Mathematics, Language, & Music: Perspectives from a Cellist and Machine Learning Researcher

MaryLena Bleile

Abstract This paper characterizes the main tenets of my philosophy on music, mathematics, and langauge, as informed by my experience in both fields. Motivated by several real-world issues, I provide examples of the relevance and importance of the theory (including a discussion of an interesting mathematical result in the field of statistics and PDEs), and conclude with a discussion of how these issues can be resolved.

Introduction

My name is MaryLena and I am a 22-year-old Ph.D candidate from the Statistical Science department at Southern Methodist University. While my current work is highly technical, including dynamical systems modelling in tandem with Deep Q-Networks in the context of Multi-Agent Reinforcement Learning for a medical physics lab, my background is perhaps unusual: my primary undergraduate degree was a B.Mus. in cello performance, and I didn't start any math or science at all until my junior year, when I added a major in statistics.

Over the years, I have noticed some patterns, similarities between music and math, as well as some key deviances which are complementary in a very specific way. This essay characterizes my theory of mathematics and music, as well as describing some of my experiences with the transition from music and mathematics, and how it was inherently informed by my background and identity.

I will never forget the conversation I had which sparked the first draft of what is now the section of this work on Science, Art, and Commmunication. It was in 2017, at Dr. David Huron's empirical musicology workshop. The seminars for that day had concluded, and I was hanging out with my two colleagues, Matt Chiu and Sid Rankaduwa. My experience with these two had been one of those magical moments when you meet someone and just kind of *click*, like you've always known each other; you've always been friends. And while none of us have seen each other since, the conversations I had with them that weekend affected my life in a deeply profound way. We engaged in a lively debate on the ontology and purpose of music, with myself

Journal Name 2021, 20XX, page 1 of 19

doi:10.1017/xxxxxxx xxxx

arguing vehemently that it was *not* a language. However, that conversation left me with a lot to think about, and on the plane ride home - I changed my mind, opening my laptop to write down my arguments on why music *was* a language. This, over the years, eventually evolved into the first section of the present document.

Science, Art, and Communication

On the Different Branches of Language

My thesis for this section is as follows: language exists on a continuous scale, ranging from aesthetic language (art, music, etc.), to communal language (English, Spanish, etc.), to mathematical language (Topology, Analysis, etc.). These three areas, or “branches" on the scale do have overlap and grey areas, but in the ranges where they do not, they are differentiated by Significant Form - which is perceptually subjective (differs from person to person) due to the variance in humans’ subjective emotions.

One of the main problems when trying to produce a definition or analysis of music, is that not everyone reacts the same way to all music: people with different personalities tend to prefer different styles (Ladinig and Schellenberg 2012). Although this may at first cause aestheticians to moan in despair of ever making any judgement about music, it is actually a very important point in defining it as such - for that is that not why we love music so much? Paradoxically, the fact that it is “so cussedly subjective” (Zachary Wallmark 2016, personal communication), is why it is able to speak to each of us in such a personal way. That is, it is from the personal subjective *balance* between individuality (appealing to everyone personally) and community (appealing to common human emotions) that music obtains its power.

Music *speaks* to us, that is, it provokes a response in us, whether or not it was the intent of the composer to do so. Even if the aesthetic response is nothing but the mere mental note of a piece’s aesthetic beauty, or worth, we have still been touched in some way by said Beauty. Thus, even so-called “absolute” works like Stravinsky’s “Octet for wind instruments” (1923) are still conveying some form of aesthetic emotion, and therefore cannot be truly self-sufficient, with no reference to beauty, emotion, or otherwise. So, if one adopts the definition of *language* as a method of conveying some idea/emotion from one individual to another, then music is a language.

Aesthetic languages, such as musical languages, are however prominently different than languages such as English - musical languages can be simultaneously much more precise, due simply to their subjectivity and ambiguity, as outlined above. This is, I believe, a powerful testament to the attainability of beauty through the balance and unity of things that we tend to pit against each other as opposites (a point that I will return to later): music derives its power from from the unity of precision and generalizabilty, where some its ability to speak so precisely comes, to a degree, from its ambiguity.

Mathematics does the same thing, but orthogonally to music. Inversely to how music operates, the language of mathematics derives its power and generalizability from its *lack* of ambiguity; from the fact that many parts of it are not subjective. It is through this objectivity that we can really share in a communal knowledge - to a certain extent - of the beauty that shines through mathematics. It, like music and art, conveys profound truths about the world - but in such a way that we can share these truths with one another, due to its precision. While effective art and music touch the depths of the soul and then radiates outward, mathematics is absorbed slowly by the skin, until you are entrenched in its glory, like a hot tub that can be shared by multiple people at a time.

One objection to the idea of music as a language comes from the fact that it is *not* universal: not all people are affected or understand the same types of music,and it takes time and effort to learn how to understand and appreciate a different type. In a way I agree: music is not a singular language (it, in all generality, is certainly not "**the** universal language"), but it is a *family* of languages: perhaps "music" isn't a language - however, "post-romantic symphony" could be. Lumping all musics together and saying music isn't a language because not everyone can understand it, would be like me saying English and Japanese aren't languages because I can speak one but not the other.

This brings me to the family of languages like English, which have the most straightforward interpretation as language. I will refer to these as *communal languages* so as not to confuse them with the others. Here perhaps English is to Japanese what post-romantic symphony is to traditional Shakuhachi flute music: the grammars are totally different and no matter how much I study English (post-romantic symphony), I still won't understand Japanese (shakuhachi flute music) without studying them, even though they are both communal (aesthetic) languages.

I have presented the three branches of languages distinctly, for clarity. However, they are not dichotomous, exclusive categories - rather they are continuous ranges that blend into one another. The three completely orthogonal branches exist only in abstract, and rather form a *basis* for language space.

