcome of the experiment is deployed to facilitate causal inference via experimentation. That same inference method could be used to at least attempt to test the assumption that the number of pipettes in the drawer does not affect the outcome of the experiment; doing their best to hold all other factors equal, the experimenter could repeat the experiment with different numbers of pipettes in the drawer, and observe whether there is an effect on the results. By contrast, an anomaly detection method that assumes that life is rare within the set of observable exoplanets cannot be used to test whether life is rare in

the universe. Returning to Pritchard's understanding of a hinge proposition, the assumption that life is rare in the set of observable planets needs to be in place for rational evaluation via anomaly detection to occur, but cannot itself be rationally evaluated via anomaly detection.

Moreover, if one were to base one's astrobiological inference process on alternative assumptions (e.g., that life is common on observable exoplanets, and can be detected by observing specific phenomena), then these assumptions might still be empirically testable via the same methodology that they are used to justify. For instance, one could test whether the proposed biosignature is common on observed exoplanets; if it is not, then either the assumption that life is common or the assumption that the putative biosignature is in fact a biosignature is likely to be misguided. This is true even though these assumptions are adopted *a priori* and before the collection of data. Thus, these kinds of assumptions would *not* be hinge propositions, further illustrating the difference between *a priori* assumptions on the one hand and hinge propositions on the other hand.

There is a conceptual link between a hinge epistemological approach to astrobiological inquiry and work on the function of *speculation* in science by, among others, Achinstein (2018), Currie and Sterelny (2017), and Currie (2018, 2019, 2021). According to Achinstein, speculation in science is the practice of "introducing assumptions without knowing that there is evidence for those assumptions" (2018, p. 1). On this account, it is clear that the hinge propositions used to motivate our defense of an anomaly-based approach to the search for evidence of life on exosolar planets are speculative; we do not know that there is evidence for these assumptions, and yet we make them anyway for the sake of establishing a context of inquiry.

Speculation in science can be vicious or virtuous. Vicious speculation in science, Currie and Sterelny (2017) argue, occurs when said speculation "is pointless: when it cannot or does not productively direct further inquiry; when it is not used to construct alternative scenarios to guide a search for evidence which would favour one at the expense of the other" (p. 16). Thus, speculation in science is vicious when it fails to serve the pragmatic end of setting up a potentially fruitful local context in which empirical inquiry can occur. By the same token, Currie (2021) argues that speculation is virtuous when it "provision[s] epistemic goods through opening new research, or scaffolding the development of theories or experiments, or generating possibility proofs, or providing epistemic links to further knowledge" (p. 16). That is, speculative hypotheses are justified to the extent that they establish a fruitful context of inquiry, much in the same way that hinge

propositions are justified. As Currie (2018, p. 287-9, 2019) argues, in scientific contexts in which severe uncertainty exists, such as paleontology or archaeology, this kind of productive speculation is necessary for the productive function of inquiry. Given the similar epistemic predicament of exoplanet astrobiologists, there is a strong case for counting astrobiology as another science in which productive speculation plays an essential role.

We conclude this section by noting that one assumption underlying our approach that *should* be subject to immediate scrutiny, with the possibility of revision, has to do with the way in which Seager et al. (2016) build their list of elementary gases. Specifically, their list of biosignature gases includes only those molecules that are unstable at standard temperature and pressure (e.g. a temperature of zero degrees Celsius and a pressure of one atmosphere). This may be appropriate for inferring that a molecule will appear in a gaseous state in Earth's atmosphere, but in light of what we have said above, it is clear that astrobiologists cannot assume that life only occurs on planets with Earth-like standards for atmospheric temperature and pressure. For instance, on Saturn's moon Titan, the surface temperature and pressure is near the triple-point of methane, allowing for liquid methane at the surface level and the possibility of methane-based life forms; this state of affairs stands in contrast with Earth, where surface temperature and pressure is near the triple-point of water, and where water is central to most biotic processes (see Hörst 2017). Further, unlike the other assumptions that are required for our inference method to be put to use, here there is a clear research program for altering this assumption so as to better facilitate astrobiological inquiry. As Seager et al. put it, "extensions [of their list of gases] beyond STP are a huge and demanding piece of research that we hope will be initiated in the future" (p. 473). If theoretical chemists are able to find a principled basis for expanding the list of biosignature gases beyond the current one, then our approach can be straightforwardly adapted to accomodate larger vectors of gas compositions, as it is explicitly designed for use in high-dimensional vector spaces.

4 Contrast with a Bayesian Approach

As mentioned in the introduction, Walker et al. (2018) propose a Bayesian approach to the epistemology of searching for life on other planets using atmospheric spectroscopy techniques. Adapting their terminology only slightly, their approach can be described as follows. Let D_i be the event

that a particular set of data about a given exosolar planet W_i is observed (while Walker et al. want to accomodate a very broad set of ways in which we might gather relevant data about a planet, for the purpose of our discussion let us confine ourselves, without loss of generality, to data from transit method observations of the atmospheric spectroscopy of an exosolar planet). Let L_i be the event that the exosolar planet W_i is a host for life. Astrobiologists are interested in the conditional probability $P(L_i|D_i)$, i.e. the probability that planet W_i is a host of life, given the data that is observed with regard to that planet. This conditional probability can calculated using Bayes' Theorem:

$$P(L_i|D_i) = \frac{P(D_i|L_i)P(L_i)}{P(D_i|L_i)P(L_i) + P_i(D_i|\neg L_i)(1 - P(L_i))} \tag{1}$$

Thus, we can indirectly estimate $P(L_i|D_i)$ by estimating three other, putatively more tractable probabilities: i) $P(D_i|L_i)$, i.e. the probability of observing the data in question, given that planet W_i hosts life, ii) $P(D_i|\neg L_i)$, i.e. the probability of observing the data in question, given that planet W_i does not host life, and iii) $P(L_i)$, i.e. the prior probability that planet W_i hosts life.

