Collaborating with other applicants on the EE is prohibited.

Using LLMs (ChatGPT, Claude, DeepSeek, etc.) or generative AI for the EE is also prohibited.

Violation of either of these rules will affect your admission.

SigmaCamp Next Entrance Exam 2025

This Entrance Exam (EE) is only one part of your application to SigmaCamp Next. For full instructions, please check https://sigmacamp.org/2025next/apply.

To help match you with the best semilab for your interests and skill level, the EE is designed to reflect the topics of this year's semilabs. Each semilab instructor has composed one problem related to their semilab and a skills questionnaire to help us understand your background. The problems are designed to give you a taste of the topics covered in the semilabs and to see whether you are excited to think deeply about the material.

How to Approach the Exam

Choose no more than five semilabs that you are interested in and solve the problems and answer the skills questionnaires for them. We recommend submitting solutions and answering questionnaires to **at least three semilabs.**

You do not need to fully solve a problem — we are most interested in seeing your thought process. Even if you only make partial progress, please submit your work!

Explain your reasoning clearly: Even if you arrive at a correct answer, showing your steps and thought process is important. We are most interested in seeing your approach and your thought process.

How Your Solutions Are Used

The Entrance Exam is not used as a strict placement test. Instead, your responses help us match you with the best semilab for your skills and interests. Demonstrating curiosity, effort, and engagement is more important than arriving at perfect solutions.

You can use the Internet, books and even help from someone (who is not another applicant), but **state precisely what sources you have used to solve each problem**. Note that you cannot post the problems to the Internet or other public forums and solicit help that way. **You are also not allowed to use any LLMs (ChatGPT, Claude, DeepSeek, etc.) to solve any of the problems.**

You cannot collaborate with other applicants.

1. Your solutions should be submitted as `.pdf` files. For EE problems that involve coding, Python and Java formats are acceptable (`.py`, `.java`).
2. Your solutions can be handwritten or typed. Handwritten solutions must be scanned as `.pdf` (not as `.jpg` or `.png`) files.
3. Each solution should be submitted in a separate file. Please don't combine the solutions for different semilabs into one file.

The application deadline is April 15, 2025 — all your materials (Entrance Exam, essay, letters of recommendation) must be submitted by that date. We will notify all applicants no later than May 5, 2025.

Good luck with your application!

Game Theory: From Nim to Surreal Numbers
Andrey Boris Khesin

Skills Questionnaire: https://forms.gle/KgfWLcreGXcZATZL9

Alice and Bob are playing the following game: There are two piles of stones. One pile has A stones and the other has B stones. Alice and Bob take turns taking **one or more** stones from **exactly one** of the two piles, with Alice going first. When there are no stones left, the player who cannot move loses. Thus, whoever takes the last stone wins.

The allowed moves are as follows. If the **current sizes** (not the starting sizes!) of the piles are X and Y, a player is allowed to take Z stones from a single pile, as long as Z is a multiple of at least one of X or Y. (Multiple, not divisor!)

For example, if the current pile sizes are 2 and 7, then the possible moves are to take either 2 stones from the first pile, 2 stones from the second pile, 4 stones from the second pile, 6 stones from the second pile, or 7 stones from the second pile. Note that making any move changes which moves are allowed!

We say that a player "has a winning strategy" if they can guarantee that they will win the game from the current position as long as they make the right moves, regardless of which moves their opponent plays. For each starting setup, exactly one player will have a winning strategy.

The question: If the starting pile sizes are A and B where $A = 8$, for which nonnegative integers B does Alice have a winning strategy?

For optional extra credit: Describe all pairs of nonnegative integers (A, B) for which Alice has a winning strategy.

Note: This question is quite tricky! Think carefully about what situations result in the current player (initially, Alice) winning versus the player who is waiting for their turn winning (initially, Bob). The case of $A = 8$ might be too hard to dive into right away. Work your way up from small cases for A and B and see if you can find a pattern! If you don't find the pattern, be sure to write up as much partial progress as you can.

Adventures in Geometry beyond Euclid
Alexander Brook

Skills Questionnaire: https://forms.gle/JFMfFe3HtVDdqcAx7

Note: This problem has 3 parts, and some parts have more than one question in them. Partial solutions answering only some of the questions are welcome. Completeness and rigor of your answers are more important than the amount of questions answered, although writing just an idea for some of the questions can also be helpful. None of the questions have a single correct answer.

In Euclidean geometry, we define the distance between a point and a (straight) line to be the shortest distance from our point to any point on the line. We then prove that this distance is realized as the length of a line segment connecting our point to the line and perpendicular to the line. If you have not seen this proof, try proving this result, and then look for a proof in a book or online, for example here or here.