Current empirical research in musicology supports the theory of a continuum, rather than a discrete division of language into branches: we cognitively process societal languages in a different way than we process music (Fedorenko et al. 2012). However, the line between these two mechanisms is blurred, as demonstrated by the *Speech to Song illusion* discovered by Diana Deutsch (Deutsch, Henthorn, and Lapidis 2011). This illusion is a remarkable phenomenon wherein a spoken phrase can be perceived as song after many repetitions. Furthermore, when the phrase is inserted back to its original context, the surrounding context is still perceived as speech, with the target phrase "jumping out" at the listener as somehow "suddenly sung" despite the fact that the entire passage was actually spoken. Viewing the three branches of language as a multidimensional continuum rather than a set of exclusive categories also removes the philosophical classification issue of lyrics in music, as well as that of poetry.

Language and Charisma

One possible objection to the idea that communication is integral to science and mathematics is the idea that while you can err in *delivery* of your results, your results may still be valid - that is, one could argue that the science lies in the objective results, regardless of delivery. However, This is simply a rephrasing of the question of the relationship between *language* (delivery) and *meaning* (scientific results): how does a sequence of sounds come to *mean* something, and are the two innately related? The answer can be discussed in general, and applied to the languages of mathematics and music, as well as communal languages.

Consider the case of communal languages, such as English. There is nothing inherent to the sounds that these words represent, that ties them to the meanings you are ascribing to them. This is quite obvious when we consider the fact that there are so many languages in existence, in which it is possible to transmit the same meaning. Thus, we can infer that language is a code, a series of symbols that we use to transmit *meaning* (and is this not exactly what mathematics is as well?).

It is obvious that things can be ambiguous. What do I mean by that? Well, I merely mean to demonstrate that there are sentences that can provoke the response of "What do you mean?"; statements that inadequately transmit their intended meaning. This happens when there is a disconnect between the process of encryption of meaning, and the process of decryption. These sorts of statements illustrate the fact that there is no one set attribution of meaning to each word or phrase; part of the brilliance of language is its multipurpose, pragmatic nature. Therefore, decryption process of language is effectively an inferential science, as opposed to a deductive method. That is to say, there has to be some movement from what we pick up from linguistic symbols, to a more generalized, larger meaning, which we can comprehend on a deeper level. This is the difference between being able to repeat a phrase back to a person, and actually having an understanding of the phrase.

It is self-evident that meaning-encryption methods (language usage) differs with individual background, culture, and life experiences, even within a specific language or dialect. For example, consider the case of the fictional language Quenya, from Lord of the Rings. Quenya spoken in the First Age in Valinor, is quite a different beast from Third-age Late Exilic Quenya, which we observe in the Lord of the Rings trilogy and thus, as a student of Third-Age Late Exilic Quenya, I would likely be much less than capable of communicating with Ingwë, high king of the Noldor, even though I would probably be able to converse with his great-grandson Finarfin.

There is an obvious correlation between those who are friends, and those who speak the same language; we naturally gravitate towards people with whom we can easily communicate - it feels good to share things, including words and, more importantly, an understanding of meaning. However, I'd also be willing to posit that we've all at some point encountered a person with whom we technically share a common language, but with whom, for some reason, it is still ever so difficult to communicate - which would be due to a difference in methods of encryption and decryption of meaning. It is reasonable to assume that this fact probably made us somewhat less likely to want to befriend this person. Conversely, it is not unreasonable to assume that people who encode meaning similarly, and thus understand each other well, are more likely to become friends.

This brings me to my definition of charisma. Charisma, as I understand it, is the ability to precisely and accurately decrypt people's coded meanings, and then to adapt their encryption methods to the other person's; people who are charismatic are, in a sense, encryptic pragmatists. An technological analog to this might be the difference between hard coding information into a program, versus using some variable that can be more easily manipulated.

The propositions of the above passages are arguably even more obvious with regards to Aesthetic languages (I will focus on music because it is what I studied for 18 years). Of course decrypting the code of music is inferential; the obviousness of which stems from the fact that since aesthetic language differs from communal language, the two cannot be directly mapped onto each other. We can teach people the grammars, the rules of different types of art (just like we can teach the grammars of different communal languages), but we can't teach the *understanding*, the intuition of the *meaning*.

There is also such a thing as "artistic charisma". However, as learning the grammars of an aesthetic language is still a real, (difficult,) and necessary process for artists, it would be wrong to attribute an artist's good performance to artistic charisma alone. Artistic charisma doesn't make you good at a particular art form any more than a +5 charisma ring in Dungeons & Dragons would give you the immediate ability to speak Quenya; it takes time and effort to become competent enough to actually *use* the charisma.

Similarly, there is a parallel phenomenon in the language of mathematics: a "mathematical charisma" if you will, which is characterized by the creative intuition to understand and connect mathematical concepts. Such a charisma is what makes one excel at what Yanai and Lercher (2020) term as "Night Science". However, it takes "Day Science", aka the learning of formalized structure in which to express these insights, for this charisma to be useful beyond personal entertainment.

Seminal aesthetic formalist Clive Bell discusses this phenomenon in depth, introducing the idea of "aesthetic emotion", which I believe to be the driving factor behind any type of charisma [1]. This is illustrative of one of the fundamental harms of the false dichotomy between logic and creativity; that is, between meaning-generation (what I have been referring to as "charisma") and meaning *delivery*, aka the grammars of mathematical language. Such a dichotomy urges students of mathematics to squash their creative charisma, their aesthetic emotion manifesting through intuitive leaps of imagination, from whence every brilliant mathematical idea comes, as discussed in great length in a book called *The Poetry and Music of Science* (McLeish 2019).

1. according to Bell, aesthetic emotion is an individual's response to the significant form of a piece of art, which characterizes the essence of an artwork

This effect is amplified, in my experience, by socio-cultural dichotmous ideas of gender, where “female" is mapped to “right-brain; creativity" (that is, charisma) and “male" is mapped to “left brain, logic". STEM gender disparities aside, this harms both men and women in arts *and* science, as well as non-binary people, who do not fit into either category. For women in STEM, particuarly in my experience, this often manifests through an intense compulsion to squash creativity and mathematical charisma in an effort to “compensate" for their gender and to fit in. Similarly, men expressing an experience of aesthetic emotion may result in a threat to their perceived masculinity. And of course, the whole dichotomization process is alienating to anyone who does not fit exclusively into one or either of the distinct categories we lay out.

