MATHEMATICS AND MEDICINE: NOT AS DISJOINT AS YOU MIGHT THINK

GENEVA LIWAG

Introduction

"Heart attacks can give you mathematics."

This quotation is the title of a public lecture on arrhythmia (or irregular heartbeats) given by James Keener, a professor of mathematical biology at the University of Utah [1]. While a few might say it should be written the other way around — mathematics can give you heart attacks — Prof. Keener suggests an intriguing yet interesting concept: medical phenomena such as cardiovascular diseases can be studied and addressed with an unlikely weapon — math.

This essay aims to dispel these two misconceptions: that mathematics and medicine are two disjoint fields, and that the medical industry does not employ math beyond Calculus I, if not no math at all.[1] Strap on your seat belts; you're in for a wild ride.

1. Enzyme Kinetics

1.1. Reaction of Interest: Metabolism

Before we talk about different biochemical processes happening in our body, we might as well delve deeper into what such a process looks like. Let's take *metabolism* for example. In its simplest sense, metabolism is the conversion of simple sugars to energy. We all know that eating food provides us with energy, among other things, to help us make it through our daily tasks. Producing energy (or *adenosine triphosphate – ATP*) is not a simple one-step process. It involves three main steps, each being a series of chemical reactions. Moreover, these individual reactions, although favoured to happen, take a long time to complete (some in the order of 10^{13} seconds) [2]. If this is the case, then why don't we end up waiting a lifetime to gain energy?

The answer lies in complex molecules called *enzymes*. These act as catalysts: speeding up the reaction by lowering the *activation energy* required for it to proceed. Think of the activation energy as a hill: the steeper the hill, the harder it takes to reach the peak. Say on a regular day, it takes five hours to bike to the peak. With enzymes (or, following the analogy, with a motorcycle or a car), it takes less than an hour to do so. Of course, the

Date: 02 April 2018.

[1] "I'm only taking this course to get into my program; I won't need math ever again!" —MAT134 students

magnitude of rate reduction is significantly greater with biological processes (typically 10^{-12} seconds [2] — almost instantaneous!), but the idea remains the same.

Now you might think, *"Oh, since enzymes hasten the reaction of interest, having an infinite amount of enzymes in my system would lead to infinite ATP! I'll never be tired anymore! I can pull all-nighters every night!"* Not so fast. Unfortunately, there is a *limit* as to how efficient an enzyme is in terms of speeding up the reaction. Moreover, there are other factors to consider: are there any competing compounds in my system that can affect enzyme efficiency? Would having a high concentration of reactants or products cause my enzyme to break down and give up? Perhaps the real question is: is there a mathematical model that can answer these questions for **any** biochemical reaction involving enzymes?

1.2. Michaelis-Menten Equation

In 1913, Leonor Michaelis and Maud Menten[2] wanted to quantitatively describe the kinetics behind enzyme-catalysed reactions, such as metabolism. Normally, substrates S are converted to products P by the following one-step equation [2]:

$$S \xrightarrow{k} P \tag{1.2.1}$$

In this case, the rate of appearance of P has a linear relationship with the amount of reactants (here, $[X]$ represents the concentration of chemical species X as a function of time), with proportionality constant (or *rate constant*) k:

$$\frac{d[P]}{dt} = k[S] \tag{1.2.2}$$

It would make sense to increase S in order to produce more P per unit time. Moreover, since (1.2.2) is linear, there is no upper bound involved; hence, on paper, it is possible to have significantly large reaction rates (tending to infinity) with large substrate concentrations. However, we mentioned in Section 1.1 that there is no such thing as infinite energy, which then means that (1.2.2) does not apply to our biochemical reaction.

