Spring 2024

# P.U.S.H. for Life Among the Stars: A Scientific and Philosophical Quest for Conceptualizing Uncertainty

Jacinta Creel Vallejo
*Bard College*

Recommended Citation
Creel Vallejo, Jacinta, "P.U.S.H. for Life Among the Stars: A Scientific and Philosophical Quest for Conceptualizing Uncertainty" (2024). *Senior Projects Spring 2024*. 53.
https://digitalcommons.bard.edu/senproj_s2024/53

# P.U.S.H. for Life Among the Stars: A Scientific and Philosophical Quest for Conceptualizing Uncertainty

A Senior Project submitted to
The Division of Science, Mathematics, and Computing
and to the Division of Social Studies
of
Bard College

by
Jacinta Creel Vallejo

Annandale-on-Hudson, New York
May, 2024

# Abstract

This senior project tackles how to deal with uncertainty in the search for life. Defining this uncertainty is tricky, and scientific efforts to do so are crucial. Such efforts include analyzing the data and biases of past, present, and future missions searching for exoplanets: planets outside our solar system. From there, the next step would be to infer what exoplanets have an atmosphere. This is a crucial, but not sufficient step, as having an atmosphere is a good sign of encountering life. However, finding an atmosphere is not an easy task, and this step will undeniably come with some amount of uncertainty. Hence, from the point of having an atmosphere, it would be essential to stop and analyze how much uncertainty comes from trying to find life at this stage, as finding an exoplanet with an atmosphere is a good step in finding life but not a sufficient one. The next step is to analyze the planet's atmospheres for their chemical composition. I will do this by using data from my advisor, Clara Sousa-Silva, on the spectra for the presumed biosignatures (RASCALL). This, as expected, also comes with an uncertainty of its own. Data on the search for exoplanet atmospheres and the detection of biosignatures within them would help us measure and navigate the uncertainty in the search for life.

Having found data, although with a fair amount of uncertainty, which will be and has been meshed along all the way, we can furthermore analyze the philosophical implications of the search for life. Uncertainty in the search for life has many philosophical implications, which I will examine as I seek to answer the following questions: On an epistemic note,

what is the weight of values and numbers in this project? Would the uncertainty attached to them make them less pure, or how can we balance the truth with uncertainty and make sure the data is reliable? I will also address a more pragmatic question dealing with uncertainty: how we should handle the risks of being wrong in our inferences, and how should we balance the risk of two kinds of errors, believing a false statement or rejecting a true one? In the context of astrobiology, this would correspond to believing we have found life when we have not, or missing life when it is present. Additionally, methodologically, is this approach, and the uncertainty it carries a good way of doing science, and what would an alternative look like?

# Contents

# Dedication

To those brave enough not only to look up at the heavens, but to be willing to persist their gaze long enough to wonder about its possible worlds.

# Acknowledgments

This journey would not have been possible without significant P.U.S.H.es.

To my family, you are everything to me, thank you for never doubting me. Thank you Ma y Pa for the constant support, even when I couldn't believe in myself, you always did. Thank you Cam, my editor-in-chief, and thank you Fer, for constantly reminding of what matters.

To my incredible advisors, thank you Clara for introducing me and guiding me in this universe of many worlds. Thank you Michelle, for navigating the wild currents of uncertainty by my side and for always pointing me towards the light.

To the entire physics and philosophy faculty, you have made my journey an extraordinary one. Thank you for your wisdom and patience.

To you Art, for showing me every day the only certainty in the expanding universe.

To my incredible friends (you know who you are), thank you for your support and constant curiosity in this journey. Thank you, Moselle-Life-Solucionator-Fredericks, for being in the star soup with me. Thank you Antonio, for your unconditional support and purposeful conversations. Thank you Andrés, for all the meaningful and intricate explanations.

To my extraordinary third advisor, thank you Zoe, for the clarity, for the excitement and for the never-ending support.

To the addition and elaboration of a very meaningful aspect in this project, thank you Professor Trudel, for introducing me to and revealing Proust's magic.

To the ones that stuck with me through it all, thank you Sasha, Luke and Chris, without you, I would have never done it.

To my squash team and friends, thank you for your endless joy and love.

Finally, to you, the reader, thank you for being curious to glimpse into what lies beyond our own planet. May you not be afraid of the uncertainty.

## 0.1 P.U.S.H. Acknowledgements

The creation of P.U.S.H. would have been impossible without the help of many incredible supporters. Special thanks to Luke Ingraham, Professor Christopher LaFratta, Yashar Khan, Farman Sayem, Clem Tarpey, Maria Inês Ferreira, and every single user.

# 1

# Introduction Part I: Context (Finding Life Outside of Earth)

> "Even those who sympathised with my perception of the truth which I wanted eventually to engrave within the temple, congratulated me on having discovered them 'with a microscope', when on the contrary it was a telescope that I had used to observe things which were indeed very small to the naked eye, but only because they were situated at a great distance, and which were each one of them in itself a world."
>
> *-Marcel Proust*
> *In Search of Lost Time: Time Regained*

The search for life comes with its uncertainty. To find life outside our own planet, we need both a location (i.e. a planet) and a sign of life (e.g. a molecule). This project investigates not only what takes for such a search, but more so, it will try to portray with the help of an algorithm where the location for life could be based on some constraints.

Before delving into the details and core of the project, it is critical to understand the human efforts that have gone into this search. Furthermore, before even regarding those, it would be helpful to illustrate this search with efforts performed regarding the question: If someone was looking for Earth, would they find us? To answer this question, let us shift to Carl Sagan's: How would we find life on Earth?[29]

## 1.1 First Attempts: How would we find life on Earth? Notes on Carl Sagan

Sagan poses a crucial question, and it is a crucial question in the search for extraterrestrial life standing from where we stand, Earth: "If the inhabitants of Mars set out to do preliminary exploration of the planet Earth, what would they have to do to detect life here?" [29] Here, we must place ourselves in the "alien's feet", could we potentially find a planet like Earth? What would the search for Earth look like? According to Carl Sagan the ways we could find our planet from far out in space, are the following.

- Ground Based Telescopes: If an extraterrestrial species were to try to look for us, using ground based telescopes would be a good starting point to try to observe and see if they can characterize the terrestrial environment. These observations would reveal temperatures, atmospheric pressures, the presence of liquid water, and also the bright and dark markings which outline the continents and oceans. All of these would immediately prompt the question: Is this environment suitable for life? Even if it is suitable, must it host life? This would probably produce some discussion among the alien astrologists particularly because these are promising signs of life however they might not be necessarily unequivocal, which would make the extraterrestrial species lay some constraints. Sagan mentions some: "There would be arguments that the great excess of oxygen in the Earth's atmosphere surely excludes the possibility of life because all organic compounds would be completely oxidized to carbon dioxide and water. There would probably be arguments that the temperatures on Earth were much too warm by Martian standards".[29]

- Radio Signals: pointing a small radio telescope at the Earth, at just the right frequency, would reveal signs of intelligent life. These signs might be unequivocal, but they are constrained by time, even if 200 years ago we were radio silent. [29]
- Photographic Search: assuming our extraterrestrial species would have developed a simple telescope, Earth could be detected going through phases, like we observe the moon or Venus. With a larger telescope, clouds and more detail of the atmosphere and Earth's surface would appear, observed directly. [29]

  The last measure mentioned in which we could find our planet from far out in space is:

- Space Vehicles: "Better resolutions could be obtained by a space-vehicle reconnaissance of the Earth".[29] What could these vehicles find? Sagan notes that even if the vehicles are able to take photographs at a resolution worse than one kilometer, it would still be quite impossible to notice life on earth.[29] Additionally, the "best" resolution may vary on each planet, and the idea of "size" could also differ for intelligent life.
- Time Constraint: A critical thing to consider is that any of these ways to detect life, are directly tied to a fundamental factor: time. Specifically, if alien astrobiologists

were to look for life before life generated here, they would not find us. If any of these were to work, when life was not developed yet, it would lead us to think that there is no life, but perhaps we are, or they are looking at the right place, just not the right *time*.

Until now, we have been talking about how a remote species would find us. Of course, one way to confirm life would be to actually go there.

“Despite widely advertised opinions that Mars is lifeless, Mariner 9 has discovered surface conditions that significantly improve the chances of life there”. Are surface conditions the only way to improve chances of life? How can we tackle this question without sending anything outside our atmosphere? Additionally, Sagan says: “Mariner 9 offers the first good chance of testing (and probably putting to rest) the persistent speculation about the existence of intelligent beings on Mars. But it is unlikely to have any direct bearing on the most fundamental issues-whether Mars can be a habitat for simpler forms of life. In my view, this question remains entirely open-at least until landing missions of the Viking-class journey to Mars in 1976.”[29]

This opportunity is very limited to only the nearest planets to Earth, leaving thousands of potential habitable planets unexplored. Sagan’s words touch upon a crucial aspect: in order for us to find life outside of Earth, we should consider the implications of how likely and difficult would it be to find *us*from far away. For example, what would aliens think of other planets, such as Venus or Mars? He concludes by saying space vehicles hold the answer when exploration cannot be done from the same planet. Further considerations are important. The shift of perspective Sagan mentions is crucial but there has to be an in between for finding life in planets that we will not be able to reach ever due to their distance, and which are likely prospects on which life could be found, in those cases we shall turn to what we have done to search for these other possible potential worlds.

Pausing here for a second, there seems to be a certain fundamental need for the extraterrestrial species to perform a type of science and to have the sufficient technology in order to perform that science. For example, Sagan notes: "Once the temperature structure of the atmosphere was determined, it would be clear that these were water clouds and not carbon dioxide clouds or dust clouds." [29] This section implies that the aliens would have to infer and realize certain things about Earth. Additionally, an important point is that for any of the methods stated, a level of technology and scientific advancement is required, this would mean, intelligent life would be required to detect any life at all. The parameters of *how* technologically advanced are not clear, but a minimum level seems to be a requirement; they would need to have the ability to detect certain frequencies or be able to tune certain resolutions for the telescopes.

Even if they did have the right technology and a similar science to us, that might not be enough to detect life: unless the resolution is so impressive that you could clearly see life, humans walking, it would be unlikely to detect life directly. Instead, they would have to infer it. What would help us determine the right parameters for a successful sighting of life?

## 1.2 Efforts to Find Life Beyond Earth

The search for life outside our own planet is one that has some history of itself. To introduce this section, it is important to consider some of the past efforts in the search for life alongside the perspective of what the "search for us" would look like.

Technologically, specifically the development of radio technology, opened up a powerful way to listen for and send messages, especially outside the solar system. The advancement of radio and optical telescopes has been key in the search for extraterrestrial life.[7]

One of the very first attempts to make our presence known in the cosmos was the Voyager record (1977). This record comprised a cosmic greeting card, with a lifetime of billions of years, and it portrays how we see ourselves in a cosmic context. [26] This record unfortunately does not offer a full perspective of our civilization, and it is controversial in regard to the idea of "space exploration" since it can be argued that we can explore and gain information without having to send anything to drift out in outer space. This theme is intricate, since dealing with exploration outside our own planet comes with philosophical questions as well regarding what is the best way to do so. What does it mean to "explore" outside our own planet?

### *1.2.1 The Equations*

Beyond sending something to outer space in the hope that it will be found, it is crucial to analyze and define how frequent life can be. Something that comes up are the statistical equations to do so. The Drake equation is a mathematical equation which predicts the number of intelligent alien civilizations in the Milky Way Galaxy. This equation is merely a guess since not all the factors in it are known, and it would be extremely difficult to know

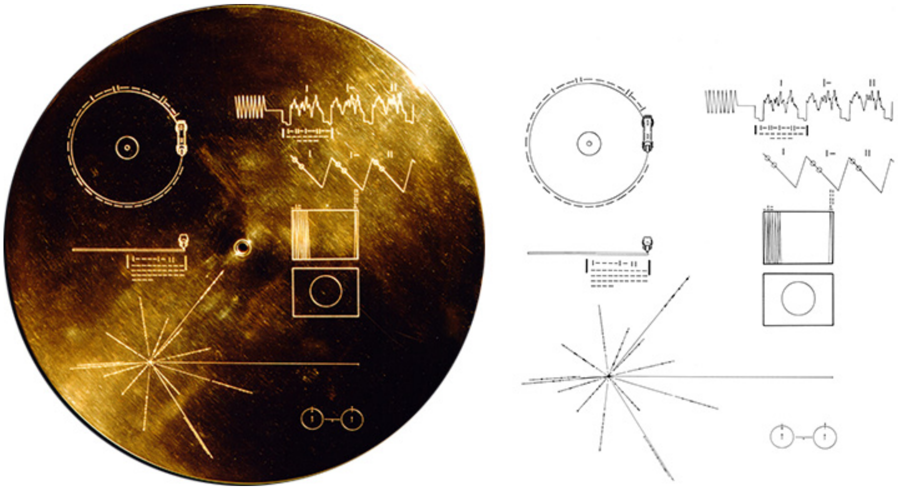


Figure 1.2.1: The Golden Record cover shown with its extraterrestrial instructions. Credit: NASA/JPL

them. The factors include: the rate of formation of stars in the galaxy **(R*)**, the fraction of those stars with planetary systems **(fp)**, the number of planets per solar system with a suitable environment for life (ne), the fraction of suitable planets on which life actually appears **(fl)**, the fraction of life bearing planets on which intelligent life emerges **(fi)**, the fraction of civilizations that develop a technology that releases detectable signs of their existence into space **(fc)**, the length of time such civilizations release detectable signals into space **(L)**. [37]The probability is calculated by a simple equation:

$$N = R* \times fp \times ne \times fl \times fi \times fc \times L,$$

[37]

where N stands for the number of intelligent aliens currently living in the Milky Way galaxy.

The Drake Seager equation expands the search for life outside our planet not only to intelligent life, but to all planets that can be detectable by bio signature gases. This broadens the scope of finding life, since it is not limited to intelligent life.[21]

The Seager Equation is formulated as:

$$N = N* \times Fq \times Fh \times Fo \times Fl \times Fs,$$

[21]

Where: N is the number of planets with detectable bio signature gases. **N*** is the number of stars observed. This number is determined by observational constraints, like the lifetime of the mission or the kind of telescope used for observation. **Fq** is the fraction of stars that are quiet, these stars are excellent for observing transiting planets, as they have less stellar activity that could mask biosignature gases. **Fh** is the fraction of stars with planets in the habitable zone. The habitable zone, the region around a star where conditions might allow liquid water to exist on a planet's surface. **Fo** is the fraction of those planets that can be observed. Since not all planets are positioned in a way that allows us to detect their bio signatures from Earth. **Fl** is the fraction of those planets that have life. Even if planets are positioned in the habitable zone, it does not mean there is life on them. **Fs** is the fraction of planets with life where the life produces a detectable biosignature gas. It is also important to note that not all life forms produce gases that we can detect. [21]

To introduce a last equation, the A Form Drake equation expands the probability of finding intelligent life, not only now but ever and anywhere in the universe.[17]. And it sets an important constraint.“Recent advances in exoplanet provide strong constraints on all astrophysical terms in the Drake equation. Set a firm lower bound on the probability that one or more technological species have evolved anywhere and at any time in the history of the observable Universe. We find that as long as the probability that a habitable zone planet develops a technological species is larger than $10^{24}$, humanity is not the only time technological intelligence has evolved. This constraint has important scientific and philosophical consequences“ [17]. This is the A Form Drake Equation:

$$A = Nast \times fbt,$$

[17] Where **A** is the number of technological species that have formed over the history of the observable universe, **Nast** stands for the number of habitable planets in a given volume of the universe and **fbt** corresponds to the likelihood of a technological species to arise on one of these planets. [37]

Although these equations may expand our understanding of our chances of finding life outside our planet by broadening the scope of what is defined as life, there is no proof this would be successful. Additionally, getting a probability would not imply we know where that life is, and if we cannot locate it, how useful is it? Finding life could look like searching where we least imagine possible, since we do not know what form a different life from our own might take. Equations like these inspire curiosity to see how far in science it takes us, but unfortunately it does not seem far at all.

Also, not all the factors in the equations are known, so this makes one consider the *use* of such equations, and how can we better them? Given that, we want a mathematical representation of finding life outside our solar system.

Overall, there seems to be some trouble when considering all of these equations, since successfully delimiting our chances for finding life in a single equation seems like an impossible task. This raises the question of what statistics are better for considering life outside our planet. What known facts can we consider to make a better approximation? The ideal would be to consider real information we can take from missions which have been tasked with finding signs of life. We must consider what we can find now and what would be possible in the future with coming missions, and from that build more concrete conclusions. The search for life, even the tiniest aspects, seems to be one in which real data is not only important, but necessary.

### *1.2.2 The Missions*

Knowledge from space missions that have as their aim finding exoplanets is crucial. Here are the past, present and future missions in the search for life outside our solar system.

- Past: Kepler & K2

  Kepler and later K2 were the first missions to search the galaxy for Earth-like planets. The main emphasis of this mission was the detection of planets around Sun-like stars. [11] Kepler detected planets by observing transits. Its lifetime was from 2009 to 2018, in which it observed more than 500,000 stars, and discovered 2779 exoplanets along with 1983 candidates, Kepler Objects of Interest (KOIs). Kepler observed about 0.25% of the sky. By such findings, Kepler and K2 also found that the universe is filled with even more planets than stars.[11] Kepler and K2 used the transit method (which will be described in detail), a method for the detection of exoplanets by its crossing (transit) across its host star. [27]

- Present: TESS

  TESS or Transiting Exoplanet Survey Satellite, was launched in 2018 and it is still functioning. TESS's main aim is to look at M type stars. TESS monitors 200,000 stars for temporary drops in brightness caused by planetary transits. It discovered transiting exoplanets by an all-sky survey. So far it has observed around 200,000 stars and has discovered around 432 exoplanets with some 7138 candidates. TESS found 6400 TOI's (Transiting Exoplanet Survey Satellite Object of Interest). It has observed around 93% of the sky. [24][28]

- Future: ARIEL & PLATO

  PLATO or (PLAnetary Transits and Oscillations of stars) is a future mission that will launch in 2026. PLATO will focus on the properties of rocky planets orbiting Sun-like stars. In particular, PLATO will discover and characterize planets in orbits up to the habitable zone. It will aim at characterizing rocky, icy and giant planets, aiding our understanding of planetary formation and evolution. [2]

  ARIEL or Atmospheric Remote-sensing Infrared Exoplanet Large-survey is a crucial mission as it differs from past missions searching for exoplanets such as Kepler or TESS, in the sense that ARIEL will be looking at the character of specific exoplanets: the composition of their atmospheres. Other missions focused more on just collecting information of how many planets are out there. ARIEL is going to launch until 2029, and it will go to the L2 Lagrange Point, 1.5 million km from Earth, opposite to the sun. It aims to observe around a thousand exoplanets, ranging all the way from Jupiter's and Neptune's down to super-Earth size planets. ARIEL will survey a diverse sample of 1000 exoplanets, simultaneously in visible and infrared wavelengths. This will gather spectra and photometric data, which will provide a catalog of planetary spectra. This catalog will then hopefully allow for the extraction of the chemical fingerprints of gases in the planets' atmospheres. This distinction from other mission, analyzing exoplanet's atmospheres, is absolutely critical in the search for life outside our solar system. Additionally, ARIEL will also study the thermal and scattering properties of the atmosphere as the planet orbits around the star. [14]

Most of these missions use as their method for detection the Transit Method. Let us get familiar with it.

- The Transit Method: This is the method in which most of the exoplanets have been discovered. This method finds exoplanets when they *transit* in front of their host star, and we are able to detect it. The way transits reveal an exoplanet is because when the planet passes in front of its host star, it produces a dim in the star's light. We can detect this dimming of the light, which can be seen in light curves – graphs showing light received over a period of time. So, when the exoplanet passes in front of the star, the light curve will show a dip in its brightness. Here is an example of a light curve.

  Why are transits useful? Transits can help us determine a lot of different characteristics of exoplanets. For example, the size of the exoplanet's orbit can be calculated from how long it takes to orbit once, in other words we can find the period. Transits can also aid us in determining the size of the planet itself. This is basically known by how much the star's brightness has lowered. An additional helpful characteristic is that we could also learn about an exoplanet's atmosphere during a transit. This can be done by the light going through the planet's atmosphere (if it has one) and that light can be analyzed to determine the atmospheric elements. We care about this aspect quite significantly, since atmospheric composition is important to determining

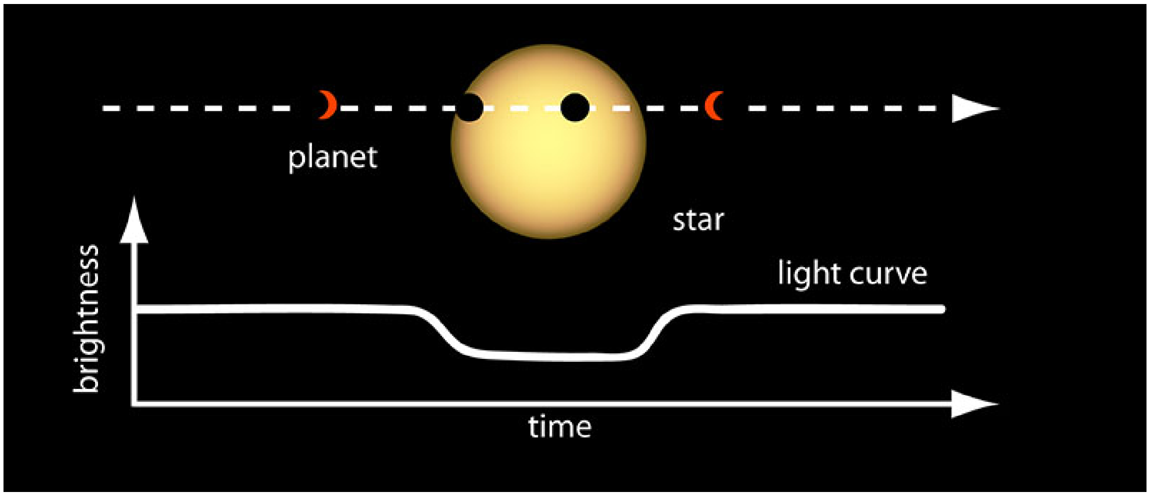


Figure 1.2.2: Light Curve of a Planet Transiting Its Star
Credit: NASA Ames

habitability. In addition, habitability could be determined by orbital size and the star's temperature, which could help us determine the temperature of the exoplanet itself, portraying whether or not the surface has a suitable temperature to host life. [5]

Aside from transits, there exist other methods for the finding of exoplanets. These include:

- Radial Velocity: Watching for Wobbles. 1089 exoplanets discovered.
- Gravitational Microlensing: Light in a Gravity Lens. 214 exoplanets discovered.
- Direct Imaging: 69 exoplanets discovered
- Astrometry: Minuscule Movements. 3 planets discovered[5]

For a visual representation of the cummulative exoplaner detections based on the different methods please see Fifure 1.2..3.

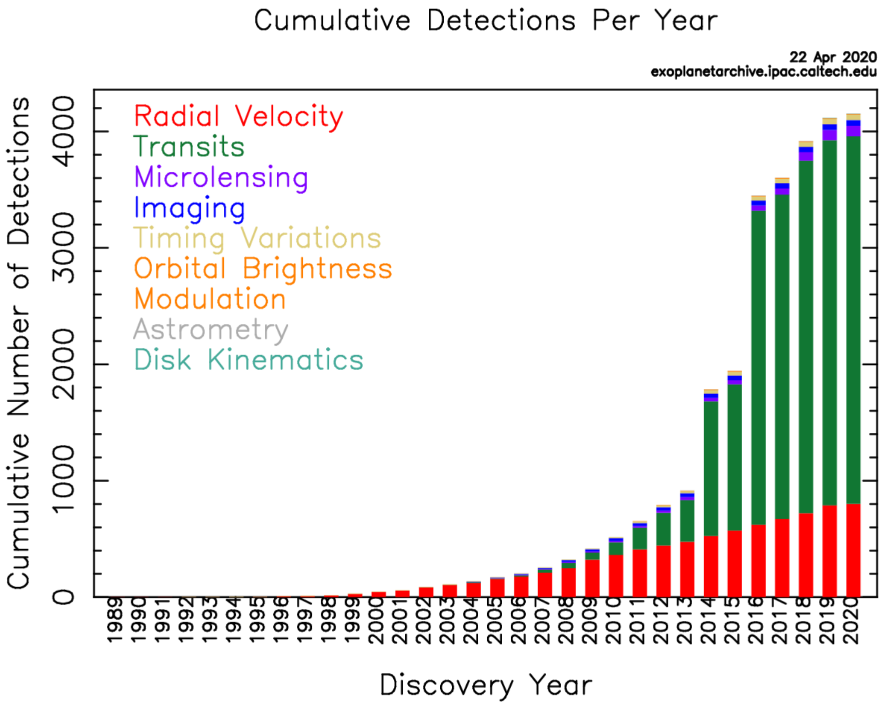


Figure 1.2.3: Cumulative Exoplanet Detections per Year
Credit: Exoplanet Archive Caltech

How can we compare knowledge from the mentioned missions with their corresponding methods to what the Drake Equation(s) and preliminary efforts to finding life convey? Why are these better statistics?

Tackling the question of how to find life outside our planet, or even further, how would aliens find us seem to be very complex ones. Such so that facing them would require more data and analysis, and putting the probability in a simple equation would not be enough. Starting the search for life based on an equation is a good start, but the equations (Drake equation, A-form Drake Equation and Drake-Seager equation) are merely a guess, and we do not count with all the information to plug into the equation to even get a reasonable answer. Data from the missions is a way to get better statistics, since from them, you can get statistics that you can trust and can deal with data that we can obtain and interpret, we delimit size, distance to nearby star, host star, and other parameters. However, there is a need for more concrete data and more specifically not only how many planets could exist as our own, but which ones would match with Earth's chemical composition, thermal structure, size, proximity to nearby star, among other characteristics. It seems crucial not only to gather all the information the missions give us, but to, furthermore, begin to set constraints on these planetary populations. This is the exact goal of the algorithm PUSH (Program for the Uncertainty in the Search for Habitability). To introduce it in detail, let us now move on to the structure of this project.

# 2

# Introduction Part II: Structure of the Project

> "...dimly realising that the universe contained innumerable elements which my feeble senses would be powerless to discern..."
>
> *-Marcel Proust*
> *In Search of Lost Time: Swann's Way*

This is a project that deals with data and uncertainty in the search for life beyond Earth. The goal of this chapter is to lay out the structure of the project. In order to tackle the concepts of *uncertainty* and *real data* from the start, it is crucial to understand where are they coming from and where are they being directed.

The search for life and additionally its uncertainty, comes with understanding *where* life could be found: a location, and *how* could we define it if we even find it. This is where, from the start, a good deal of data comes in.

To address the first part, the *location*, this project takes on its first data set: **Planetary Population Data Set**. For this, we look at the data and biases from missions that have as their goal seeking exoplanets (planets outside our solar system). Here we combine data from a couple of sources (will be described in more detail in Chapter 3), which are: Data from the Missions (what can the missions themselves tell us, what did they find or will find), Detection Biases from the Missions (understand in what way is the mission biased, is it more likely to find a certain type of planet, or star), and the last source is Debiased Population Statistics (knowing the biases of the missions allows us to infer useful data).

Now, to address the second part, *how* could we define life if found, it is important to introduce the second data set: **Molecular Data Set**. For this, we delve into the molecular data of biosignatures (molecules that signal life). The main data set for this comes from my advisor's, Professor Clara Soussa-Silva's data on the spectra for the known biosignatures named RASCALL (Rapid Approximate Spectral Calculations for ALL). [33]

Now an important question arises: We now have our sources of data, which in themselves come with a lot of uncertainty, how do they fit together with the goal of the project?

This is an excellent moment to introduce the core of the project. This is where we are actually *doing* something to the data we compiled. The goal for this project is to create an algorithm with inputs and outputs. This will be explained in more detail. For this to be understood better, the project will be divided in the following phases: Phase I: The Inputs, Phase II: The Algorithm, Phase III: The Outputs. Let us now briefly delve into each one.

## 2.1 Phase I: The Inputs

The First Phase will take as its inputs our sets of data:

- **Planetary Population Data Set**: which is comprised by *Planetary Statistics*, which we will delve deeper in Chapter 3
- **Molecular Data Set (RASCALL)**: which comprises two other data sets: *1)Biosignature Statistics* and *2)Bond Dissociation Energy Statistics.*

  And additionally a list of planetary, stellar and Chemical Constraints on both our sets of data.

### *2.1.1 Location Constraints*

The lists of constraints is as follows:

- **General Constraint:** the first constraint chosen was that the Planetary Data Set includes planets the size of Jupiter or smaller. It is unlikely to find life in a planet bigger than Jupiter. [16]
- **Star Type:** K, G, B, F, M, White Dwarf (WD), or the user will be able to choose any combination of these.
- **Star Mass (SunMass):** the user will be able to choose how many solar masses to set their constraints as. Ranging approximately from 0 to 10.
- **Planet Mass (EarthMasses):** the user will be able to choose the range of planetary mass using Earths mass. Ranging approximately from 0 to 500.
- **Planet Radius (EarthRadii):** the user will also be able to choose the range of the planet's radius, if this is perhaps more intuitive. Ranging from 0 Earth's radii to 500.