There is also an interesting irony at play: we, as mathematicians, deify the grammatical, delivery process of mathematics over the charismatic, aesthetic-emotion filled process of ingenuity. However, when it comes to delivery via communal language, we (as a collective field) balk, overwhelmingly dismissing it as unimportant. I theorize this is because there is an inherent bypass of precision that must take place in effective grammatical delivery, as loosely illustrated by Figure 1.

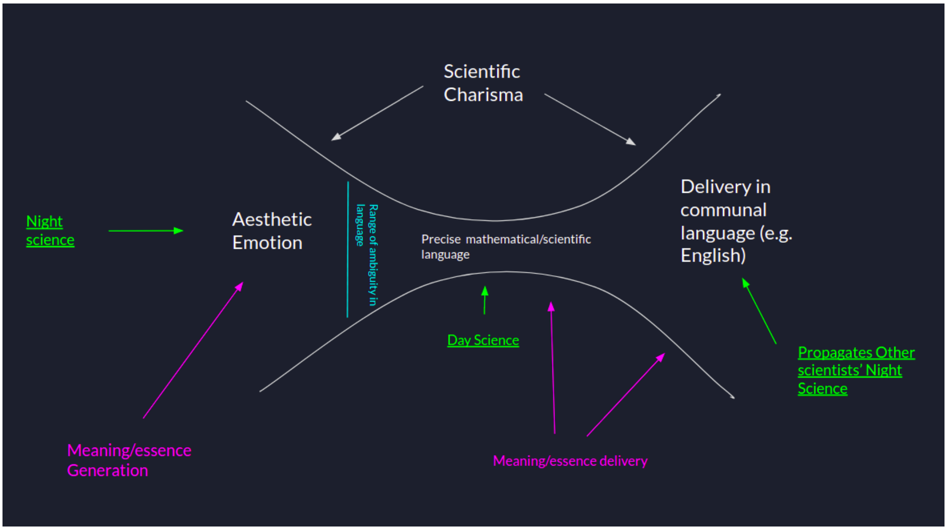

FIGURE 1. *Illustration of the scientific process*

Music, on the other hand, follows a process more closely resembled by Figure 2; here, similarly to the case of science, artistic charisma begins with a conceptualization; an imaginative visualization of something which has not yet been willed into existence, as discussed in great length *via* the concept of the “creation story" in *The Poetry and Music of Science*. However, in order for this idea to be brought into existence, a strict method of execution must take place. McLeish discusses the case of the writer, who is compelled to write and rewrite a piece until it is exactly right; similarly with music, in order for the piece to be brought into reality the way you imagine it, a very strict, constrained, co-ordinated physical process must take place. This is where G.K.

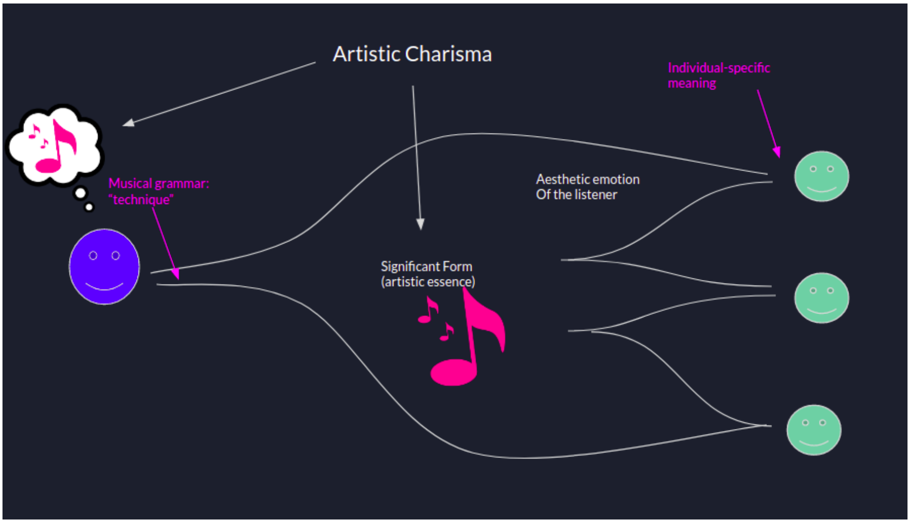

FIGURE 2. *Illustration of the process of musical communication*

Chesterton, in his book *Orthodoxy* is in error: Chesterton claims "But the point is that a story is exciting because it has in it so strong an element of will, of what theology calls free-will. You cannot finish a sum how you like. But you can finish a story how you like. When somebody discovered the Differential Calculus there was only one Differential Calculus he could discover. But when Shakespeare killed Romeo he might have married him to Juliet's old nurse if he had felt inclined". This is untrue; when bringing a work of art into fruition, there is really only one "correct" outcome: the perfect expression of the artist's conceptualized work. Of course, the process of bringing the work into fruition can actually *change* this conceptualization, as pointed out by Del Cook (personal communication, 2021), which parallels the research process wherein performing an experiment can cause you to redefine your initial research question. Furthermore, there is (in my opinion) an inherent loss of generality in the process of bringing such a work to fruition (as Aquinas points out in "Being and Essence", the coercion of a phenomenon from potency to act is ultimately reductive), and since art gleans its power from its generality, this perfect expression will never be attainable - much like the inherent imperfect efficiency of any mechanical process, where some energy will always be wasted as heat.