Instead, Michaelis and Menten set up the problem as follows. Metabolism — or any biochemical reaction for that matter — is composed of a series of elementary reactions like (1.2.1), as mentioned in Section 1.1. First, the substrate attaches itself to the enzyme E at its active site, forming an enzyme-substrate complex ES. This process is reversible: since the reactants and products exist in *equilibrium*, the substrate is able to fall on and off the enzyme freely, with corresponding rate constant k_1 (for the forward reaction) or k_{-1} (for the reverse reaction). Once on the enzyme active site, the substrate then is converted to products: at this point, it is unfavoured for P to revert to S, nor is it possible to form an EP complex. Hence, this second step is simply a forward reaction, with its own rate constant k_2. The new chemical equation is then [2]:

$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} P + E \tag{1.2.3}$$

[2]She was a UofT alumna, earning her MD in 1911.

Now, the rate of appearance or disappearance of each component with respect to time can be determined from (1.2.3), as follows [3]:

$$\frac{d[S]}{dt} = k_{-1}[ES] - k_1[S][E] \tag{1.2.4}$$

$$\frac{d[E]}{dt} = k_2[ES] + k_{-1}[ES] - k_1[S][E] \tag{1.2.5}$$

$$\frac{d[ES]}{dt} = k_1[S][E] - k_2[ES] - k_{-1}[ES] \tag{1.2.6}$$

$$\frac{d[P]}{dt} = k_2[ES] \tag{1.2.7}$$

These four equations are derived in the same way as (1.2.2) — the change in concentration of each compound X is given by each equation in which it shows up; if X is being produced in the reaction, then the rate component is positive, and if X is being consumed, then the rate component is negative. Take (1.2.4) for example: for the forward reaction $S + E \rightarrow ES$, S is being consumed, while for the reverse reaction $ES \rightarrow S + E$, S is produced, both with corresponding rate constants k_1 and k_{-1}. Hence, the change in $[S]$ is the sum of the contributions of these reactions: $k_{-1}[ES] - k_1[S][E]$. Note that we have to multiply both concentrations for the forward reaction since the rate is linearly related to all reactants involved, as per (1.2.2).

Another interesting observation about these four equations is that (1.2.5) and (1.2.6), when added, result to zero. This makes sense in terms of the biochemical reactions since these describe the concentrations of the free unbounded enzyme and the enzyme-substrate complex. Given a fixed amount of enzyme in the system E_o, a fraction of this will be free enzymes ($[E]$), while the remaining fraction will be bounded with substrate ($[ES]$). Hence, the total enzyme concentration will remain constant throughout the reaction, and can be written as follows:

$$[E_o] = [E] + [ES] \tag{1.2.8}$$

Now, going back to the original question: we want to be able to obtain an equation that describes the rate of appearance of P, using quantities that we can measure. In this case, it would be beneficial to have (1.2.7) written in terms of $[S]$ and $[E]$, which are our starting compounds. Because the reactions $S + E \rightarrow ES$ and $ES \rightarrow S + E$ are in equilibrium, we can assume that $[ES]$ does not change with time (or is in *steady state*); hence:

$$\frac{d[ES]}{dt} = k_1[S][E] - (k_{-1} + k_2)\,[ES] = 0 \tag{1.2.9}$$

Rewriting (1.2.9) in terms of (1.2.8) yields:

$$\begin{aligned} k_1[S][E] - (k_{-1} + k_2)\,[ES] &= 0 \\ k_1[S]\,([E_o] - [ES]) &= (k_{-1} + k_2)\,[ES] \end{aligned}$$

With some rearrangements and algebraic manipulation, we can obtain another expression for $[ES]$:

$$[ES] = \frac{k_1[S][E_o]}{k_{-1} + k_2 + k_1[S]} = \frac{[S][E_o]}{\frac{k_{-1}+k_2}{k_1} + [S]}$$

The above expression can now be used to rewrite (1.2.7):

$$\frac{d[P]}{dt} = k_2 \left(\frac{[S][E_o]}{\frac{k_{-1}-k_2}{k_1} + [S]} \right) \tag{1.2.10}$$

Finally, to have a more compact equation, we recall that $\frac{d[P]}{dt}$ is the rate (or speed v) at which the entire metabolism reaction — including the intermediate steps — produce P. Moreover, if all of the enzyme present in the system is present as ES ($[E_o] = [ES]$), then (1.2.7), being similar to (1.2.2), would reach its maximum rate as $[E_o]$ is fixed.