Walker et al. go into considerable detail as to how astrobiologists might estimate these three probabilities. They argue that the probability $P(D_i|L_i)$ will be able to be accurately estimated once astrobiologists have a better understanding of the full range of ways in which a living organism can leave biosignatures in its environment, such as the changes in the chemical composition of a planet's atmosphere discussed above. Similarly, $P(D_i|\neg L_i)$ will be able to be estimated once there is a better sense of the kinds of conditions, atmospheric and otherwise, that are typically present when life is absent. Finally, estimating the probability $P(L_i)$ requires an understanding of how common or rare life is in the galaxy. As Spiegel and Turner (2012) note, the existing evidence is consistent with life being extremely rare or extremely common. One way of estimating this prior, they argue, would be to determine how common abiotic conditions like those found on early Earth are within the galaxy as a whole (p. 399). However, a decided strength of the Bayesian approach proposed by Walker et al., compared to ours, is that it does not, in principle, require any specific assumptions about the relative frequency with which a given planet hosts life.

Nevertheless, we believe that our approach, as detailed the previous section, has certain advantages over the Bayesian approach to astrobiological discovery. To be clear, we do not mean

to attack Bayesianism *itself*, which we take to be a normatively well-grounded epistemological approach in a wide range of cases. Rather, we argue that the Bayesian approach requires highly detailed assumptions about the probability of observing possible spectroscopic data sets, whereas our argument depends on a smaller set of more coarse-grained assumptions. Our argument begins by noting that any rigorous application of probability theory must begin by defining the probability space to be used for probabilistic modelling. A probability space is a triple $(\Omega, \Sigma_\Omega, P)$, where Ω is a set of possible worlds, or sample space, Σ_Ω is a an *algebra* on Ω, i.e. a collection of subsets of Ω that is closed under complement, union and intersection, and $P : \Sigma_\Omega \to [0, 1]$ is a function that satisfies Kolmogorov's (1933) axioms of non-negativity, normalization, and countable additivity. Suppose that A and B are both elements of the algebra Σ_Ω. When Σ_Ω is generated via a partition of Ω with countably many elements, the conditional probability $P(A|B)$ is calculated via the following formula:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \text{ where } P(B) \neq 0. \tag{2}$$

This "ratio formula", along with its implication that $P(A \cap B) = P(B|A)P(A)$, allows us to derive Bayes' theorem. Adding the law of total probability as a way of expanding the denominator allows us to derive the version of Bayes' theorem shown in equation (1).

What probability space could be used to specify the probabilities used in equation (1)? A naive attempt to answer this question might proceed as follows. For a given planet W_i, let us define a probability space $\mathcal{P}_i = (\Omega_i, \Sigma_{\Omega_i}, P)$, where the sample space is defined as follows:

$$\Omega_i = \{(\vec{w}_i, l_i), (\neg\vec{w}_i, l_i), (\vec{w}_i, \neg l_i), (\neg\vec{w}_i, \neg l_i)\} \tag{3}$$

The pair $(\vec{w}_i, l_i)$ denotes a world in which the data vector $\vec{w}_i$ describes the abundance of gases on the atmosphere on planet W_i, and there is life on W_i. The set of possible worlds Ω_i is closed under negation of either element of this pair. The algebra Σ_{Ω_i} is the power set of Ω_i. Let $D_i = \{(\vec{w}_i, l_i), (\vec{w}_i, \neg l_i)\}$, $L_i = \{(\vec{w}_i, l_i), (\neg\vec{w}_i, l_i)\}$, and $\neg L_i = \{(\vec{w}_i, \neg l_i), (\neg\vec{w}_i, \neg l_i)\}$. Thus, D_i, L_i, and $\neg L_i$ are all elements of the algebra Σ_{Ω_i}, such that they and any possible intersections of them have well-defined probabilities in the probability space $\mathcal{P}_i$. Thus, this probability space is sufficient for

us to define the terms in equation (1) in the case of the planet W_i. In effect, $\mathcal{P}_i$ amounts to a very minimal theory about biosignatures on planet W_i. It tells us the probability that planet W_i hosts or does not host life, the probability of observing or not observing a particular data set with respect to planet W_i, the probability of observing the specific data set when the planet does or does not host life, and the probability of the planet hosting or not hosting life when the data set is observed or not observed.

While this minimal theory might render equation (1) tractable, it is not what astrobiologists want from a theory of biosignatures. The theory sketched above is a planet-specific theory, meaning that it says nothing about the probability of observing the data set in question on *any* planet, both on the condition that the planet hosts life and on the condition that it does not. It is also a dataset-specific theory, meaning that it says nothing about the probability of observing any data set other than the one that happens to be observed on planet W_i. What astrobiologists want is a more general theory, one that specifies the probability of observing any data set on any given planet, conditional on whether or not that planet is a host for life.