(a) How would you rephrase the definition of distance and the theorem about perpendicularity for the distance from a point to a circle? Prove your theorem based on your definition. Make sure to consider all possible cases.

(b) Pick another class of "nice" curves; it can be as general (for example, some version of "smooth") or as specific (for example, ellipses) as you like. Provide the exact definition of "nice curve" that you are going to use. Define the distance between a point and a curve to be the shortest distance from the point to any point on the curve. State and prove a theorem connecting this distance to the length of a "perpendicular" segment (you will probably need to define what "perpendicular" means exactly for your curves).

Try to make sure that your class of "nice curves" is as general as your theorem allows; if it is not, describe what you think the most general class can be.

(c) Give an example of a "curve" and a point so that the distance from the point to the curve is not the length of a perpendicular segment.

Can you provide a definition of "curve" that will include your example but be substantially more restrictive than "any set of points in the plane"?

MatheMagics: Relating Advanced Math and Magic Tricks
Ricardo Teixeira

Skills Questionnaire: https://forms.gle/vAMv2saMMdzEsvD29

Question 1

Ricardo is twice as old as Jeffrey was when Ricardo was half as old as Jeffrey will be when Jeffrey is three times as old as Ricardo was when Ricardo was three times as old as Jeffrey. Their combined age makes forty-eight. How old is Ricardo now?

Question 2

Four friends, Amy, Danny, Jeff, and Park, are playing cards. There are 20 different cards, each one has one colour (blue, green, yellow, or red) and a number from 1 to 5. Each friend received five cards, so each card belongs to someone. They say:

Ali: "I have four cards with the same number."

Brad: "All my cards are red."

Jeff: "I don't have any blue or green cards."

Park: "I have a full house, that is, 3 cards with the same number, and 2 cards with some other number."

Only one person is lying. Who lied?

Intro to Complex Systems
Lilianne Mujica-Parodi

Skills Questionnaire: https://forms.gle/XB86tS8po5HeFc9r8

Consider a network of users of a newly-released app, where each **new** user is connected (e.g. uses the app to communicate with) to two randomly chosen **existing** users.

1. (a) Draw an example network configuration when the app is used by 10 people.

   (b) Estimate the number of connections that the most highly connected user might have when the app scales to all of Earth's population.

2. Now consider a similar app, where each new user is also randomly connected to two existing users. These two existing users are chosen randomly with probability proportional to their number of existing connections.

   (a) Draw an example network configuration when the app is used by 10 people.

   (b) Estimate the number of connections that the most highly connected user might have when the app scales to all of Earth's population.

3. The market value V of the network to each new user is given by

$$V = An^2,$$

   where n is the number of existing users and A is the affinity of existing connections.

   Let's define the cost to join the network for a new user as

$$C = Bn,$$

   where B is the cost of establishing a connection to each existing user in the network.

   Find the ratio of A to B if this network became attractive (more value than cost) to new users only after after reaching the size of $n = 100$.

DNA detectives: Cracking the Code from Genes to Traits
Vladislava Sokolova

Skills Questionnaire: https://forms.gle/VhyiK7UTjTv1HE1u8

A graduate student in the plant biology group was working on a genome organization project. The student extracted a fragment of the Arabidopsis chromosome as part of their work. The student was concerned about the possible failure of the pilot experiment, so as a precaution, he decided to replicate his sample using PCR (we will assume the existence of the polymerase capable of replicating very long DNA, such as plant chromosomes). Sometime after that, the student realized that he did not label neither his original sample nor the PCR reaction tube. He is now in possession of two identical tubes, one with the fragment of the chromosome, one with its PCR copy.