Thus, the **Michaelis-Menten equation** is given as follows [2, 3]:

$$v = \frac{v_{max}[S]}{K_m + [S]} \qquad \text{where } v = \frac{d[P]}{dt}, \;\; v_{max} = k_2[E_o], \;\; K_m = \frac{k_{-1} + k_2}{k_1} \tag{1.2.11}$$

From (1.2.11), we can see that for small values of $[S]$ (or when $K_m >>> [S]$), the rate of production follows a linear relationship with $[S]$, just like in (1.2.2). However, when substrate concentration is sufficiently large (i.e. $K_m <<< [S]$), the infinite limit of (1.2.11) would then tend to $v = v_{max}$. Thus, there is indeed an upper bound as to how fast (and how much) P is produced, which relates to our earlier hypothesis that we cannot have infinite energy from eating infinite amounts of food.

2. Matters of the Heart

2.1. ECG: Modelling Heartbeats

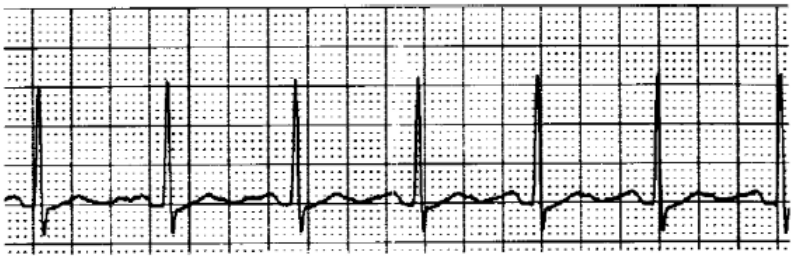

Figure 1. A normal scalar electrocardiogram reading [3].

When we talk about electrocardiograms (or *ECGs*), the first thing that comes to mind is the zig-zag graph of a person's heartbeat that shows up on a small screen. This graph (call it a *scalar ECG*) is periodic[3], with one cycle lasting about 300 milliseconds [3]. Potential heart problems can be discovered on a patient with the aid of an ECG, such as

[3]Of course, if the patient is close to death, then it won't be *perfectly* periodic. Beep beep beeeeeeeep.

abnormalities in heart rhythm: irregularities in the periodicity of each heartbeat at a certain time interval are possible signs of fibrillation (or arrhythmia), while significant differences in amplitude are signs of tachycardia (rapid beating of the upper/lower heart chambers) [3].

Although the scalar ECG is already helpful on its own, more information can be unearthed if we deal with the heart as a *volume* and not simply as a point. More importantly, it would be more advantageous health-wise if we can determine at which side of the heart the irregularities occur, without being invasive. This is where *vector ECGs* come in handy.

There are two unsolved problems in electrocardiography that are motivated by the relationship between the heart (as an electric current source) and the body. The entire body — including the heart — is electrically conductive on its entire surface, which allows us to measure its potential differences or *voltage* (independent of direction) given an applied external current (a vector). However, the most important current source occurs in the heart at every heartbeat [3].[4] Because current I is related to potential gradient $\nabla\phi$ as follows, with conductivity constant σ:

$$I = -\sigma\left(\nabla\phi\right)$$

and current in the entire body is conserved, the sum of all current sources, S, can the be written as the following Poisson equation[5]:

$$S = \nabla \cdot I = -\nabla \cdot [\sigma(\nabla\phi)] \tag{2.1.1}$$

The first problem, called the *forward problem of electrocardiography*, seeks a way to determine a solution to (2.1.1). Being able to find this solution ϕ, which gives the change in potential in the body caused by heartbeats *at any given time*, requires the exact location, direction, and strength of the current, as well as the conductivity of the entire body, which varies depending on the type of tissue [4].