Consider what such a theory might look like, represented as a probability space $\mathcal{P} = (\Omega, \Sigma_\Omega, P)$. Let l be the proposition that a given planet hosts life and let the sample space Ω be defined as follows:

$$\Omega = [0,1]^n \times \{l, \neg l\} \tag{4}$$

That is, the sample space is the set of all possible pairs $(\vec{w}, l)$ and $(\vec{w}, \neg l)$, where $\vec{w}$ is an n-entry vector whose entries are each in the unit interval and represent the relative abundance of each gas in the atmosphere of a given planet. Let $\mathcal{B}_{[0,1]^n}$ be the set of all Borel-measurable subsets of the space $[0,1]^n$. The event algebra Σ_Ω is defined as follows:

$$\Sigma_\Omega = \{\emptyset\} \cup \{B \times \{l\} | B \in \mathcal{B}_{[0,1]^n}\} \cup \{B \times \{\neg l\} | B \in \mathcal{B}_{[0,1]^n}\} \cup \{B \times \{l, \neg l\} | B \in \mathcal{B}_{[0,1]^n}\} \tag{5}$$

This entails that the probability distribution P assigns a probability to every element of $\mathcal{B}_{[0,1]^n}$ when there is and is not life on a given planet, and when it is not specified whether the planet hosts life. Equipped with this algebra, we can use Bayes' theorem to calculate the probability

that a given planet hosts life, conditional on our observations of the chemical composition of that planet's atmosphere. Let $D = B \times \{l, \neg l\}$ be the event that a particular Borel-measurable subset $B \in \mathcal{B}_{[0,1]^n}$ of the set of possible data vectors is observed on a given planet. Let $L = [0,1]^n \times \{l\}$ be the event that a given planet hosts life, and let $\neg L = [0,1]^n \times \{\neg l\}$ be the event that a given planet does not host life. Note that L and $\neg L$ are in Σ_Ω since $[0,1]^n \in \mathcal{B}_{[0,1]}$. If the probability space $\mathcal{P} = (\Omega, \Sigma_\Omega, P)$ is defined in the manner specified above, then we can use Bayes' theorem to calculate the conditional probability $P(L|D)$, i.e. to calculate the probability that a given planet hosts life, conditional on the observed data. Note that the subscript i has been dropped since we are now discussing *any* planet, rather than a particular planet W_i.

It is remarkable how much information such a probabilistic model contains. Consider the set of all possible chemical compositions of an atmosphere of any planet. The theory not only requires us to specify the probability of each element of this set under the condition that a planet does or does not host life, it also requires us to specify the probability of all measurable subsets of this set under the condition that the planet does or does not host life. Compared to our proposed methodology for anomaly detection in astrobiological inquiry, applying the Bayesian approach will require a significantly greater level of sophistication with respect to our understanding of planetary chemistry. To see why this is the case, note that our method requires that scientists can make reliable estimates of the relative abundance of a large number of gases in the atmospheres of exosolar planets. The Bayesian approach requires that we not only have this ability to measure chemical abundances in exosolar atmospheres, but also that we have a rigorous understanding of how those abundances correlate with the presence or absence of life on a given planet, so that every element of the set of possible atmospheric chemical compositions, and every measurable subset thereof, can be assigned a probability. As Kishimoto et al. (2018) point out, the combinatorial space of possible chemical compounds has a cardinality of roughly 10^{60}, such that large regions of this space remain unexplored. So, while the Bayesian approach allows us to integrate into our epistemic framework a wide array of scientific knowledge, by giving us control of the prior probabilities that we assign to a wide swathe of chemical and biological possibilities, it does so largely at the expense of applicability and tractability.

Even if we restrict ourselves to the space of possibly observable compounds at a given time, the size of the algebra over which probabilities must be defined in order to perform Bayesian inference

in astrobiology may soon become very large. This means that assembling the algebra described above may be well beyond the capabilities of scientists even as the spectroscopic investigation of exoplanets begins to generate workable data. Similarly, one might attempt to define a manageable number of coarse-grained intervals $\mathcal{I}$ into which some measure of the chemical disequilibrium of a given planet's atmosphere might fall into, and then use the power set of the cross product $\mathcal{I} \times \{l, \neg l\}$ to define the algebra over which probabilities are defined in a Bayesian approach to astrobiology. While this strikes us as a promising approach to rendering the Bayesian approach to astrobiology more tractable (although it still requires a significant quantitative articulation of the probabilistic relationship between life and atmospheric disequilibrium), we note that it still faces some foundational problems, including the specification of a specific prior probability that a given planet does or does not host life. Thus, our proposed approach to astrobiological inference is likely to require strictly less scientific progress than the Bayesian approach, and is therefore likely to be more fruitful in the near-to-medium term.

These difficulties with a probabilistic, Bayesian approach to astrobiological inference speak to the advantages of our approach to anomaly detection, which does not make any explicit assumptions regarding the underlying probability distribution from which observations are drawn. In this way, our approach incorporates a statistical agnosticism about the distribution of chemical compositions of atmospheres, making only the assumption that life is a rare phenomenon in the set of observed planets, and directly inferring statistical parameters from available data. This statistical agnosticism is in keeping with our broader commitment to agnostic approaches with respect to the definition of life and with respect to the particular kinds of biosignatures that we believe astrobiologists ought to be looking for.