- **Host Star Distance to Us (Light Years):** there will be a constraint that includes the distance the planet's host star is to us. Ranging from 0 to 1000 Light Years.
- **Known Habitable Zone or not:** this is mainly part of the future work of the algorithm, to also filter the planets to see if they reside in the known habitable zone. This is determined by a set of parameters based on the planets' mass, type of host star, among others. [16]

This list of constraints will be applied to the main tabulated list of data of the Planetary Population Data Set. Mainly of which is taken from the Exoplanet Encyclopaedia. [15] More details about this will be seen in Chapter3. The key aspect of this list of constraints is that they can apply to any list of exoplanets, they are not tied to only the Exoplanet Encyclopaedia one.

Now let us look at the list of Chemical Constraints.

### *2.1.2 List of Chemical Constraints*

The lists of constraints is as follows:

- **Molecular Weight (g/mol):** how big or small the molecule is, according to its weight.
- **How easy are the molecules to destroy? Bond Disassociation Energy of Molecules (kJ/mol):** The energy that makes a specific atom in a bond of a molecule disassociate or jump off in such energy.

This list of constraints will be applied to the main list of the molecules and functional groups for 16,367 biosignatures from RASCALL. [33]

## 2.2 Phase II: The Algorithm

The algorithm for this project has as its goal, to take in the inputs, which as it was stated in the section before includes the data from both the data sets: Planetary Population Data Set and Molecular Data Set, plus both the Lists of Constraints: Location and Chemical Constraints. To then create a *filtered list* of the planets and the molecules.

The **filtered lists** are the following:

- **Possible Planetary Locations for Life**
- **Possible Biosignatures**

This is something that can be tuned by the user, to choose what constraints matter in each case and their scale. The algorithm in this stage is just a constrained list of potential locations for life. However, it is a very important one, in which the user can note how the constraints match the data sets, and change when the list of constraints is tuned for each user. The user ultimately gets a list of planets and molecules based on the constraints they chose, they get a narrower population to where life could potentially exist.

Additionally, in this stage, comes a **third list of constraints** which becomes an input in the algorithm. In which the algorithm takes in both the subgroups : Possible Planetary Locations for Life and Possible Biosignatures tuned in by the user and incorporates this third list. The goal for this third list of constraints is to unite both outputs. It looks something like this:

- **The Final Output: Potential Biosignature Targets Given Location** taking the two outputs, the goal for this constraint is to determine what locations (the constrained list of exoplanets) and signs of life (constrained list of molecules) the user will get. This will of course come with its uncertainty, which will be greatly analyzed further on in Chapter 8: Philosophical Implications.6 The output here is simply both lists of constraints.

- **Dealbreakers:** for future work, dealbreakers can be further applied to this output. For example, a planet having a "surface" could be a dealbreaker.

## 2.3 Phase III: The Outputs

The outputs are the *Location and Chemical Constraints* themselves, but with some physical and philosophical interpretation.

Once the user sets the constraints, they become: *the Possible Planetary Locations for Life* and *the Possible Biosignatures.* Furthermore, if the algorithm puts these outputs together it could provide two things: 1)"The Final Output: Potential Biosignature Targets Given Location" and 2) a "philosophical interpretation" of what this output means, specifically in terms of uncertainty, or it could also provide nothing at all, which in itself will be highly philosophical. Here, we can analyze further if the constrained list of potential locations for life is more "sensitive" to a specific type of constraint. The point here is to also, as mentioned in Phase II, to go further and be able to "tune" the constraints in the first inputs to then have different outputs and uncertainties. For example, if we are extremely conservative in our choices then we would expect nothing to come out in our second inputs, but if we are too open, then everything would come out, so the goal for this part of the algorithm is for the user to be able to "tune" it to the constraints they wish. The algorithm is highly adaptable, and as mentioned before it can take a more updated version of exoplanets, or debiasing factors, which will be mentioned in Chapter 3. 3

As mentioned before, the goal of the results for these outputs is to in any way (which can be tuned by the user) combine the previous outputs and come out with:

- **"The Final Output: Potential Biosignature Targets Given Location"**: Whatever comes out in this output would be very rich in the sense that it would give the user an idea of the "likelihood of life". This would be greatly dependent on the conditions the user chose.

  These conditions are the ones the user tunes in the first inputs (stated above as **Location Constraints** and **Chemical Constraints**). Thus, if one was too conservative in their choices, it would be logical for the output to be: none. If on the other hand one is too open, it is possible to get a huge amount as the output on both cases, and the output here for "The Final Output: Potential Biosignature Targets Given Location" could include either all the planets or molecules, none, or some. It will depend greatly on the choice tuned by the user.

- **Interpretation**: If the output is something, it will tell the story of connection between the places, and that will determine as well something about its uncertainty.

- **Nothing**: To get nothing out is also possible, since to a good amount this output is going to be decided by conditions as well. Additionally, to get "none" would not exactly count as a failure on the output, it would first tell us some information about the conditions we chose, and the connections between the data and second, it would also have huge philosophical implications.

- **Future Outputs**: for future work, the algorithm would be able to take both of the outputs as inputs and apply further constraints on them. Some of the future constraints could be “Stellar Radiation on Molecules”: taking the two constrained list of potential locations for life, the goal for this constraint is to determine a process of elimination/constrained list of potential locations for life, in which the algorithm can tell what molecules will be destroyed by sunlight/which ones survive. Here, knowledge of the wavelength of the UV rays will be useful. This also gives us information, since some molecules by “dying” protect others which survive. The output here which will be explained briefly is to portray how the biosignature atmospheric composition interacts with stellar radiation. This is a very interesting part, in which the basis of this project could extend to.

**Note:** On the following page, a visual diagram (Figure 2.3.1) is presented to showcase the structure of the project.

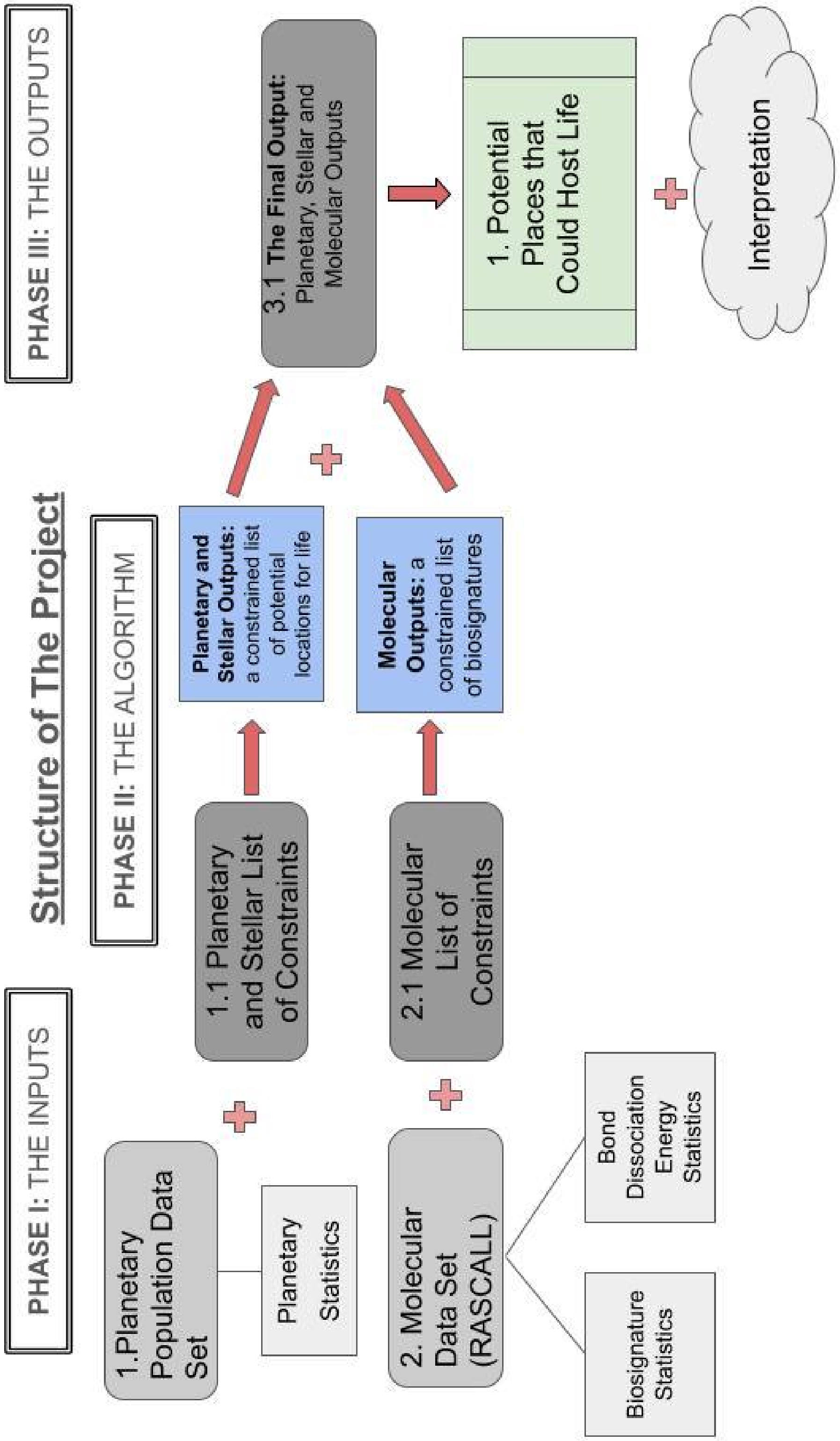


Figure 2.3.1: Structure of the Project

# 3
# Data Sets

> "But when from a long-distant past nothing subsists, after the people are dead, after the things are broken and scattered, still, alone, more fragile, but with more vitality, more unsubstantial, more persistent, more faithful, the smell and taste of things remain poised a long time, like souls, ready to remind us, waiting and hoping for their moment, amid the ruins of all the rest; and bear unfaltering, in the tiny and almost impalpable drop of their essence, the vast structure of recollection. "
>
> ---
>
> *-Marcel Proust*
> *In Search of Lost Time: Swann's Way*

The main goal of this project is to ultimately explore and conceptualize the uncertainty in the search for life. This goal is accompanied by many questions, such as: How many planets that could host life are in our galaxy? How many lie in the habitable zone? How many planets have atmospheres? If we find a planet with the same size of the Earth, does that immediately symbolize that it has life? How can we match this with the information we know about biosignatures? To answer questions such as these, it is crucial to infer from solid data.

The goal of this section is to get you, the reader, familiar with where the data the algorithm is taking in is coming from, and furthermore for you to understand how tricky and uncertain the data in astrobiology can be.

It will be extremely useful if we take a look at our missions that have as their objectives finding exoplanets, and notice what we can extrapolate from them.

Now, let us delve into the data.

### *3.0.1 Data From Missions*

- Data from Kepler and K2: Both Kepler and K2 observed more than 500,000 stars, and discovered 2779 exoplanets along with 1983 candidates. Kepler observed about 0.25% of the sky. [11] Although Kepler did not cover the entirety of the sky, it was assumed that what it saw, and found, could potentially be found in the rest of the sky as well. Kepler was also faced with some engineering limitations that had to make its search more narrowed. However, as we will see in the Debiased Population Data, a lot can be inferred from what was not found. Kepler Objects of Interest (KOIs) candidates were identified during the first four months of operation of the spacecraft in its quest to determine the frequency of Earth-size planets around Sun-like stars. Kepler rapidly became an invaluable resource for statistical investigation of the properties and distributions of exoplanets. It is important to know that by analyzing the information Kepler provided, by the finding of its planets using the transit method, it was noticed that not all photometric signals were caused by planets. [18] This emphasizes that false positive contamination is typically the main concern in transit surveys, including Kepler. This is due to there being a large array of astrophysical phenomena that can produce small periodic dimmings in the light of a star which can be virtually indistinguishable from those due to a true planetary transit. An example of this relating to Kepler is a background eclipsing binary falling within the photometric aperture of a Kepler target.[18]

What is a false positive? In very simple terms, a false positive occurs when something is said to be true when it is actually false. On the other hand, a false negative happens when something is claimed to be false when it is actually true. In astrobiology, and science in general, scientists are faced with these types of errors in their data, therefore it is crucial to keep their meanings in mind. [3]

- Data from TESS: TESS or Transiting Exoplanet Survey Satellite has observed around 200,000 stars and has discovered around 432 exoplanets with some 7138 candidates, and it is still active. In contrast to Kepler and K2, it has observed around 93% of the sky. TESS monitors 200,000 stars for temporary drops in brightness caused by planetary transits. [24] TESS stars are 10-100 times brighter than those surveyed by Kepler and K2 missions. This will make TESS candidates easier to characterize with follow-up observations. TESS is expected to find more than a thousand planets smaller than Neptune, including dozens that are comparable in size to the Earth. TESS aims to be a catalog of the nearest and brightest stars hosting transiting planets, which will therefore provide highly favorable targets for detailed investigations.[24] Some TESS findings are: total number of TOIs (TESS Objects of Interest) with Radii $< 4$ Earth Radii: 1367. Total number of Confirmed Planets: 329. Total number of False Positives: 1701

- Data from Plato: Plato or (PLAnetary Transits and Oscillations of stars) as we have seen, is a future mission that will focus on discovering and characterizing planets in orbits up to the habitable zone the zone where the temperature is just right for liquid water to exist on a planet's surface. In its lifetime Plato will characterize hundreds of rocky (including Earth twins), icy or giant planets by providing very exact measurements of their radii (up to 3% precision), their masses (more than 10% precision) and their age (up to 10% precision). Plato will observe more than 200 000 stars. It will acquire images every 25 seconds, and every 2.5 seconds for the two 'fast' cameras, for at least two years per target star. The main takeaway from this mission is that it will aid in our understanding of planetary formation, evolution and potential habitability of diverse exoplanets.[2]

  Plato will also analyze the host stars of these planets. From the data collected in this mission, scientists will hopefully perform stellar seismology, gathering evidence of 'starquakes' in the imaged stars. This would potentially give insight into the characteristics and evolution of the stars, and as a result it will better our understanding of planetary systems.[2]

- Data from ARIEL: Another future mission, with an expected launch date in 2029. It aims to observe around a thousand exoplanets, ranging all the way from Jupiter's and Neptune's down to super-Earth size planets. It will explore the properties (chemical composition, temperature) and atmospheres (weather, thickness, pressure) of approximately 1000 transiting planets. It will observe these planets in the visible and the infrared wavelengths, using both transit and eclipse spectroscopy (in near infrared wavelengths) and photometry (in visible wavelengths). This will gather spectra and photometric data, which will provide a catalog of planetary spectra. It will carry out long-term observations of the same exoplanet system for a duration of between 10 hours and three days. It will observe the dimming of a host star by a planet with a precision of 10–100 parts per million relative to the star. This catalog will then hopefully allow for the extraction of the chemical fingerprints of gases in the planet's atmospheres. This distinction from other mission, analyzing exoplanet's atmospheres, is absolutely crucial in the search for life outside our solar system. Additionally, ARIEL will also study of thermal and scattering properties of the atmosphere as the planet orbits around the star. [14]

### *3.0.2 Detection Biases*

These missions have and will provide us with a lot of data on potential places that could host life. Although all this data comes straight from the telescopes, the missions come with their biases. The next steps, are recognizing the biases and the efforts to debias them. We need a reliable population, and since we are not able to look everywhere, this is where knowing the mission's biases are crucial, to debias and infer from actual data. The biases will introduce some consequences in terms of uncertainty, but they will also allow us to get to the "Debiased Population Statistics", which portrays a bigger planetary population, hence expanding our sample size for a search for life.

In this next section, we will explore the mission's biases.

### Detection Biases from Kepler and K2

Getting to know the biases from the Kepler missions is essential to later be able to understand how getting debiased population statistics works. Most of the biases from Kepler are either taken from the mission itself or from Debiased Population Statistics Papers.[27] Some of the main biases from Kepler are related to the planets it detected.As Petigura et al. notes: "The most easily detectable planets in the Kepler survey are those that are relatively large and orbit close to their host stars, especially those stars having lower intrinsic brightness fluctuations (noise). These large, close-in worlds dominate the list of known exoplanets." Petigura et al. in their paper, "Prevalence of Earth-Size Planets Orbiting Sun-Like Stars" [27] aim at debiasing to include smaller planets. They mention: "the Kepler brightness measurements can be analyzed and debiased to reveal the diversity of planets, including smaller ones, in our Milky Way Galaxy."[27]. Another bias that Kepler had was that it was also imperfect at detecting Earth-like planets. Since these would fall into the "smaller" category. [27] Additionally, Kepler was biased due to the brightness of stars it observed, which, for example, the brightest Sun-like stars observed by Kepler(Kp=10–15 mag). [27]

### Detection Biases from TESS

Although TESS is still active, some of its biases are known. Even debiased population statistics have been created using these biases. Some of the biases TESS has are:By using the transit method in exoplanet detection, TESS in a way is biased. By using other methods, more exoplanets could be discovered. [5] TESS is also in a way biased by having at its mission objective the detection of M type stars. [23] By only having one type of planet as its mission's goal, it also allows for only certain types of planets.

### Detection Biases from PLATO

Although PLATO and ARIEL are future missions, and they do not have data yet to infer more population, we can already know some of their biases from what their mission's goals are.

- PLATO's first biased is the star it is mainly solar-type stars, relatively bright stars (between magnitudes 4 and 11)[2]

- Additionally, based on that bias, this mission is also biased on the planets it will be looking for. Which are, small planets orbiting in the habitable zone of stars. If life were to be present on a much bigger planet, PLATO would not find it.

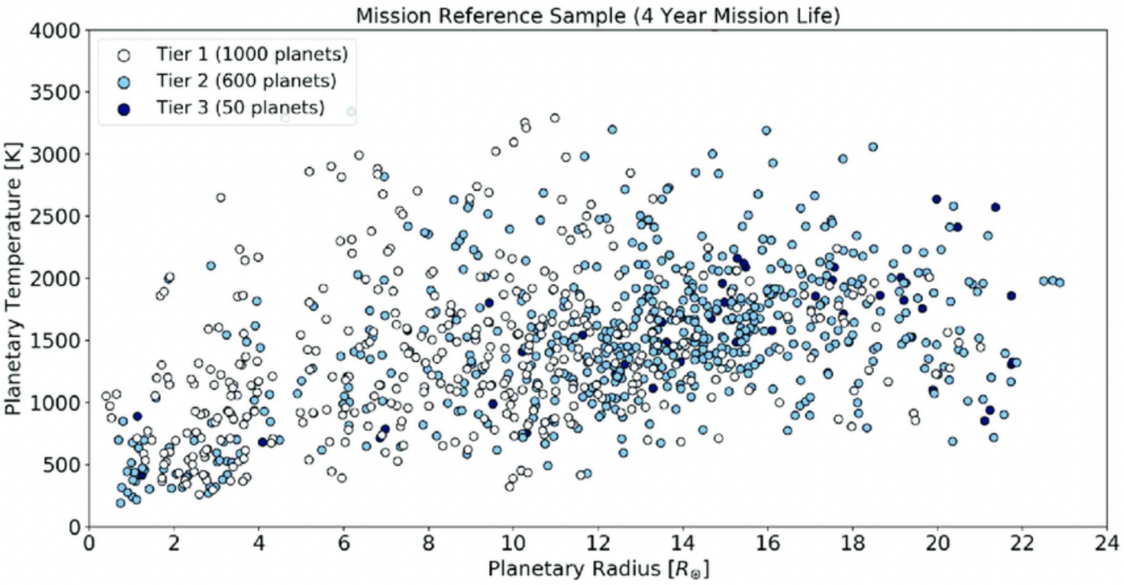
### Detection Biases from ARIEL

We are also able to find some biases from ARIEL based on its mission's goals. Much of this is described in a paper titled: "An Updated Study of Potential Targets for Ariel" by Edwards et al. A lot of biases could be inferred from this paper, and some of them include:

- Phase-curve targets for Ariel. Ariel's spectroscopic phase curves should be possible for Jupiter-sized planets, while smaller planets are suitable for multiband photometric observations. [13]

- This paper also includes the currently known exoplanets which are considered to be potential targets for Ariel. Although this list is broad, it well could happen life is actually present in an exoplanet that Ariel is not looking at. This is extremely hard to know, since by knowing we would already have found life. Therefore, as much as this is a biased, it seems to be a necessary one.[13]

- Another bias this mission has is based on the temperature of the exoplanet's Ariel aims at finding. We know the mission's goal for planetary radius and temperature distribution of a potential mission reference sample.[13] For this, it will be most helpful to see a graph. Please see Figure 3.0.1.



Planetary radius and temperature distribution of a potential Ariel mission reference sample from ArielRad.

Figure 3.0.1: Planetary radius and temperature distribution of a potential Ariel mission reference sample. Credit: ArielRad.[13]

### Detection Biases from the Transit Method

As great as this method is, there are a few drawbacks. First of all, it has a low probability of properly aligned planets. The probability of transit for Earth-like planets around Sun-like stars is around 0.5%. Additionally, it is harder to detect smaller planets and other astrophysical phenomenon can also mimic transits, so there could be some confusion. Even though this method has been successful, there is still an abundance of information hidden in the data sets. (Personal communication with Daniel Yahalomi, April 2024).

### *3.0.3 Debiased Population Statistics*

Knowing some of the mission's biases will be crucial in inferring to create a bigger population data set: Debiased Population Statistics. These are statistics that take the bias of past or present missions in order to portray some information. So, it uses the lack of information or the biases, to get more information. These are efforts taken from real papers to understand the biases each mission has (technological, goal-oriented) and to infer more data, hence, more population from them. It is a fascinating method, and in this next section we will observe the extrapolated data from such biases. It is important to note, as we learned from the biases in themselves, they were adding more uncertainty, therefore these new, debiased population statistics will also have uncertainties. However, they are very useful, since we cannot travel to every exoplanet one by one, we must infer.

#### Debiased Population Statistics Data from Kepler and K2

To gather the debiased information regarding the Kepler missions, it will be necessary to draw upon a couple of papers that have exactly that as their aim. This section will provide data from the following papers: “The Occurrence of Potentially Habitable Planets Orbiting M Dwarfs Estimated from the Full Kepler Data Set and an Empirical Measurement of the Detection Sensitivity” [11], “Prevalence of Earth-size planets orbiting Sun-like stars” [27] and “The False Positive Rate of Kepler and the Occurrence of Planets”[18]. What new data do they offer?

Dressing et al [11] searched the full four-year Kepler data set for transiting planets and focused on small planets orbiting small stars. They were also interested to know how planet occurrence depends on factors such as planet radius, orbital period, stellar insolation, and host star properties based on this, they were able to find:

- 156 planet candidates plus 1 not identified KOI. They added more population to those directly found by Kepler!
- Later on, by analyzing the planet's periods shorter than 50 days, they found "0.56+0.06-0.05 Earth-size planets (11.5 RadiusEarth ) and 0.46+0.07-0.05 super-Earths (1.52 RadiusEarth) per M dwarf."[11] This again added more planets to those just found by the mission. They were able to analyze more in depth the data taken from Kepler for such findings.
- More general claims found by Dressing et al. were that in total, they estimated a cummulative planet occurrence rate of 2.5±0.2 planets per M dwarf with radii 1-4 Earth Radius and periods shorter than 200 days. [11]
- Further on, their population got more specific. Once they shifted their focus not to the general locations of planets, but to those residing in the habitable zone, where within a conservatively defined habitable zone (HZ) (which is based on the moist greenhouse inner limit and maximum greenhouse outer limit) they estimate an occurrence rate of "0.16+0.17-0.07 Earth size planets and 0.12+0.10-0.05 super-Earths per M dwarf HZ."[11]
- Dresing et al. went even further to habitability and adopting the broader insolation boundaries, of the recent Venus and early Mars limits, they suggest that "the nearest

potentially habitable non transiting and transiting Earth size planets are 2.6±0.4 pc and 10+1.6-1.8 pc away respectively." [11] This is incredible! They estimated the distance of the nearest potentially habitable planets.

- Further on they included super-Earths, and these distances are lowered to 2.1±0.2 pc and 8.6+0.7-0.8pc. [11]

- In astrobiology, as it is the case with science in general, teamwork is essential. Therefore, Dressing et al. also mention that other scientists acquired spectra for a subset of stars without detected planets. This is also essential in inferring more population, since if we know the stars, we can infer the planets. They did this in efforts to recognize the importance of characterizing the full target sample as well as the planet host stars in order to constrain the planet occurrence rate, (Mann et al. 2012) [11]

- Some of their finding were that they found that "the majority of bright (Kp ¡ 14) Kepler target stars are giant stars (several hundred of which were classified as dwarfs in the KIC) but that 93% of fainter stars are correctly classified as dwarfs."[11] Knowing these details about the stars is essential in later inferring details about the planets that orbit these stars.

- Additionally, Dressing et al mention another group of scientists (Howard et al. 2012) [11] focused on another set of stars. They worked with the first three quarters of Kepler data and were able to estimate the frequency of planets around main sequence GK stars. They found that the occurrence rate of planets "increased sharply with decreasing planet size and moderately with increasing orbital period."[11]

Pausing here for a second, it is crucial to take a look back and notice, the new data papers such as this one [11] are taking in. Debiased Population Statistics is able to give us more data in relation to what we already have. We now have much more information than what was simply found in the mission.

We are now introducing another paper that focused on the Kepler missions:“Prevalence of Earth-size planets orbiting Sun-like stars” by Petigura et al [27]. Petigura et al. also analyzed the data from the Kepler missions and provided interesting findings. The most easily detectable planets in the Kepler survey are those that are relatively large and orbit close to their host stars, especially those stars having lower intrinsic brightness fluctuations (noise). These planets dominate the list of known exoplanets. However, the Kepler brightness measurements can be analyzed and debiased to reveal the diversity of planets, including smaller ones, in our Milky Way Galaxy. These previous studies showed that small planets approaching Earth size are the most common, but only for planets orbiting close to their host stars. Petigura et al. extend the planet survey to Kepler's most important domain: Earth-size planets orbiting far enough from Sun-like stars to receive a similar intensity of light energy as Earth. The key for their work was to find Earth like planets that receive almost the same light intensity as the Earth. This paper debiases the sample by extending the planet survey. [27]

- Petigura et al's. search for planets was further restricted to the brightest Sun-like stars observed by Kepler(Kp=10–15 mag). These 42,557 stars have the lowest photometric noise, making them susceptible to the detection of Earth-size planets.[27]

- After searching for planets around 42,557 relatively quiet GK stars, Petigura et al. found that “7.7 ± 1.3% of GK stars host small planets (1–2 Earth’s Radius) in periods between 25 and 50 days.

- They also extrapolated to predict that “22% ± 8% of GK stars host 1–2 Earth’s Radius planets receiving between 1/4 and 4 times the insolation received by the Earth”. “Their calculation incorporated a 10% correction for false positives. [11][27]

- Petigura et al. were interested in determining whether Earth-like planets are common or rare. They searched for Earth-size planets that cross in front of their host stars by examining the brightness measurements of 42,000 stars from the Kepler mission. They found 603 planets, including 10 that are Earth size (12 Earth’s radius).Already adding even more to the exoplanet population.

- Petigura et al. utilized an interesting method, they noticed Kepler was imperfect regarding the detectability of such planets (603) and accounted for it by “injecting synthetic planet–caused dimmings into the Kepler brightness measurements and then recorded the fraction detected”[27].

- With this interesting method, they found that “11± 4% of Sun-like stars harbor an Earth-size planet receiving between one and four times the stellar intensity as Earth”. Additionally, by extrapolating, they found “5.7(+1.7 2.2)% of Sun-like stars harbor an Earth-size planet with orbital periods of 200–400 days”. [27] By their findings, more Earth-size planets were inferred, with more information than just the planet’s presence, as it was in this case, the stellar intensity.

As mentioned before, teamwork and collaboration are essential in looking for life outside of Earth. Therefore, it is important to draw your attention to efforts to showcase how these scientists have collaborated.

- Dressing et al. mentioned that for a follow-up study, Foreman-Mackey et al. (2014) used the reported search completeness and planet candidates from Petigura et al. to rederive the planet occurrence rate using a hierarchical Bayesian model. "The Foreman-Mackey et al. analysis differed from the Petigura et al. analysis in two key aspects: (1) Foreman-Mackey et al. (2014) considered measurement errors in the stellar and transit parameters and (2) they did not assume that the planet occurrence rate was flat in log period, instead using a flexible Gaussian process to model the occurrence rate assuming a smooth functional form. *As a result, Foreman-Mackey et al. (2014) found an occurrence rate of potentially habitable Earth-size planets three times lower than the Petigura et al. (2013a) estimate.*"[11] This is essential. Not only are Debiased Population Statistics inferring more locations, but based on their methods and findings, even more knowledge can come out from them.