The other problem, called the *inverse problem of electrocardiography*, follows from the previous problem by supposing that there already exists a solution for (2.1.1), which we call the *transfer function* T. T yields the surface potential ϕ_B of the entire body at any point in time:

$$\phi_B(t) = T \cdot S(t) \tag{2.1.2}$$

By knowing the explicit equation for T, and using potentials measured in the body, all current sources at play could then be obtained by inverting (2.1.2) [3]:

$$S(t) = T^{-1} \cdot \phi_B(t) \tag{2.1.3}$$

At this point, since these problems are yet to be solved, we can simplify some parts of the problem to have an adequate workable approximation. The easiest one is to take surfaces or regions in the heart with different σ values and represent them altogether in

[4]Not a coincidence that this subsection is also at the *heart* of this essay.

[5]We need not know more about this equation (nor ∇) for this essay, though it wouldn't hurt to learn something new in our spare time.

order to work with only one value. We call this value the *heart dipole vector* **H**: the sum of the current dipoles[6] experienced by the heart at every point (here, **J** is the current dipole density at a given point of the heart, and V is heart volume) [3,4]:

$$\mathbf{H}(t) = \int_V \mathbf{J}\, dV \tag{2.1.4}$$

We then say that (2.1.4) is fixed in space: the value only changes in terms of orientation and dipole strength; this allows us to take the dipole origin (or where we place the arrow's tail) to be the centre of the heart for convenience. Thus on the surface of the body, assuming σ on the surface is the same as σ on the inside of the body at a particular point, the experienced potential that came solely from the heart can be written in a similar form as (2.1.2):

$$\phi_B(x,t) = l_x \cdot \mathbf{H}(t) \tag{2.1.5}$$

wherein l_x is the *lead vector* associated with the electrode attached to a part of the body during ECG measurements [3]. Note that by taking the dot product of l_x and $\mathbf{H}(t)$, we are assured that ϕ describes the potential at a specific point x at a given time t.

Another simplification that we could make is to fix the number and locations of ECG measurements in order to describe **H** accurately using (2.1.5). While there are 12 locations currently used, the first three were discovered by Willem Einthoven[7] in 1913 [5]. He proposed that since **H** is a three-dimensional vector (corresponding to the heart as a three-dimensional surface), three lead vectors are sufficient. Moreover, these vectors form an equilateral triangle given by unit lead vectors $l_I = (1,0,0)$, $l_{II} = (\frac{1}{2}, \frac{\sqrt{3}}{2}, 0)$, and $l_{III} = l_{II} - l_I$ [3,5].

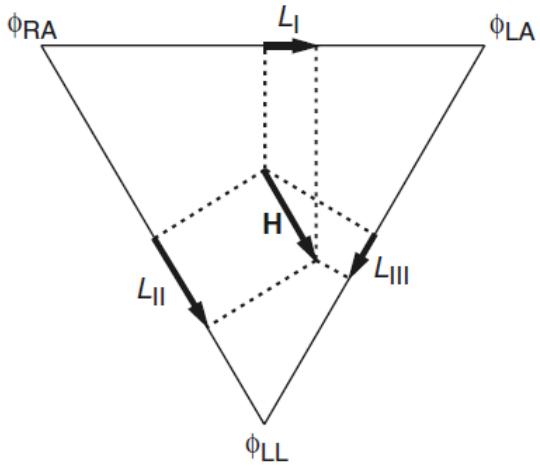

FIGURE 2. The Einthoven triangle, showing the heart vector **H** and lead vectors L_I, L_{II}, and L_{III}. ϕ_{RA}, ϕ_{LA}, and ϕ_{LL} denote the potential difference in the right arm, left arm, and left leg, respectively [3].

[6]Think of this as an arrow – a visual representation of a vector.