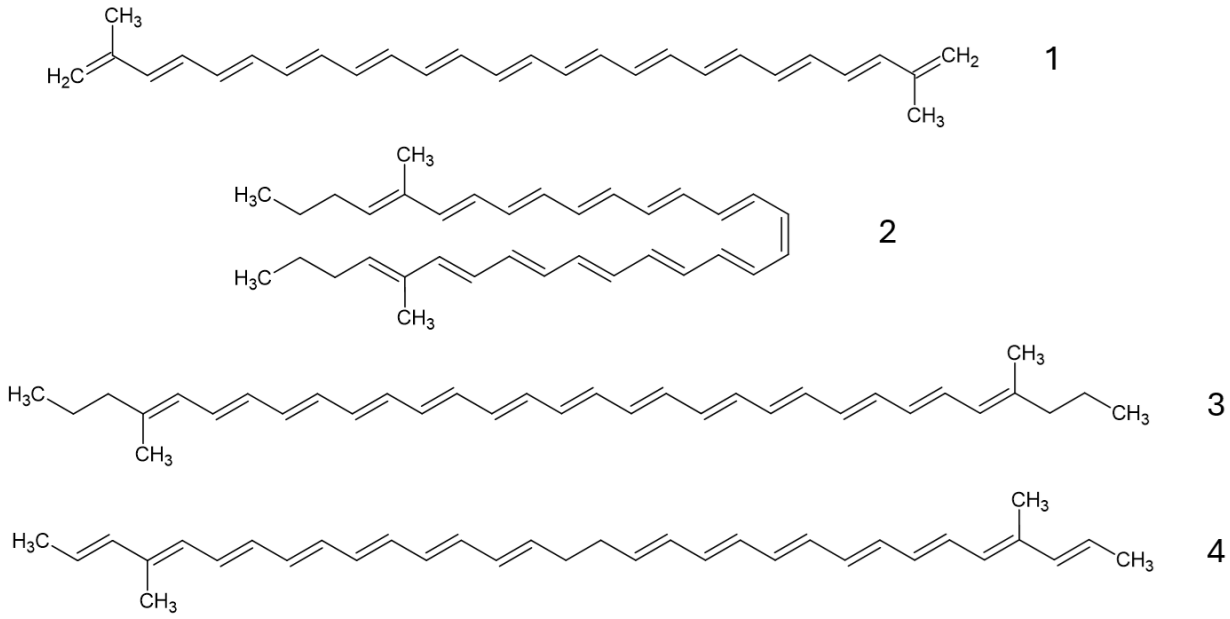
(a) How can the student determine which tube is which? He can use any biophysical and biochemical methods that you can think of. Name as many options as possible and explain the determination process.

(b) How could you use Tube 1 (chromosome) and Tube 2 (PCR)? Plan several experiments with them. What kind of biological problem can you elucidate with these experiments?

Molecular Structure, Spectra, and Spectroscopy
Mark Lukin

Skills Questionnaire: https://forms.gle/YGmfdrHQLj2Y2hTA8

The figure below displays the structures of four organic compounds (labelled 1 to 4). Please rank them according to their absorption maximum in the visible spectrum, beginning with the compound that has its absorption maximum nearest to the "red" end of the spectrum. Explain your answer.



If you're unsure of how to respond, please clarify what specifically is causing the most difficulty.

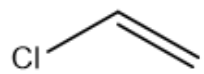
Inside the World of Polymers
Alina Khankin and Victor Paiva

Skills Questionnaire: https://forms.gle/xnv3G7Fjmpjf58pZ9

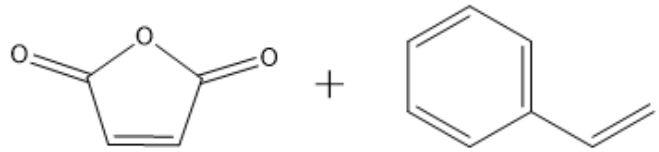
Polymerization is a process in which small molecules called monomers join together to form a long-chain molecule or polymer. This reaction can occur through various mechanisms, such as addition or condensation, depending on the type of monomers and reaction conditions. The resulting polymers are the building blocks of many everyday materials like plastics, synthetic fibers, and rubbers, making polymerization a key process in both industrial production and everyday life.

(a) Draw the structures of the polymers obtained by using the following monomers. Indicate if more than one structure is possible and, if so, draw one more possibility.

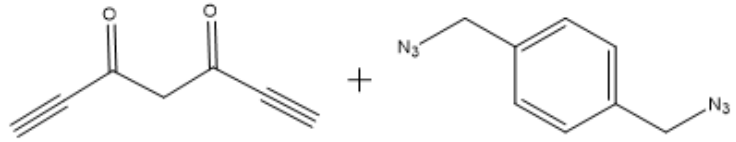
i)



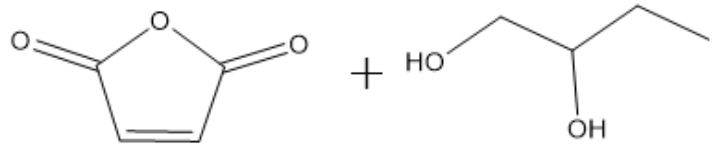
ii)



iii)



iv)



(b) Suppose that the reaction (iv) was repeated, replacing some of 1,2-butanediol with butanol. What is going to be the effect of this replacement on the physical properties of the obtained polymer?