Nevertheless, it is important to note that the distinction between our approach and the Bayesian approach is ultimately a distinction of degree rather than kind. The proponent of the Bayesian approach could ultimately argue that their particular choice of probabilistic model is itself a hinge proposition on which their broader scientific project rests, such that Bayesian reasoning in astrobiology is ultimately an instantiation of the kind of hinge epistemology that we endorse throughout this paper. As such, we clarify that we do not believe that the Bayesian approach is fundamentally misguided with respect to the broad structure of its mode of inquiry. Rather, we point out that the Bayesian approach requires a rich set of precise assumptions about the probability of each

measurable subset of a large space, whereas our approach requires a small set of coarse-grained assumptions. Thus, we take our approach to compare favorably with the Bayesian framework on the basis that our approach is more tractable. However, both approaches are ultimately well-motivated from their respective theoretical starting points.

Finally, we note that, in practice, the data to which we envision applying our proposed measure of anomalousness will likely be incomplete, with missing data points for some gas abundances for some observed planets, but not for others. To solve this missing data problem, we may need to use Bayesian techniques in which a prior distribution over possible values for missing data points is updated based on the evidence that we do have, in order to arrive at an estimate of the values of the missing data points. Thus, the preceeding arguments should not be read as a wholesale rejection of Bayesian inference, but rather as an argument against a particular application of Bayesian inference. Nevertheless, we note that the missing data problem described here will typically require assigning probabilities to a much more manageable space (namely, a set of plausible values for a particular set of data points) than the more general Bayesian approach to astrobiology advocated by Walker et al. (2018), which, as described above, requires an assignment of prior probabilities to a rich set of both possible observations *and* the possibility that a planet is a host of life.

5 Relationship to Cleland's Approach

In the philosophy of science literature to date, the most extensive discussion of life detection in astrobiology is due to Carol Cleland, especially in Cleland (2019a) and Chapter 8 of Cleland (2019b). Specifically, Cleland argues that anomalous observations from other planets should be treated as potential evidence for the existence of life on those planets. In this respect, her fundamental epistemic commitments are consistent with our own. Moreover, she notes, quite rightly, that in previous attempts at astrobiological research, such as the analysis of Martian soil samples recovered during NASA's 1976 Viking mission, scientists' commitment to a particular biological paradigm hindered their ability to observe potentially anomalous phenomena, whether those anomalies are biological or abiological in nature. To briefly summarize this case study, when Martian soil samples were injected with nutrients, they released $^{14}CO_2$ gas, in a manner consistent with microbial life on Earth (Cleland 2019b, p. 173). However, on the second injection of nutrients, the amount of

$^{14}CO_2$ gas decreased, in a manner inconsistent with what one would expect from the perspective of terrestrial biology. However, to date, no decisive abiotic explanation has emerged for the release of $^{14}CO_2$ gas on the first nutrient injection, followed by a diminished release of $^{14}CO_2$ gas on the second injection. NASA accepted the conclusion that the observed $^{14}CO_2$ data had an inorganic explanation in large part because it was more consistent with the failure of gas chromatograph mass spectrometer (GCMS) observations to find any organic compounds in Martian soil, despite the fact that GCMS observations were never intended to be used as part of Viking's life detection experiments (2019b, p. 174). The upshot, for Cleland, is that the results of soil sample analysis after NASA's Viking mission "represented a potentially (but not definitively) biological anomaly," but that it can be difficult, when one is committed to a particular biological paradigm, for scientists to recognize these kind of anomalies as such (2019b, p. 176).

In this context, our proposed methodology can be understood as complementing Cleland's arguments. As our approach is completely agnostic with respect to the specific molecules that might be the output of biological processes like metabolism, we do not run the risk of prematurely concluding that evidence like the Viking soil sample experiments are not indicators of biological activity. Instead, our criterion for identifying particularly anomalous exosolar atmospheres depends on the *statistical* anomalousness of an observation, rather than its coherence, or lack of coherence, with a given body of theory, or with other observations. This feature of our approach reflects an underlying conceptual overlap between our approach to astrobiological inquiry and Cleland's; both approaches are intended to overcome the scientific tendency to commit to a particular observational paradigm in a way that limits our ability to correctly classify certain phenomena as genuinely anomalous.

Cleland's positive proposal is that, in keeping with the anti-definitionist approach to life detection, astrobiologists should search for evidence of extraterrestrial objects that satisfy various "tentative criteria" for being a living organism. These criteria are not meant to provide conclusive evidence for life, nor are they meant to constitute a set of necessary and sufficient conditions which jointly define life. Rather, they are examples of what *could* be included in an ever-changing list of pieces of evidence that would be intriguing and potentially probitive, from a life detection perspective. Four specific examples of these kinds of tentative criteria are as follows:

1. Sedimentary structures of the sort produced by microbes on Earth (2019b, p. 188),
2. extremely large and complex polymers of any chemical composition (2019b, p. 188),
3. unusual concentrations of rare Earth minerals (2019b, p. 190), and
4. evidence of characteristics that are rarely, but still sometimes, associated with life on Earth. For example, magnetite crystals, which are rarely produced by organisms on Earth but which are interpreted by paleomicrobiologists as fossils, and which have been found on meteorites, could provide tentative evidence of possible life (2019b, p. 190).