[7]Known as "the father of electrocardiography", he won the Nobel in physiology in 1924 [3].

H can then be broken down into its projections on the sides of the equilateral triangle (Figure 2), which are the lead vectors Einthoven mentioned (here, L_i and l_i correspond to the ith lead vector (or projection) and the ith unit vector (or triangle side), respectively):

$$L_i = \left(\frac{l_i \cdot \mathbf{H}}{\|l_i\|^2} \right) l_i = (l_i \cdot \mathbf{H})\, l_i \qquad \text{since } l_i \text{ is a unit vector, } \|l_i\| = 1$$

Finally, in order to bridge the gap between measurable and computed quantities, Einthoven defined the lengths of each lead vector to be the potential differences measured from both arms (RA and LA, for right and left) and the left leg (LA) during electrocardiography:

$$V_I = \phi_{LA} - \phi_{RA} = \|L_I\| \tag{2.1.6}$$

$$V_{II} = \phi_{LL} - \phi_{RA} = \|L_{II}\| \tag{2.1.7}$$

$$V_{III} = \phi_{LL} - \phi_{LA} = \|L_{III}\| \tag{2.1.8}$$

Thus, **H** can be determined through the vector ECG given by (2.1.6) to (2.1.8). A change in amplitude of part of the vector ECG would be caused by a change in either amplitude or orientation of **H**. Moreover, the direction in which **H** shifts away from corresponds to the side of the heart in which the irregularity lies, as in Figure 2 [5]. For example, if **H** shifted towards L_{III}, then the lead vector contributed by RA is relatively weak compared to LA and LL; this means that the right side of the heart experiences dysfunction.

2.2. Arrhythmias

As mentioned in the introduction, arrhythmias are irregular heartbeats, or disruptions of normal cardiac electrical cycle [3]. There are two types of arrhythmias: mild temporal disruptions, and re-entrant arrhythmia — caused by spatial distribution and deformation of cardiac tissue, and thus can lead to serious health risks.

Re-entrant arrhythmias occur in an analogous way as the "wave" [3]: a stadium crowd phenomenon wherein rows of spectators would suddenly rise from their seats, wave their arms up and cheer, then sit down again. No one would dare attempt to do the wave *alone*; there must always be a cheerleader or pep squad initiating this collective motion — an external stimulus. Moreover, the direction at which the wave travels, despite the stadium housing a huge crowd, always ends up being unidirectional: even though there is a possibility of the wave starting at two different points, the crowd would find a way to adjust and return to only one starting point.

We can then describe re-entrant arrhythmias as being triggered by an initial stimulus, occurring at a certain time and position in the heart such that there is non-symmetric propagation and possibility of initiation back to the starting point [3]. While the medical challenge would be modelling this accurately without the need for invasive techniques, the mathematical challenge then would be creating a phase singularity that satisfies these conditions.

Suppose there is a one-dimensional closed pathway of length L (or our *stadium* in this case), which allows only unidirectional flow (our *wave*), and is stimulated by a pacemaker (our *cheerleader*). This pacemaker fires electric pulses at time t with period T, and does so $n+1$ times. Then the stimulus period for this pathway would be $\Delta T_{n+1} = t_{n+1} - t_n$. If we take the speed of pulse propagation to be dependent on frequency (i.e. speed $= c(\Delta T)$), then the travel time around the pathway would be $\frac{L}{c(\Delta T)}$. Moreover, even though there is a delay for cardiac cells (our *spectators*) to positively respond between stimuli (call this the *refractory period* T_r), the time it takes for the pulse to move completely around the pathway is shorter than T (i.e. time $= \frac{L}{c(\Delta T)} < T$).