Now is time to introduce another Debiased Population Statistics paper, that uses not only Kepler's data but the rate of their false positives. As we saw earlier in the chapter, 3.0.1 a false positive occurs when something is said to be true when it is actually false. Using this logic, our last group of scientists; Fressin et al. [18] in their paper, "The False Positive Rate of Kepler and the Occurrence of Planets"[18] conducted a follow-up study of the Kepler planet occurrence rate incorporating both contamination from false positives and a more sophisticated model of pipeline sensitivity. They employed a hierarchical approach by first estimating the population of Jupiter-size planet candidates that might be astrophysical false positives. They then iteratively determined the occurrence rate of small planets by modeling the fraction of larger planet candidates that might be hidden as smaller planet candidates in diluted transit events.[18] Based on this:

- Fressin et al. performed numerical simulations of the Kepler targets and of stars in the background to predict the occurrence of astrophysical false positives detectable by the mission. [18]

- “They found a global false positive rate of 9.4 ± 0.9%” [11] and noted that considering false positives is particularly important when calculating the occurrence rate of giant planets “(622R, FP rate = 17.7% ± 2.9) and Earth-size planets (0.8–1.25R , FP rate = 12.3% ± 3.0).” [11] Knowing the rate of false positives is also essential to infer what could be true.

- Using this false positive data, Fressin et al. quantified and characterized the distribution and rate of occurrence of planets down to Earth size with no prior assumptions on their frequency by subtracting from the population of actual Kepler candidates their simulated population of astrophysical false positives.

- By doing this, they found that “16.5% ± 3.6% of main-sequence FGK stars have at least one planet between 0.8 and 1.25 Earth’s Radius with orbital periods up to 85 days.”

- Why is this important? Fressin et al. mention the significance of this step: “this result is a significant step toward the determination of eta-earth, the occurrence of Earth-like planets in the habitable zone of their parent stars.”[18]

As we can see, huge amounts of data can be extrapolated from methods such as debiased population statistics. It is a collaborative effort, which involves knowing the biases of the missions, but it portrays much more information than what comes directly from the telescope. Finding life outside of Earth is definitely, as we can see, not an easy task. However, having a larger pool of candidates is a good idea. All the stated data from Kepler's debiased population statistics, adds to that pool and even gives us more detail about the mission, such as the rate of false positives.

Now, what about other missions? Do they have similar statistics? We can only run these methods on missions that are or have been active and therefore portray results, such as Kepler, K2 and TESS. For future missions, such as ARIEL and PLATO, these statistics can be created, once the data from the telescopes is collected.

For now, let us shift to the other mission that is active and has this type of statistics: TESS. For this, we are going to focus on two papers: "The Occurrence Rate of Terrestrial Planets Orbiting Nearby Mid-to-late M Dwarfs from TESS Sectors 1–42"[23] and "Predicted Number, Multiplicity, and Orbital Dynamics of TESS M-dwarf Exoplanets" [4]

### Debiased Population Statistics Data from TESS

Starting with the first paper, "The Occurrence Rate of Terrestrial Planets Orbiting Nearby Mid-to-late M Dwarfs from TESS Sectors 1–42"[23]. TESS provided us with a lot of information, specifically because, in contrast to Kepler, TESS was able to look at most of the sky. With the information we got, and the biases from the mission, we are able to create more statistics. Let us delve into them. Ment and Charbonneau took a sample of stars from TESS, and present an analysis of 363 mid-to-late M dwarfs within 15 parsecs of the Sun with masses between 0.1 and 0.3 solar masses observed by TESS within sectors

1–42. The median stellar mass of the sample is 0.17 solar masses. They search the TESS light curves for transiting planets with orbital periods below 7 days, and recover all six known planets within the sample. [23]

- An interesting finding was that each of these planets was consistent with a terrestrial composition, counting with a planet radii between “0.91 and 1.31 Earth’s radii”. [23] This finding was pretty significant since compatibility with Earth could be essential for the search for life, since we know there is life on Earth.
- Ment and Charbonneau obtained a cumulative occurrence rate of “0.61+0.19-0.24 terrestrial planets per star with radii above 0.5 R and orbital periods between 0.4 and 7 days.”[23]
- They also found that, planets larger than 1.5 Earth’s radii (sub-Neptunes) are significantly less abundant around mid-to-late M dwarfs compared to earlier-type of stars, while the occurrence rate of terrestrial planets is comparable to that of more massive M dwarfs. This was also an important finding, since knowledge about not only the planet but its relation to its star is crucial.[23]
- Overall, they estimated that terrestrials outnumber sub-Neptunes around mid-to-late M dwarfs by 14 to 1, in contrast to GK dwarfs, where they are roughly equinumerous. Regarding standard deviation, they placed a 1 upper limit of 0.07 planets larger than 1.5 Earth’s radii per star within the orbital period range of 0.5–7 days. The concluding finding was that for a downturn in occurrence rates for planet radii below 0.9 Earth’s radii, suggesting that Earth-sized and larger terrestrials may be more common around mid-to-late M dwarfs. [23]

As we can see, debiased population statistics from the mission TESS allow for a lot of crucial conclusions. Such as around mid-to-late M dwarfs type stars, terrestrial planets were more abundant than sub-Neptunes. All this data helps build our understanding of where we should look for life.

What have other findings regarding the biases and population of the TESS mission? Shifting our attention to the second paper: "Predicted Number, Multiplicity, and Orbital Dynamics of TESS M-dwarf Exoplanets". [4] Ballard et al. presents a study of the M-dwarf exoplanetary systems from the TESS mission. They extracted ensemble completeness functions that recover the planet detections from previous work for stars between 3200 and 4000 Kelvin. Therefore, they mainly focused on the temperature of the stars TESS observed. They predicted four main findings.[4]

- The first finding was that TESS will likely detect more planets orbiting M dwarfs than previously predicted. Around stars with effective temperatures between 3200 and 4000 K, they predict TESS will find "1274 ± 241 planets orbiting 1026 ± 182 stars, a 1.2-fold increase over previous predictions".[4]

- The second finding related to the transiting planets, mentioning that was TESS will find two or more transiting planets around 20% of these host stars, similar to Kepler.[4]

- The third finding was that TESS light curves in which one or more planets are detected will often contain transits of additional planets below the detection threshold of TESS. "Among a typical set of 200 TESS hosts to one or more detected planets, 93 ± 17 transiting planets will be missed." However, transit follow-up efforts will result in additional planet discoveries using the photometric sensitivity to detect an Earth or larger around a mid-M dwarf, even with very modest period completeness. [4]

- The final finding was that the strong preference of TESS for systems of compact multiples indicates that "TESS planets will be dynamically cooler on average than Kepler planets, with 90% of TESS planets residing in orbits with e ¡ 0.15." For this, Ballard et al. included both (1) a predicted sample of planets detected by TESS orbiting stars between 3200 and 4000 K, including additional non transiting planets, or transiting and undetected planets orbiting the same star and (2) sample completeness functions for use by the community. [4]

This last point "(2) sample completeness functions for use by the community" [4], is essential since the community, either scientists or regular citizens using the data can contribute in a lot of ways to the search for extraterrestrial life. In fact, citizen help is encouraged for such a task, since looking for life outside our planet is not an easy one. Therefore, it is crucial to expose the data to the public.

Hopefully, getting some direct data from multiple debiased population statistics aided the reader in understanding the process of what does behind a planetary population sample. As you can see, there is much more information than what the telescope gives us. With technology, and collaborative efforts, we can start to build good population statistics that can be used both by scientists and civilians to keep exploring the quest for extraterrestrial life.

Now, it is imperative to introduce the actual data sets that were used in this project and that are going into our algorithm. As mentioned before, this project is based on data from two main sources, and ultimately creating Possible Planetary Locations for Life from them. Now is the time to delve our data sets and understand the actual data this project is dealing with to be able to comprehend what exactly is the algorithm taking in as its inputs.

The data sources and the goals they aim at achieving are the following:

### *3.0.4 Planetary Population Data Set*

It is important to mention that a lot of effort went into building and choosing this data set, and there were two different attempts for it. There was a preliminary data set that aimed to collect all the data we just saw in the previous section relating to the data from the missions and the debiased population statistics. As ambitious as it sounds, given the scope of this project, it became really difficult to elaborate, and this first trial will be exhibited, but it was unfortunately revised since it did not encompass all the information. However, inserting into the algorithm a data set such as this one would be a future goal.

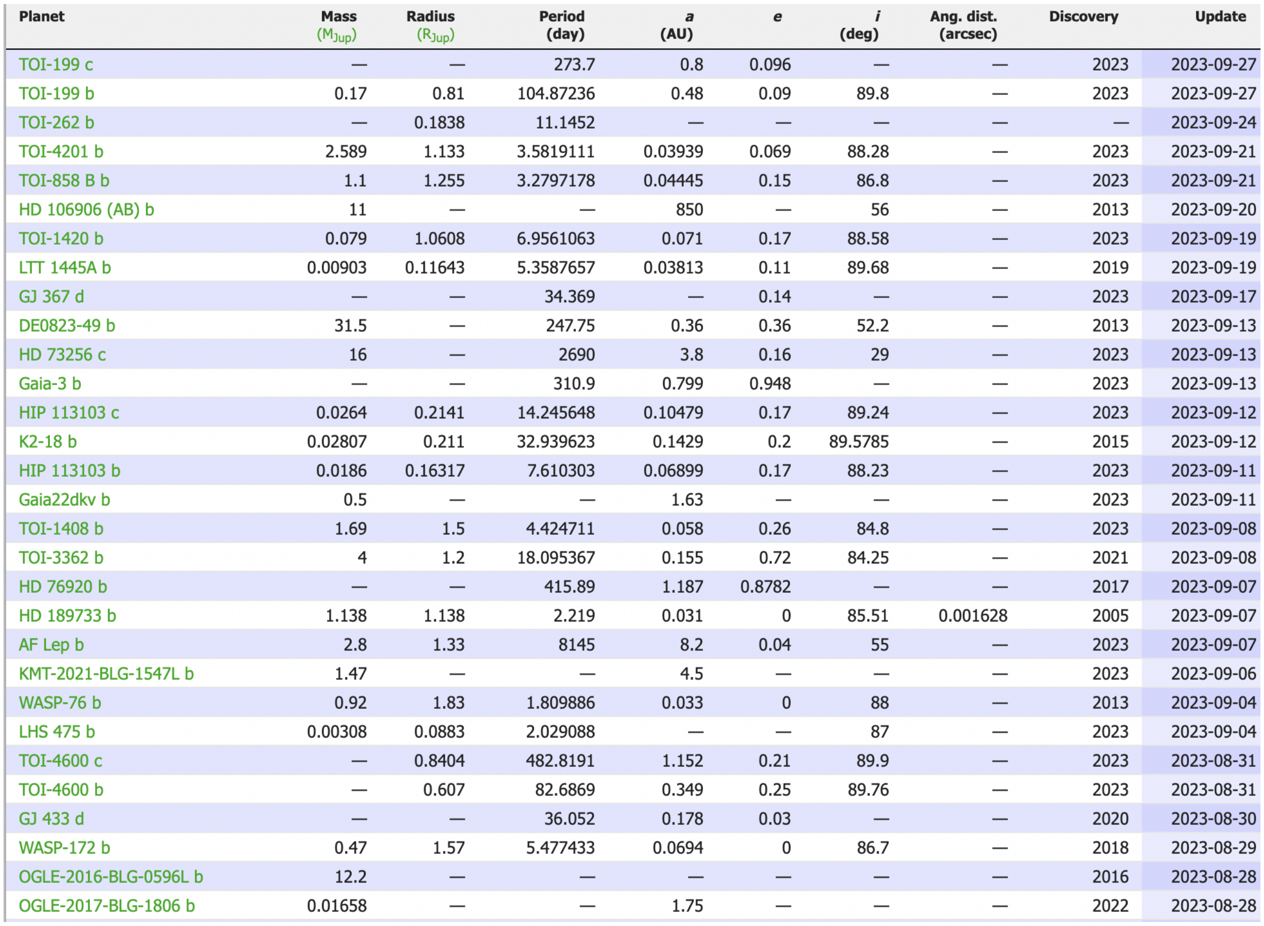
The second data set was successful, since it fits the scope of the project. It was a collaborative effort between institutions, and this (as we will see in the next section) was the main input for the algorithm. Let us delve into it.

- **Planetary Statistics:** The goal for this data set is to get a grasp of a number of possible habitable planets with specific characteristics such as: Planet Name, Planet Mass(MJup), Planet Mass (MEarth), Planet Radius(RJup), Planet Radius (REarth), Star Name, Star Distance to Us (pc), Star Distance to Us (LY), Star Mass (SunM), Star Radius, Star Type, Has Known Atmosphere. To obtain this number and further information about these planets it is necessary to make assumptions from several sources: data from missions (Kepler, K2, PLATO and TESS), Detection Biases from the missions (transiting time variations, telescope position and mapping area, tendency to find only a certain type of planet or star). The third type of data is Debiased Population Statistics. The goal is to study all of these data sets and make an informed estimate of the potentially habitable planets, with specific information about the location of the estimated planets. This would output the number of "nearby opportunities for life".

This data is coming from The Extrasolar Planets Encyclopaedia[15]. It is an encyclopedia that since its last update which was September 2023 the catalog shows 5506 confirmed planets, 2638 candidates and 821 multiple planets systems. These included planet detections published or submitted to professional journals or announced by professional astronomers in professional conferences. [15] Acknowledgement: "This research has made use of data obtained from or tools provided by the portal exoplanet.eu of The Extrasolar Planets Encyclopaedia."

Additionally, for the exact input that goes into the first part of my algorithm, "Planetary Population Data Set", what was done was taking the whole catalog and setting a general constraint: reducing the planets to only those that have the size of Jupiter or smaller. This was done mainly due to the fact that life is more likely to be found in a Jupiter size planet or smaller. [16] Once we had that constraint set, it reduced significantly our Planetary Population Data Set from 5506 planets, 4064 planetary systems and 878 multiple planet systems to 118 planets, 852 planetary systems, and 171 multiple planet systems. This constrained list was what became our Planetary Population Data Set, and was the main input for this section of the algorithm, in which the user would set more constraints, as will be explained in detail in the next chapter. 4

For actual data from the Encyclopaedia, look at Fig 3.0.1.



| Planet | Mass ($M_{Jup}$) | Radius ($R_{Jup}$) | Period (day) | *a* (AU) | *e* | *i* (deg) | Ang. dist. (arcsec) | Discovery | Update |
|---|---|---|---|---|---|---|---|---|---|
| TOI-199 c | — | — | 273.7 | 0.8 | 0.096 | — | — | 2023 | 2023-09-27 |
| TOI-199 b | 0.17 | 0.81 | 104.87236 | 0.48 | 0.09 | 89.8 | — | 2023 | 2023-09-27 |
| TOI-262 b | — | 0.1838 | 11.1452 | — | — | — | — | — | 2023-09-24 |
| TOI-4201 b | 2.589 | 1.133 | 3.5819111 | 0.03939 | 0.069 | 88.28 | — | 2023 | 2023-09-21 |
| TOI-858 B b | 1.1 | 1.255 | 3.2797178 | 0.04445 | 0.15 | 86.8 | — | 2023 | 2023-09-21 |
| HD 106906 (AB) b | 11 | — | — | 850 | — | 56 | — | 2013 | 2023-09-20 |
| TOI-1420 b | 0.079 | 1.0608 | 6.9561063 | 0.071 | 0.17 | 88.58 | — | 2023 | 2023-09-19 |
| LTT 1445A b | 0.00903 | 0.11643 | 5.3587657 | 0.03813 | 0.11 | 89.68 | — | 2019 | 2023-09-19 |
| GJ 367 d | — | — | 34.369 | — | 0.14 | — | — | 2023 | 2023-09-17 |
| DE0823-49 b | 31.5 | — | 247.75 | 0.36 | 0.36 | 52.2 | — | 2013 | 2023-09-13 |
| HD 73256 c | 16 | — | 2690 | 3.8 | 0.16 | 29 | — | 2023 | 2023-09-13 |
| Gaia-3 b | — | — | 310.9 | 0.799 | 0.948 | — | — | 2023 | 2023-09-13 |
| HIP 113103 c | 0.0264 | 0.2141 | 14.245648 | 0.10479 | 0.17 | 89.24 | — | 2023 | 2023-09-12 |
| K2-18 b | 0.02807 | 0.211 | 32.939623 | 0.1429 | 0.2 | 89.5785 | — | 2015 | 2023-09-12 |
| HIP 113103 b | 0.0186 | 0.16317 | 7.610303 | 0.06899 | 0.17 | 88.23 | — | 2023 | 2023-09-11 |
| Gaia22dkv b | 0.5 | — | — | 1.63 | — | — | — | 2023 | 2023-09-11 |
| TOI-1408 b | 1.69 | 1.5 | 4.424711 | 0.058 | 0.26 | 84.8 | — | 2023 | 2023-09-08 |
| TOI-3362 b | 4 | 1.2 | 18.095367 | 0.155 | 0.72 | 84.25 | — | 2021 | 2023-09-08 |
| HD 76920 b | — | — | 415.89 | 1.187 | 0.8782 | — | — | 2017 | 2023-09-07 |
| HD 189733 b | 1.138 | 1.138 | 2.219 | 0.031 | 0 | 85.51 | 0.001628 | 2005 | 2023-09-07 |
| AF Lep b | 2.8 | 1.33 | 8145 | 8.2 | 0.04 | 55 | — | 2023 | 2023-09-07 |
| KMT-2021-BLG-1547L b | 1.47 | — | — | 4.5 | — | — | — | 2023 | 2023-09-06 |
| WASP-76 b | 0.92 | 1.83 | 1.809886 | 0.033 | 0 | 88 | — | 2013 | 2023-09-04 |
| LHS 475 b | 0.00308 | 0.0883 | 2.029088 | — | — | 87 | — | 2023 | 2023-09-04 |
| TOI-4600 c | — | 0.8404 | 482.8191 | 1.152 | 0.21 | 89.9 | — | 2023 | 2023-08-31 |
| TOI-4600 b | — | 0.607 | 82.6869 | 0.349 | 0.25 | 89.76 | — | 2023 | 2023-08-31 |
| GJ 433 d | — | — | 36.052 | 0.178 | 0.03 | — | — | 2020 | 2023-08-30 |
| WASP-172 b | 0.47 | 1.57 | 5.477433 | 0.0694 | 0 | 86.7 | — | 2018 | 2023-08-29 |
| OGLE-2016-BLG-0596L b | 12.2 | — | — | — | — | — | — | 2016 | 2023-08-28 |
| OGLE-2017-BLG-1806 b | 0.01658 | — | — | 1.75 | — | — | — | 2022 | 2023-08-28 |

Figure 3.0.2: Exoplanet Encyclopaedia Sample Credit: The Extrasolar Planets Encyclopaedia[15]

Before moving into our other main data set, it will become useful to the reader to get a sense of a sample of both 1) the preliminary Planetary Statistics 2) the actual Planetary Statistics. The following tables aim at providing this visual aid.

### Sample of the Preliminary Planetary Statistics

Here is a sample of what was included in this data set. This sample was going to be elaborated into the main data set, but it ended up being beyond the scope of this project. However, in this sample one can notice that it incorporates most of the data sets previously mentioned, with a specific highlight to findings in the Debiased Population Statistic papers we just learned about. It is important to note that whatever was left blank does not correspond to zero, it simply has not been determined yet.

**Preliminary Planetary Statistics Sample**

| **Missions** | Population Sample | Number of Super Earths | Period | Radii (EarthR) | Stars Number | Star Type | Potentially Habitable Transiting Earth size planet distance(pc) | Potentially Habitable Non-Transiting Earth-size planet distance(pc) |
|---|---|---|---|---|---|---|---|---|
| *Kepler*[11] | 2.5± 0.2 | | >200 days | 1-4 | | M dwarf | | |
| *Kepler (Habitable Zone)* [11] | | 0.12 | >200 days | | | M dwarf HZ | | |
| *Kepler (broad insolation boundaries of the recent Venus and early Mars)* | | 0.21 | >200 days | | | | 10.6 | 2.6 |
| *Kepler*[27] | 603 | | >200 days | | 42557 | GK | | |
| *K2* | | | | | | | | |
| *TESS* [23] | 6 | | one | 0.91-1.31 | | | | |
| *TESS*[28] | 0.61 terrestrial planets | | 0.4-7 days | >0.5 | | | | |

Table 3.0.1: Table to portray the Preliminary Planetary Statistics

## Sample of the Planetary Statistics Sample

Now is time to show you a glance into the actual data set that was imported into the algorithm. Please note this is simply a sample, since the main data set has around 1000 planets. The goal of this table is to simply give you some idea of the data the algorithm is taking in. Again, whatever is left blank, has simply not been assigned, and it is part of future work.

**Planetary Statistics Data Set Sample**

| **Planet Mass** | **Planet Radius** | **Star Name** | **Star Distance (pc)** | **Star Mass** | **Star Radius** | **Star Type** | **Has Atmosphere** |
|---|---|---|---|---|---|---|---|
| 0.00315 | 0.0892 | Sun | 0.00000484 | 1 | 1 | G | YES |
| 0.00017 | 0.0341 | Sun | 0.00000484 | 1 | 1 | G | YES |
| 0.00256 | 0.0846 | Sun | 0.00000484 | 1 | 1 | G | YES |
| 0.00034 | 0.0474 | Sun | 0.00000484 | 1 | 1 | G | YES |
| 1 | 1 | Sun | 0.00000484 | 1 | 1 | G | YES |
| 0.47 | 1.9 | 51 Peg | 14.7 | 1.11 | 1.27 | G2 IV | |
| 0.84 | | 55 Cnc | 12.34 | 1.015 | 0.98 | K0 IV-V | |
| 0.027 | 0.174 | 55 Cnc | 12.34 | 1.015 | 0.98 | K0 IV-V | |
| 0.0142 | 0.141 | TOI-469 | 68.192 | 0.91 | 1.0 | | |
| 0.116 | 0.457 | AU Mic | 9.79 | 0.5 | 0.75 | M1 V | |
| 0.101 | 0.28 | AU Mic | 9.79 | 0.5 | 0.75 | M1 V | |
| 0.0032 | | AU Mic | 9.79 | 0.5 | 0.75 | M1 V | |
| 0.07 | 0.23 | BD+20 594 | 180.39 | 1.67 | 1.08 | G | |

Table 3.0.2: Table to portray a sample of the actual Planetary Statistics

### *3.0.5 Molecular Data Set*

This next data set is comprised mainly of two sets of data which we will explore more in detail: *Biosignature Statistics* and *Bond Dissociation Energy Statistics.* This data set was important to the project since, once we know the possible locations for life, how are we going to determine if life resides there or not. This data set, has as its goal to input into the algorithm both the known biossignatures (RASCALL) [33] and even more information about how likely they are to be destroyed: Bond Dissociation Energy. To create this data set, the help of both Professor LaFratta, Professor Sousa-Silva and Maria Inês Alves Ferreira. Let us analyze each section of it.

- **Biosignature Statistics:** Once there is a grasp of the “possible locations for life” which as we recently saw came from our Planetary Population Data Set, the next step is matching that to the list of biosignatures (RASCALL) we could look for in those locations. Additionally, a goal is to infer, in the locations of the estimated planets, what could be a possible, likely or even unlikely bio signature to be present? What does the presence or lack of certain biosignatures mean? These questions will also be explored in the following chapter. For now let us return to the first part of this data set, Molecular Data Set, has a twofold goal. On one hand, the first part containing, Bio-Signature Statistics, is simply to input into the algorithm all the molecules that count as biosignatures, 16,367 molecules [33], with specific aspects to them such as: Molecular Formula, Molecular Weight, SMILES name (molecular name that is friendly to the program we are using: Jupiter Notebook).

As mentioned, RASCALL includes all 16,367 biosignatures and can be found here. [33]

### Sample of the Biosignature Statistics

Here is a glimpse into the actual Biosignature data set that was imported into the algorithm. Again, please note this is simply a sample, since the main data set has around 1600 molecules. Note:"SMILES" and "rdkit SMILES" correspond to a notation for RASCALL.

**Biosignature Statistics Sample**

| Molecular Formula | rdkit_SMILES | SMILES | Chemical Name | Molecular Weight (g/mol) |
|---|---|---|---|---|
| H2 | [H][H] | [H][H] | dihydrogen | 2.016 |
| CH4 | C | C | methane | 16.043 |
| H3N | N | N | ammonia | 17.031 |
| H2O | O | O | water | 18.015 |
| C2H2 | C# | C# C | ethyne | 26.038 |
| CHN | C# N | C# N | hydrogen cyanide | 27.026 |
| N2 | N# N | N# N | dinitrogen | 28.014 |
| C2H4 | C=C | C=C | ethene | 28.054 |
| CH2O | C=O | C=O | formaldehyde | 30.026 |
| C2H6 | CC | CC | ethane | 30.07 |
| HNO | [N]=O | [N]=O | nitrosyl hydride | 31.014 |
| CH5N | CN | NC | methanamine | 31.058 |
| O2 | O=O | O=O | dioxygen | 31.998 |
| CH4O | CO | OC | methanol | 32.042 |
| H4N2 | NN | NN | hydrazine | 32.046 |
| H3NO | NO | NO | hydroxylamine | 33.03 |
| H3P | P | P | phosphane | 33.998 |
| H2O2 | OO | OO | hydrogen peroxide | 34.014 |
| CH3F | CF | C(F) | fluoromethane | 34.033 |
| CH3F | S | S | hydrogen sulfide | 34.076 |

Table 3.0.3: Table to portray a sample of the actual Biosignature Statistics

- **Bond Dissociation Energy Statistics:** The second part of this data set, has as its goal to then assign a certain energy, "Bond Dissociation Energy" to molecules in RASCALL. This was mainly achieved with the help of Professor Christopher LaFratta in the Chemistry Department at Bard College. With his help, my advisor, Professor Clara Sousa-Silva and myself were able to assign a bond dissociation energy to molecules in RASCALL.

  Before moving forward, it is important to ask, what is bond dissociation energy? According to the CRC Handbook of Chemistry and Physics 95th Edition[38], The Dissociation energy (De)* for a diatomic molecule, has to do with "the difference between the energies of the free atoms at rest and the minimum in the potential energy curve. The term bond dissociation energy (BDE), which can be applied to polyatomic molecules as well, is used for the difference between the energies of the fragments resulting when a bond is broken and the energy of the original molecule in its lowest energy state."[38] Additionally, the term bond strength implies differences in enthalpy rather than energy.

  In other words, it is the amount of energy needed to break apart one mole of covalently bonded gases into a pair of radicals. The SI units used to describe bond energy are kiloJoules per mole of bonds (kJ/Mol). It indicates how strongly the atoms are bonded to each other.[38]

  What was done with this energy was that the book, CRC Handbook of Chemistry and Physics 95th Edition[38] [38] provided us with a certain "Bond Dissociation Energy" for specific bonds which corresponded to specific functional groups which then corresponded to specific molecules. What we did was, with the immense help of RASCALL [33], assign the energy that a certain bond has, to a bio signature in RASCALL that also has that bond. This would then allow for, later on, in the algorithm to set constraints to this list and obtain (based on the constraints) a more narrowed list of bio signatures based on how likely they are to survive, how weak their bonds are, based on the corresponding bond dissociation energy.

Each bond dissociation energy was taken from: The CRC Handbook of Chemistry and Physics 95th Edition[38]ñ The full Molecular Data Set then includes both the Bio-Signature Statistics and the Bond Dissociation Energy Statistics, and they are the inputs for the algorithm.

Now, a sample of these statistics will be shown.

### Sample of the Bond Dissociation Energy Statistics

This sample portrays a few examples of the Bond Dissociation Energy Statistics. This was a laborious process involving RASCALL [33] and the CRC Handbook of Chemistry and Physics 95th Edition
citeCRC. This sample table portrays the lowest Bond Dissociation Energies that were assigned to Molecules in RASCALL (Biosignatures) with the same functional groups as their corresponding bonds.

Again, please note this is simply a sample, since the main data set has around 1600 molecules.

Table 3.0.4: Bond Dissociation Energy Statistics Sample

| **Type Of Bond** | **Bond** | **BDE of bond** | **Molecules with bond** | **Molecules in RASCALL with bond** |
|---|---|---|---|---|
| CH CH double bond | CH=CH | 464.8 | CH3CH=CH-H/ (C3H6) | CC=C, C[H], [H]C([H])=C([H])[!#1], C=C, [H]C([!#1])=C([H])[H], [H]C([H])([H])[!#1] ... |
| BrC C single bond | BrC-C | 356.9 ± 12.6 | Br3C-CH3/ (C2H3Br3) | C[H], CBr, [H]C([H])([H])[!#1] ... |
| Cl C single bond | Cl-C | 422.6 ± 8.4 | Cl-CN/ (CClN) | CCl, C#N ... |
| HO O single bond | HO-O | 169.9 ± 2.1 | HO-OC(O)CH3/ (C2H4O3) | [H]OC(C)=O, [H]OC([H])([H])[!#1], C[H], [H]OC(=O)C, [H]C([H])([!#1])[!#1], [H]OC, OC(=O)C, CC(O)=O ... |
| CHO O single bond | CHO-O | 167.4 ± 6.3 | CH3O-OCH3/ (C2H6O2) | [H]OC([H])([H])[!#1], C[H], [H]C([H])([!#1])[!#1], [H]OC, [H]C([H])([H])[!#1], [H]OC([H])([H])[!#1], C[H], [H]C([H])([H])[!#1], C[H], COC, [H]OC, [H]C([H])([H])[!#1] ... |
| CH O single bond | CH-O | 351.9 ± 4.2 | CH3-OCH3/ (C2H6O) | [H]OC([H])([H])[!#1], C[H], [H]C([H])([!#1])[!#1], [H]OC, [H]C([H])([H])[!#1], C[H], COC ... |

Table 3.0.4: Table to portray a sample of the Bond Dissociation Energy Statistics

**From both**: It is important to note that many of the planets might not be able to exhibit the presence of bio signatures until much later. Hence, adapting the goal of displaying the concerns and problems in detecting unambiguous signs of life (and uncertainty) of different types of planets, might be more achievable. For example, Earth-like planets that orbit Sun-like stars or super-Earths around M-dwarfs. Ultimately aiming at making a conceptual connection to infer assumptions about: size, host star, and linking that information to what we know about the biosignatures (RASCALL). The next step after is, as we have seen in Part II of the Introduction, the creation of the algorithm that takes in the first set of data, Planet Population Data Set ("nearby opportunities for life") and all the information that can be inferred about the biosignatures (molecular weight, bond disassociation energy), in the second data set, Molecular Data Set, and output a constrained list of potential locations for life of possibilities about what biosignatures on the estimated planets could portray a sign for life.