This phenomenon is consistent with both my own experiences as an artist, as well as my conversations with others. There will be a moment of clarity, of "illumination" in the words of McLeish (2019), wherein you conceptualize an entire, complete work; you audialize exactly how you want to play something. You have an artistic *message* to share; an overwhelming aesthetic emotion (which is maybe tied to other emotions as well) that you want - no, that you *need* to express. But in order to do this, you have to go through a process of disciplined, constrained practice and use the formulaic rules to your advantage - in a way, in order to express the emotion, you have to become detached from it, which lessens the impact to a certain degree. And the degree to which this sacrifice is required varies between genres: I brought this concept up with a close musician friend with whom I attended college, who pointed out that “that’s why [they] love grunge music": grunge, through its rawness and simplicity, allows for a quicker passage from conception to reality, requiring a relatively lesser degree of emotional inhibition and constraint to execute (Dorian Stewart, personal communication, 2021). That is where its beauty comes from - a stark contrast to the more “technical" genres such as classical which, while it perhaps involves a more expansive emotional palette, involves an exponentially greater loss of aesthetic emotional information in transfer. Personally, I love heavy metal - particularly blackened melodic/symphonic deathcore - because it combines the rawness of punk with the harmonic complexity and technique of classical in a balance which is exquisitely sublime.

Implications of Mathematics as Language

Mathematics is a language; a formalized set of precise terms which we use to describe the world. But all too often, we get caught up in a false dichotomy that we try to impose between STEM and the humanities. So many people have this idea that being better at STEM means worse at things like writing and communication, which leads to a gross lack of attention paid to self-improvement with regards to communication.

In order to better emphasize this point, I will turn to an example. The Central Limit Theorem (CLT) is one of the foundational pillars of statistics and data science. Consequently, it is of particular interest in the field of Mathematical Statistics to generate new and improved versions of the CLT for different cases, as well as deriving novel intuitive methods for its proof. In particular, a recent paper (Zong and Hu 2013) developed a proof of the CLT which takes inspiration from the theory of Partial Differential Equations, relying heavily on the known form of viscosity solutions (that is, solutions to the heat equation). While their proof is interesting and creative, in order for the proof to be useful as an intuitive explanation of the CLT as advertised in the paper’s motivation, further explanation is required. Additionally, I believe the notation used in their Theorem 6 is confusing and misleading; in order to further elucidate their statements, revision of notation is required. In order to supplement this, I provide a high-level proof sketch using the proposed notation, as well as a simple visualization application which illustrates the connection between the dispersion of heat and the CLT.

Preliminaries

This section consists of much of the same information in the "Preliminaries" section of Zong and Hu (2013). Define $C_{b,Lip}$ as the class of bounded functions f such that $f(x) - f(y) \leq C|x-y|$, for some $C \epsilon \mathbb{R}^+$ and $\forall x, y \epsilon \mathbb{R}^n$. We refer to $f \epsilon C_{b,Lip}$ as "bounded and Lipschitz continuous". The following result is taken from basic probability: Suppose X is a random variable with probability distribution defined as $P(X < x) = V(x)$. Then, $\forall f \epsilon C_{b,Lip}(\mathbb{R})$, $\exists E[f(X)] := \int_{\mathbb{R}} f(x) dV(x)$. We refer to $E[f(X)]$ as the expected value of $f(X)$.

Results

Statement of Results

Zong and Hu (2013) prove the following:

A normally distributed random variable Z with $E[Z] = 0$ and $E[Z^2] = \sigma^2 > 0$ is characterized by the following PDE defined on $[0, \infty) \times \mathbb{R}$:

$$\begin{cases} \partial_t v + \frac{1}{2}\sigma^2 \partial^2_{xx} v = 0 \\ v(x, 0) = f(x) \end{cases}$$

This PDE is called the *Heat Equation* and the set of functions v that satisfy it are called *Viscosity Solutions*. [Central Limit Theorem]

Suppose $\{Y_j\}_{j=1}^n$ is a sequence of independent and identically distributed random variables with mean $\mu < \infty$, variance σ^2 s.t. $0 < \sigma^2 < \infty$, and finite 3rd moment $E[X^3]$. Then, as $n \to \infty$, $\frac{1}{\sqrt{n}} \Sigma_{j=1}^n (Y_j - \mu) \xrightarrow{d} N(0, \sigma^2)$. For notational simplicity, let $S_n = \Sigma_{j=1}^n (Y_j - \mu)$.

The real “magic" of the proof happens in the proof of the following lemma: $\forall$ bounded, continuous functions f, as $n \to \infty$, $E[f(\frac{S_n}{\sqrt{n}})] \to E[f(Z)]$, where $Z \sim N(0, \sigma^2)$. The main difference in notation lies within the choice of symbols Y_j and Z. Note that each element of the Y_j sequence has a distribution that is not necessarily Gaussian, whereas $Z \sim N(0, \sigma^2)$. My notation distinguishes between these two by utilization of Y and Z, in contrast to Zong and Hu (2013), who use X for both of these. This choice of notation is commonplace when writing about convergent sequences. However, using the same symbol for both here is disingenuous given that the whole point of the proof is that Y_j and Z (X_i and X in Zong and Hu (2013)) do not necessarily have the same distribution, nor (most notably) do the X_j themselves converge to X as the notation is commonly used to denote. Furthermore, given the number of nested expectations taken in the proof: using X for sequence elements as well as the “target" normal random variable means unnecessarily leaving it to the reader to figure out what variable each expectation is being taken with respect to. Further, x is re-used later in the proof as one of the two arguments to the viscosity solution function $v(x, t)$, which greatly adds to the confusion since x is conventionally used in statistics as a particular value of random variable X. Since the motivation for this proof was that it should be more intuitive to statistics students, an intuitive choice of notation is of great importance, particularly in light of the fact that mathematics is a language.

Proof Sketch

As mentioned, the main part of the proof is found in the proof of Lemma 5. We include a brief explanation of their methods below.