In Cleland (2019a), we are given a more general overview of the kinds of tentative criteria for life detection that she has in mind. Astrobiologists, Cleland writes, should generally be searching for two kinds of phenomena when searching for extraterrestrial life. First, they should be looking for "phenomena that 'shouldn't be there' given our current, Earth-centric understanding of both biological and abiological phenomena" (2019a, p. 726). An example of this kind of phenomenon would include the magnetite crystals described above. Second, astrobiologists should search for evidence of phenomena which, "while not viewed as 'essential' to life, are nonetheless universally found in association with Earth life" (2019a, p. 725). Examples of this kind of phenomenon would include the large and complex polymers and unusual concentrations of rare Earth minerals included in Cleland's list of specific examples of tentative criteria.

Comparing Cleland's approach to our own, we begin by stressing that, in principle, *any* of her tentative criteria listed above *could* qualify as good evidence for the presence of life on our approach, as long as they can be represented as elements of a vector space, which we hold here to be possible in principle. Indeed, Cleland's commitment to both an anomaly-based approach to life detection and her inclusion of these examples in her list of possible tentative criteria suggest that each of these phenomena are likely to be statistically anomalous within a representative sample of observations of non-Earth planets, such that both our approach and Cleland's would yield the result that the presence of any of these tentative criteria would be positive indication of the possible presence of life.

At the same time, there are points of contrast between our approach and Cleland's, specifically with respect to the kind of inferential process that both approaches recommend. Cleland's approach

consists in carefully considering the sorts of observations that *could* indicate the presence of life. Coming up with a list of such tentative criteria requires careful and creative consideration of life on Earth (often with special attention to highly unusual organisms), as well as theoretical consideration of how life *could* be organized chemically, even if we have never observed life in that particular chemical form on Earth. Once such a list is compiled, one can begin examining extraterrestrial observations in an attempt to detect the presence of tentative criteria. By contrast, our approach begins by observing data from extraterrestrial planets, and then uses that data to define what counts as a relevant anomaly for the sake of life detection. In other words, our approach recasts the process of determining tentative criteria as a process of statistical inference. This allows for greater agnosticism with respect to how life could be realized (because there is no input to our model in which an astrobiologist could specify that they are looking for something specific like sedimentary structures or large polymers), providing one way to deal with a vast possibility space in the absence of a complete theory for life within that space (see also the discussion in Kempes and Krakauer 2021). However, this agnosticism comes with a trade-off; unlike Cleland's proposal, our approach does not allow astrobiologists to bring important work on unusual forms of life on Earth, or work in theoretical chemistry on possible chemical realizations of life not found on Earth, to bear on the search for extraterrestrial life. Also, Cleland's proposed use of theoretical considerations allows us to specify particular measurements that constitute severe tests of the theory in question and provide the experimenter with a significant amount of information. For instance, the Michelson-Morley experiments used a theoretically predicted effect size for the aether wind to provide evidence against aether theories in physics. Theoretically agnostic approaches such as our anomaly detection method do not allow for the possibility of such severe tests.

6 Conclusion

We conclude by emphasizing that the various perspectives that we have discussed here each come with potential advantages and potential pitfalls. For example, known scientific principles can be more readily incorporated into the Bayesian approach by setting the probabilistic inputs of Bayes' rule in accordance with said principles. This is an advantage of the Bayesian approach, but grappling with the size of the parameter spaces may lead to intractably large algebras over which

Figure 1: Diagram illustrating the trade-offs associated with different approaches to inquiry in astrobiology. An approach is represented as a point in the triangle, with points near the far left corner being closest to an approach based entirely on Bayesian inference, approaches in the far right corner being closest to an approach based entirely on unsupervised anomaly-detection algorithms, and points near the top-middle corner being closest to an approach based entirely on a specific theory of life. The strengths and weaknesses of each of these varieties of approach are listed in the diagram, and points inherit these strengths and weaknesses to the extent that they are situated in proximity to the corner in question. Our approach is closest to the algorithmic extreme, but does incorporate some biological theory due to its reliance on metabolism, and also incorporates the transparency of Bayeisan approaches. By contrast, we view Walker et al.'s approach as an amalgamation of the Bayesian formalism, but with a prior that reflects a commitment to a particular theory.

probabilities are defined. Our data-driven method for anomaly detection can be tractably applied to any data and allows us to identify which planets are promising candidates for future astrobiological study under certain assumptions, but it does not provide definitive detection of life and does not allow us to represent the specific assumptions of any particular theory of life or biosignatures. The gold standard for life detection would be a complete theory capable of defining all possibilities for life in a particular planetary environment, and mapping these possibilities uniquely to specific observable biosignatures. There is some scientific precedent for such a theory to first be proposed and then empirically confirmed later on; in particle physics the Higgs boson was theoretically proposed forty years before its experimental discovery, and physical theory that assumed the existence of the particle guided the development of instrumentation for detecting the particle. However, empirically confirming such a theory for life poses a particularly hard challenge and hybrid methods are likely to be useful in the interim. For example, generating estimates of Bayesian inputs based on current astrobiological knowledge is a worthwhile endeavor within the theory of astrobiology, though it is subject to the problems highlighted above. On the other hand, finding ways to relax aspects of the Bayesian framework and combine it with algorithmic methods like our SVM approach is similarly worthwhile, as is taking aspects of scientific theory and incorporating them into Bayesian or algorithmic perspectives. Figure 1 provides a conceptual diagram of these trade-offs along with the possibilities for combinations.