## 3.1 Oxygen: Is it necessary for life?

Since in this project we are dealing with a key aspect in the search for life, which is, what would *determine* life: bisosignatures, and much of them were mentioned in this chapter, it becomes imperative to pause for a moment and consider the most obvious one: oxygen. Is oxygen necessary for life? What makes it a good (or bad) biosignature? What would aliens think of Oxygen? This section briefly explores these questions. We will be pulling from a few sources, mainly from Meadows et al. in the paper "Exoplanet biosignatures: understanding oxygen as a biosignature in the context of its environment"[22] [30][34]

As we have mentioned, “biosignatures” are one of the most important aspects in the search for life outside our planet. They are life’s fingerprint on the atmosphere or surface of a planet. This search is a scientific challenge that pushes the boundaries of observational astronomy and requires careful consideration of the signs of life that we will best be able to detect, in other words, it requires careful considerations of the biosignatures.[22]

According to Meadows et al., “the most useful biosignatures meet three criteria: reliability, survivability, and detectability, which together enhance the probability that the biosignature can be detected and interpreted as being due to life.” [22] Together, they enhance the probability that the biosignatures can be detected and interpreted as being due to life. Reliability: Is it reliable? Answer whether an observed feature of the planet has a biological origin. Is it more likely to be produced by life or by planetary processes like geology or photochemistry? Survivability: Can it survive? determines if the biosignature can avoid the normal sinks in planetary environment and build up to detectable levels. What are the sinks? These sinks include, but are not limited to, destruction by photolysis and photochemistry, reactions with volcanic gases and the surface, and (for soluble gases) dissolution in the ocean. Detectability: Can we detect it? Is it spectrally active? Is it clear of overlap with other chemical species in the observed wavelength region? Is it accessible in high enough abundance and at appropriate regions in the planetary atmosphere to be observed by techniques like transmission, secondary eclipse, phase curves, and direct imaging.[22]

Meadow notes: “With these three criteria, Earth’s abundant molecular oxygen (O2) has been identified as the strongest biosignature for terrestrial planets and was also initially thought to be straightforward to interpret as having a biological origin."[22] Is that it? Is oxygen the best sign of life we could find?

As optimistic as this might sound, as a biosignature, O2 faces two major challenges: (1) it was only present in high abundance for a relatively short period of Earth's history and (2) we now know of several potential planetary mechanisms that can generate abundant O2 without life being present.[22] Therefore if alien astrobiologists were to rely on finding us just using oxygen as a sign of life, they could have missed life altogether.

To understand oxygen better, we need to understand the environmental context.[22] For this, it becomes important to look back in time and examine the "coevolution of life with the early Earth's environment to identify how the interplay of sources and sinks may have suppressed O2 release into the atmosphere for several billion years, producing a false negative for biologically generated O2." [22]

Regarding false positives, recent research indicates that there are several mechanisms that could produce abiotic O2 and O3 in a planet's atmosphere, with each presenting a potential false positive to different degrees.meadows2018exoplanet

Thinking about Oxygen forces upon us again the question of thinking about uncertainty. How can we understand uncertainty in biosignatures? What about the False Positives?

As we just saw, some papers [22] [30][34] comment on how oxygen is the strongest biosignature for terrestrial planets. What are the implications of this claim? Does this mean that given the fact we do find oxygen in an atmosphere of a planet, does it immediately mean there is life? The certainty of such claims matter. What are the exact parameters for life? This one is definitely a tricky and complex question, and one for which the answer might need a philosophical approach. If we are able to infer from life on Earth certain parameters, then we could take the search for life further. Thinking about life on Earth makes us think that any planet which would have life has to have, at least: an atmosphere and some form of chemistry. How does oxygen help us? Well, noting the presence of oxygen is not merely enough for establishing life, therefore other inferences and other molecules might be necessary.

# 4

# Phases of the Project P.U.S.H.: Program for the Uncertainty in the Search for Habitability

> "What an abyss of uncertainty, whenever the mind feels overtaken by itself; when it, the seeker, is at the same time the dark region through which it must go seeking and where all its equipment will avail it nothing. Seek? More than that: create."
>
> *-Marcel Proust*
> *In Search of Lost Time: Swann's Way*

Now that the Data Sets have been laid out and hopefully understood, it is time to immerse ourselves deeply into the inner working of the project: the algorithm. Now is time to shift our language from simply using "the algorithm" to its actual name, **PUSH: Program for the Uncertainty in the Search for Habitability.** In this chapter, we will not only introduce but explore PUSH at its fullest to see what it can actually do and more so; what its outputs mean physically and philosophically.

Here you have it: the main layout of PUSH.

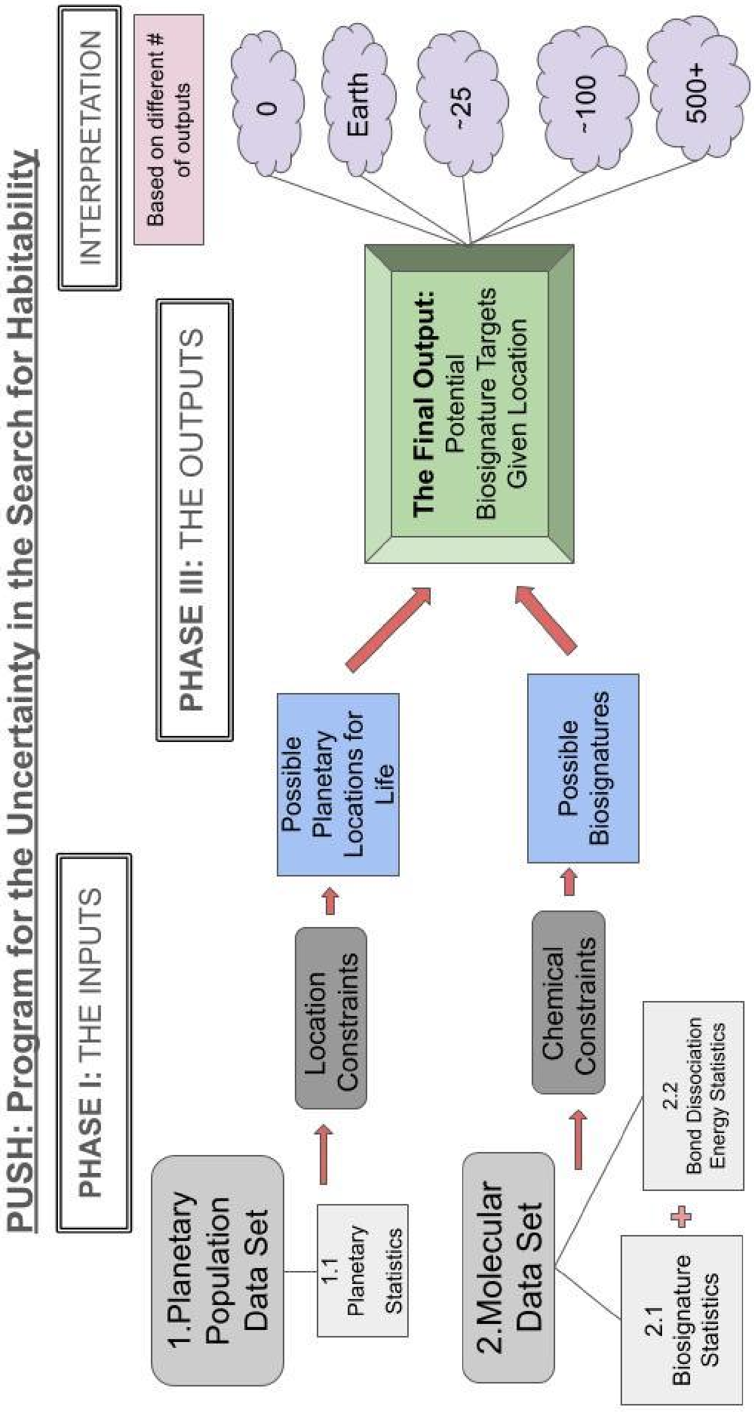


Figure 4.0.1: Structure of PUSH

## 4.1 Phase I: The Inputs of PUSH

Let us start by the inputs of PUSH. As you might be very anxious to meet and explore PUSH, this section will be short. Most has been put forward and seen in detail in the previous chapter, "Data Sets", 3 hence we will keep it brief.

PUSH has two main data sets from where its information comes in:

- Planetary Population Data Set (for more information return to Chapter 3 3)
- Molecular Data Set (for more information return to Chapter 3 3) which is in itself comprised of two other data sets:
  - Biosignature Statistics
  - Bond Dissociation Energy Statistics

It is important to note that these inputs are inserted into PUSH as "csv" files, therefore as PUSH is highly adaptable, they can be modified and re-entered.

## 4.2 Phase II: PUSH

It is finally time to get to know P.U.S.H.: (Program for the Uncertainty in the Search for Habitability). Welcome! Although its goal might sound simple, its construction was definitely not. PUSH was built using Anaconda Jupiter Notebook and with a lot of help. Special acknowledgements go to: Luke Ingraham, Yashar Khan, Farman Sayem, Clem Tarpey, and Maria Inês Ferreira The main goal of PUSH, as mentioned multiple times before, might sound straight forward: provide the user with *Possible Planetary Locations for Life*. However, its inputs (as we have seen) and its outputs (as we will learn) are not only scientifically challenging, but additionally, philosophically intricate. When the user operates PUSH, they will be faced with a challenge: *setting the constraints.* PUSH takes in the mentioned inputs: Planetary Population Data Set and Molecular Data Set, and based on those asks the user to provide a list of constraints in order to supply the outputs.

What are these constraints? Let's explore them!

### 4.2.1 The Location Constraints

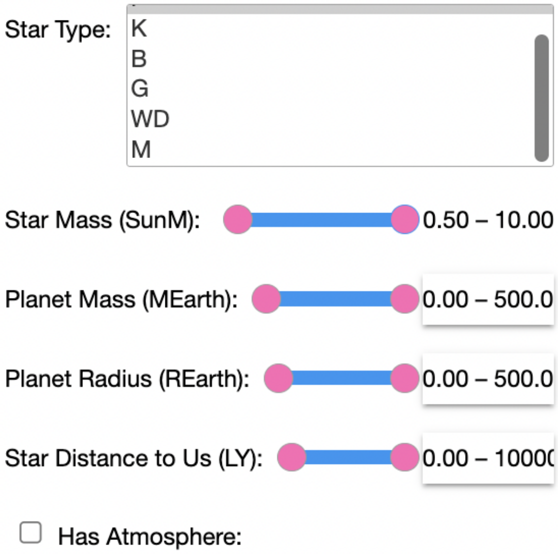
The Location Constraints are imposed on the first data set PUSH takes in: Planetary Population Data Set. These constraints can be tuned by the users as many times as they wish. They include the following:

- **Star Type**

  K, B, G, WD (White Dwarf), M. These types of stars come from the stellar classification which includes seven types or classes of stars: O, B, A, F, G, K and M. Stars in the 'O' class are the most massive and hottest, stars in the 'M' class are the smallest and coolest. For a more visual interpretation of the star types, see Figure 4.2.2. [6]

- **Star Mass (SunM)** ranges from 0 up to 10 sun masses.
- **Planet Mass (EarthM)** ranges from 0 up to 500 Earth masses.
- **Planet Radius (REarth)** ranges from 0 up to 500 Earth radii.
- **Star Distance to Us (LY)** ranges from 0 up to 10,000 light years
- **Has Known Atmosphere** this is mainly a constraint for a future version of PUSH. But, since we know of a planet that definitely has an atmosphere, Earth, it was important to acknowledge it.

A sample of the Location Constraints taken directly from PUSH looks like this:



Star Type:
K
B
G
WD
M

Star Mass (SunM): 0.50 – 10.00

Planet Mass (MEarth): 0.00 – 500.0

Planet Radius (REarth): 0.00 – 500.0

Star Distance to Us (LY): 0.00 – 1000(

☐ Has Atmosphere:

Figure 4.2.1: PUSH Location Constraints

This list of constraints can be tuned as many times as the user wishes, as it is refreshed every time the constraints are re-applied.

### *4.2.2 The Chemical Constraints*

Now is time to consider the second list of constraints PUSH takes in. These are applied to the second data set, Molecular Data Set, and include:

- Molecular Weight (g/mol)
- How easy are the molecules to be destroyed? This is based on the lowest Bond Dissociation Energy (BDE) assigned. Therefore, once the user selects the range, PUSH will filter through the data set and output the selected range. If two molecules were to have different (BDEs) PUSH automatically selects the lowest one.

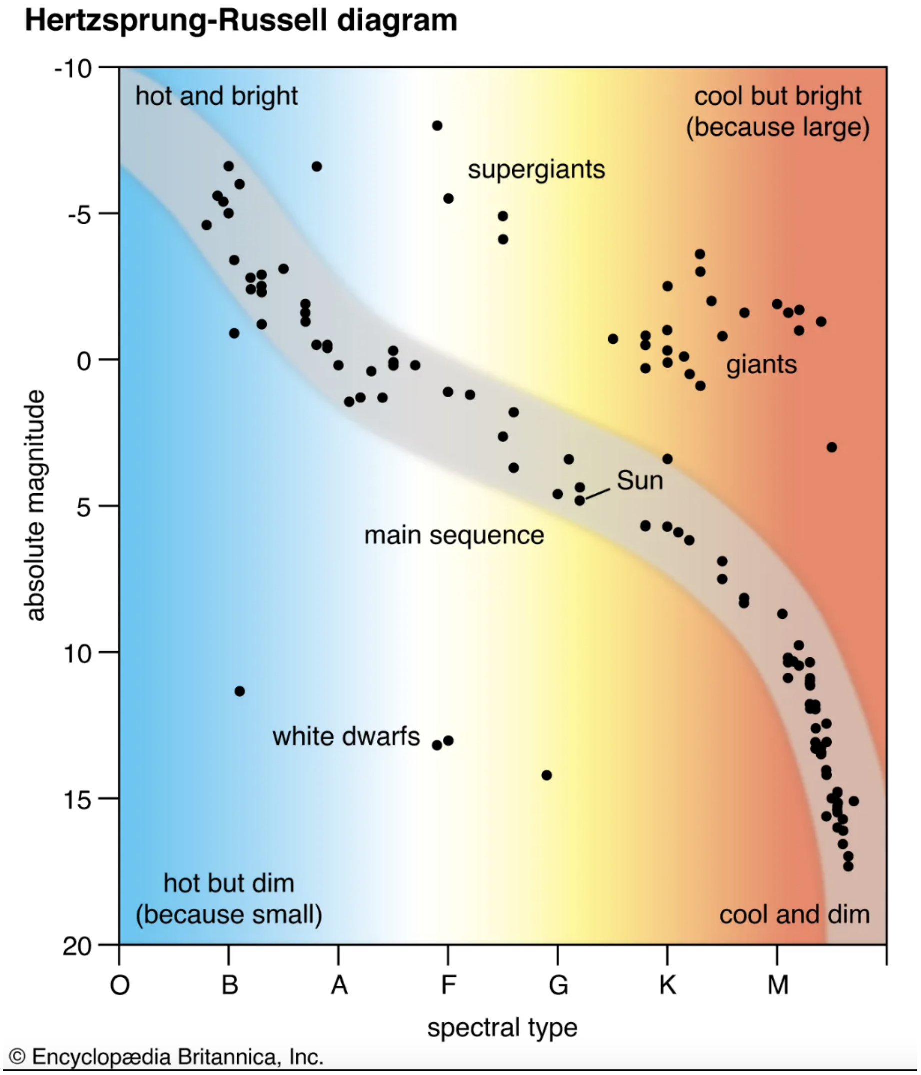


Figure 4.2.2: Hertzsprung-Russell diagram for Stellar Classification. Credit: Encyclopædia Britannica, Inc.

More constraints can be added to this list, specially as PUSH grows this will become essential.

These constraints look like this in PUSH:

Molecular Weight: 2.0 – 52.6

Dissociation Energy Range: 200.00 – 40

Figure 4.2.3: PUSH Chemical Constraints

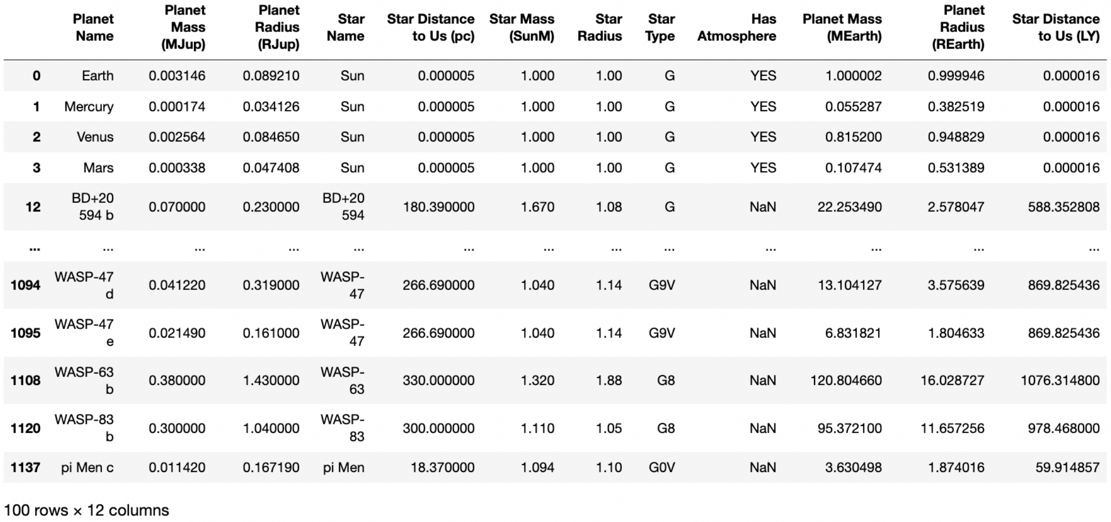
Again, these constraints, can be set by the user as many times as they wish, refreshing automatically. Both these lists are crucial for this project, since here is where we are actually *doing* something to the data. A future PUSH, would additionally rank the outputs based on these constraints. Talking of which, what are the outputs? Let us introduce them.

## 4.3 Phase III: The Outputs

The much anticipated moment is here! You arrived to the actual results. By this stage, PUSH has successfully been utilized by the user and would have portrayed outputs. Let's check them out!

Once the user has set both Location and Chemical Constraints, PUSH will have ready both a "Possible Planetary Locations for Life" and a "Possible Biosignatures".

How would this look like? It largely depends on the list of constraints, of which we'll see some examples further on, but for now, a rough Planetary Output would mostly look like:



| | Planet Name | Planet Mass (MJup) | Planet Radius (RJup) | Star Name | Star Distance to Us (pc) | Star Mass (SunM) | Star Radius | Star Type | Has Atmosphere | Planet Mass (MEarth) | Planet Radius (REarth) | Star Distance to Us (LY) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | Earth | 0.003146 | 0.089210 | Sun | 0.000005 | 1.000 | 1.00 | G | YES | 1.000002 | 0.999946 | 0.000016 |
| **1** | Mercury | 0.000174 | 0.034126 | Sun | 0.000005 | 1.000 | 1.00 | G | YES | 0.055287 | 0.382519 | 0.000016 |
| **2** | Venus | 0.002564 | 0.084650 | Sun | 0.000005 | 1.000 | 1.00 | G | YES | 0.815200 | 0.948829 | 0.000016 |
| **3** | Mars | 0.000338 | 0.047408 | Sun | 0.000005 | 1.000 | 1.00 | G | YES | 0.107474 | 0.531389 | 0.000016 |
| **12** | BD+20 594 b | 0.070000 | 0.230000 | BD+20 594 | 180.390000 | 1.670 | 1.08 | G | NaN | 22.253490 | 2.578047 | 588.352808 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **1094** | WASP-47 d | 0.041220 | 0.319000 | WASP-47 | 266.690000 | 1.040 | 1.14 | G9V | NaN | 13.104127 | 3.575639 | 869.825436 |
| **1095** | WASP-47 e | 0.021490 | 0.161000 | WASP-47 | 266.690000 | 1.040 | 1.14 | G9V | NaN | 6.831821 | 1.804633 | 869.825436 |
| **1108** | WASP-63 b | 0.380000 | 1.430000 | WASP-63 | 330.000000 | 1.320 | 1.88 | G8 | NaN | 120.804660 | 16.028727 | 1076.314800 |
| **1120** | WASP-83 b | 0.300000 | 1.040000 | WASP-83 | 300.000000 | 1.110 | 1.05 | G8 | NaN | 95.372100 | 11.657256 | 978.468000 |
| **1137** | pi Men c | 0.011420 | 0.167190 | pi Men | 18.370000 | 1.094 | 1.10 | G0V | NaN | 3.630498 | 1.874016 | 59.914857 |

100 rows × 12 columns

Figure 4.3.1: Planetary and Stellar Outputs:Possible Planetary Locations for Life. Credit: PUSH

### 4.3.1 The Planetary and Stellar Outputs: Possible Planetary Locations for Life

Additionally, alongside this output, there will be provided a philosophical reflection/interpretation of what precisely your constrained list means. For the Planetary and Stellar List of Constraints the user chooses, these are :

### 4.3.2 Interpretation for the Planetary and Stellar Outputs:Possible Planetary Locations for Life

:

- If based on the user's constraints, PUSH has no outputs. The following paragraph is displayed: "It seems like you missed life altogether :( However, one thing we are certain of is that there is one place: "Earth". Your constraints are too limited. Don't be afraid! We know life is clever and resourceful, so open the possibilities. Try running a less conservative search. It's okay to have more data, don't be scared of it! Science works with uncertainty! Try again!

- If based on the user's constraints, PUSH has only Earth as its output. The following paragraph is displayed: "You were right! You found the place we are sure life exists! Safe answer! Consider expanding your search, to allow for more possibilities, be willing to explore the galaxy to see what it can offer! Do not be afraid of allowing for more data in, even if it carries uncertainty, it is needed in order to have more locations and ultimately find life! Even if you think you are straying from the truth, it's okay if you're not sure, uncertainty is essential in such explorations. Try again, explore our galaxy!"

- If based on the user's constraints, PUSH has around 25 outputs. The following paragraph is displayed: "You only got 25 potential possibilities for life! You might have missed life. However, you have the same ideas as NASA, (Decadal Survey) your output portrayed around what NASA considers a good sample of locations. The ideal location for life is unknown, and it is why the search for it is filled with uncertainty, therefore it might be useful to expand your search to allow for more opportunities of life. Try running your constraints again, maybe broadening your scope. Know that having more possibilities can serve useful in the search for life."

- If based on the user's constraints, PUSH has around 100 outputs. The following paragraph is displayed: " You got around 100 potential possibilities for life! You know what that means? You might be on the right track! Chances are one of them might actually host life. Your sample is large enough to have statistical significance.

You would be able to have a good sample for the frequency of life. Although, life is rare and might not be among your outputs. For this, further analysis has to be done, but you have allowed a narrowed enough and large enough search for habitability! Good job! Keep in mind, the quest for life is one that carries a lot of doubts, hence these locations are a good place to start. I invite you to analyze them further! For a more specific search, set your constraints again."

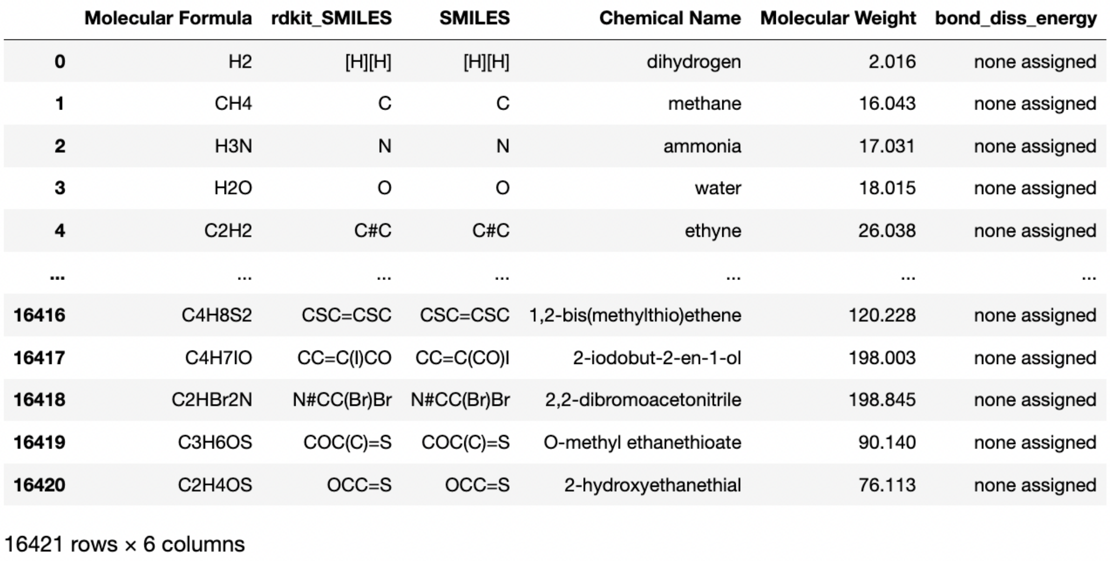
- If based on the user's constraints, PUSH has 500 or more outputs. The following paragraph is displayed: "Congratulations! Good news is: perhaps one of these locations might host life. Bad news: the pool you picked is a little too large, and we might miss it altogether. You have allowed for too many locations! Life could be present in one of them, but your search is too broad. There's too much uncertainty involved. Recommendation: try being more specific with your constraints! The search for life is not an easy task, but we have to start somewhere. Try again!"

These paragraphs are simply interpretations of the results PUSH will output. Anyone using PUSH (students to young children to astrobiologists) should be able to extrapolate something from it, and should feel that they can contribute to the search for life, because they are. All the way from an astrobiologist looking for a specific exoplanet to a student who wants to learn more. PUSH, and in particular its interpretations, should aid the user to obtain something from it.

Now is time to shift based on the Chemical Constraints, PUSH will output something like:

### *4.3.3 The Molecular Outputs: a constrained list of biosignatures*

Also, as with the Planetary and Stellar Outputs, PUSH will provide a philosophical reflection/interpretation of what precisely your constrained list means. For the Chemical Constraints the user chooses, these are :



| | Molecular Formula | rdkit_SMILES | SMILES | Chemical Name | Molecular Weight | bond_diss_energy |
|---|---|---|---|---|---|---|
| **0** | H2 | [H][H] | [H][H] | dihydrogen | 2.016 | none assigned |
| **1** | CH4 | C | C | methane | 16.043 | none assigned |
| **2** | H3N | N | N | ammonia | 17.031 | none assigned |
| **3** | H2O | O | O | water | 18.015 | none assigned |
| **4** | C2H2 | C#C | C#C | ethyne | 26.038 | none assigned |
| **...** | ... | ... | ... | ... | ... | ... |
| **16416** | C4H8S2 | CSC=CSC | CSC=CSC | 1,2-bis(methylthio)ethene | 120.228 | none assigned |
| **16417** | C4H7IO | CC=C(I)CO | CC=C(CO)I | 2-iodobut-2-en-1-ol | 198.003 | none assigned |
| **16418** | C2HBr2N | N#CC(Br)Br | N#CC(Br)Br | 2,2-dibromoacetonitrile | 198.845 | none assigned |
| **16419** | C3H6OS | COC(C)=S | COC(C)=S | O-methyl ethanethioate | 90.140 | none assigned |
| **16420** | C2H4OS | OCC=S | OCC=S | 2-hydroxyethanethial | 76.113 | none assigned |

16421 rows × 6 columns

Figure 4.3.2: Sample of Possible Biosignatures

Note: The column for Bond Dissociation Energy portrays the result “None Assigned”. This is not due to no energy being attributed from the molecules, this is mainly due to the fact that RASCALL is such a big data set that most of the molecules do have a BDE, but this sample does not portray all of them. The goal for future work would be to closely analyze the bonds in the molecules in RASCALL and assign an energy to all of them.

### 4.3.4 Interpretation for the Possible Biosignatures

: As with the Possible Planetary Locations for Life interpretations, a corresponding one appears when the user sets the Chemical Constraints.

- Based on the specific constraints the user chooses, alongside with the output, PUSH will include the following paragraph: 'You found (#) and (future percentage) potential successful biosignatures. How often they are found comes from a variety of parameters.'Based on your constraints you narrowed the search, therefore you made a more targeted search for life! Good job! You can always run your constraints again to see what you can find!'