Trick 1: *"Reference function".* Find a convenient reference function ψ. Show that as $n \to \infty$, both $E_Y[f(\frac{S_n}{\sqrt{n}})]$ and $E_Z[f(Z)]$ get arbitrarily close to ψ

We do this by defining a small deviance h, and showing that as $n \to \infty$, both $|E_Y[f(\frac{S_n}{\sqrt{n}})] - \psi|$ and $|\psi - E_Z[f(Z)]|$ are bounded above by $C\sqrt{h}$ (some $C \epsilon \mathbb{R}$). Then take $h \to 0$ since h can be whatever we want.

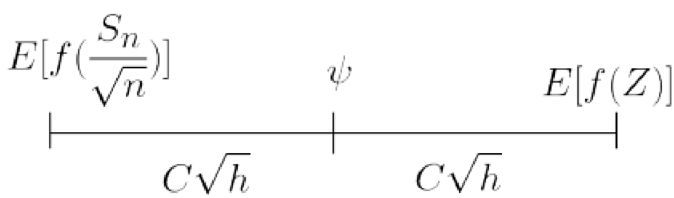

FIGURE 3. *Illustration of the reference function technique*

In order to prove lemma 5, the idea is to choose ψ that "mixes" Z and $\frac{S_n}{\sqrt{n}}$. That is, take $\psi = E_Y[E_Z[f(\frac{S_n}{\sqrt{n}} + \sqrt{h}Z)]]$ (note, when $h = 0$ this is just $\psi = E_Y[f(\frac{S_n}{\sqrt{n}})]$)

Trick 2: *"Pseudo-Method of Characteristics"*

So, $\psi = E_Y[E_Z[f(\frac{S_n}{\sqrt{n}} + \sqrt{h}Z)]]$. Consider everything inside the $E_y[]$. Note that the only stochasticity involved comes from the normal random variable Z. Also recall that the non-homogeneous heat equation is solved by what is equal mathematically to an expectation of a transformed normal random variable.

$$\rightarrow \text{Let } v(x,t) = E_Z[f(x + \sqrt{1+h-t}Z)] = \int_{\mathbb{R}} \Phi(z, \sigma^2) f(x + \sqrt{1+h-t}Z) dz$$

Note that $v(\frac{S_n}{\sqrt{n}}, 1+h) = E_Y[f(\frac{S_n}{\sqrt{n}})]$ and $v(0,h) = E_Z[f(Z)]$.

Also, v solves the well-known PDE which describes the dispersion of heat:

$$\begin{cases} \partial_t v + \frac{1}{2}\sigma^2 \partial_{xx}^2 v = 0 \\ v(x, 1+h) = f(x) \end{cases}$$

on $x \epsilon \mathbb{R}, t \epsilon [0, 1+h]$. Note that the function of t in v is reversed from the classical heat dispersion PDE; while in the original formulation of the heat equation, dispersion (variance) increases in time t, here we reparametrize in terms of $1+h-t$ which means that the distribution actually gets *sharper* over "time" (t) to match the behavior of the distribution of the mean of a random sample with increasing sample size n (this follows from the fact that the variance of a random variable Z multiplied by a constant

a is $V(aZ) = a^2V(Z)$, so $\sqrt{1+h-t}Z$ has variance $(1+h-t)\sigma^2$ which is decreasing in time).
The idea of the proof is to use the properties of v to show that $E[f(\frac{S_n}{\sqrt{n}})]$ and $E[f(Z)]$ asymptotically approach ψ.

Perhaps the most important property of $v(x,t)$ is as follows:

$$|v(x,t') - v(x,t)| \leq C\sqrt{t-t'} \; (t,t'\epsilon[0,1+h]) \quad (1)$$

and

$$|v(x+h,t) - v(x,t)| \leq C\sqrt{h} \quad (2)$$

That is, $v(x,t)$ is Lipschitz continuous in both the x and the t parameter (note that this is useful because we will eventually need to have Z in the "t" position of v, and $\frac{S_n}{\sqrt{n}}$ in the "x" position.) Proposition 2 follows from the fact that v is an expectation of a Lipschitz-continuous function f whose argument is the random variable.

Web Application

At this point, the aspiring student of statistics or physics may question, why does this approach work? Is it pure serendipity that the solution to a heat dispersion problem is tied innately to the Central Limit Theorem?

In order to propagate the exploration and deeper understanding of both phenomenon, we present a simple web application (illustrated in Figure 4) which allows the user to see the inverse relationship between the behavior of heat density over time, and the behavior of the empirical density of a sample mean with increasing sample size n. It is available at this link.

In fact, the intuition behind why heat density is Gaussian as it disperses, can be built upon one's understanding of the Central Limit Theorem, which is helpful to the physics student who comes from a statistically-informed background. One of the main "tricks" employed in the classical derivation of the solution to the heat equation, is to notice that in order for the PDE to be solved, the solution must possess a property of radial symmetry: that is, while the solution may flatten, it cannot change its overall *shape* with increasing time (it is always Gaussian). But in light of the illustration of the same behavior of an expected value of the mean of a sample of normal random variables (taken as a function of its sample size, which drives the variability of the distribution of the mean), the solution to the heat equation *must* be Gaussian: the distribution of the mean of an i.i.d. sample of any other distributions will necessarily change shape with n, since the distribution of any such sample mean converges to a Gaussian with increasing sample size.