If the pulse travels at a shorter time than T_r, then it will have no effect on the cardiac cells, and thus cannot initiate arrhythmia. However, if its travel time is larger than T_r, then it can travel around in a wave without any obstructions until it reaches the initial point (same as the wave analogy earlier). Thus, the stimulus period can be written as:

$$\Delta T_{n+1} = t_{n+1} - t_n = \begin{cases} T, & \text{for } \frac{L}{c(\Delta T_n)} < T_r \\ \frac{L}{c(\Delta T_n)}, & \text{for } T_r < \frac{L}{c(\Delta T_n)} < T \end{cases} \tag{2.2.1}$$

We can thus map $\Delta T_n \mapsto \Delta T_{n+1}$ (Figure 3) and analyse its fixed points (where the solid graphs intersect the dashed line), which represent a periodic stimulus [3]. In the first case, wherein L is small and T is large, there are two fixed points corresponding to each case in (2.2.1); this represents a healthy heart as this follows the normal pulse movements. However, when T is sufficiently small as in the second case, there is only one fixed point, which occurs on the graph corresponding to the re-entrant stimulus. This then means that arrhythmia occurs when the difference between the pulse period and refractory period is small: the path will experience continuous pulses at a quicker pace, leading to irregularity in heartbeat.

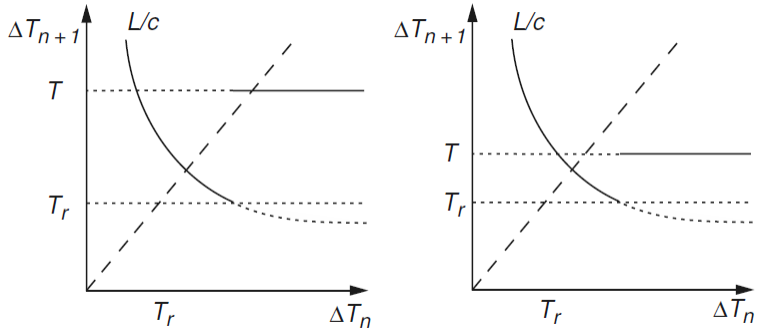

FIGURE 3. Next-interval maps of stimulus frequency that correspond to (2.2.1). The left graph illustrates normal conditions, while the right graph shows signs of re-entrant arrhythmia [3].

Looking at this at a physiological perspective, an increase in L without change in T could correspond to a growth in an already-existing diseased region in the heart. On the other hand, T could decrease with fixed L during periods of strenuous activity like

exercise; the heart itself does not grow in size permanently, but the greater need for oxygen in the body requires it to pump more blood per unit time. Both of these cases cause T to be relatively lower than L, which then leads to re-entrant arrhythmia.

3. Urine Analysis

3.1. Problem of Interest: Water and Solutes

We have been told at least once in our lives that we should drink plenty of water after eating salty food in order to prevent kidney or bladder problems.[8] But how much water is *plenty*? Say, for an interesting and realistic word problem [3], that we consumed a 1.5-ounce (or 42.5-gram) bag of potato chips. If salt is 58.5 grams per mole (or unit amount), and it dissociates in water into 2 osmoles (or 2 salt molecules per litre), then how much water[9] should we drink in order to flush all the salt out of our body?

First, we need to find out exactly how much salt in moles (n) is in our potato chips:

$$n = \frac{42.5 \text{ grams of chips}}{58.5 \text{ grams of salt per mole}} \approx 0.7265 \text{ moles of salt}$$

Then, since the body can concentrate a maximum of 1200 milli-osmoles (or 1.2 osmoles) of solutes per litre of urine, we only need to find the minimum volume of urine required to flush n moles out — exactly the same volume of water we need to drink:

$$V = \frac{0.7265 \text{ moles of salt } \cdot 2 \text{ osmoles per mole of salt}}{1.2 \text{ osmoles/L}} \approx 1.21 \text{ L}$$

That was an easy calculation, although hard to do in reality. Perhaps the real question is, how does the body concentrate all these unwanted solutes into urine without causing complications?