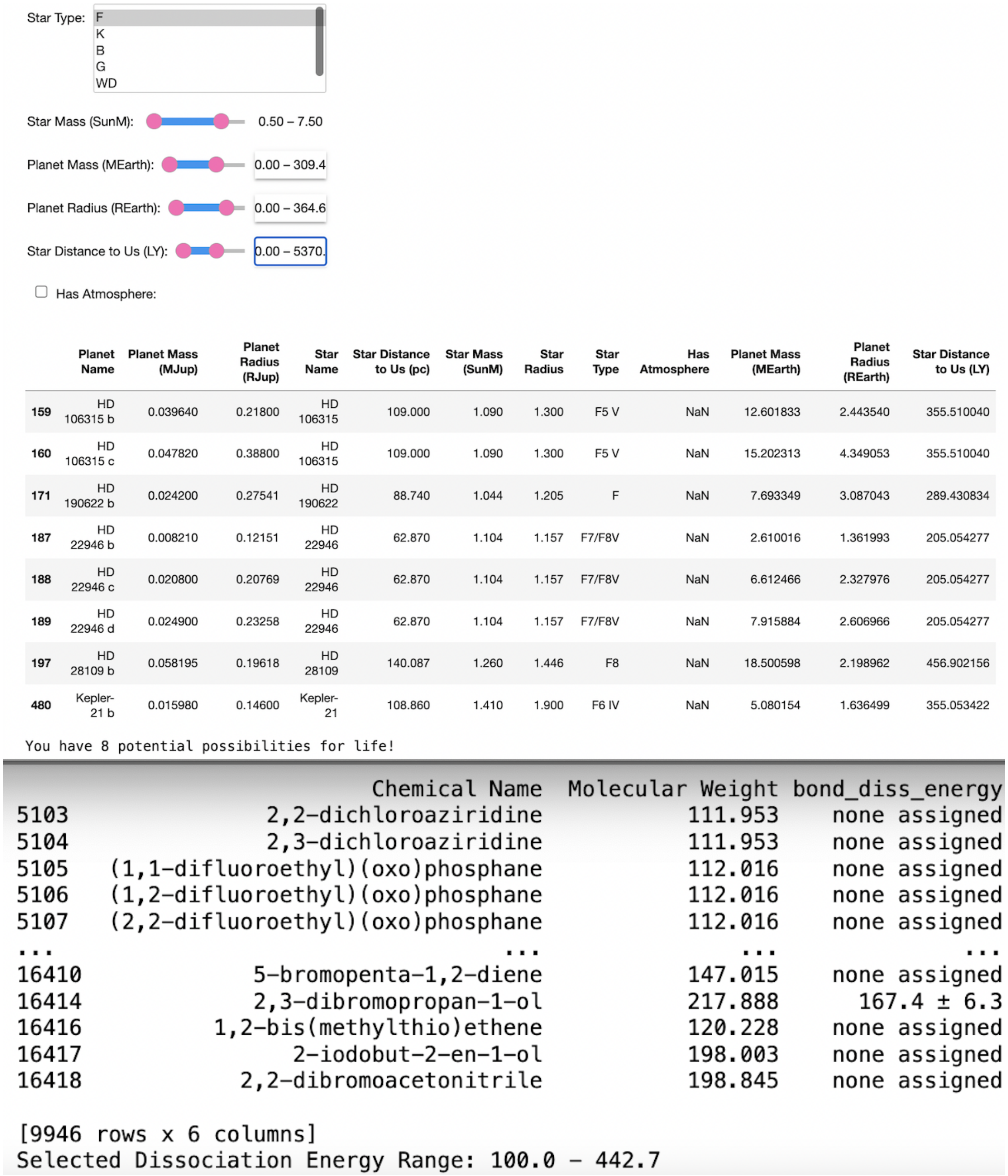
Given that we now have a pretty good idea of what PUSH inputs and outputs, it is time to transfer to the final output, which is nothing more than both of the ones we have seen combined.

### 4.3.5 The Final Output: Potential Biosignature Targets Given Location

The final output is here. This last output would correspond to both Possible Planetary Locations for Life and the Possible Biosignatures. What does this look like? The best way to figure this out is to discover it for yourself by using PUSH. If you are unable to do so at the moment, here is a sample final output:

And the corresponding interpretation:

This output is provided based on the constraints we can see in Figure 4.3.3: “F type stars, with star masses ranging from 0.5 to 7.5 Sun masses, with a planet mass ranging from 0 to 309.4 Earth masses,0 to 364.6 Earth Radii, with a distance constraint of a range between 0 to 5370 light years away. This outputs exactly 8 potential locations for life. Additionally,



Star Type:
F
K
B
G
WD

Star Mass (SunM): 0.50 – 7.50

Planet Mass (MEarth): 0.00 – 309.4

Planet Radius (REarth): 0.00 – 364.6

Star Distance to Us (LY): 0.00 – 5370.

☐ Has Atmosphere:

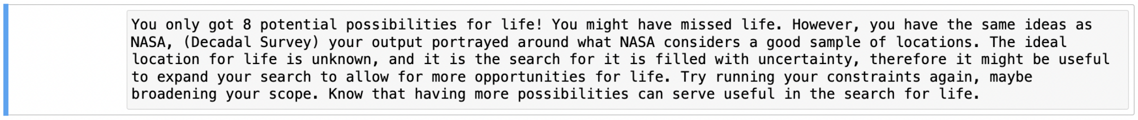
| | Planet Name | Planet Mass (MJup) | Planet Radius (RJup) | Star Name | Star Distance to Us (pc) | Star Mass (SunM) | Star Radius | Star Type | Has Atmosphere | Planet Mass (MEarth) | Planet Radius (REarth) | Star Distance to Us (LY) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **159** | HD 106315 b | 0.039640 | 0.21800 | HD 106315 | 109.000 | 1.090 | 1.300 | F5 V | NaN | 12.601833 | 2.443540 | 355.510040 |
| **160** | HD 106315 c | 0.047820 | 0.38800 | HD 106315 | 109.000 | 1.090 | 1.300 | F5 V | NaN | 15.202313 | 4.349053 | 355.510040 |
| **171** | HD 190622 b | 0.024200 | 0.27541 | HD 190622 | 88.740 | 1.044 | 1.205 | F | NaN | 7.693349 | 3.087043 | 289.430834 |
| **187** | HD 22946 b | 0.008210 | 0.12151 | HD 22946 | 62.870 | 1.104 | 1.157 | F7/F8V | NaN | 2.610016 | 1.361993 | 205.054277 |
| **188** | HD 22946 c | 0.020800 | 0.20769 | HD 22946 | 62.870 | 1.104 | 1.157 | F7/F8V | NaN | 6.612466 | 2.327976 | 205.054277 |
| **189** | HD 22946 d | 0.024900 | 0.23258 | HD 22946 | 62.870 | 1.104 | 1.157 | F7/F8V | NaN | 7.915884 | 2.606966 | 205.054277 |
| **197** | HD 28109 b | 0.058195 | 0.19618 | HD 28109 | 140.087 | 1.260 | 1.446 | F8 | NaN | 18.500598 | 2.198962 | 456.902156 |
| **480** | Kepler-21 b | 0.015980 | 0.14600 | Kepler-21 | 108.860 | 1.410 | 1.900 | F6 IV | NaN | 5.080154 | 1.636499 | 355.053422 |

```
You have 8 potential possibilities for life!
```

```
                    Chemical Name  Molecular Weight bond_diss_energy
5103          2,2-dichloroaziridine         111.953    none assigned
5104          2,3-dichloroaziridine         111.953    none assigned
5105   (1,1-difluoroethyl)(oxo)phosphane         112.016    none assigned
5106   (1,2-difluoroethyl)(oxo)phosphane         112.016    none assigned
5107   (2,2-difluoroethyl)(oxo)phosphane         112.016    none assigned
...                               ...             ...              ...
16410        5-bromopenta-1,2-diene         147.015    none assigned
16414        2,3-dibromopropan-1-ol         217.888      167.4 ± 6.3
16416    1,2-bis(methylthio)ethene         120.228    none assigned
16417           2-iodobut-2-en-1-ol         198.003    none assigned
16418      2,2-dibromoacetonitrile         198.845    none assigned

[9946 rows x 6 columns]
Selected Dissociation Energy Range: 100.0 - 442.7
```

Figure 4.3.3: Sample Final Output: Potential Biosignature Targets Given Location



```
You only got 8 potential possibilities for life! You might have missed life. However, you have the same ideas as
NASA, (Decadal Survey) your output portrayed around what NASA considers a good sample of locations. The ideal
location for life is unknown, and it is the search for it is filled with uncertainty, therefore it might be useful
to expand your search to allow for more opportunities for life. Try running your constraints again, maybe
broadening your scope. Know that having more possibilities can serve useful in the search for life.
```

Figure 4.3.4: Corresponding interpretation for output of 8 locations.

with a chemical constraint set: Selected Dissociation Energy Range: 100.0 to 442.7 g/mol. This portrays 2,3-dibromopropan-1-01 molecule, $C3H4Br2O$, with a molecular weight of 217,888 g/mol and a Bond Dissociation Energy of 167.4 ‡ 6.3 (kJ/mol)".

Additionally, it portrays the interpretation, as we can see in Figure 4.3.4.

Now, these are very rich and dense ideas that were just introduced. Getting a working algorithm in itself is a lot of work, and requires a lot of interpretation. We must step back and reflect. What does this algorithm mean? How close is it from the truth? What would the next steps (pushes) for PUSH be?

### *4.3.6 A Glimpse into the Future: Habitable Worlds Observatory and Decadal Survey*

As mentioned before, this is very much a first attempt of PUSH. It started with a very ambitious goal: a ranking for potential locations for life, and to be adapted. But that is the good news with PUSH. It is highly adaptable! Indeed, it is meant to be adapted in order for it to grow. For the scope of this project many things such as more ambitious data sets had to be adapted, but once those exist, PUSH can easily take them in. Moreover, we can make PUSH more advanced.

In the large branch of astrobiology, does PUSH portray any impacts? Does this project mean and contribute to something bigger?

For this, it is essential for the reader to get acquainted with The Habitable Worlds Observatory and the Decadal Survey.

The Habitable Worlds Observatory (HWO) is a large infrared/optical/ultraviolet space telescope recommended by the National Academies' Pathways to Discovery in Astronomy and Astrophysics for the 2020s. [25] HWO would be designed specifically to identify potentially habitable planets around other stars, closely examining their atmospheres to determine if life could possibly exist. It has as its main objective to identify and directly image at least 25 potentially habitable worlds. It would then use spectroscopy to search for chemical "biosignatures" in these planets' atmospheres, including gasses such as oxygen and methane, which could serve as critical evidence for life. The observatory would introduce new capabilities to study the universe with unprecedented sensitivity and resolution, giving us important new insights into the evolution of cosmic structures, including how galaxies form and develop over time. [25]

Specific emphasis needs to be given to "the mission's main objective is to identify and directly image at least 25 potentially habitable worlds." This is already what PUSH is able to do! Therefore, PUSH had a wider impact than this project.

Furthermore, to add to the bigger picture, it is important to introduce the Decadal Survey. This is a report prepared by the National Academy of Sciences, Engineering, and Medicine at the request of NASA every 10 years. The National Academies convenes a committee of leading planetary scientists to write the report, which gives their consensus opinion of the most important scientific questions facing the community and a priority list of missions to answer them.[1]

The current decadal survey is Origins, Worlds, and Life: A Decadal Strategy for Planetary Science and Astrobiology 2023-2032. This projects already contributes to this goal.

It has as its recommendation a balanced program of solar system exploration. No one program or destination should grow too large at the cost of others. [1]

## 4.4 Installation

Information on how to install PUSH can be found on the platform, where it is published publicly as a README file. To access this and try PUSH please go to GitHub. Here is the link for it:

https://github.com/jacintacreel/PUSH.git

# 5
# Analysis

> "Such is the new and perishable universe which has just been created."
>
> ---
>
> *-Marcel Proust*
> *In Search of Lost Time: The Guermantes Way*

To draw on some limitations, the boundaries of life can be drawn as follows (personal communication with my advisor Clara Sousa-Silva): life must exist somewhere, there must be a place. It must be able to interact with its environment. Here we are not assuming any perfect system, as there is no such thing (there is no 100 percent efficiency, will not interact perfectly) But by imposing this condition we are simply looking for something that will ultimately release something else and which we can look for and detect. At the end, we want:evidence of metabolism.

Imposing some constraints moves us further from the "forms" that life could take. So this is a risk, is it a risk worth taking? Here, are we willing to leave some possibilities out? Well, as long as it does allow some in. We want our search to be successful.

As mentioned before, another way in which this project tackles such a risk is that in order to portray how difficult the task of finding life beyond Earth is, a critical aspect might be its frequency: how much life is out there, is it just Earth? When we look for life, what are we looking for? What are the empirical considerations?

On the Drake equation(s) and missions, philosophically, why are equations, such as the Drake Equation and its variations less effective than say Debiased Population Statistics? What method (in this area of science and science in general) makes the best measurement? And more specifically, how can we determine it? The search for life comes with many risks.

Here I want to introduce two different ends for this problem. On one end we have what seems like a beautiful, obvious and pure approach: The Drake Equation. The reality of this equation(s) is that we will never be able to get these numbers, so it is useless in our search. On the other end, we have a more practical solution, 1) looking at planetary detections and their biases actually provides numbers, although this numbers carry uncertainties; 2) it simultaneously provides potential targets for looking for life, though not in a systematic or representative way. Taken to an extreme, in this approach we could look for everything and find ourselves lost with information, so we have it: a risk. My project tackles this and aims to find a middle ground. How much certainty are we willing to sacrifice in delimiting the uncertainty?

Regarding the data, as you might have imagined, here we must take a very necessary pause and consider: what about the uncertainty in all these values? Each value included in the data set table is either taken directly from the data that the mission provides or from assumptions based on that bias. Does the data become less accurate? Additionally, how can we interpret the exact uncertainty if the data is taken from different missions each with their own biases and uncertainties, is there a way to combine it? To swim in uncertainty is a tricky task, and finding certainty in all of it is crucial, that is why some limitations had to be implemented. Limitations such as the data inscribed and interpreted accounts only for the search of planets that "could have life" and those have a certain distance from their star, a certain size, thermal structure and such. Although messy, data is data and the numbers are justified from actual missions. Teasing the values and deciding how progress looks in data sets such as these is key, but at the same time some flexibility needs to be allowed to gather any interpretative data at all. One must gather what data is relevant and justify, and lie on the small questions that carry the uncertainty of each of the values in the data sets.

What has been presented thus far is a preliminary PUSH for the search for life. It is highly adaptable for any future efforts which could include but are not limited to the following: an updated Planetary Population Data Set that includes aspects such as atmospheric constraints, an updated Molecular Data Set with a revised Bond Dissociation Energy using RASCALL. The goal for PUSH is to grow and be further adapted. Until now, it serves us as an instrument to try it out, to think and to explore all of the questions that have been presented.

# 6
# Philosophical Implications Introduction

> "... it often happened that, in my brief spell of uncertainty as to where I was, I did not distinguish the various suppositions of which it was composed..."
>
> *-Marcel Proust*
> *In Search of Lost Time: Swann's Way*

## 6.1 Central Issues in the Philosophy of Astrobiology

Looking for life outside our own planet incorporates a number of parameters. On one end, the search can be done by a definitional basis; taking the life we know here and looking for a similar one. One could argue this search is biased since we are only searching for life that in any way resembles our own, and we are leaving behind other "areas of life that are unknown". If, in contrast, we take a much broader perspective and open up our search expecting to find any sort of life, we are swimming in a vast ocean of possibilities, without any success and are faced with a scientific paralysis. A set of clear constraints is then critical: a point where we can get some answers but are still open to finding anything. This project will tackle these constraints, proposing a philosophical way to conceptualize the uncertainty this project carries.

How is astrobiology related to philosophy? Here it is crucial to draw from a couple of different sources that have attempted to define this question. The one that should come up first is one that had a great impact while investigating this project, a paper from Kristina Sekrst, "Philosophy in Astrobiology/Astrobiology in Philosophy"[32] She immediately admits the following statement: "finding life outside of Earth would be philosophically profound."[32] Drawing from Jakosky's Philosophical aspects of astrobiology in "A New Era in Bioastronomy"[32]

Building from Sekrst, finding life outside of Earth, would not only be philosophically profound, but it would touch upon a lot of philosophical notions and implications. However interesting those notions and implications might be, the focus of this project lies in the epistemological considerations. Nonetheless, before delving deeper into that area, it is key to proportionate the reader with some context of where the field of philosophy of astrobiology comes from.

Sekrst in her paper highlights that the philosophy of cosmology and the philosophy of astronomy have not been well-developed fields, as it has been the case with other disciplines such as philosophy of physics and philosophy of biology. [32]

In Dick's book "Space, Time and Aliens"[8] he emphasizes six areas of philosophy of astrobiology:

1. “ the nature of reasoning,
2. the problematic nature of evidence and inference,
3. the influence of metaphysical preconceptions and non-scientific worldviews,
4. the epistemological status of astronomy and its concepts,
5. the role of technology in shaping astronomy,
6. mutual interactions of astronomy and cosmology with society over time.”[8]

“Astrobiology can be regarded as a specific case study of these issues.”[32]

With what would the philosophy of astrobiology deal with? Persson[12]has made a big step in emphasizing philosophical questions in astrobiology. These include: “justification of resources and connection with ethics, the question of life in general and its definition, the (im)possibility of knowledge of being alone or not in the universe, our stance and ethics towards extraterrestrial life, issues with governing uninhabited worlds, along with ecological questions”.[32]

For a broader perspective on this rather novel discipline, Duner sheds some new light as well. Sekrst notes: “For Duner, the philosophy of astrobiology is “an ongoing existential exercise in individual and collective self-understanding”, i.e. finding the human place in the universe. Duner also raises ethical questions regarding resource mining, possible contamination of life, or money spent on various programs, epistemological questions concerning the limits of astrobiological knowledge, linguistics and semiotic questions regarding interstellar communication, and issues in cognition, i.e., the definition of intelligence and cognition in general.”[32]

Apart from the ethical stance, metaphysics portrays an important and known role. “The definition of life itself also has metaphysical consequences, since it is also an ontological question, giving us either categorical similarity or dissimilarity, and establishing a new kind, which is of metaphysical importance. Such questions are intriguing but are being posited as philosophical questions in astrobiology, and not a part of philosophy of astrobiology”.[32] Sekrst agrees with Persson that “astrobiological questions can give new perspectives to old questions and even pose new questions”[32]

Questioning the very name of this discipline bears philosophical importance. Astrobiology is not the same as astrophysics, therefore its corresponding philosophy will be different. The very name *astrobiology* tries to talk about not just biology. There is a bit of uncertainty already present in the name. According to Sekrst “If biology was not Earth-centric, as other sciences definitely are, then astrobiology would just be called biology. By employing terms such as “astrophysics” or “astrobiology”, we are either 1) hinting towards different physics and different biology 2) talking about specialized areas: physics applied to astronomical phenomena and biology applied to extraterrestrial biology 3) talking about a more generalized approach to usually terrestrial science.”[32]

Sekrst notices a shift in the philosophical questions that shape astrobiology. Saying that if life was found in the universe, it shifts the question from an ontological and metaphysical one (Is there life in the universe? Are we the only ones in the universe?) to a question of epistemology and philosophy of mind (Is it intelligent? If so, how do we revise human epistemology and philosophy of mind to fit the new intelligence?). Even if life is not intelligent, it still bears importance, and epistemological questions can and should be asked: what counts as life? How do we know we have encountered life? What is the uncertainty in the process?[32] Astrobiology and philosophy fit well together since we are not using astrobiological research for a jump from philosophy to science, but in contrast, we are using it to resolve previous questions and ask new ones.

Hopefully some of these passages have motivated the reader into some questioning and understanding how these two disciplines, philosophy and astrobiology could potentially coincide. To put it in another way let us go back to Sekrst : "if astrobiology really aims to explain life and connected concepts in a universal way, along with questions of ethics, politics, epistemology, metaphysics, cognition, etc., then it has all the goals of what we know as philosophy, but from a large-scale point of view."[32]

For the scope of this project, "astro-philosophy" would not go as far as to delve into the different kinds of intelligent minds. Since here, our objective is not as complex. This project is concerned with the *uncertainty in the search for life*, not even finding life per se. Thus, as much as it is important to get to know further philosophical implications of the search and discovery of extraterrestrial life, it is also crucial to stay grounded to the main objective of the project: **conceptualize the uncertainty and try to understand the epistemic boundaries regarding astrobiology**.

## 6.2 Epistemological Dilemmas in Astrobiology

Let us now briefly introduce and explore the epistemological implications of astrobiology. What does a discovery of extraterrestrial life mean? Or even a constrained list of potential locations for life? More specifically, if there is uncertainty in such a discovery or constrained list of potential locations for life, what does that tell us about knowledge. In other words, what are the epistemic dilemmas of astrobiology?

As much as we will go into detail into these implications, a short definition of epistemology might be useful. Epistemology is the theory of knowledge, including its nature, extent and justification; its practitioners undertake to determine “the standards to which genuine knowledge should conform”. [32] Returning to Dick, in his book “Space Time and Aliens Chapter 37. The Philosophy of Astronomy, Cosmology, and Astrobiology: A Preliminary Reconnaissance” he raises some interesting epistemological considerations. [8]

Some of the main claims are that within philosophy of astrobiology, the question of the nature of reasoning and reliability of observations could be epistemologically rich. What is the nature of reasoning in astronomy and cosmology? And what are the limits of reasoning? By the nature of reasoning, Dick mentions “the relative roles of observation and theory, as well as the role of arguments such as simulation, modeling, and analogy. This is essentially a question of epistemology, a question of what are the legitimate sources of knowledge in astronomy.” [8]

Dick also notes that the question of reliability and interpretation of observation through epistemological lenses would be a major area of research in a philosophy of astronomy”.[8]

Shifting to another very important author, Duner, on “The History and Philosophy of Astrobiology: Perspectives on Extraterrestrial Life and the Human Mind”[12] he examines fascinating notions in relation to the epistemological dialogue.

“The search for life in the Universe touches on fundamental hopes and fears, on the essence of what it means to formulate a theory, grasp a concept, and have an imagination.” [12]Furthermore, he mentions that: “The philosophy of astrobiology is an ongoing, existential exercise in individual and collective self-understanding. That is, what it means to be human, where humankind’s place in the Universe is, and how both of these conceptions have inevitably evolved."[12] It raises, among others, very important epistemological questions.

Touching upon briefly on the methodology, are the methods in astrobiological empirical ones? According to Dick, one could say the search for gravitational waves, the Higgs boson, or planetary systems, are, or were, "sciences without a subject." However, he notes: "every science is looking for a subject until it finds it or thinks it may have found it".[8] Later on, he makes a key point: "from an epistemological point of view, the methods of astrobiology are as empirical as in any historical science such as astronomy or geology, though it is true that astrobiological observations and experiments are often especially difficult, and the inferences more tenuous". [8] Therefore, this idea of 'science without a subject' should be retired, and we should view it in a different way.

Additionally, further critical issues in the philosophy of astrobiology include: epistemological issues such as the status of astrobiology as a science, the problematic nature of evidence and inference, and the limits of science.[8] Now is time to jump to some of those.

Furthermore, in astrobiology there is a problematic nature of evidence and inference. These problems remain whether we are studying the nearest planets of the Solar System, further planetary systems or theories and experiments about the origin and evolution of life. And it is present whether we are using ground-based telescopes or space vehicles that land in other worlds. "It applies equally to the visual, photographic, electronic, photometric, and spectroscopic techniques of astronomy, as well as to the variety of techniques employed in studies of the origin and evolution of life." [8] Grounding this problem to my project: Possible Planetary Locations for Life in our galaxy seems to only make this harder. As

Dick himself states: “problems of evidence and inference are redoubled in the elusive search for planetary systems beyond our own.”[8]

As we have learned throughout this project, most of the exoplanets (planets that reside outside our Solar System) are not directly observed, but inferred from their host star. [5] This makes the problem of evidence and inference even harder. And even when we are able to observe them, there will still be aspects of them which we will not be able to know by pure observation due to their distance from us. Such aspects include, for example, the planet’s exact mass, which we are not able to know without uncertainty, or even more crucially, if life resides in it.

It is important to note that all these issues of observation and inference may also be discussed in relation to the philosophical problem of epistemological realism and anti-realism. [8]

A concluding question regarding Dick is the following: “how have observation, evidence, and inference been deployed in arguments in astrobiology, and how does this deployment compare to their use in other sciences?”[8] This remains a rich area for further research, and can be further pursued through many avenues. [8]

To conclude this section, let us consider; what is the epistemological aspect to this discipline? Building off of Duner, “epistemology here portrays a rigorous consideration of what is known, what is knowable in practice or in principle, and what is knowably unknown”. [12] As we have seen the philosophy of astrobiology has not become a very explored subject yet, hence if we were to explore it deeply, the epistemology of astrobiology is a less explored philosophical territory concerning the limits of astrobiological knowledge. Duner is rightfully troubled by questions in this category. For example: “how long should we search without positive results before we give up? What is possible and not possible? *"The epistemological problems of astrobiology are somewhat similar to those of other branches of science, but with the exception that the limits of our astrobiological knowledge seem to be much more uncertain."* [12] This here is essential to the project. "The limits of our astrobiological knowledge seem to be much more uncertain."[12] This is exactly what this project tackles, not only the epistemological aspect of astrobiology, but furthermore how uncertain are its boundaries, and how can we deal and conceptualize this uncertainty? The goal of the following chapter will be to face these issues, hopefully providing some light in this somewhat unexplored territory.

## 6.3 Sources of Uncertainty in Astrobiology and in this Project

Not only would finding extraterrestrial life be philosophically profound, but additionally, the uncertainty in such a discovery touches upon key philosophical inquires and implications as well. Ethical, epistemological, pragmatic and methodological questions come up.

On an epistemic note, it becomes important to also consider what is the weight of values in this project; would the uncertainty attached to the data and theories make them less pure, or how can we balance the truth with uncertainty and make sure the data is reliable? How can we know if the parameters we are choosing are good ones?

A more pragmatic question dealing with uncertainty is: how should we handle the risks of being wrong in our inferences, and how should we balance the risk of two kinds of errors: believing a false statement or rejecting a true one?

Additionally, methodologically, is this approach, and the uncertainty it carries a good way of doing science, and what would an alternative look like?

This project is flooded with uncertainty. The goal of it is not to reduce this uncertainty, but to conceptualize it. To understand where this uncertainty is coming from and what it means.

To grapple with these questions, we might need the aid of another science: climate science. In order to understand the sources of uncertainty in astrobiology, it is critical to invoke another author, Eric Winsberg. In his paper: "Values and Uncertainties in the Predictions of Global Climate Models", he mentions there are two main sources of uncertainty:"structural model uncertainty and parameter uncertainty". [39] Aside from these two sources of uncertainty, Winsberg also identifies a third source, "data uncertainty". [39] We shall start by defining this one. This uncertainty comes in when evaluating a particular climate model, including both its structure and parameters, when we compare the model's output to real data. Winsberg makes the crude assumption that "the data against which climate models are evaluated are known with certainty." However, he urges us to notice that data uncertainty is part of parameter uncertainty and structural uncertainty, "since it acts by affecting our ability to judge the accuracy of our parameters and our model structures.".[39] Mentioning this one first is essential since this project deals with a lot of data. 3 There is direct data obtained from the missions that have as their goal to look for exoplanets (planets outside our solar system). Let us start in Phase I of the project: The Inputs, first data set: "Planetary Population Data Set“. This data that contains, for example, how many exoplanets each telescope has found, the distance to their nearest star, their size, and so on, carries an amount of uncertainty. Additionally, each mission has certain biases: are they more likely to find a specific type of planet? Or star? These biases affect each mission and therefore the uncertainty in their results? But, how can we comprehend the uncertainty of different missions? Can we add them together? What is the best method to do so? These are all challenging questions to answer and as you can see from now, it is from the very beginning that this project deals with uncertainty and the way of even understanding it is tricky.