Perhaps one might brush off this revision as an inconsequential detail, unimportant to the content itself. However, what scientists who adopt this stance (consciously or inadvertently) fail to realize, is that **your scientific and mathematical results are useless if nobody understands them but you**; that is, delivery is an inherent attribute of mathematics, since it is a language. Similarly, the utility and scientific value of one's results are positively correlated with the number of scientists who are able to

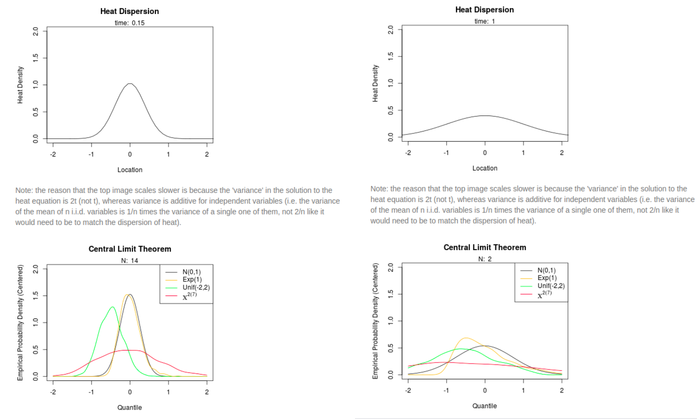

FIGURE 4. *Output from the designed application with two different values of user input*

understand them - in the absence of interaction (that is, *communication*, by definition) with the world and with other humans, you aren't really doing science or mathematics, you are just having thoughts in your own head. Of course I am aware that there is a certain threshold of domain knowledge that may be necessary to assume, and this threshold may inevitably prevent some people from understanding your work. But that is entirely beside the point; the point is that making our results clear and understandable is not just a "side thing" we tack on at the end, but a crucial, fundamental part of the work. Since mathematics is a language, content and communication cannot really be separated. Just like logic and emotions, they must be maximized synergistically in balance with each other.

Obstructions to Language: Dichotomistic Bias

Mathematics, as a language, is really just a formalized version of philosophy: a method of communicating ideas about the underlying phenomena of the universe. So, too, is music a language, and using that language, art conveys profound truths about the world. Thus, art and math share the same common goal: they complement each other. All languages are inherently flawed, but when mathematics breaks down, we can turn to art for the answers and vice versa, like a sort of way of piecing together the profile of beauty, bit by bit, like a Principal Components Analysis - and in my experience, the world is best understood when one has a solid knowledge of both. Furthermore, the false dichotomy that we commonly impose upon them is, I believe, a special case of a more general phenomenon: dichotomistic bias. Dichotomistic bias is also exemplified by the common false dichotomy between logic and emotions, which I will return to in the next section, as well as the false dichotomy between theory and application in mathematics; between wonder and relevance, as illustrated in Grant Sanderson's

TEDx talk on what compels people to engage with mathematics (Sanderson 2020).

Statistician Stephen Senn popularized the term *dichotomania* for falsely dichotomizing continuous variables (Senn 2005). In his 2017 paper *Statistical Errors in the Medical Literature*, statistician David Harrell voices his affirmation for the existence of this phenomenon, and outlines several prominent examples where chopping up dichotomous variables like this in an effort to "simplify" them by making them discrete, has led to disastrous effects on statistical analyses in various papers - and not just any papers; articles where lives are potentially contingent on the results (Harrell 2017). One compelling example of this is the well-known discrepancy in power between the (continuous) t-test and the (categorical) sign test: that is, the efficiency of the sign test with respect to the t-test is a meagre 64%. I hold the papers by Senn (2011) and Harrell (2017) in high regard, however, "dichotomania" has the connotations of a mental illness (mania), which is not only clinically inaccurate, but also insensitive towards that portion of the disabled community, many of whom may excel at mathematics and statistics (Mildred Boveka & Madeleine Jennings 2021, personal communication). Cognitive scientist and statistician Sander Greenland has identified the phenomenon as a cognitive bias (Greenland 2017), hence, I will henceforth refer to the phenomenon as "dichotomistic bias".

I remember the moment I decided that I *had* to pursue mathematics: it was a pure moment of *wonder*, and emotion. I was watching the video series by Welch labs, "Imaginary Numbers are Real" - in the first of which the author introduces the Fundamental Theorem of Algebra, which states that any function has a number of roots (points at which the function is equal to zero) which corresponds to its power. So for example $f(x) = x$ has one root, $f(x) = x^2$ has two, and so on- which makes intuitive sense. But, the author notes, $f(x) = x^2 + 1$ (visualized in Fig. **??**) appears to have no roots. What does this mean? Is the theorem wrong?

And that's when it happened; the graphic that changed my life. While I sadly cannot include a screenshot of it here for copyright reasons, I encourage you to go

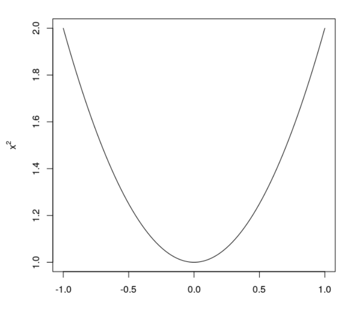

FIGURE 5. *Visualization of the function* $f(x) = x^2 + 1$

look at it at time 1:43 in the video *Imaginary Numbers are Real [part 1: Introduction]* since no description does it justice (Labs 2015).

And like.... WHOA there's a whole other WORLD of new NUMBERS I just diDn'T knOW ABOUT???!??

And so it was.

Now, of course the mathematics elitists would be very quick to jump on the fact that, "aren't you a statistician though?? Statistics isn't *math*". But honestly, I think that that is actually another great illustration of the malevolent phenomenon I'm trying to illustrate: the whole "is statistics math" question is stupid, since the field of mathematics does not have well-defined boundaries; the scale of math to science to engineering is continuous, and trying to forcibly impose a boundary like that, is an example of dichotomistic bias. Statistics is in the ambiguous grey overlap area between math and science, and no amount of statistical apologetics nor mathematics gatekeeping will force it exclusively into one section or the other. It lives happily in the ambiguity, in harmony with measure-theoretic probability (the corresponding branch of pure mathematics on which statistics rests): measure theory is what infuses statistics with mesmerizing wonder, while statistics breathes a refreshing air of relevance and applicability to measure and probability theory.... and the way they intimately interlace is a profound expression of Beauty.