Astrobiological inquiry is conducted at the bleeding edge of our scientific understanding of both biology and geochemistry, and has the potential to overturn many deep assumptions in both of these fields. It is also conducted under conditions of severe uncertainty, using novel observational techniques. As such, it provides fruitful material for epistemologists interested in both the possibility and the limitations of reliable inquiry under these kinds of conditions. In this paper, we have proposed a methodology for astrobiological inquiry, at least insofar as that inquiry is conducted via spectroscopic inference of the chemical composition of exosolar planets. This methodology is underwritten by an epistemological approach that defends the claim that beliefs can be warranted even if they are subject to irrefutable skepticism. Thus, while our approach is in some respects specific to its astrobiological applications, it also depends crucially on classic arguments in epistemology. At the same time, we are explicit about the fact that our proposed methodology is itself defeasible; it is subject to revision if future techniques for anomaly detection with different assump-

tions or formalisms prove themselves more fruitful. Finally, and as discussed above, we recognize that the choice of approach with respect to astrobiological life detection requires astrobiologists to make tradeoffs between various desiderata for such an approach. Thus, our arguments here can be viewed as an opinionated invitation to explore more fully this epistemic and scientific landscape.

Acknowledgements

We are very grateful to Artemy Kolchinsky for detailed feedback on an earlier draft of this paper, to David Wolpert and Natalie Grefenstette for helpful discussions, and to audiences at Philosophy of Biology at the Mountains hosted by the University of Utah, an online audience at the meeting of the European Philosophy of Science Association, and the Uncovering the Laws of Life Workshop in Grindavik, Iceland. Christopher Kempes' work on this project was supported by CAF Canada and the National Aeronautics and Space Administration (Grant No. 80NSSC18K1140).

Appendix A

Our measure of anomalousness is defined precisely as follows. Let $\vec{w}_i$ be a vector with n entries, where each entry w_{ij} is the abundance of gas j from some subset of Seager et al.'s list in planet i's atmosphere (namely, some subset of the set of gases that can be measured in a given context). Let the set of observed composition vectors be $\mathbf{W}$. Our proposed anomalousness measure draws on support vector machine (SVM) based approaches to anomaly detection in high-dimensional space, as proposed by Banerjee et al. (2006). Specifically, we propose to measure the anomalousness $\mathcal{A}(\vec{w}_i, \mathbf{W})$ of the atmospheric composition of a given planet W_i, in the context of a set of planets $\mathbf{W}$ with cardinality m, using the following equation:

$$\mathcal{A}(\vec{w}_i, \mathbf{W}) = 1 - 2 \sum_{j|j\neq i}^{m} \alpha_j \exp\left(\frac{-||\vec{w}_i - \vec{w}_j||^2}{\sigma^2}\right) + \sum_{j|j\neq i}^{m} \sum_{k|k\neq i}^{m} \alpha_j \alpha_k \exp\left(\frac{-||\vec{w}_j - \vec{w}_k||^2}{\sigma^2}\right) \tag{6}$$

See Banerjee et al. (2006) for a detailed derivation of this particular measure of anomalousness. Each α_j and α_k is a positive, real-number entry in a vector $\vec{\alpha}$ in which each entry is a specific weighting parameter for each planet. The norm $||\vec{w}_i - \vec{w}_j||$ denotes the Euclidean distance between the vectors $\vec{w}_i$ and $\vec{w}_j$. The scale parameter σ sets the shape of the distribution of the anomalousness

of the planets in the dataset. Thus, if $\vec{w}_i$ is a large distance from all of the other data vectors in the set, then the second term of the equation will be smaller, and so $\mathcal{A}(\vec{w}_i, \mathbf{W})$ will be larger, all else being equal. This justifies the claim that higher the value of $\mathcal{A}(\vec{w}_i, \mathbf{W})$, the greater the anomalousness of the atmospheric composition of the atmosphere of the planet W_i.

The scale parameter σ and the weight vector $\vec{\alpha}$ can be estimated from the available data. To estimate σ, we use a technique from Ghafoori et al. (2018). Let δ_{min} be the minimal Euclidean distance of any vector $\vec{w}_i$ from its nearest neighbor in the data set $\mathbf{W}$. Let δ_{avg} be the average Euclidean distance of all vectors in $\mathbf{W}$ from their nearest neighbor, excluding the vector that is δ_{min} from its nearest neighbor. The scale parameter σ is estimated as follows:

$$\sigma = \frac{-\ln(\delta_{min}/\delta_{avg})}{\delta_{avg}^2 - \delta_{min}^2} \tag{7}$$

This method for estimating σ has proven successful at detecting anomalies in a series of empirical tests, using standard machine-learning data sets (Ghafoori et al., 2018, pp. 5064-5069). Following Schölkopf et al. (2001), we can set the weight vector $\vec{\alpha}$ by solving the following optimization problem:

$$\min_{\vec{\alpha}} \sum_{j=1}^{m} \sum_{k=1}^{m} \alpha_j \alpha_k \exp\left(\frac{-||\vec{w}_j - \vec{w}_k||^2}{\sigma^2}\right), \text{ s.t. } \sum_{i=1}^{m} \alpha_i = 1 \tag{8}$$

This ensures that $\mathcal{A}(\vec{w}_i, \mathbf{W})$ assigns as high anomalousness as possible to each planet W_i.