Moving on from those values from the missions, where else can we find uncertainty? Our second main data set, "Molecular Data Set" is covered with uncertainties itself. This uncertainty comes from two main sources, regarding RASCALL, there is uncertainty because a lot of the molecular spectra are estimates in themselves. Not all the spectra for all the biosignatures exist, therefore an estimate such as this data set (RASCALL) is essential. Regarding the Bond Dissociation Energy assigned to the molecules includes uncertainty as well, since it is using RASCALL. Measuring these uncertainties is also tricky, and might be different for each molecule. That is where the list of constraints of each data set become meaningful. Perhaps we do not have to comprehend the uncertainty of all the molecules and all the planets, but only those that will be relevant to our list of constraints. This narrows the uncertainty a little bit, however, combining that uncertainty would still be tricky. Here is where 'data uncertainty" may place useful. Finally regarding "data uncertainty", has to be included since we are dealing as we just saw with the other two sources of uncertainty, and it compares to real data, say: if we have a Planetary Population Data Set for exoplanets, we can compare that data to real data from our own planet. Or even those in the solar system. This uncertainty is present not only because it is part of parameter uncertainty and structural uncertainty, but because it affects our ability to judge the accuracy of our parameters and our model structures.[39]

Next, in order to explain the second source of uncertainty, “structural model uncertainty”, Winsberg pulls from climate models and how their construction is guided by “basic science” which he defines as “science in which we have a great deal of confidence”. Winsberg also notes that climate models involve assumptions, approximations, parameterizations. All of these contribute to a degree of uncertainty about the predictions of these models. Each model, with a certain basic structure, produces substantially different predictions. Winsberg names this source of uncertainty as “structural model uncertainty.”[39] How is this seen in my project? There seems to be “structural model uncertainty”, since the whole project, from the data sets in particular the Debiased Population Data Sets[11][17][23], involve assumptions, approximations and parameterizations, which contribute to a degree of uncertainty in the models of finding exoplanets or regarding the second main data set: Molecular Data Set, establishing the spectra for the biosignatures[33] and their Bond Dissociation Energy. [38]. Each model, or mission [11][17][23][24][15], produces different predictions.

To finalize with the sources of uncertainty, later on, Winsberg introduces the third type of uncertainty in climate science, “parameter uncertainty”. This source of uncertainty will be key for understanding the uncertainty in this project, since we are pulling from a lot of parameters. This uncertainty comes from complex models involving large sets of parameters or aspects of the model that have to be quantified before the model can be used to run a simulation of a climate system. [39] Since scientists are usually highly uncertain about what the best value for many of these parameters is, even if there existed a model with ideal (or perfect) structure, there would still remain uncertainty regarding the behavior of the model, since because the same model structure will make different predictions for different values of the parameters. [39]

Thinking about this type of uncertainty, “parameter uncertainty”, immediately seems very relevant to my project.

Now, in order to analyze this type of uncertainty in Phase II: The Algorithm, we must briefly describe what this phase is doing. As seen in Chapter 4,4 the algorithm, PUSH, essentially takes the two data sets, the list of constraints (tuned by the user), and creates a constrained list of potential locations for life of both. Here what was mentioned earlier of combining the uncertainty becomes key. This is because, this constrained list of potential locations for life is going to have some uncertainty in itself, which most possible will have to carry exactly from that one in the data sets, and the list of constraints. The algorithm then does a more complex task: it takes those two outputs: “Planetary and Stellar Outputs: Possible Planetary Locations for Life” and “Molecular Outputs: a constrained list of biosignatures” and combines them (in a way combining the uncertainty in some way, which is not defined at the moment) and produces the Final Output: Planetary, Stellar and Molecular Outputs, Potential Locations that Could Host Life.

No matter what the outcome is, the outputs of the algorithm are going to entail uncertainties. Even if the outcome does not depict any results at all, that result perhaps in itself showcases how much uncertainty was present in the first two phases, that it was perhaps too much to produce a result. Additionally, since the First Phase is tuned by the user, the uncertainty in the outputs could vary every time.

It is important to pause here for a moment and notice why “parameter uncertainty” is relevant. Anyone who uses the algorithm, from astrobiologists to students, they are forced to make judgement calls and make out the output of PUSH. What does too narrow or too broad a list of locations mean specifically to *each* user? PUSH can be tested for the implication of parameters.

As mentioned earlier, uncertainty permeates this project since the very start, now how do we *grapple* with this uncertainty is the real question. The way to accumulate the uncertainties into a single value or definition is not a clear one. How do you combine uncertainties from complete different data sets, constraints and also different goals and objectives every user of the algorithm might tailor? What this project could do if combining it into a single value becomes too challenging is to perhaps simply *conceptualize* the uncertainty. This might be still a tricky, but a more realistic task this project can take on. Now is time to face that challenge. We shall dive straight into this goal in the following chapter.

# 7
# Conceptualizing Uncertainty in Astrobiology

> "So that the first sort of beauty ceases to astonish us as soon as we have reached the things themselves, but the second is something that we understand only when we have passed beyond them."
>
> ---
>
> *-Marcel Proust*
> *In Search of Lost Time: Time Regained*

## 7.1 Thoughts on Uncertainty and Inductive Risk

For this next section, in which I will delve deeper on the ideas, problems and conclusions drawn from some papers and further writings dealing with inductive risk, uncertainty and the values in science, I believe it crucial for the reader to begin by becoming familiar with some useful definitions. The literature around the topics of "inductive risk", "uncertainty" and "values," all require the reader to understand the difference between concepts such as: the philosophical field of *epistemology* along with more technical terms such as *epistemic* and *non-epistemic*values in science.

This is a project that tackles broadly and narrowly the concept of epistemology. What is epistemology? According to the Standford Encyclopedia of Philosophy, it "seeks to understand one or another kind of cognitive success (or, correspondingly, cognitive failure)"[36], and it gives a brief summary of its history through some important philosophers: "Plato's epistemology was an attempt to understand what it was to know, and how knowledge (unlike mere true opinion) is good for the knower. Locke's epistemology was an attempt to understand the operations of human understanding, Kant's epistemology was an attempt to understand the conditions of the possibility of human understanding, and Russell's epistemology was an attempt to understand how modern science could be justified by appeal to sensory experience. Much recent work in formal epistemology is an attempt to understand how our degrees of confidence are rationally constrained by our evidence, and much recent work in feminist epistemology is an attempt to understand the ways in which interests affect our evidence, and affect our rational constraints more generally." [36]

Having a clearer definition of the term, epistemology, was essential since it is important to note that this project will mainly deal with issues having to do with "epistemology". Moving on from that, it is crucial in order to introduce other interesting topics that relate to epistemology such as : "epistemic" and "non-epistemic values", to define briefly both Hume's problem of induction and the concept of "inductive risk".

### 7.1.1 Inductive Risk and The Value-Free Ideal

To even begin to discuss and draw conclusions from the term “inductive risk”, it is important to briefly note what the “Problem of Induction” is. This problem is first encountered in Book 1, part iii, section 6 of A Treatise of Human Nature by David Hume, published in 1739[19]. Hume was concerned with what is known as “inductive inferences”, those inferences that enable us to justify predictions about future observations based on known information. These are inferences from the observed to the unobserved, and which generate general laws.

Hume analyses the notions of cause and effect in order to grapple with the problem of induction. He believed that ultimately all our ideas could be traced back to the “impressions” of sense experience, noting a relationship between causation and ideas. Hume proposes that we are not able to make a causal inference by merely a priori means, only based on experience. Hume believed there to be no logical reason why induction should occur.[19] To aid us with a known and useful example: for Hume, if we know by our senses the sun has risen today, there is no logical way to infer (a priori) that it will rise tomorrow. Even though the sun has always risen, and we are able to confirm this every day with our senses, there is no logical way of inferring it will rise tomorrow as well. We do not have the sense of confirmation for that event until it occurs.

Building from Hume, it becomes essential for the reader to also understand what exactly does “inductive risk” mean and how is it related to this project. Let us draw directly from the papers which we will be mainly focusing on this chapter, starting by “Inductive Risk and Values in Science” by Heather Douglas. [9] The term “inductive risk” was first used by Carl Hempel and is defined as: the chance that one will be wrong in accepting (or rejecting) a scientific hypothesis."[9] There are other voices involved regarding this debate. For example, another side is portrayed by Churchman and Rudner, stating that the risk of inductive error meant that values must play a role in science generally, while on another end, Jeffrey and Levi engaged in limiting the influence of *non-epistemic values* in science. [9] A term which we will closely analyze in the next two sections.7.1.1 Furthermore, Hempel claims that accepting a hypothesis will carry in that decision the 'inductive risk' that the hypothesis might be incorrect. [9] Therefore, inductive risk encompasses the *risk of error* in accepting or rejecting hypotheses.

Douglas also challenges traditional views of science and the scientific method. As she mentions from as early as her abstract: "I argue that because of inductive risk, or the risk of error, non-epistemic values are required in science wherever non-epistemic consequences of error should be considered". [9] Douglas, from the very start, is bringing in the concept of "epistemic and non-epistemic values" in relation to *inductive risk* to portray an intriguing claim. These crucial terms will play an important role in the discussion around inductive risk and science in general, and will be explored in detail. For now, let us just have present that the manner in which she is challenging what is "required" in science, is thought-provoking, and its understanding will become crucial as we move on.

How are all of these ideas reflected in the scope of my project? They are directly linked to my project in the sense that not only the hypothesis or main goal: (conceptualizing the uncertainty in the search for life) comes with a risk of error, but every value in every data set, input and output of the phases of this project, carry a level of uncertainty or margin of error. In every single value and as early as the hypothesis, there is a risk. The important question relating to it is, how do we define or conceptualize this uncertainty?

Can we know how much this risk is? And, will it be able to alter the "truthfulness" of the outcomes/results of my project? What does it depend on? It implores the follow-up questions: what are epistemic and non-epistemic values, and what weight do they carry?

### The Value-Free Ideal

Before mentioning and delving into epistemic and non-epistemic values, it is key to introduce what is called "The Value-Free Ideal". As we noticed with inductive risk, making judgement calls in science is tricky, and error can be everywhere, hence some values are needed. According to Douglas, The value-free ideal rests on at least one distinction for values in science, between *acceptable* and *unacceptable* values. Acceptable values became known as "epistemic," meaning related to knowledge, whereas the rest became known as "non-epistemic," a category that includes social, ethical, and other values, all the "forbidden" values.

In this view of science, non-epistemic values such as social values do matter, just not in the internal aspects of science, they lie more in the outskirts of science, leaving the internal parts of it to the epistemic values that hold a higher stand.

Douglas makes a crucial point, criticizing this ideal. She mentions that if this distinction fails, between epistemic and non-epistemic, not only is there another reason to reject the value-free ideal (for one of its foundational distinctions is faulty), but also she comments on how one should not rely upon that distinction in forging a new ideal. [10] This point is key since it imposes a reconsideration of a critical claim: upon building new science, this is not an ideal we might want to consider as our starting point.

That said, we are now able to jump ahead to the inside workings of mentioned definitions such as "epistemic and non-epistemic values" which are crucial for the portrayal and the understanding of the relevant papers that will help in analyzing and comprehending the data and the role of uncertainty in this project.

### *7.1.2 Epistemic and Non-Epistemic Values*

Going back to Douglas' paper: "Inductive Risk and Values in Science" it imposes an emphasis on non-epistemic values relegated to the 'external' parts of science, while "epistemic values" have become widely accepted as part of scientific reasoning, and that "only epistemic values have a legitimate role to play in science". [9] What are these "external parts of science" and why do they matter?

It is pressing to define and think more about these values specifically in relation to "inductive risk". The term "epistemic values" as defined by McMullen are those "presumed to promote the truth-like character of science" [20] While non-epistemic values are simply the "rest"[9]. Douglas mentions the position of these values in relation to science, epistemic values being located at the center or core of science, and non-epistemic, conversely, situated at the boundaries or outskirts of science. What are each, then? And how are they present in my project? These are very challenging questions, for which I believe a lot of examples are necessary. Epistemic values sound as values that rely solely on "numbers", that may not always be the case. Epistemic values encompass any value that, as the definition of the word epistemology means, *contains truth.* Therefore, non-epistemic ones would be those that not necessarily contain "truth". However, are they less important? If non-epistemic values do not contain truth, what are they? They could be seen as the values related to the decision-making process in scientific endeavors. The choices each scientist takes to what is "relevant" for the goals of the project. Something that becomes interesting to consider is that perhaps it is crucial to delve more in depth to the epistemological aspect of it. Epistemic values are based on some "truth". That "truth" is defined by a decision-making process in which the standard of decisions is based on certain assumptions about "truth". Could it be the case that epistemic values have non-epistemic characteristics to them?

Grounding this to my own project, what would each part of my project, each value, be considered as? This is a complex question. My first thought would be to assign the name of epistemic values to all the values which seem to contain truth. Since my project grapples with so much uncertainty, this decision may be very ambiguous. An obvious example of what could perhaps be defined as an epistemic value in my project, could be the first data set "Planetary Population Data Set", in which the values are quite literally either taken from the direct missions, or from scientific papers. Does the name scientific paper, make its values epistemic? Boundaries and limitations seem critical to be drawn in order to define each value.

### *7.1.3 The Kinds and Roles of Values in Science*

This subsection has as its goal to go deeper into Inductive Risk and the goal of values in science beyond the Value-Free Ideal and notions of considering values in science as solely epistemic and non-epistemic. This section will be drawing a lot of the claims from Douglas' book, "Science, Policy, and the Value-Free Ideal, Chapter 5: The Structure of Values in Science". [10]

It was essential for the reader to first understand where the term inductive risk was coined, but additionally, to comprehend where did the values in science, known as epistemic and non-epistemic, were coming from and what they meant. Now is time to go beyond again and, as the title of the chapter suggests: understand the “structure” of the values in science. Meaning to understand the *kind* of values and their*roles*. This is key, since we will shift from our previous understanding of *epistemic* and *non-epistemic*, into a whole new realm of comprehending values in science.

The very first shift in our perspective is on *science* itself. Douglas claims science is a “value-laden process”. [10] According to Douglas, values have a role to play *throughout* the scientific process. They play a role from the very decision to do any science, to engage in a particular project, all the way to the specific choice of what methods are being used, to the interpretation of found data, even to the very final results and conclusions. [10] As we can see, the scientific process as described by Douglas is one which is completely covered by the *role* of values. Here, it is important to simply mention once more that a very big shift has occurred from the “Value-Free Ideal” which as we noticed earlier relies on one distinction for values in science: those that are accepted or *epistemic* and the unacceptable or forbidden, that are all the rest, the *non-epistemic* values. She mentions this earlier in the chapter (page 89). [10] This distinction shifting from “epistemic” and “non-epistemic” to what Douglas proposes which is simply to consider the *kinds* and the*roles* of values in science is key for understanding the leap that takes place from these more traditional view of values regarding science, to values that can be more useful for science. And additionally to go further and evaluate the consequences of error.

Let us now understand in detail the *kinds* of values we can find in science. There exist, according to Douglas three types of values: cognitive, ethical, and social. [10]

- **Cognitive Values**: these values are more precise than the vague notion of acceptable values in science which the Value-Free ideal takes on, often in the same ground as epistemic values. Instead, cognitive values are: "aspects of scientific work that help one think through the **evidential and inferential aspects** of one's theories and data. Cognitive values embody the goal of assisting scientists with their cognition in science."[10] To put it simply, cognitive values are concerned with : "the possibilities of scientific work in the immediate future."[10] Some examples are: **simplicity, explanatory power, scope, consistency predictive precision, fruitfulness,**[10]

**simplicity** is a cognitive value because complex theories are more difficult to work with, and the full implications of complex theories are harder to unpack. **Explanatory power** is a cognitive value because theories with more explanatory power have more implications than ones that do not, and thus lead to more avenues for further testing and exploration. (Explanatory theories structure our thinking in particular but clearly articulate ways, and this allows one to draw additional implications more readily.) **Scope** is a cognitive value because theories with broad scope apply to more empirical areas, thus helping scientists develop more avenues for testing the theories. The **consistency** of a theory with other areas of science is a cognitive value because theories consistent with other scientific work are also easier to use, allowing for applications or extensions of both new and old theories, thus again furthering new research. **Predictive precision** is a cognitive value because making predictions with precision and testing to see if they are accurate helps scientists hone and refine theories more readily. And **fruitfulness** is an obvious cognitive value because a productive theory provides scientists with many avenues for future investigation. Fruitfulness broadly construed may be considered the embodiment of cognitive values—the presence of any cognitive value should improve the productivity of an area of science. It should allow for more predictions, new avenues of testing, expansion of theoretical implications, and new lines of research. [10]

Examples of cognitive values in astrobiology: Definitely scope would be a cognitive value present in this project, since as we can already see this being multidisciplinary, astrobiology can be analyzed philosophically. Fruitfulness would definitely be another one, since any advancement in the search for life that adds to another researcher's work, is useful and encouraged. Cognitive Values in this project: Again here it would be the data sets, which provide evidence for building the algorithm, but they are again flooded with uncertainty.

- **Ethical Values:** important when considering the consequences of error for the general public. Focuses on the good or the right. These values aid us in weighting if potential benefits are worth potential harms, or if say some harms are worth no price. [10] Some examples are: the rights of human beings not to be used for experimentation without fully informed consent, the consideration of sentient beings for their pain, concern for the death and suffering of others, whether it is right to pursue research for new weapons of mass destruction, and whether an imposition of risk is ethically acceptable. [10]

  Examples of ethical values in astrobiology : The first example that comes to mind is considering if we should even be looking for life? Does it pose a "good" outcome? Could there be a potential harm in looking for life, either for us or for something out there? In case our search is not fruitful, would we add to skepticism in the general public? Ethical Values in this project: pragmatic view in searching for life, we have to be willing to miss life in order to find it.

- **Social Values:** they arise from what a particular society values. Can overlap with or be opposed to ethical values. Some examples are:[10] **justice, privacy, freedom, social stability, or innovation**[10]

Examples of social values in astrobiology: Do we have a responsibility as a species to look for extraterrestrial life? Would searching for life, affect the social stability we have on Earth, given the search is successful? Social Values in this project: the relationship astrobiology has to policy and the larger community, we have to acknowledge that in our inferences we are pulling from a certain level of uncertainty in which the potential targets for life are not assured, they are simply estimated.

Now let us turn to the *role* these values take. These place values in an option or outcome, and values can play either an "indirect" or a "direct" role. However, values playing an "indirect role" in science are not less important, they are indeed very useful. We should be able to find these values everywhere, and most importantly, analyze them for the benefit of each scientific project. Let us analyze each role in detail.

1. **Indirect role:** they are responsible for weighting the uncertainty, when considering the consequences of error.(They are present usually in the later stages of science or throughout the stages). They come up after research. This role (whether ethical/cognitive/social) has no threat to the *integrity* of science. This role is relevant when uncertainty is present. Values act to weight the importance of uncertainty.

Values determine the importance of the inductive gaps left by the evidence. The more evidence, the less uncertainty and the fewer values in this role. [10]

2. **Direct role:** this role is present in the choices in the *early* stages of science. As Douglas puts it, they "determine our decisions in and of themselves, act as stand-alone reasons to motivate our choices". The value in this role accepts or rejects. Uncertainty is irrelevant in this role. Asking if it is the choice what we want, there is a threat to integrity of science. Values in this role act as evidence to accept a claim. [10] It is important to note that the direct role must only be considered to the decisions having to do with the early stages of science. Since at this point, there is no deciding from derived conclusions. Douglas makes a key point: "once a scientist has embarked on the course of research, the indirect role for values should be the only role for values, whether social, ethical, or cognitive. To admit of a direct role of values in these later stages is to violate the good reasoning practices needed to obtain reliable knowledge about the world." [10] Obtaining "reliable knowledge about the world" seems a good goal for science, hence not only should we include both values in our scientific process, but we should know at *which stage* of this process to use them. Why is the direct role constrained to the choices in the early stages of science? As mentioned before, values in this role act as reasons to *motivate* our choices. [10]

Going back to the main goal of this section, which was to introduce a divide from the Value-Free ideal to the Kinds and Roles of values in science, let us simply pause and consider: why is a new understanding and break from the Value-Free Ideal needed? Highlighting from Douglas herself: **"We need social, ethical, and cognitive values to help weigh the importance of uncertainty and to evaluate the consequences**

**of error.”** [10] This is extremely important for the sake of this project. These values carry a great responsibility, “weighing the importance of uncertainty and evaluating the consequences of error”. In this project, both responsibilities, are absolutely crucial. Not only is this project invaded with uncertainty, but knowing how to weigh it, and going further than knowing there is error to actually evaluate it, are essential tasks in this project. From every data set, to every result, there is a great deal of uncertainty present. When is this uncertainty more relevant than in other points of the project? For example, let us think, would the uncertainty in knowing the distance the exoplanets are positioned to their host stars carry the same weight as the uncertainty in claiming there is life in a known exoplanet? Social, ethical and cognitive values help us weigh this importance.

Further on, not only are the *kinds* of values relevant to the uncertainty a scientific endeavor carries, but the *roles* they are in is of importance as well. **In this direct role, uncertainty is irrelevant to the importance of the value in the judgment**[10] **(uncertainty does not matter, what matters is if it is worth pursuing).** The issue is not whether the choice will somehow come out wrong in the end, but whether the choice, if it comes out as expected, is what we want. This role for values in science is crucial for some decisions, but it must be restricted to certain decisions made in science and excluded from others.[10] This makes sense, since there is a point in science where as Douglas mentions, the indirect role is much more important, one cannot solely depend on whether we want to pursue something. The indirect role comes out with ethical, social and cognitive values and asks us to reassess what is important at other stages. “The integrity of the scientific process cannot tolerate a direct role for values throughout that process.”[10]

**The indirect role, in contrast, can completely saturate science, without threat to the integrity of science.**[10] This role arises when there are decisions to be made but the evidence or reasons on which to make the decision are incomplete, as it often occurs, and therefore there is uncertainty regarding the decision.[10]

At what point are the roles for values important regarding the uncertainty in the scientific process? How do they help us accept or reject a claim? And with how much uncertainty? To answer this, it is crucial to go back to Douglas: "The values can act as reasons in themselves to accept a claim, providing direct motivation for the adoption of a theory. Or, the values can act to weigh the importance of uncertainty about the claim, helping to decide what should count as sufficient evidence for the claim. In the first direct role, the values act much the same way as evidence normally does, providing warrant or reasons to accept a claim. In the second, indirect role, the values do not compete with or supplant evidence, but rather determine the importance of the inductive gaps left by the evidence. More evidence usually makes the values less important in this indirect role, as uncertainty reduces. Where uncertainty remains, the values help the scientist decide whether the uncertainty is acceptable, either by weighing the consequences of an erroneous choice or by estimating the likelihood that an erroneous choice would linger undetected. A direct role for values at this point in the scientific process is unacceptable, but an indirect role is legitimate."[10] This understanding of the roles of values is absolutely crucial for our understanding of science as a whole. Specifically and related to my project is the importance of *uncertainty.* How, mostly, it is dealt with by indirect values becomes crucial to the understanding of this project. If the indirect role values "determine the importance of the inductive gaps left by the evidence"[10], there is much work to be done here regarding the indirect role of values in this project, and in the consideration of the whole of science.

Douglas mentions that this "new ideal" should cover all science. "The importance of values in the heart of science decreases with decreasing uncertainty, but as long as science is inductively open, uncertainty is ineliminable, and thus so are values." [10] Then, a crucial question arises, with this "new ideal" can science still be objective? "... the objectivity of science does not fall with the rejection of the value-free ideal."[10]

As we learned from Douglas, in science, there is no 'Cartesian Certainty'. Therefore, we need values, it is simply a necessity. Doing science immediately brings questions such as: is the uncertainty acceptable, are the inferences acceptable? We can now understand why epistemic constraints do not help with answering those questions; we need ethical and social values. To simply judge the uncertainty, we need those values. This strengthens the argument that science cannot be value free. Also, as stated before, it is in the internal aspect of science, where crucial inferences occur, where we need to consider the values. We are limited 'epistemic actors' [10], and therefore we need values to compensate for these limitations. We must always consider the impact of scientific work on society. Scientists are active actors in society, therefore we have to consider the impact of their work in the society and impact of error and of uncertainty. Social and ethical values matter because of these reasons. We need scientists to also make judgments (policy). The values that matter are the ones that framed and carried out the work, they all matter. Therefore, not only values are important but the role they play: direct and indirect also are crucial.

### 7.1.4 An Example of the kinds and roles of values in science: Rat Livers

Before further delving ourselves in other interesting discussions, it is essential for the reader to apply these past terms, and situate them in an actual example to illustrate how inductive risk entails non-epistemic value judgments.

In this section, I aim to provide an example of how non-epistemic values shape an "internal" aspect of the research process. Here, I am going to draw from Douglas' paper on "Inductive Risk and Values in Science". From section five, "Inductive Risk in Evidence Characterization: Rat Liver Tumors´´there becomes clear internal aspects can be shaped by non-epistemic values. This experiment doses rats for dioxin cancer studies. Philosophically, the goal is to portray that there exists an inductive risk in making choices, and additionally one has the responsibility of deciding first what kind or of inductive risks are acceptable but additionally what level/amount of inductive risks are acceptable. The crucial aspect being, some of the consequences of the risks are non-epistemic, and thus non-epistemic values are needed to weigh the consequences and to make the choices. [9]

This relates to Douglas' observations in Jeffrey's response to Rudner, where scientists often have to make methodological choices that do not lie on a continuum. [39] Let us imagine we are investigating the possible hypothesis that substance "S" causes disease "D" in rats. We then proportionate a specific experimental group of rats a large dose of "S" and then perform biopsies to determine what percentage of them has disease "D". To perform the biopsy, imagine there are two staining techniques that could be used. One is more sensitive and the other is more specific. One actually produces more false positives and the other more false negatives. Which technique should we choose? Douglas notes that which one I choose will depend on my inductive risk profile. If we weigh more heavily the consequences of saying that the hypothesis is false if it is in fact true, we will choose the stain with more false positives, and vice versa. That ultimately depends on our social and ethical values. Social and ethical values therefore play an inevitable role in science.[39][9]

### *7.1.5 Inductive Risk in Theoretical Physics*

How are both inductive risk and the values in science regarded if we begin to shift away from policy? If we aim at an area of science in which policy does not apply for the sake of science in itself, as could be the case with an elementary particle from the Standard Model (SM) of particle physics: the Higgs Boson. Let us consider Kent W. Staley Chapter 3 Decisions, Decisions: Inductive Risk and the Higgs Boson"Exploring Inductive Risk: Case Studies of Values in Science"[35] The most important claim on Staley's view is a grounded backup from Douglas, that "eschews the classification of value judgments into epistemic and non-epistemic."[35] Staley advocates for Inductive Risk to be present in the discovery of the Higgs boson, by returning to the mid-twentieth-century roots of the literature on the argument from inductive risk (AIR). He extrapolates from them an argument about how to apply a broadly" pragmatic philosophical orientation to the interpretation of statistical inference".[35] By doing so, he ends up emphasizing the central role of practical decisions in the production of theoretical knowledge.[35]

He establishes: "The drawing and reporting of inferences in the context of scientific inquiry is not merely a matter of forming beliefs. It is instead a practical matter, and as such is open to the full range of value considerations that bear on our decisions in every domain of activity. If we choose to build scientific knowledge in a way that preserves epistemic autonomy, it will be because we think doing so will deliver the greatest good, all things considered."[35]

Staley offers a technical example of how values in science, even in the corners of science that are not linked to policymaking at all, will and should end up getting entangled with all types of values. Social, ethical and cognitive ones. Specifically, where inferences happen, as is the case, urges our "value considerations". [35]

The example that Staley portrays is the discovery of the Higgs Boson and the standard of certainty attached to this discovery. How can the $5\sigma$ standard help us understand values in science and uncertainty? What is $5\sigma$? According to CERN, The European Organization for Nuclear Research, "five sigma is considered the 'gold standard' in particle physics because it guarantees an extremely low likelihood of a claim being false."[31] Staley opens up the discussion with questioning: for what is the $5\ \sigma$ standard a standard? He states: "it is a standard governing the decision of how to report the outcome of the experimental search for the Higgs boson. This decision concerns not only cognitive but also communicative actions."

The main take-away from this paper is that as much as you are dealing with a science that is "policy-free", the way you convey your findings, in terms of their "certainty or uncertainty" matter to yourself (scientist), your fellow scientists and institution and to the general public. This might be the sort of pathway that could lead into discussions of social, ethical and cognitive values in "policy-free" science, but for now, let us simply try to understand what this paper means by $5\sigma$. Staley mentions that the criterion for the Higgs boson's discovery, remaining at $5\sigma$, has to do with an important aspect: "*the impact of the discovery.*" The impact of discovery is broadly talked about in this discussion. There is an impact related to the 'value of the discovery claim itself' as well as 'the potential harms caused by making a discovery claim that turns out to be erroneous'.

Therefore, what is probably one of the most important aspects of this reading is the different ways in which the Higgs boson's "discovery" had *impacts*. Staley focuses on impacts that fall into two main categories: First the impacts that relate in a direct way to argumentation in future physics inquires, or to put it in another way, the impact that is directly related to the ones doing the experiment, the physicists. **(Impacts in relation to the physicists)**. Additionally, there is an indirect impact that goes beyond the direct one, into the external community, impacting (in this case) the broader goals of the A Toroidal LHC Apparatus (ATLAS) and the Compact Muon Solenoid (CMS) collaboration groups, the High Energy Physics (HEP) community, and science generally **(Impacts in relation to broader research goals, and in relation to the scientific community)**.[35] In the following section we will state and narrow these impacts and further on apply them to astrobiology.