3.2. The Loop of Henle: Natural Filter

The loop of Henle, located in our excretory system, serves as the main diluting and concentrating site for unwanted excess solutes before its release as urine. Since our body is composed of several fluids, all containing solutes that need to be removed, the challenge is to be able to transfer as much solute into the loop of Henle as possible, given biological and physical constraints. How can we model this?

To maintain a large concentration difference between the different fluids in the body and urine, it is important to have a *countercurrent mechanism* that deals with the different rates of solute diffusion between these two fluids [3]. Suppose concentration is a function of space and time (i.e. $C(x,t)$), and we have two liquids with concentrations C_1 and C_2 and initial concentrations C_1^o and C_2^o flowing into two tubes with length L

[8] That or limit our junk food intake, but do we *really* want to say goodbye to tasty unhealthy snacks?

[9] The easiest answer would be 8 glasses of water, but we'd like to know the bare minimum.

separated by a permeable membrane. Then the concentrations in these tubes are given by:

$$\frac{\partial C_1}{\partial t} + q_1 \frac{\partial C_1}{\partial x} = d(C_2 - C_1) \tag{3.2.1}$$

$$\frac{\partial C_2}{\partial t} + q_2 \frac{\partial C_2}{\partial x} = d(C_1 - C_2) \tag{3.2.2}$$

We assume that *steady state* exists between these two tubes (a similar assumption was made in Chapter 1.2), so that the time derivative of both liquid concentrations is zero. Thus, combining and integrating (3.2.1) and (3.2.2), we find that the sum of concentrations, no matter which point in space, is constant along L [3]. Thus, we can find an ODE that describes the change of concentration (say C_1) with respect to position:

$$\frac{dC_1}{dx} = \frac{d}{q_1 q_2}(k - (q_1 + q_2))C_1) \tag{3.2.3}$$

and integrating this would give us:

$$C_1(x) = \frac{k}{q_1 + q_2} + (C_1^o) - \frac{k}{q_1 + q_2} e^{-\lambda x} \qquad \text{where } \lambda = d\left(\frac{q_1 + q_2}{q_1 q_2}\right) \tag{3.2.4}$$

The same derivation can be made for C_2. Thus, we can say that the output concentration from liquid 1 is an exponentially decreasing function of position; the longer L is, the harder it is for the solutes to become concentrated into the loop of Henle.

Conclusion

As we have seen, mathematics appears in different forms in different aspects of physiology and medicine, with complexity ranging from simple ordinary differential equations to more complex systems. Moreover, these applications allow medical practitioners and researchers to come up with discoveries and draft hypotheses without requiring numerous living test subjects, making it cost-efficient and ethical. So while the rest of the world still look at mathematics as arrhythmia-inducing[10], Prof. Keener is right: arrhythmias — or biological processes in general — cause mathematics.

References

[1] J. P. Keener. *Arryhthmias by Dimension (Lecture),* `http://www.math.utah.edu/~keener/lectures/AMS/Arrhythmias_by_dimension.pdf`

[2] R. H. Garrett and C. M. Grisham, *Biochemistry, Fourth Edition.* Brooks/Cole, Boston, 2010.

[3] J. P. Keener and J. Sneyd, *Mathematical Physiology.* Springer-Verlag, New York, 1998.

[4] R. M. Gulrajani, *The Forward and Inverse Problems of Electrocardiography,* IEEE Engineering in Medicine and Biology Magazine, vol. 17, No. 5 (1998), pp. 84–101. `doi:10.1109/51.715491`

[10] This is why we need better mathematics outreach. Math should be easily accessible and not scary!

[5] W. Einthoven, G. Fahr and A. de Waart, *On the Direction and Manifest Size of the Variations of Potential in the Human Heart and on the Influence of the Position of the Heart on the Form of the Electrocardiogram*, trans. by H. E. Hoff and P. Sekelj, Am. Heart J., vol. 40, No. 2 (1950), pp. 163–211. doi:10.1016/0002-8703(50)90165-7

E-mail address: geneva.liwag@mail.utoronto.ca