Appendix B

To demonstrate the potential applicability of our proposed anomaly-detection method, we tested our method on simulated data. The simulation proceeded in the following steps:

1. Generate 90 gas abundance vectors by repeatedly sampling from a Dirichlet distribution $\text{Dir}(1000, \vec{u})$, where $\vec{u} = [u_1, \ldots, u_{1000}]$ is a concentration parameter such that $u_i = 10$ for $i \leq 100$ and $u_i = 1$ for $i > 100$. This ensures that abundance vectors sampled from this distribution are expected to have abundances concentrated near the first 100 entries. This is meant to simulate observations from a typical, lifeless planet, on which certain gases are much more likely to be common than others. Call this sample of abundance vectors $\mathbf{W_A}$.

2. Generate 10 gas abundance vectors by repeatedly sampling from a Dirichlet distribution $\mathrm{Dir}(1000, \vec{v})$, where $\vec{v} = [v_1, \ldots, v_{1000}]$ is a concentration parameter such that $v_i = x$ for $i \leq 100$ and $v_i = 1$ for $i > 100$. For higher values of x, abundance vectors sampled from this distribution are expected to have abundances concentrated near the first 100 entries. This is meant to simulate observations from a typical, lifeless planet, on which certain gases are much more likely to be common than others. For lower values of x, all possible abundance vectors become increasingly likely. This is meant to simulate observations from planets that are possible hosts of life, where many more combinations of gas abundances are assumed to be possible. Call this sample of abundance vectors $\mathbf{W_B}$.

3. Form the full data set $\mathbf{W} = \mathbf{W_A} \cup \mathbf{W_B}$.

4. Compute the parameters σ and $\vec{\alpha}$ of the anomalousness measure for the data set $\mathbf{W}$.

5. Calculate the anomalousness measure $\mathcal{A}(\vec{w}, \mathbf{W})$ for each element of $\mathbf{W}$.

6. Identify the ten most anomalous abundance vectors in $\mathbf{W}$.

7. Record as y the proportion of the ten most anomalous vectors that are elements of $\mathbf{W_B}$.

8. Repeat steps 1-7 for all possible values of x in the interval $(1, 10)$, in increments of .2.

9. Repeat steps 1-8 fifty times, for a total of 460 data points.

Recall that lower values of the parameter x, which determines the concentration parameter for the distribution from which elements of $\mathbf{W_B}$ are sampled, are meant to simulate cases in which the elements of $\mathbf{W_B}$ represent planets that are possible hosts of life. Thus, we expect that as x increases, the proportion of abundance vectors with top-ten anomalousness in $\mathbf{W}$ that are also elements of $\mathbf{W_B}$ should decrease.

Indeed, this is what we observe in the results of our simulation. Figure 2 shows the proportion of top-ten anomalies from the set $\mathbf{W}$ that are also elements of the set $\mathbf{W_B}$, for each value of x in each simulation. For values of x lower than or equal to 6, the proportion of top-ten anomalous vectors that are elements of $\mathbf{W_B}$ is tightly clustered around 1. For values of x greater than 6, this same proportion is declining linearly in x. We take this to be positive evidence for the in-principle utility of our measure for detecting the kinds of anomalies that we are interested in. It should be noted

Figure 2: Plot showing the relationship between the value of x, which determines the concentration parameter for the distribution from which elements of $\mathbf{W_B}$ are sampled, and the proportion of abundance vectors with top-ten anomalousness in $\mathbf{W}$ that are also elements of $\mathbf{W_B}$.

that while there is a fairly sharp transition in correctly identifying W_B, knowing the value of x (or the analogous parameter(s) in a given setting) at which this transition occurs in an applied setting will require careful thought. As pointed out above, correct identification depends on the degree of anomalousness and thus, such methods are expected to fail when life is not very anomalous. Code for this simulation is available at `https://github.com/davidbkinney/astrobiologyappendixB`.

References

P. Achinstein. *Speculation: within and about science*. Oxford University Press, 2018.

F. A. Anet. The place of metabolism in the origin of life. *Current opinion in chemical biology*, 8 (6):654–659, 2004.

A. Banerjee, P. Burlina, and C. Diehl. A support vector method for anomaly detection in hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 44(8):2282–2291, 2006.

L. Bich and S. Green. Is defining life pointless? operational definitions at the frontiers of biology. *Synthese*, 195(9):3919–3946, 2018.

A. S. Burrows. Spectra as windows into exoplanet atmospheres. *Proceedings of the National Academy of Sciences*, 111(35):12601–12609, 2014.

J. Chela-Flores. *The science of astrobiology: a personal view on learning to read the book of life*, volume 20. Springer Science & Business Media, 2011.

C. E. Cleland. Moving beyond definitions in the search for extraterrestrial life. *Astrobiology*, 19(6): 722–729, 2019a.

C. E. Cleland. *The Quest for a Universal Theory of Life: Searching for Life as we don't know it*, volume 11. Cambridge University Press, 2019b.

C. E. Cleland and C. Chyba. Defining 'life'. *Origins of Life and Evolution of the Biosphere*, 32(4): 387–393, 2002.

C. E. Cleland and C. Chyba. Does 'life' have a definition? In *The nature of life: classical and contemporary perspectives from philosophy and science*. New York: Cambridge University Press, 2007.

A. Currie. *Rock, bone, and ruin: An optimist's guide to the historical sciences*. MIT Press, 2018.

A. Currie. Epistemic optimism, speculation, and the historical sciences. *Philosophy, Theory, and Practice in Biology*, 11, 2019.