### *7.1.6 Expanding connections between astrobiology and the role of values in determining significant thresholds (5$\sigma$)*

Let us go in detail into the first impact that encompasses the **"impacts in relation to the physicists"**. It is going to be useful to know the examples to then translate them into astrobiology. These impacts mentioned by Staley are:

- By accepting the existence of a new particle, scientists would therefore use this acceptance as a presupposition for the pursuit of further scientific inquires. It would lay the basis for what a successful discovery would entail to then follow suit. Because of this, a discovery would affect the work of ATLAS and CMS. In this case, it would force them to shift from hypothetical experiments towards measuring the properties of the discovered particle to have more evidence to secure the theoretical interpretation of the discovery (to analyze if it is actually the Higgs boson). [35]

- Additionally, regarding other scientists, more specifically physicists working with the Standard Model (SM) and Beyond-SM problems, "the announcement by ATLAS and CMS has the consequence of changing the *logical terrain.*" [35] By announcing the discovery of the Higgs boson, each scientist/investigator will have to decide (as an individual or as a member of a group) whether the evidence offered by the group of physicists suffices to warrant accepting the discovery of the new particle as a premise or assumption in their future work. [35]

- The burden of proof is now positioned on whoever rejects or disagrees with the claim. They will be faced with the responsibility to explain their reasons of disagreement. Staley finds this crucial since he mentions: "These considerations contribute to our understanding of the $5\sigma$ standard for the Higgs search by highlighting the importance, for the pursuit of physics inquiries, of guarding against an erroneous discovery claim, while also pointing toward the tremendous value of that discovery claim, as it enables the pursuit of new inquiries that, prior to discovery, had to wait offstage."[35]

Regarding the **"impacts in relation to broader research goals, and in relation to the scientific community"**, Staley comments on a few:

- Expectation: given the great expense of building the LHC and operating the CMS and ATLAS experimental programs, it is not surprising that success at achieving this goal was highly valued. [35]
- Reputation: there is an impact on reputation since the discovery was very anticipated, and it was communicated in scientific talks but moreover it was conveyed to the media, which made it to the Internet worldwide, and were featured in the news. Therefore, as Staley puts it: "to get things wrong would have been tremendously embarrassing." [35]
- Funding: in case the discovery were to be erroneous, a comprehensive assessment of all risks of errors would be necessary, which would potentially add a political dimension with possible negative consequences for the funding of High Energy Physics (HEP) research. [35]
- Responsibility: there is an impact having to do with responsibility in the announcement of the discovery of the Higgs boson, since a cautious attitude is needed regarding how this is conveyed to the public, and more specifically, how this can affect the public's perception of science. [35]
- Opportunity: due to intense public spotlight that the LHC had since 2008, this portrayed how there was an impact on the opportunity to showcase very high quality science to the general public. Specifically notes Staley: "in an environment where there was public skepticism about some scientific claims." [35]
- Skepticism: if the discovery turns out to be a false one, this would only contribute to mentioned skepticism about scientific claims. [35]

Pausing here for a moment is imperative in order to understand how this impacts relate to the discovery of the Higgs boson and therefore how it relates to the $5\sigma$ standard. All the impacts mentioned above are factors that should be of relevance in the determination of the standard of evidence, so we should not take them lightly. Why are they relevant? They help us analyze the broader picture and goals of such a claim. They portray how if the discovery is successful or erroneous it would not only impact the HEP community, but it would potentially create or cease funding, it could motivate claims or create more skepticism. These impacts must be considered of extreme relevance to the determination of standards, specifically such an important one as the $5\sigma$ one. They are in a way embedded decisions that must be considered in science, and as we have seen not in an immediate way, just to the scientists, but more broadly, to the institutions and general public.

Now, this project deals not with theoretical physics but with astrobiology, and although there is no standard of certainty as clear as the $5\sigma$ one, there are impacts that must be considered to know how the decision process in a discovery in astrobiology would entail. Mentioning the word "discovery", suggests that in this next section we must be extremely careful to note that the impacts that will be mentioned are a direct adaptation to those in the Higgs boson discovery, therefore, they relate to a "discovery" in astrobiology, hence a potential finding of extraterrestrial life. My own project focuses on a " constrained list of potential locations for life" not a "discovery", however first getting an idea of the impacts on a "discovery" in this field will be useful to later shift to my project, a constrained list of potential locations for life.

If we take the same methodology, and divide the impacts relating to a "discovery" in astrobiology into the same two categories: **"impacts in relation to the physicists"** and

**"impacts in relation to broader research goals, and in relation to the scientific community"**, the first category would then have to do with the impacts directly related to the subjects doing the astrobiological experiments and measurements, the scientists (astrobiologists). What could these impacts be?

- A discovery of extraterrestrial life, would also potentially involve a commitment or a certain type of license to adopt statements entailing the finding of such life as premises in the pursuit of further inquiries. Hence, a discovery of extraterrestrial life would then position that finding as a given, and would orient the future search for extraterrestrial life. All this to say, the discovery would directly impact the methodology and route astrobiologists take after such efforts.

- There would also be a great impact for other astrobiologists, working on the finding of extraterrestrial life and astrobiologists in other fields, by announcing that a discovery of life has been made, it will most likely again change the logical terrain.[35] Each scientist/investigator will have to decide (as an individual or as a member of a group) whether the evidence offered by the group of astrobiologists suffices to warrant accepting the discovery of extraterrestrial life as a premise or assumption in future work.

- Whoever disagrees (in the realm of astrobiology) will be faced with the responsibility to explain their reasons of disagreement, since they would go against accepted claims.

- Another impact in this first category would be that such a discovery claim, as is the discovery of extraterrestrial life, enables the pursuit of new astrobiological inquires that perhaps prior to these findings had not been deemed as important.

Now, apart from these impacts that translate pretty smoothly from particle physics into the field of astrobiology, there are others that fall into this first category. Let's explore them.

- There would have to be a justification for resources, the discovery of extraterrestrial life would potentially need a lot of funding, hence this institutional/governmental justification would be perhaps necessary.

Next, what about the second category, the **"impacts in relation to broader research goals, and in relation to the scientific community"** in astrobiology? How would those translate? Before delving fully into those impacts, it is important to make a crucial distinction. There will be impacts concerning broader research goals, and in relation to the scientific community, but these could be for a number of reasons. I am specifically interested in: (i)the impacts of a successful discovery of life, (ii)the impacts regarding an incorrect claim, and (iii) the impacts regarding an unsuccessful discovery (not finding life). The following impacts aim at addressing all three.

- Expectation for success: such a discovery would entail much expectation from the scientists and the general public since success at achieving the discovery of extraterrestrial life would be highly valued.
- Reputation and Responsibility: there is a huge indirect impact and responsibility related to the broader society (could be deeply linked to social values) in communicating the anticipatory claims in scientific talks. Such talks, not only aimed at the astrobiological community but to the media and the world, would make an incorrect claim of the discovery of extraterrestrial life embarrassing. This would compromise

credibility and status. Perhaps aside from social values, there would also need to be considered ethical values regarding reputation and credibility.

- Responsibility and Skepticism: additionally, a broader sense of responsibility toward the public perception of science in general could play a role in any successfully or unsuccessful discovery or incorrect claim of extraterrestrial life announcement. It could either increase or decrease existing scientific skepticism accordingly.
- Funding: another risk in a promised answer being wrong, (ii)regarding an incorrect claim, and (iii) regarding an unsuccessful discovery (not finding life). Both would portray negatively to the funding of astrobiologists in future projects.
- Opportunity: there is also an impact in the sense of 'opportunity'. Such a finding would have an intense public spotlight, hence there would be a crucial opportunity to show the astrobiologial efforts and such level of science to the general public, where as mentioned, there could be some public skepticism about astrobiological claims.

Again, what other impacts not directly taken from theoretical physics that fall into this second category?

- Ethical: the most pressing impact is the ethical impact not only on the science community but on a much larger scale. Claiming to have found life could make society feel like they are under threat, hence there would have to be careful implication in a given announcement that carries such news.
- Ethical: could our impacts in the successful search for extraterrestrial life potentially cause harm either to us or to the life we found? Are we intervening in a harmful way

in this search? Here it is important to consider the ethical impacts to the broadest reach, the life itself we find. By attempting to and successfully finding life, could we possibly cause a bigger harm, say by contamination? And consequently there could be the impact where by finding life they could cause harm to us, in an offensive way.

- Biological, metaphysical and ontological : a potential successful search for life would perhaps have further biological, ethical, metaphysical and ontological impacts in the sense that the *definition of life* would perhaps need reformulation. The question: *what counts as life?* would need to be reassessed, and redefined based on such a discovery.

- Personal: a potential successful search for life would pose existential questions about the self, would make us reconsider what our role is if we are not the only life in the cosmos. It would potentially create a sort of existential crisis revolution, in which we would have to give some serious thought to our very own definition of self.

- Epistemological: a really crucial impact to both scientists and the general public would be that such a discovery of extraterrestrial life, would mean a lot epistemologically. Having successfully found life, would mean accepting a claim. This means that the other boundaries of uncertainty in the project would not shatter, but new kinds of questions would arise, particularly epistemological ones. This is only due to what we mentioned before, a discovery in the search for life would have to pretty much include a very high standard of certainty, perhaps even more than five sigma. This is mainly due to the huge impacts (as have been stated) such a discovery would pose. This is crucial, since in a way such a standard does not quite exist on the astrobiology side of science. Perhaps this is mainly due to how such a discovery has not been made yet. However, it is important to consider that it would shift the terrain, and it would definitely stir crucial and interesting questions.

Furthermore, impacts related to both categories:

- Credibility and Certainty: the first impact that would definitely affect both categories would be the impact that the credibility or certainty of the successful or unsuccessful find would have. A crucial aspect is that as mentioned before to have claimed to have found life solicits a perhaps an even stronger standard than $5\sigma$ is needed. Since a discovery in astrobiology would have huge ethical and social implications, the credibility terrain of such finding would have to be extremely high, (this is crucial as perhaps none exists). To direct this to a lot of areas developed in this project, something as serious as a *discovery* in astrobiology, would have less inductive risk. There would be more evidence to warrant the inference. However, a constrained list of potential locations for life could have plenty. It is worth to recall the fact that accepting or rejecting a claim does not change the truth of that claim. It changes how we *regard* it. To the claim in itself, it makes no difference.

Hopefully, haven gotten a sense of the impacts of a future “discovery” in astrobiology aided the reader in terms of context, however, it is time to shift our attention to the actual impacts on this project: which is a constrained list of potential locations for life of potential possibilities for life, not a “discovery”. It is important to recall that a constrained list of potential locations for life in a way is a *claim*.

Let us jump right in into the first category, **“impacts in relation to the physicists”**

- A constrained list of potential locations for life of potential locations in the search for life, would also (although in a smaller scale) potentially involve a commitment or a certain type of license to adopt statements entailing the constrained list of potential locations for life of potential possibilities for life as premises in the pursuit of further inquiries. Hence, a constrained list of potential locations for life of potential possibilities for life would then position that method as a given, and would orient the future search for extraterrestrial life. All this to say, possibly, not so much as with the discovery, but a potential constrained list of potential locations for life would directly impact the methodology and route astrobiologists take after such efforts.

- A constrained list of potential locations for life of potential locations for life, would pose an impact for other astrobiologists by again, possibly in changing the logical terrain.[35] Each scientist/investigator will have to decide (as an individual or as a member of a group) whether the evidence offered by the group of astrobiologists suffices to warrant accepting the constrained list of potential locations for life of potential locations for life as a premise or assumption in future work.

- Epistemological: there could be an impact of reliability and interpretation of observations. This could impact the immediate science community and institution, in the sense that the constrained list of potential locations for life has to have a good foundation, since this would potentially lay a good ground for similar future efforts in the field. To put it another way, perhaps the most certain the constrained list of potential locations for life is, the more it can be replicated and perfected.

Regarding the second category: **"impacts in relation to broader research goals, and in relation to the scientific community"** in a constrained list of potential locations for life of potential possibilities for life, some of the impacts could be:

- Epistemological: a constrained list of potential locations for life is not a discovery and hence then it would not be a successful search for life, instead a potential possibility for life. This would very much include uncertainty and would invoke notions concerning the limits of astrobiological knowledge. How certain is this constrained list of potential locations for life? Can we trust it? What is possible to find and not find? It is important to note that **the boundaries on the epistemological impacts of an astrobiological constrained list of potential locations for life are uncertain.**

- Epistemological and reliability: impact of reliability and interpretation of observations. This could impact the broader scientific community in the sense that the constrained list of potential locations for life has to be well-supported enough not to create further skepticism.

Now that we have gone very much in detail into the exact impacts a potential "discovery" of extraterrestrial life and a potential " constrained list of potential locations for life" of possible locations for life would present, it is essential to question, what does this all mean? Going back to Staley might be useful.

Staley rightfully notes that the impacts just mentioned include both "the costs of error and the benefits of getting it right". [35] Both, which are crucial to consider in the discovery or constrained list of potential locations for life of extraterrestrial life. However, he asks if these are epistemic considerations? (As opposed to say ethical or social). He brings up Daniel Steel, mentioning they proposed that what is distinctive of epistemic values is that they “promote the attainment of truth,” either*intrinsically*, in that “manifesting that value constitutes an attainment of or is necessary for truth” or *extrinsically*, in that “manifesting the value promotes the attainment of truth without constituting the attainment of or being necessary for truth”.[35]

If we rely on Steel’s proposed criterion, the impacts described both in the Higgs boson case and in astrobiology, could seem to qualify as epistemic, but only in the *extrinsic* sense. ”The benefits of correctly accepting the claim to have discovered a new boson in no way make that claim more likely to be true. Neither are they prerequisites for its being true”.[35] It is important to note that there is a significant difference between an incorrect constrained list of potential locations for life as a claim and an incorrect discovery as a claim.

Later on, Staley also mentions some further epistemological considerations, the costs and benefits of having a relationship with the truth. An important concept that stands out is ”epistemic autonomy”. Staley now brings up Levi, mentioning that ‘epistemic autonomy’ is: parts of physics or science autonomous of epistemic concerns, parts only interested in truth. [35] ”Levi contends that such epistemic autonomy is preserved so long as whatever value judgments the scientist makes exert their influence only via the determination of an investigator’s degree of caution.”[35]

This introduces the role of scientists relating to value judgments and criteria of degree of caution. In a way, the impacts that we went over, “the costs of error and the benefits of getting it right”. [35] are not simply epistemic considerations. This would very much be aligned with the Value-Free Ideal, and would provide a very closed in view of science, which might not be aligned with the scientist's goals. The impacts, both on the case of the Higgs boson, but additionally in the potential discovery of extraterrestrial life or a potential constrained list of potential locations for life, might be closer related to the kinds and roles of values in science, as introduced by Heather Douglas. [10]

To strengthen this, we might want to go back to Staley once more with regard to the idea and goals of scientific knowledge and science in general. He notes: "allowing, therefore, that Levi's cognitive decisions occur as the outcome of attempts to seek the truth and nothing but the truth, they are insufficient for the production of scientific knowledge. For that, investigators must decide on the most beneficial action to take in response to the results in hand. It may well be that, as in the Higgs search, the benefits that are relevant to that decision are not practical in the narrow sense, but accrue to science understood as a knowledge-generating enterprise. Nonetheless, it is a practical decision— i.e., one regarding not just what to believe but what action to take."

The impacts might aid the scientists in determining what kinds of values and roles in science matter and at what stages. Evaluating the impacts of your project, as a scientist, therefore force you to evaluate the values you are using and the roles they are in.

### *7.1.7 Uncertainty Quantification*

This next section portrays an analysis of the concept of "Uncertainty Quantification' (UQ)' regarding Winsberg's paper: "Values and Uncertainties in the Predictions of Global Climate Models" [39]. As this project aims at conceptualizing uncertainty, how does this approach work? Winsberg's main claim about UQ is that it cannot work well in climate science. Could it be a useful resource for astrobiologists?

The problem of Uncertainty Quantification (UQ) might be relevant to this project. We will jump now to analyze this problem in relation to climate science. Specifically, the problem arises in giving quantitative estimates of the degrees of uncertainty associated with the predictions of global and regional climate models. Why study Uncertainty Quantification in relation to climate science? As stated by Winsberg, "UQ, I would like to argue, is first and foremost a tool for communicating knowledge from experts to policymakers in a way that is meant to be free from the influence of social and ethical values" [39] Pausing here for a moment is essential. Astrobiologists might not communicate with policymakers directly, but could still find UQ appealing.

From here we should sense there is a problem. Simply *freeing* ourselves from social and ethical values, is pushing us further away from Douglas and an all-encompassing redefining view of science. As we say with Staley, these values matter.

Let us entertain the argument to climate models. Winsberg notices that "the standard ways of using probabilities to separate ethical and social values from scientific practice cannot be applied in a great deal of climate modeling". [39] He mentions this is mainly because "the roles of values in creating the models cannot be discerned after the fact- the models are too complex and the result of too much distributed epistemic labor." [39] He then argues that typical approaches for handling ethical/social values in science do not work well here.[39] Or should not work in any area of science.

It is important to understand where the need to produce quantitative estimates of uncertainty comes from in the first place. Winsberg suggests that UQ, is "first and foremost a tool for communicating knowledge from experts to policymakers. Experts, in this case, climate scientists and climate modelers, have knowledge about the climate."[39] Here, it is key to understand the role that astrobiology should play if, say, no policy is involved. How could this tool be used? Who makes the decisions, who are the experts? "In one sense, they are the people who ought to be considered best situated to make decisions about what we ought to do in matters related to climate. But in another sense, they are not."[39]

There is, from this very first point, a social aspect to the experts themselves. "One clear motivation for solving the problems of UQ, in other words, is to maintain this division of labor between the epistemic and the normative—between the people who have the pure scientific expertise and the people with the legitimate ability to represent the values of the relevant stakeholders. And so if we want to understand where the need to produce quantitative estimates of uncertainty comes from, we need to delve into the role of social values in the administration of scientific expertise."[39]

Winsberg does not distinguish the difference between social and ethical values, and recognizes Ernan McMullin's use of the word "epistemic values" to flag the difference to what we have seen Douglas mentions in the Value-Free Ideal as "non-epistemic values". For Winsberg it goes beyond that. "Setting constraints on experimentation, for example, deciding which projects to pursue and which projects to ignore, are choices that uncontroversially reflect social values."[39]

To deal with these social and ethical values, Winsberg proposes to look back to Rudner's argument all the way from Inductive Risk: "1. The scientist qua scientist accepts or rejects hypotheses. 2. No scientific hypothesis is ever completely (with 100 percent certainty) verified. 3. The decision to either accept or reject a hypothesis depends upon whether the evidence is sufficiently strong. 4. Whether the evidence is sufficiently strong is 'a function of the importance, in a typically ethical sense, of making a mistake in accepting or rejecting the hypothesis.' 5. Therefore, the scientist qua scientist makes value judgments."[39]

Rudner's conclusion was that "how sure we need to be before we accept a hypothesis will depend upon how serious a mistake it would be to accept it and have it turn out false"[39]. How does this relate to particular sciences? Winsberg offers an example in climate science. "Consider a prediction that, given future emissions trends, a certain regional climate outcome will occur. Should we accept the hypothesis, say, that a particular glacial lake dam will burst in the next 50 years? Suppose that if we accept the hypothesis, we will replace the moraine with a concrete dam. But whether we want to build the dam will depend not only on our degree of evidence for the hypothesis, but also on how we would measure the severity of the consequences of building the dam, and having the glacier not melt, vs. not building the dam, and having the glacier melt."[39]

This strengthens Rudner's argument in the sense that, that as long as the evidence is not one hundred percent conclusive, we cannot justifiably accept or reject the hypothesis without making reference to our social and ethical values. [39]

How would this look in astrobiology? The extreme case being, we cannot accept there is life, or a signal of life without the evidence being conclusive, and for it to be, would social and ethical values suffice? We certainly cannot claim life unless we are epistemically certain, but again, as in climate science, there is and will be uncertainty. Can we ever be certain? Therefore, does this make Rudner's claim to say, we must pull from those ethical and social values to accept life? Because in that case, life might not be found at all. Additionally, are social and ethical values needed in order to conceptualize the uncertainty in the search for life? There seems to be the pretty obvious answer, which is, the constraints on what life is, can draw on these values, but a balance has to exist between those and the “epistemological” ones. All coming with Rudner’s mention of risk.

Richard Jeffrey responds to Rudner saying his first premise is wrong. “The proper role of scientists, he urged, is to assign probabilities to hypotheses with respect to the currently available evidence.”[39]

Finally, we arrive at UQ, why is UQ of relevance? Winsberg states it is a tool for dividing our intellectual labor. “If we were entirely comfortable simply letting experts qua experts decide for us how we should act, then we would not have such an acute need for UQ.”[39]

This urges to desperately return to Douglas. Regarding, Douglas contra Jeffrey. “Jeffrey’s goal of separating the epistemic from the normative cannot be achieved using UQ based on statistical ensemble methods. But Heather Douglas’s (2000) discussion of the debate about science and values should have made this clear from the beginning.”[9]

As we have seen from the examples of inductive risk from Douglas [9] “By factoring the specificity and sensitivity of the method into their degrees of belief, they are essentially eliminating or “screening out” the influence of the social or ethical values that otherwise would have been present And if they could do this, social and ethical values, at least the kind that normally play a role in the balance of inductive risks, would not to play a role in their assessments of the probabilities. Let us call this the Bayesian response to the Douglas challenge (BRDC).”[39]

Relating it to Douglas, you are kidding yourself if you think UQ is value free. Astrobiology is another example of considering inductive risk based on how we weigh data and models. In UQ, epistemic labor widely distributed, UQ no sense entrenchment of value decisions. UQ not objective, still value laden.

As mentioned before, I believe this paper is helpful in returning to Douglas, and noting that in any case, policy related or not, we are doing science. In this doing of science, all types of values must come into play. The only thing worth considering is *when* and *where* these values come into play. For astrobiology, a more direct lens is needed to regard these values in a way that allow for the best science, and to understand where the uncertainty is coming and what we can do with it. There is more than one lens pointing at the sky, and there might be a few needed to point to us.

### *7.1.8 Values, their Roles and Uncertainty in the Inner Workings of My Project*

This project is flooded with uncertainty, in which the most obvious statement to make is that the role that values are in that actually matter is the Indirect Role. This is crucial since it emphasizes that the decisions that will be made require values to weight the uncertainty and determine the importance of the inductive gaps left by the evidence. The values that are going to matter here are in the indirect role. What does this mean? It means that since most of the stages in this project have uncertainty, the ethical, social and cognitive values need to be oriented to weighting the uncertainty. How are we going to weigh it? Well, what becomes useful at this stage is to use the provided examples (Higgs boson[35] and Uncertainty Quantification (UQ) [39]) to establish that as theoretical as the science is, there needs to be importance assigned to social and ethical values as well. Therefore, to “analyze” and “conceptualize” the uncertainty we need to pull from the ethical and social as well as the cognitive.

How can this project help us conceptualize uncertainty? UQ tells us that science is full of value judgements in models using constantly inductive risk. The epistemology of the algorithm that was developed in this project takes in parameters. It is hard to build a new model that does not have inductive risk already "baked in", therefore, all we can do is really analyze what we are going to do with this risk. UQ does not work since the sources of uncertainty resist quantification. When evaluating evidence, Douglas becomes meaningful, since it is more about the roles values are in science than the kinds.

Given the methods of conceptualizing uncertainty, we can conclude:depending on your goals as a scientist, climate science (Winsberg), theoretical high energy physics (Staley), and inductive risk (Douglas) help you determine parameters.

# 8

# Embracing Uncertainty: starlights

> “He knew that his memory of the piano falsified still further the perspective in which he saw the music, that the field open to the musician is not a miserable stave of seven notes, but an immeasurable keyboard (still, almost all of it, unknown), on which, here and there only, separated by the gross darkness of its unexplored tracts, some few among the millions of keys, keys of tenderness, of passion, of courage, of serenity, which compose it, each one differing from all the rest as one universe differs from another, have been discovered by certain great artists who do us the service, when they awaken in us the emotion corresponding to the theme which they have found, of shewing us what richness, what variety lies hidden, unknown to us, in that great black impenetrable night, discouraging exploration, of our soul, which we have been content to regard as valueless and waste and void.”
>
> *-Marcel Proust*
> *In Search of Lost Time: Swann’s Way*

Portrayed here is a short chapter dedicated to another humble take on uncertainty. Once uncertainty goes beyond what is possible to conceptualize, we ought to take another route: embracing it.

With the help of a very esteemed composer, Artemy Mukhin, we take on this task. For this quest, a musical piece has made it to this project. Before immersing ourselves in its embrace, let us hear from the composer himself.

## 8.1 A Note from the Composer

Note written for the piece “starlights” composed by Artemy Mukhin in 2024.

“Uncertainty is a crucial part of music and is present in my composition, ‘starlights’ as well as in my process of composing. In the most obvious sense, composing music is not a linear and clear process, and I am often unaware of what will be the next note or phrase. I rely on my musical intuition (i.e sense of harmony and melody, where does the music ‘want to go’) as well as a careful analysis of what material I have used previously in order to lessen this uncertainty. The latter is a more logical and analytical approach, which when combined with intuition and emotion can hopefully create an interesting and coherent piece of music. When writing ‘starlights’ I tried to be very cautious and frugal about the material I would use – I outlined several “profiles”, or small musical ideas, which I would use all throughout the piece. These musical ideas would then be transposed, expanded and modified – this was done in hope of creating a global narrative and allow the listener to not get lost in the music.”

A performance of “starlights” can be found here (https://www.youtube.com/watch?v=fOsaHX9cby0).

This was performed at Bard Conservatory by Chris Gross (cello) and Christopher Oldfather (piano) in April, 2024 as part of “Da Capo Chamber Players: Premieres by Student Composers” The sheet music can be found in the appendix.

# 9
# Discussion and Conclusions

> “We passionately long that there may be another life in which we shall be similar to what we are here below. But we do not pause to reflect that, even without waiting for that other life, in this life, after a few years we are unfaithful to what we have been, to what we wished to remain immortally.”
>
> *-Marcel Proust*
> *In Search of Lost Time: Sodom and Gomorrah*

Uncertainty is a challenging concept to conceptualize, but not an impossible one. Hopefully this project has aided the reader in such efforts. As much as there is more work to be done both in the physical pursuits and philosophical investigations, this is hopefully a preliminary trial in such a quest. The elaboration and functioning of P.U.S.H. allowed for a user to set a constrained list of planetary locations as well as a constrained list of molecular characteristics to finally provide*potential biosignature targets given location.* It does this by taking in intricate sets of planetary and chemical data sets. In essence, it provides a humble P.U.S.H. for efforts in the search for life outside our home, Earth. However, this is very much a preliminary algorithm, that is deeply adaptable, and one that in fact should be adapted to output more complex and intricate results. Nonetheless, it hopefully provided with a way to think deeply about the connections of parameters and locations.

Touching upon this, the philosophical investigations of this project and the findings were the result of a deep scientific and philosophical analysis. Key aspects of it include a thorough investigation and understanding into inductive risk. With the help of Heather Douglas, we were able to notice that for a conceptualization of uncertainty, the Value-Free Ideal needs some serious reconsideration and where rich and meaningful work in astrobiology is present is when we are able to characterize the *roles* values are playing in our scientific process.

As much as in policy oriented research such as climate science, astrobiologists are constantly trying to navigate and make decisions pertaining to Inductive Risk, such as the risk of missing a good location or choosing an unpromising one. Astrobiologists might be driven to think this is solely an epistemic problem, however it is permeated with other values such as social and ethical, more so it relies on the role these values are in. Hopefully by now, more considerations than simply ethical ones should come to mind when considering astrobiology, particularly an image is portrayed with values. And more specifically, the way it weighs evidence supports empirical claims.

Staley and the Higss boson aided us immensely in portraying a crucial example from high energy physics where the *impacts* of scientific endeavors should apply to our value-judgements while doing science. Specifically influential was the application of such impacts into astrobiology, highlighting with extreme precision where a *discovery* versus a *constrained list of potential locations for life* in this field matter, and to what extent.

Finally, Winsberg and the concept of Uncertainty Quantification brings us back to the idea of inductive risk. Through climate science and its models we can extrapolate how these inductive risk considerations which are baked into the nooks and crannies of the field. UQ showcased that the sources of uncertainty resist quantification.

Proust and music aided us as well, in portraying how aside from the sciences, uncertainty can and will be present throughout the arts. This take on uncertainty did not aim to conceptualize it, since overall, we cannot get rid of this uncertainty, all we can do is perfect our measure at conceptualizing it, or *embrace* it.

Overall, some astrobiologists can aim to reduce the uncertainty, but philosophers can remind you it is here, and offer a way of thinking about how to weight the risks of inductive leaps. Specifically, how to think clearly and honestly about how inductive risk can aid us in our considerations to weight evidence.

The lens of philosophers, can point towards what resources you may have in order to navigate it, it portrays a perspective to think about inductive risk and uncertainty honestly and communicate with the public.