But of course, we hate that. Humans - particularly humans who work in STEM fields - can't *stand* ambiguity, especially when we expect something to fit distinctly into one of two exclusive sets. If you are ever in doubt about this, just ask an androgynous-presenting person how frequently random people will rudely inquire about what genitals they have. This is, I think, also one of the reasons why there is such major resistance to the idea of sexuality and gender as a spectrum; we hate the ambiguity, and prefer when things fit into nice little boxes with labels. I theorize this is what motivates a good part of biphobia, as well.

My personal experience of dichotomistic bias runs very deep into my childhood. My parents, strict conservative Christians, held very firmly to a dichotomy between men and women, wherein women were associated with emotion, creativity, art, etc., and men with logic and other stereotypical phenomena. My experience of a deeply traumatizing event as a child was dismissed and forgotten about, and when the emotions resurfaced in my adolescence, I was told that I was just an "emotional teenage girl" and that "all teenage girls are depressed". Furthermore, my struggles to overcome the emotions associated with said event - which often became apparent through my deep love for art and music - were taken as an indication that I would not be good at science or anything technically related (once I realized - my junior year - that their assessment of my capabilities was incorrect, statistics was the closest field I could still fit in as a second major). When I was twelve, I remember distinctly my mother laughing off my aspirations of being an astrophysicist with "[Dead name redacted] is careless; [she has] to go into music because if you play a wrong note onstage, nothing bad happens, but if you build a bridge wrong, people will die". Similarly, I recall my father proposing to me the idea that “women [were] just too emotional to have careers", and that “it’s a good thing that men find women sexually attractive, otherwise the species would die off. Because if you think about it, what other use would a man have for a woman?"

Sweeping generalizations and sexism aside, this is a peak example of the false dichotomy between emotions/creativity and logic: the question of whether women are more emotional than men is irrelevant to the question of whether they belong in science, because emotions and creativity are not only compatible with but *integral to* good science. McLeish (2019) discusses this at length, illustrating Max Planck’s quote, “imaginative vision and faith.... are indispensible - the pure rationalist has no place in science". I would even go so far as to say that the two aren’t opposites; each is maximized in harmony with the other.

Obstructions to Language: The Need for Emotional IQ in Science

I propose that good emotional IQ - i.e. knowing how to appropriately process emotions - is *imperative* to good science.

Mental capacity is like a stack[2], the elements of which can be emotional or intellectual. Heavy emotional elements in this stack are simply nuisance items when we are trying to do intellectual tasks, and continuously repressing them - shoving them back at the bottom of the stack when they surface - only serves to decrease the amount of space in the stack left for intellectual tasks. In order to remove them and use one’s full capabilities, one has to allow these emotional items to surface and be processed. Furthermore, if you leave heavy elements in the system too long they will build up and eventually make the stack overflow, usually at the most inopportune of moments (of course, this does not necessarily mean always giving your emotions top (or real-time, if you prefer) priority - but rather that you need to develop a balanced process *scheduler*, that allows these things resurface in a safe, appropriate context).

As an aside, in *the Poetry and Music of Science*, McLeish (2019) points out how according to the ancient philosophers “wonder (thaumazein) is the beginning of philosophy", going on to argue that it is the beginning of science and art as well. I agree with this, but the ability to *wonder* about math to the best of one’s capacity requires a freedom of mind on which the emotions and worries associated with things like abuse and poverty place ineffable burden, making uninhibited peaceful, playful mathematical wonder a state of immense privilege. On the other hand, some of these heavy elements can provide an intimate connection, in my experience, to the *wonder* component of art, the caveat being that it becomes more difficult to force yourself to disseminate your ideas and express them through constrained execution of artistic language. It is thus a solitary state; no one else will be able to understand your art until

2. in computer science, a stack is an abstract data type that holds objects with a last-in, first-out order. Like a stack of plates: the last one you put on top is the first one you take out.

you disseminate it through a constrained process that facilitates general precision. In this regard it is the opposite of math, in my experience: if you have heavy emotional elements, you can still aggressively compartmentalize in order to force yourself to do the mechanistic part of math, but the intuitive wonder part of it becomes nearly impossible.

The Antidote: Balance

As a possible objection to my criticism of dichotomistic bias is found in the acknowledgement of the statistical counterpart to Senn (2011)'s “dichotomania": *continuitis*. That is, “...[dichotomistic bias affects] medics. Statisticians suffer from continuitis" (Senn 2014). Here, “continuitis" refers to the propensity of statisticians to make everything continuous, even when it should truly be discrete. But I argue that continuitis, or *continuistic bias* as I prefer, is not the opposite of dichotomistic bias. Rather, it propagates a new dichotomy: the dichotomy between dichotomous and continuous. The antidote is not aggressively continuizing all variables, but rather it is found in *balance*. Similarly with the false dichotomy between genders, the answer is not to insist that no-one use labels to describe themselves or put anyone in categories ever: rather, the answer is to find a balance between avoiding unnecessary assumptions based on a person's appearance, and reading non-verbal communication and self-expression.

Of course, perfect balance is impossible. Our balance will always be lopsided; in some way, we have all favoured logic and blunt truth at the expense of healthy emotions, and we have all sacrificed logic to indulge our emotions at some point. Perhaps some people tend to try to favour one unhealthy unbalance over the other (I know for years I extensively deified and prioritized logic over emotions), but we have all done both I think, and ironically, disbalancing the two in one direction seems to cause one to swing and disbalance it in the other direction as well, like the oscillations of an undamped system. For example, cognitive science itself tells us that the idea of objectivity serves to increase bias: a large scale 2007 study (Uhlmann and Cohen 2007) showed that self-perceived objectivity increases hiring bias against women, even after controlling for stereotypic beliefs. It is thus that the irony of the situation is apparent: leaning heavily on one side of a dichotomy does not even serve to maximize that end (that is, the illusion of objectivity breeds subjective bias) - rather, the two ends can only be maximized synergistically, in balance with one another.