A. Currie. Science & speculation. *Erkenntnis*, 2021. doi: https://doi.org/10.1007/s10670-020-00370-w.

A. Currie and K. Sterelny. In defence of story-telling. *Studies in History and Philosophy of Science Part A*, 62:14–21, 2017.

S. J. Dick. Critical issues in the history, philosophy, and sociology of astrobiology. *Astrobiology*, 12 (10):906–927, 2012.

F. Dyson. *Origins of life*. Cambridge University Press, 1999.

I. Fry. The emergence of life on earth: a historical and scientific overview. 2000.

Z. Ghafoori, S. M. Erfani, S. Rajasegarar, J. C. Bezdek, S. Karunasekera, and C. Leckie. Efficient unsupervised parameter estimation for one-class support vector machines. *IEEE transactions on neural networks and learning systems*, 29(10):5057–5070, 2018.

S. M. Hörst. Titan's atmosphere and climate. *Journal of Geophysical Research: Planets*, 122(3): 432–482, 2017.

S. M. Hörst, V. Vuitton, and R. V. Yelle. Origin of oxygen species in titan's atmosphere. *Journal of Geophysical Research: Planets*, 113(E10), 2008.

S. Kauffmann. *The origins of order*. Oxford University Press, 1993.

C. P. Kempes and D. C. Krakauer. The multiple paths to multiple life. *Journal of molecular evolution*, 89(7):415–426, 2021.

A. Kishimoto, B. Buesser, and A. Botea. Ai meets chemistry. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

T. Knuuttila and A. Loettgers. What are definitions of life good for? transdisciplinary and other definitions in astrobiology. *Biology & Philosophy*, 32(6):1185–1203, 2017.

A. N. Kolmogorov. *Foundations of the theory of probability: Second English Edition*. Courier Dover Publications, 1933.

L. Laudan. A confutation of scientific realism. *Philosophy of Science*, 48:19–49, 1981.

A. Léger, M. Pirre, and F. Marceau. Search for primitive life on a distant planet: relevance of 02 and 03 detections. *Astronomy and Astrophysics*, 277:309, 1993.

K.-P. Lv, L. Norman, and Y.-L. Li. Oxygen-free biochemistry: The putative chn foundation for exotic life in a hydrocarbon world? *Astrobiology*, 17(11):1173–1181, 2017.

E. Machery. Why i stopped worrying about the definition of life... and why you should as well. *Synthese*, 185(1):145–164, 2012.

N. Narita, T. Enomoto, S. Masaoka, and N. Kusakabe. Titania may produce abiotic oxygen atmospheres on habitable exoplanets. *Scientific reports*, 5:13977, 2015.

B. H. Patel, C. Percivalle, D. J. Ritson, C. D. Duffy, and J. D. Sutherland. Common origins of rna, protein and lipid precursors in a cyanosulfidic protometabolism. *Nature chemistry*, 7(4):301, 2015.

D. Pritchard. Wittgenstein on hinge commitments and radical scepticism in on certainty. In *A companion to Wittgenstein*, pages 563–575. Blackwell, 2017.

M. Ralser. An appeal to magic? the discovery of a non-enzymatic metabolism and its role in the origins of life. *Biochemical Journal*, 475(16):2577–2592, 2018.

H. Reichenbach. *Experience and prediction: An analysis of the foundations and the structure of knowledge*. University of Chicago Press, 1938.

B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson. Estimating the support of a high-dimensional distribution. *Neural computation*, 13(7):1443–1471, 2001.

E. Schrödinger. *What is life?* Cambridge University Press, 1944.

S. Seager. The future of spectroscopic life detection on exoplanets. *Proceedings of the National Academy of Sciences*, 111(35):12634–12640, 2014.

S. Seager and W. Bains. The search for signs of life on exoplanets at the interface of chemistry and planetary science. *Science advances*, 1(2):e1500047, 2015.

S. Seager, W. Bains, and J. Petkowski. Toward a list of molecules as potential biosignature gases for the search for life on exoplanets and applications to terrestrial biochemistry. *Astrobiology*, 16 (6):465–485, 2016.

D. S. Spiegel and E. L. Turner. Bayesian analysis of the astrobiological implications of life's early emergence on earth. *Proceedings of the National Academy of Sciences*, 109(2):395–400, 2012.

M. Swain. Exoplanet spectroscopy: a bright present, a brilliant future. In *EGU General Assembly Conference Abstracts*, volume 12, page 7631, 2010a.

M. R. Swain. Finesse-a new mission concept for exoplanet spectroscopy. In *Bulletin of the American Astronomical Society*, volume 42, page 1064, 2010b.

D. B. Vance and J. A. Jacobs. Water, bacteria, life on mars, and microbial diversity. *Water Encyclopedia*, 4:746–748, 2005.

S. I. Walker, W. Bains, L. Cronin, S. DasSarma, S. Danielache, S. Domagal-Goldman, B. Kacar, N. Y. Kiang, A. Lenardic, C. T. Reinhard, et al. Exoplanet biosignatures: future directions. *Astrobiology*, 18(6):779–824, 2018.

L. Wittgenstein. *On certainty*, volume 174. Blackwell Oxford, 1969.

C. Wright. Warrant for nothing (and foundations for free)? In *Aristotelian Society Supplementary Volume*, volume 78, pages 167–212. Wiley Online Library, 2004.