However helpful this lens is, it is clear it cannot get rid of this uncertainty, all we can do is grab another lens, by try to perfect our measure at conceptualizing it, and keep aiming.

music

# starlights

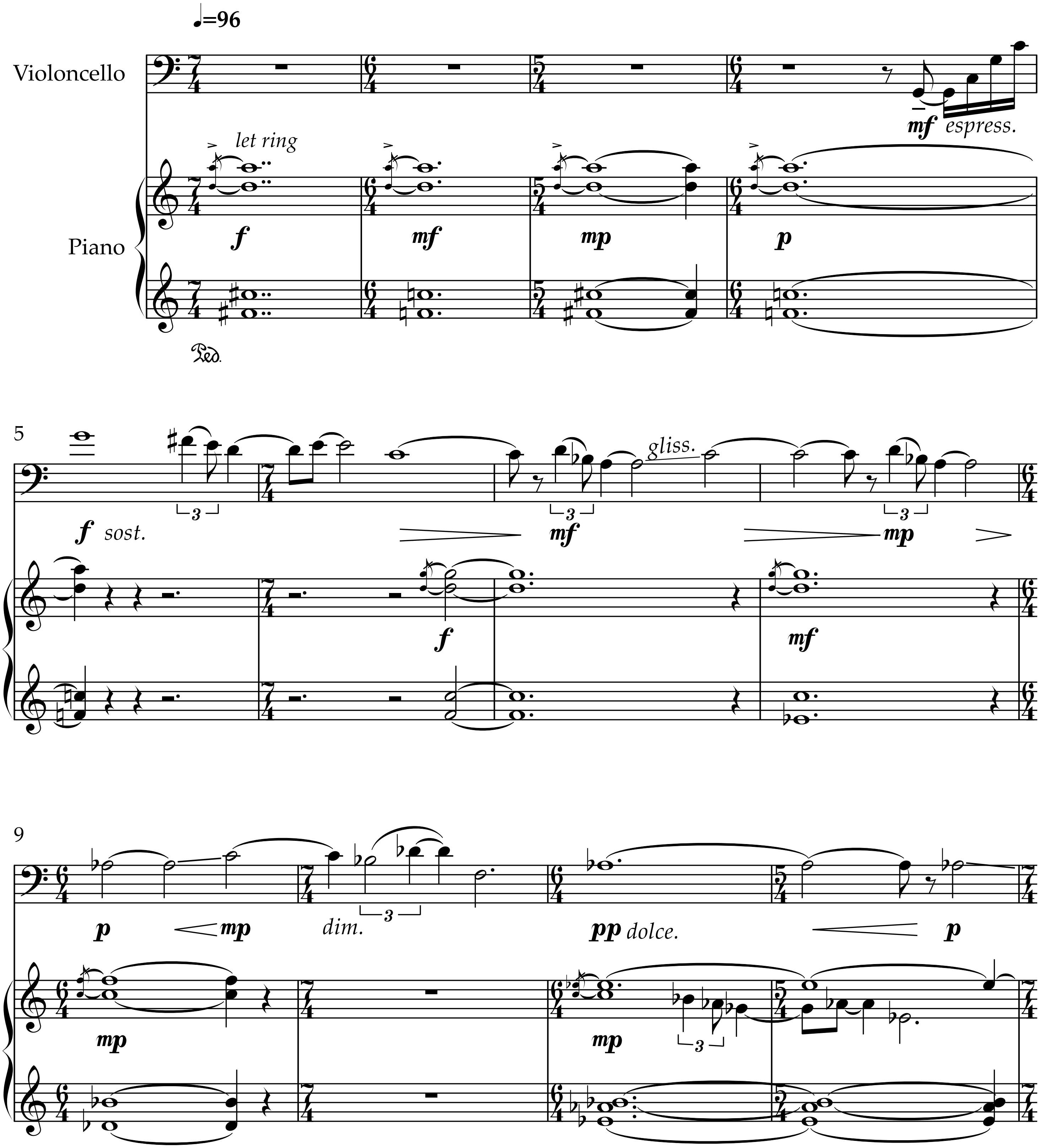
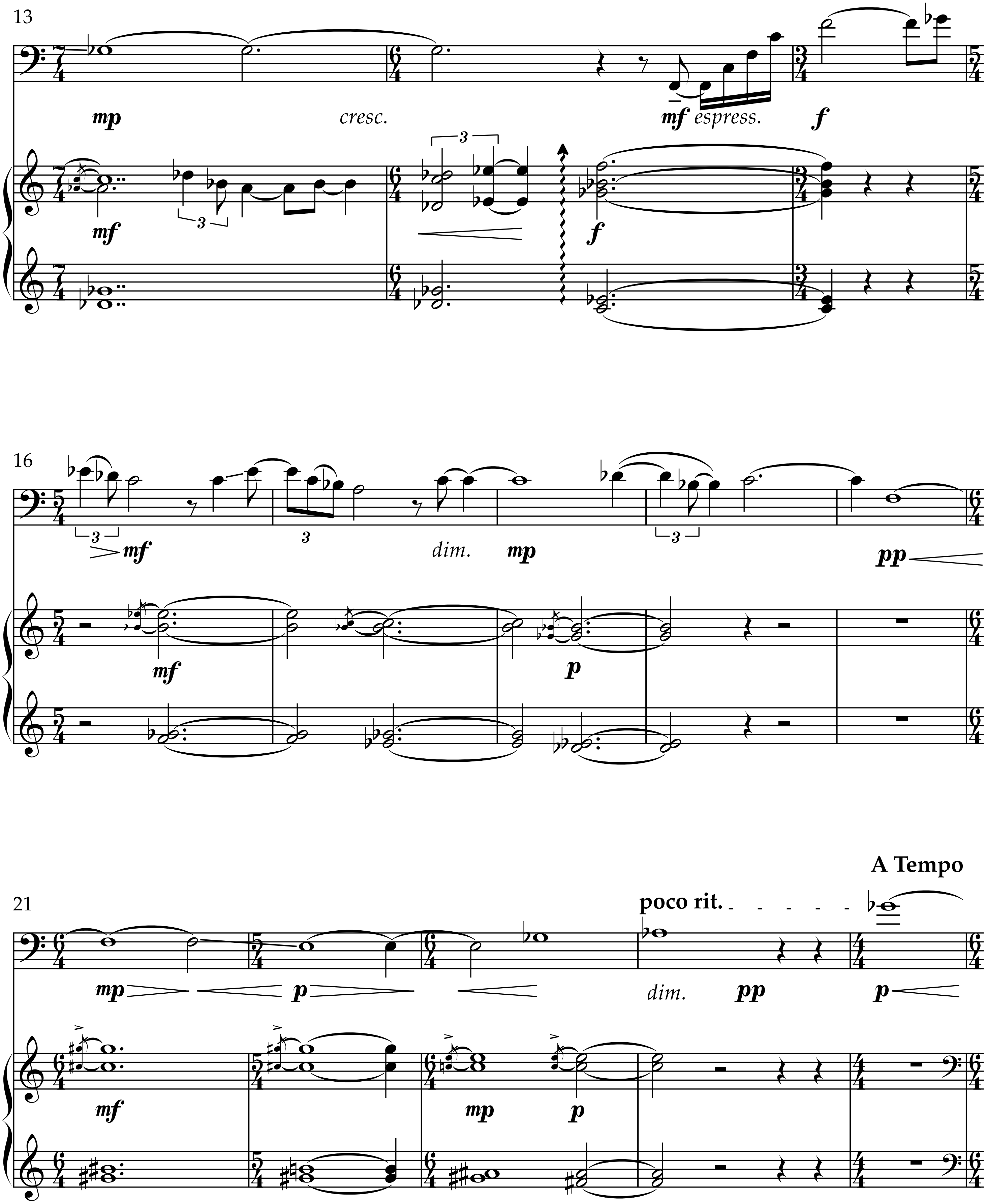
## For cello and piano

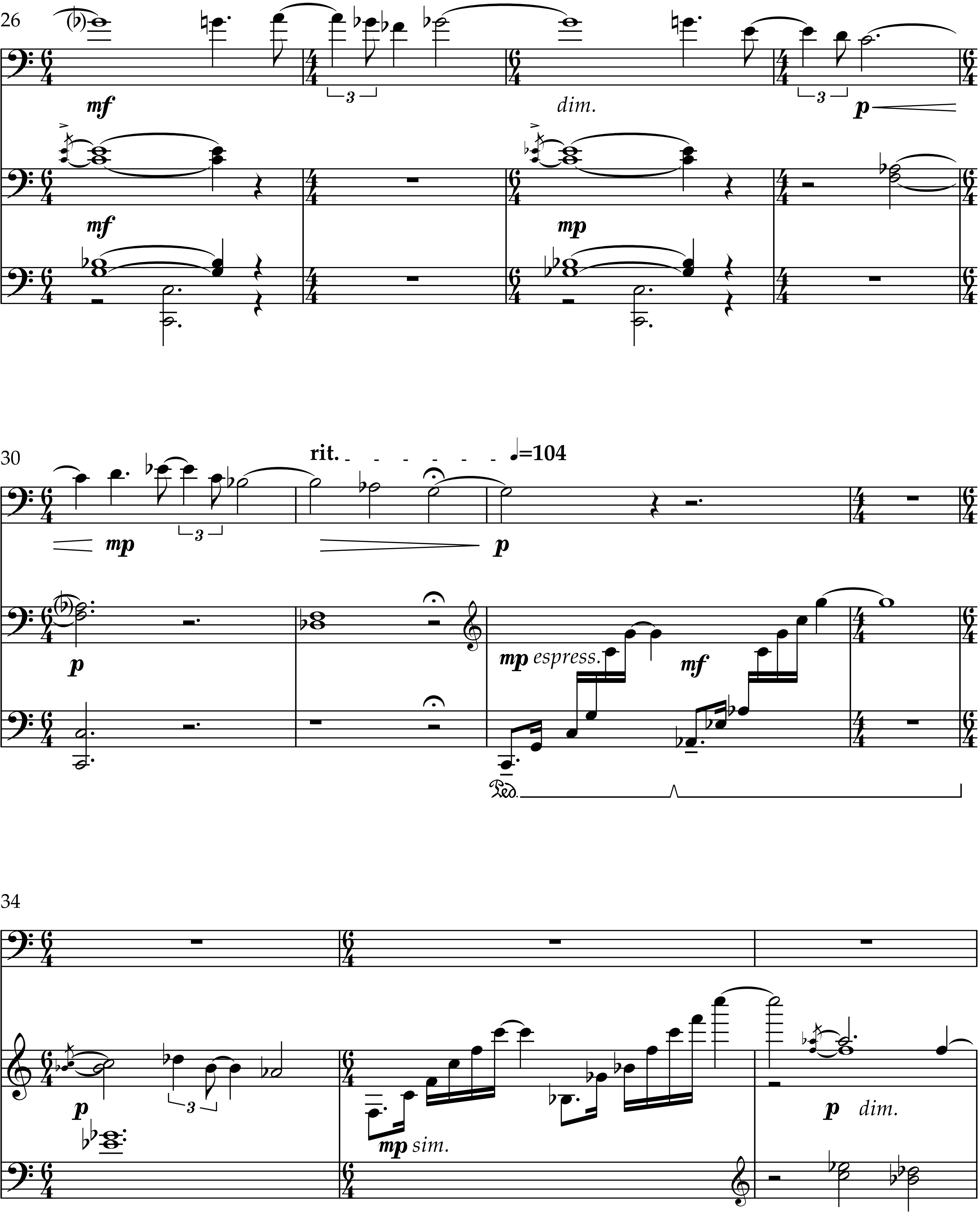
By: Artemy Mukhin

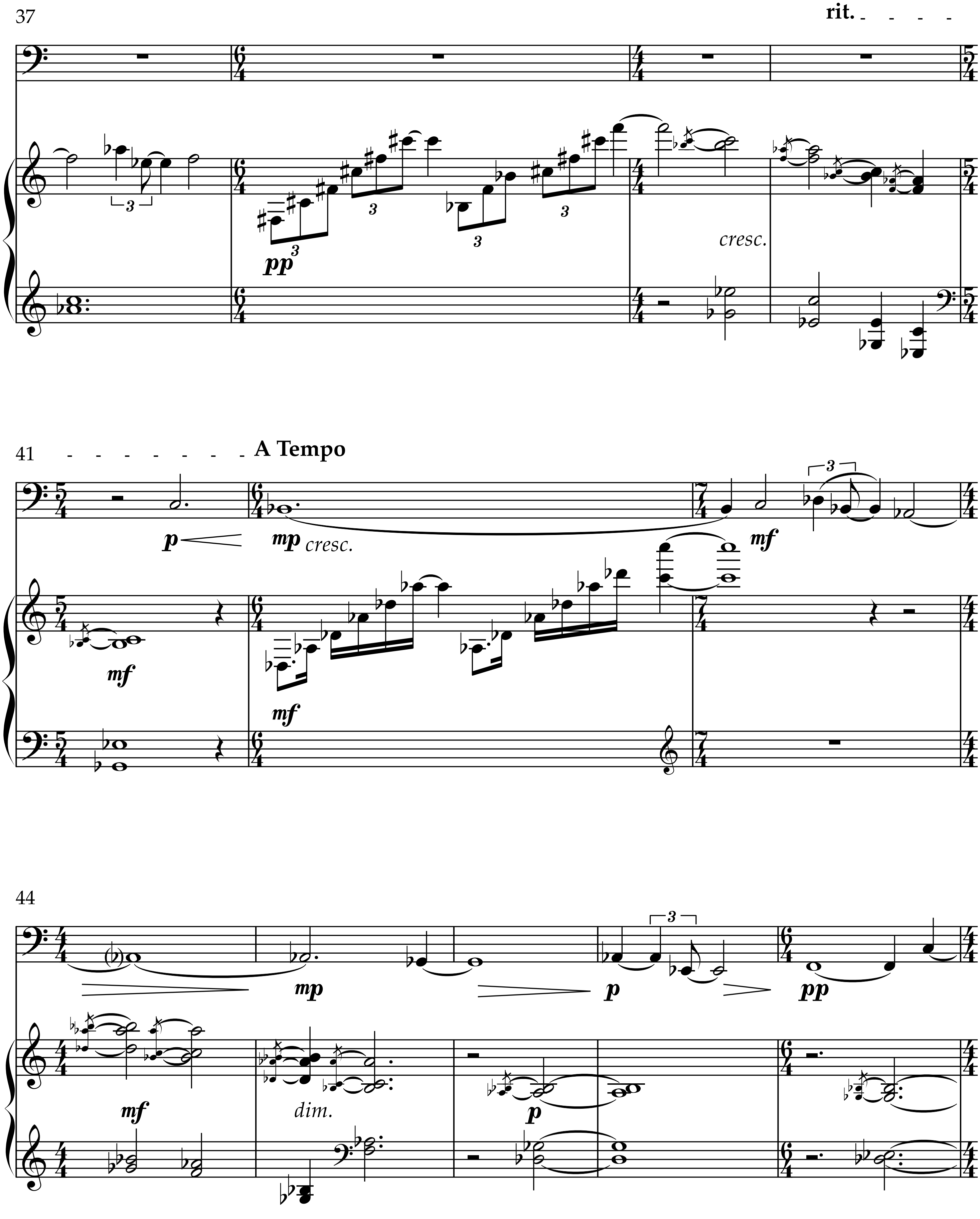
*Written for the Da Capo Chamber Ensemble*

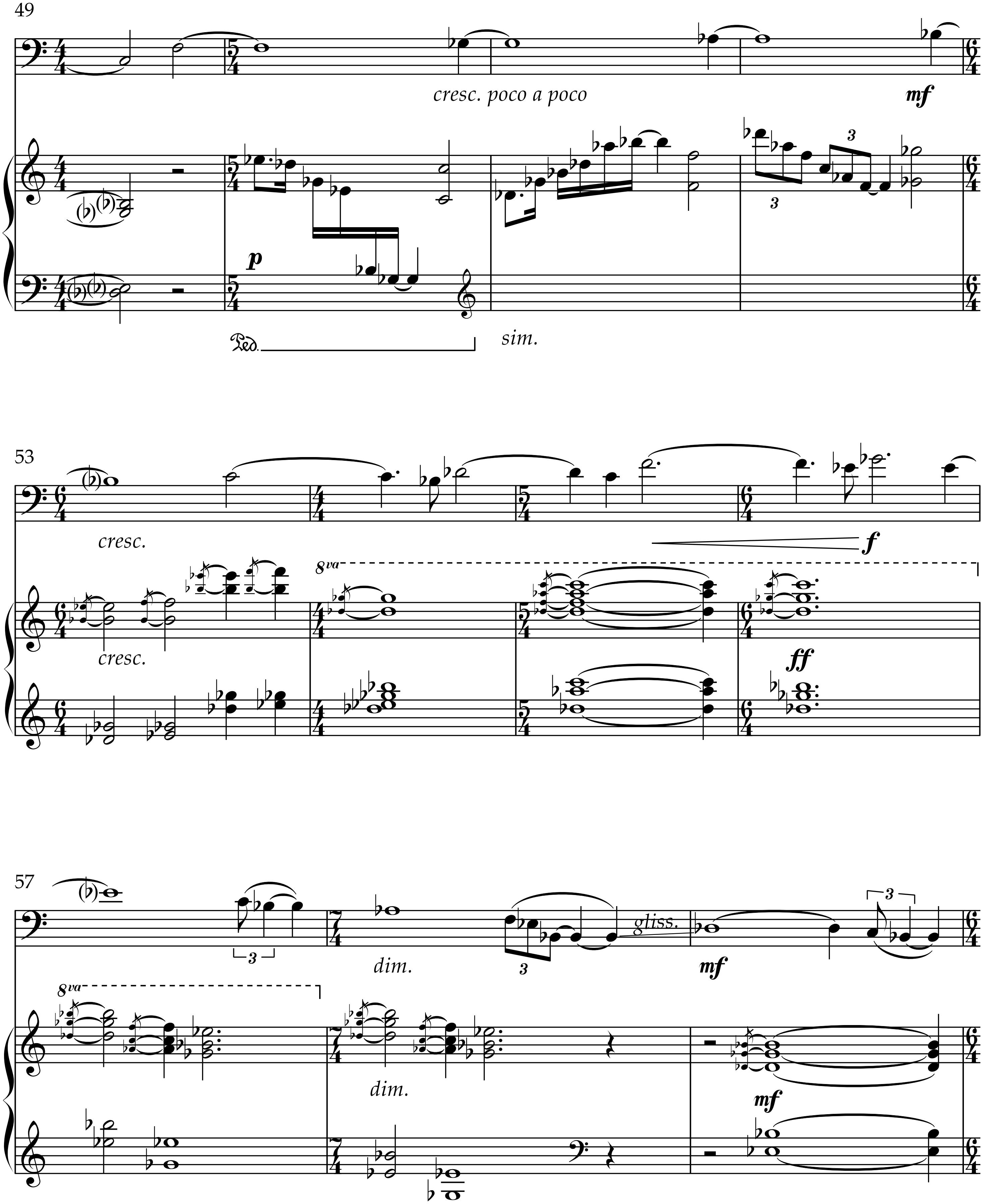
# starlights

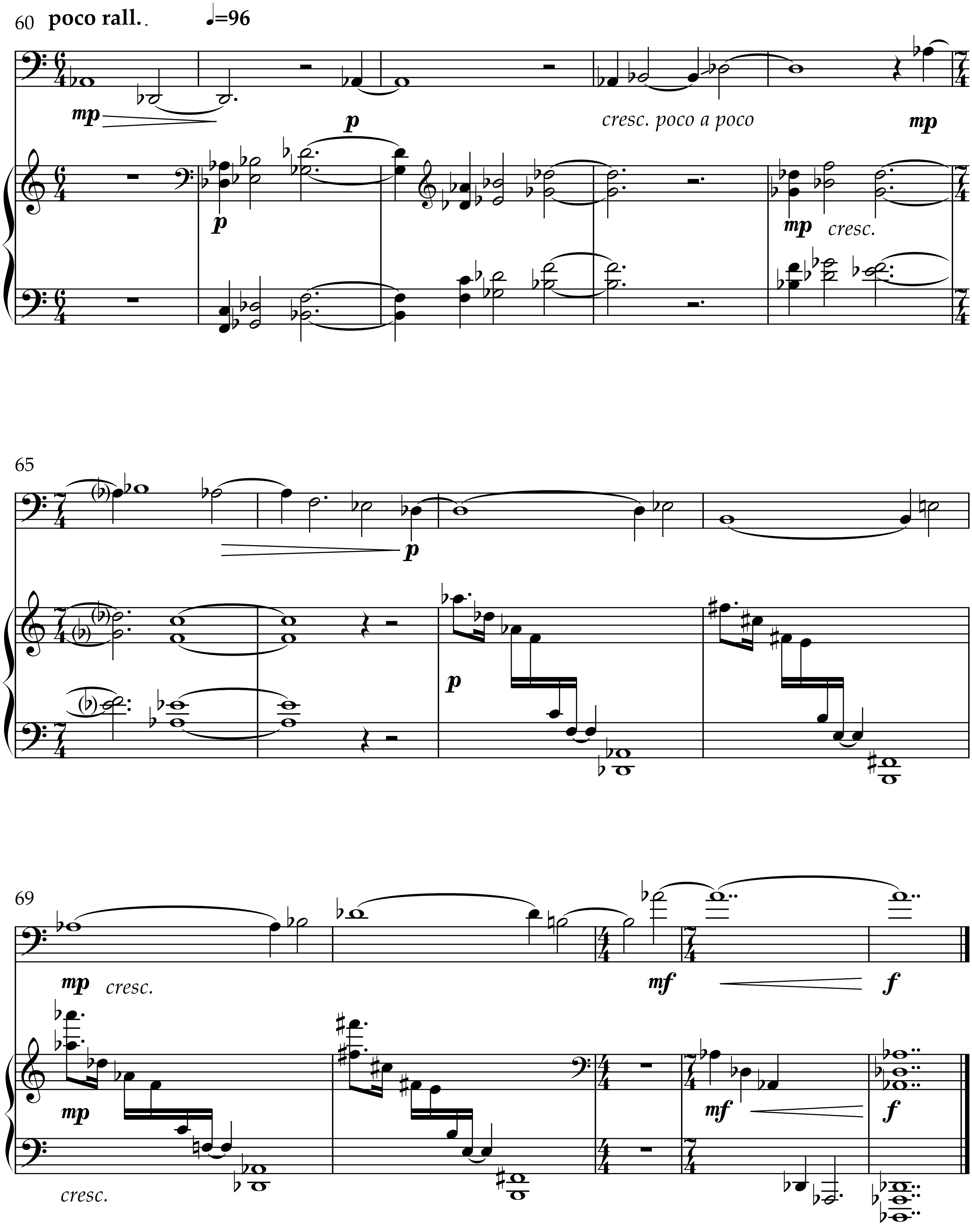
for piano and cello

**Artemy Mukhin**


13
mp
cresc.
mf espress.
f
mf
3
3
f
16
3
mf
3
dim.
mp
3
pp
mf
p
A Tempo
21
poco rit.
mp
p
dim.
pp
p
mf
mp
p


26
mf
dim.
p
mf
mf
mp
30
rit.
♩=104
mp
p
p
mp espress.
mf
Ped.
34
p
mp sim.
p
dim.


37
rit. - - - -
3
pp
3
3
3
3
cresc.
41
- - - - - - - -
A Tempo
p
mp cresc.
mf
mf
3
mf
44
mp
3
p
pp
mf
dim.
p


49
cresc. poco a poco
mf
p
sim.
53
cresc.
cresc.
f
8va
ff
57
8va
3
dim.
3
gliss.
mf
dim.
mf


60
poco rall.
♩=96
mp
p
p
cresc. poco a poco
mp
mp
cresc.
65
p
p
69
mp
cresc.
mp
cresc.
mf
f
mf
f

# Bibliography

[1] National Academies, *Origins, worlds, and life: A decadal strategy for planetary science and astrobiology 2023-2032*, 2024.

[2] The European Space Agency, *Plato factsheet*, 2024.

[3] NASA Astrobiology, *False positives, false negatives; the world of distant biosignatures attracts and confounds*, 2024.

[4] Sarah Ballard, *Predicted number, multiplicity, and orbital dynamics of tess m-dwarf exoplanets*, The Astronomical Journal **157** (2019), no. 3, 113.

[5] Pat Brennan, *What's a transit?*, 2024.

[6] Encyclopaedia Britannica., *Stellar classification*, 2024.

[7] Keith Cooper, *Search for intelligent aliens explores new radio-frequency realms*, 2023.

[8] Steven J Dick, *Space, time, and aliens*, Springer, 2020.

[9] Heather Douglas, *Inductive risk and values in science*, Philosophy of science **67** (2000), no. 4, 559–579.

[10] ______, *Science, policy, and the value-free ideal*, University of Pittsburgh Pre, 2009.

[11] Courtney D Dressing and David Charbonneau, *The occurrence of potentially habitable planets orbiting m dwarfs estimated from the full kepler dataset and an empirical measurement of the detection sensitivity*, The Astrophysical Journal **807** (2015), no. 1, 45.

[12] David Dunér, Gustav Holmberg, and Erik Persson, *The history and philosophy of astrobiology: Perspectives on extraterrestrial life and the human mind*, Cambridge Scholars Publishing, 2013.

[13] Billy Edwards, Lorenzo Mugnai, G. Tinetti, Enzo Pascale, and Subhajit Sarkar, *An updated study of potential targets for ariel*, The Astronomical Journal **157** (201905), 242.

[14] Billy Edwards, Lorenzo Mugnai, Giovanna Tinetti, Enzo Pascale, and Subhajit Sarkar, *An updated study of potential targets for ariel*, The Astronomical Journal **157** (2019), no. 6, 242.

[15] exoplanet TEAM, *The extrasolar planets encyclopaedia*, 2023.

[16] Responsible NASA Official for Science: Dana Bolles, *The habitable zone*, 2024.

[17] Adam Frank and Woodruff T Sullivan III, *A new empirical constraint on the prevalence of technological species in the universe*, Astrobiology **16** (2016), no. 5, 359–362.

[18] François Fressin, Guillermo Torres, David Charbonneau, Stephen T Bryson, Jessie Christiansen, Courtney D Dressing, Jon M Jenkins, Lucianne M Walkowicz, and Natalie M Batalha, *The false positive rate of kepler and the occurrence of planets*, The Astrophysical Journal **766** (2013), no. 2, 81.

[19] Leah Henderson, *The Problem of Induction*, The Stanford encyclopedia of philosophy, 2022.

[20] Greg Lusk and Kevin C Elliott, *Non-epistemic values and scientific assessment: an adequacy-for-purpose view*, European Journal for Philosophy of Science **12** (2022), no. 2, 35.

[21] Claudio Maccone, *Statistical drake–seager equation for exoplanet and seti searches*, Acta Astronautica **115** (2015), 277–285.

[22] Victoria S Meadows, Christopher T Reinhard, Giada N Arney, Mary N Parenteau, Edward W Schwieterman, Shawn D Domagal-Goldman, Andrew P Lincowski, Karl R Stapelfeldt, Heike Rauer, Shiladitya DasSarma, et al., *Exoplanet biosignatures: understanding oxygen as a biosignature in the context of its environment*, Astrobiology **18** (2018), no. 6, 630–662.

[23] Kristo Ment and David Charbonneau, *The occurrence rate of terrestrial planets orbiting nearby mid-to-late m dwarfs from tess sectors 1–42*, The Astronomical Journal **165** (2023), no. 6, 265.

[24] MIT, *Tess mission*, 2024.

[25] NASA, *Habitable worlds observatory*, 2024.

[26] Jon Nelson, *The golden record*, 2023.

[27] Erik A Petigura, Andrew W Howard, and Geoffrey W Marcy, *Prevalence of earth-size planets orbiting sun-like stars*, Proceedings of the National Academy of Sciences **110** (2013), no. 48, 19273–19278.

[28] George R Ricker, DW Latham, RK Vanderspek, KA Ennico, G Bakos, TM Brown, AJ Burgasser, D Charbonneau, M Clampin, LD Deming, et al., *Transiting exoplanet survey satellite (tess)*, American astronomical society meeting abstracts# 215, 2010, pp. 450–06.

[29] Carl Sagan, *Is there life on earth?*, Ballantine Books Edition **1.0** (1997), 193.

[30] Edward W Schwieterman, Nancy Y Kiang, Mary N Parenteau, Chester E Harman, Shiladitya DasSarma, Theresa M Fisher, Giada N Arney, Hilairy E Hartnett, Christopher T Reinhard, Stephanie L Olson, et al., *Exoplanet biosignatures: a review of remotely detectable signs of life*, Astrobiology **18** (2018), no. 6, 663–708.

[31] CERN Accelerating science, *Why do physicists mention "five sigma" in their results?*, 2024.

[32] Kristina Šekrst, *Astrobiology in philosophy or philosophy in astrobiology?* (2022).

[33] Clara Sousa-Silva, Janusz J Petkowski, and Sara Seager, *Molecular simulations for the spectroscopic detection of atmospheric gases*, Physical Chemistry Chemical Physics **21** (2019), no. 35, 18970–18987.

[34] Clara Sousa-Silva, Sara Seager, Sukrit Ranjan, Janusz Jurand Petkowski, Zhuchang Zhan, Renyu Hu, and William Bains, *Phosphine as a biosignature gas in exoplanet atmospheres*, Astrobiology **20** (2020), no. 2, 235–268.

[35] Kent W Staley, *Decisions, decisions*, Exploring Inductive Risk: Case Studies of Values in Science (2017), 37.

[36] Matthias Steup and Ram Neta, *Epistemology*, The Stanford encyclopedia of philosophy, 2024.

[37] NASA Science Editorial Team, *Are we alone in the universe? revisiting the drake equation*, 2024.

[38] Ph.D. W. M. Haynes, *Crc handbook of chemistry and physics 95th edition*, CRC Press, Boca Raton, FL.

[39] Eric Winsberg, *Values and uncertainties in the predictions of global climate models*, Kennedy Institute of Ethics Journal **22** (2012), no. 2, 111–137.