We see the phenomenon of balance come up frequently in science. For example, one of the major wars in statistics consists of the debate between frequentists, who model parameters as fixed but unobservable values, and Bayesians, who model underlying parameters as random variables with distributions. The frequentist p-value, in particular, is an issue of grave concern and importance. My own recommendations can be found in my papers on the Bi Error Method (Bleile 2021). Within the Bayesian camp, there is one issue in particular that is of philosophical and practical interest: the choice of priors. Stephen Senn provides an insightful commentary in his paper titled “You may believe you are a Bayesian but you are probably wrong" (Senn 2011).

The general mathematical framework for Bayesian statistics goes like this: you start with a prior set of beliefs, which are expressed in a model. As you add data, the model gradually updates its “beliefs" (parameters) to reflect consistency with the data in light of its prior values. This is very similar to how knowledge acquisition works in humans: we start with a set of beliefs (biases), we receive new information, and then we change our beliefs accordingly (unless, of course, you’re a bigot, in which case you put probability 1 on your prior values and probability 0 on the data). In terms of Bayesian statistics, it is generally agreed that non-informative priors - that is, priors that give equal weight to every possibility - are preferable in most if not all circumstances (indeed, informative priors are one of the main arguments *against* Bayesian statistics). However, the catch is that in order to be truly noninformative, often we have to choose values that are not *proper*, that is, they induce a prior distribution which is not well-defined. Philosophically, the purist Bayesian would say that these are unacceptable; only internal consistency (having a well-defined prior distribution) matters, and we should use proper priors even if they cause the outcome to be influenced by this bias. However, while this is perhaps the most defensible approach in theory, Senn (2011) points out that it is often just not reasonable in practice: this problem can be thought of as a tradeoff (balance) between internal consistency (having a coherent prior distribution) and external consistency (a coherent model in practice). Of course this is a simplistic analogy, however, I believe the point still stands. Similarly, executing a Bayesian approach at all may not always be practically feasible, and balanced eclecticism may occasionally dictate that frequentist statistics are the way to go for an analysis.

Were I to discuss in depth all of the mathematical and scientific examples of where extremes are found in balance, I would likely never move on from this paper. However, a few that I have made particular note of and/or studied include the bias and variance tradeoff (Bleile 2021), as well as the balance between internal and external validity in musicology experiments (that is, the tradeoff between proper context and experimental control), and the balance between quantum uncertainties necessitated for the attainment of the lower bound expressed in the Heisenberg Uncertainty Principle (which, since the Heisenberg Uncertainty Principle can be reframed as an entropic lower bound (Bourret 1958), I theorize may be the same - in some mathematical sense - as the bias/variance tradeoff). In all of these cases, balance is the key to acheiving extremities; optimal results.

Conclusion

In summary, I have presented my philosophy of language, which extends to language of the aesthetic as well as mathematical branches. Further, I have outlined some key components of my experience transitioning from music to mathematics, and pointed out several key issues as well as potentials for their resolution: effective communication is often inhibited by dichotomistic bias, which is only resolved through balance.

References

Bleile, MaryLena. 2021. Youden's J and the Bi Error Method. *Forthcoming in Advances in Intelligent Systems and Computing: Proc. Computing Conference.*

Bourret, Richard. 1958. A note on an information theoretic form of the uncertainty principle. *Information and Control* 1 (4):398–401.

Deutsch, Diana, Trevor Henthorn, and Rachael Lapidis. 2011. Illusory transformation from speech to song. *The Journal of the Acoustical Society of America* 129 (4):2245–2252.

Fedorenko, Evelina, Josh H McDermott, Sam Norman-Haignere, and Nancy Kanwisher. 2012. Sensitivity to musical structure in the human brain. *Journal of Neurophysiology* 108 (12):3289–3300.

Greenland, Sander. 2017. Invited commentary: the need for cognitive science in methodology. *American journal of epidemiology* 186 (6):639–645.

Harrell, Frank. 2017. Statistical errors in the medical literature, https://doi.org/https://www.fharrell.com/post/errmed/#catg. Available at <https://www.fharrell.com/post/errmed/#catg>.

Labs, Welch. 2015. "Imaginary Numbers are Real [part 1: Introduction]". Youtube. Available at <https://www.youtube.com/watch?v=T647CGsuOVU&ab_channel=WelchLabs>.

Ladinig, Olivia, and E Glenn Schellenberg. 2012. Liking unfamiliar music: Effects of felt emotion and individual differences. *Psychology of Aesthetics, Creativity, and the Arts* 6 (2):146.

McLeish, Tom. 2019. *The poetry and music of science: comparing creativity in science and art.* Oxford University Press, USA.

Sanderson, Grant. 2020. "What makes people engage with math". Youtube. Available at <https://www.youtube.com/watch?v=s_L-fp8gDzY&ab_channel=TEDxTalks>.

Senn, Stephen. 2005. Dichotomania: an obsessive compulsive disorder that is badly affecting the quality of analysis of pharmaceutical trials, 13. June.

Senn, Stephen. 2011. You may believe you are a Bayesian but you are probably wrong. *Rationality, Markets and Morals* 2 (42):48–66.

Senn, Stephen. 2014. Twitter. Available at <https://twitter.com/stephensenn/status/438397623360565248>.

Uhlmann, Eric Luis, and Geoffrey L Cohen. 2007. "I think it, therefore it's true": Effects of self-perceived objectivity on hiring discrimination. *Organizational Behavior and Human Decision Processes* 104 (2):207–223.

Zong, Zhaojun, and Feng Hu. 2013. A new proof of central limit theorem for iid random variables. Paper presented at the Abstract and Applied Analysis, vol. 2013. Hindawi.

Acknowledgements

We thank Dr. Zachary Wallmark, Del Cook, Dorian Stewart, Matt Chiu, Sidath Rankaduwa, Madeleine Jennings, and Dr. Mildred Boveka for compelling conversation and insights which informed this paper.

Key Words

Partial Differential Equations; CLT; Music Theory; Language.

Dummy dates; please ignore. Date received: MMMM DD, YYYY; Date accepted: MMMM DD, YYYY.