(c) Define the polydispersity index and compare this parameter for polymerizations (i) and (iii).

Analytical Mechanics
Alex Frenkel

Skills Questionnaire: https://forms.gle/k5XXzKnaf8WQdJTz6

Note: Parts (b) and (c) may be quite challenging. Partial solutions explaining your strategy and thought process are encouraged. Clearly written solutions and clearly drawn diagrams are especially encouraged.

Consider Figure 1 below. A bead of mass m slides along a ring of radius R. The ring is very massive, and rotates (much like how a spinning coin rotates) around the vertical axis (represented by the dashed line) with angular frequency Ω. The gravitational acceleration is g.

(a) **(Easy)** At what angle θ will the bead be in stable equilibrium? At what angle θ will the bead be in unstable equilibrium?

(b) **(Medium)** If you slightly displace the bead by an angle $\delta\theta$ from a stable equilibrium, it will oscillate around the equilibrium with some period T. Find this T in terms of the given parameters.

(c) **(Hard)** Repeat the exercise for the system in Figure 2. Now, instead of just a ring and a bead, we have a smaller ring of mass M_2 and radius $R_2 < R_1$ that may slide freely along a big ring of mass M_1. The smaller ring rotates with angular frequency Ω_2 around the line connecting its attachment points to the big ring (this line is a diameter of the small ring). The big ring is constrained to rotate around the vertical axis with frequency Ω_1.

Find all necessary relations between the parameters M_1, M_2, Ω_1, Ω_2, R_1, R_2, θ_1, θ_2 and m in order for the system to be in stable or unstable equilibrium. Find the period T of small oscillation around the stable equilibria if the bead is displaced by $\delta\theta_2$.

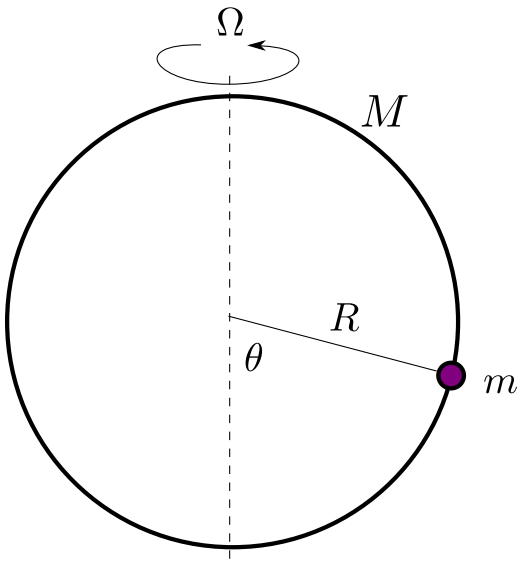


Figure 1

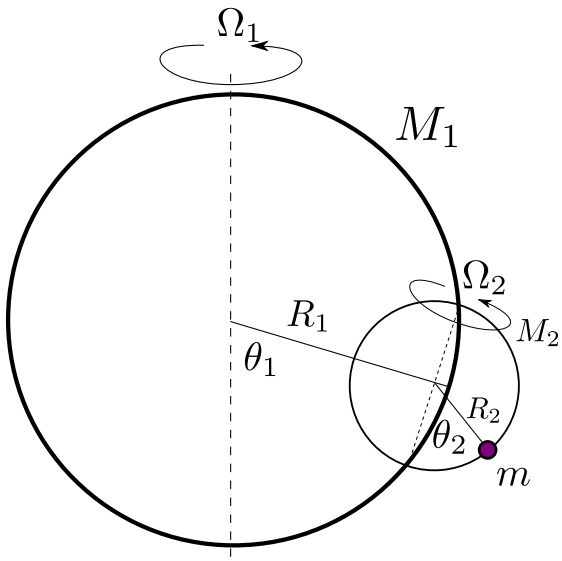


Figure 2

Ocean Physics
Valentin Skoutnev

Skills Questionnaire: https://forms.gle/FvG4rKWBtzbRrLyz5

A cylinder of ice is submerged vertically into a large bath of fresh water initially at some temperature $T > 0°$ C and allowed to melt, as shown on the left in Figure 1. Three such experiments are carried out at different initial temperatures and the photographs of the melting ice, labeled (1), (2), and (3), are shown on the right in Figure 1. The three temperatures were $T = 4°$, $7°$, $10°$ C.

Question 1: Which photograph corresponds to which experiment and why?

Question 2: How would the experimental outcome change if the bath water was instead very salty?

Hint: The density of liquid water is highest at $4°$ C and lowest near $100°$ C.

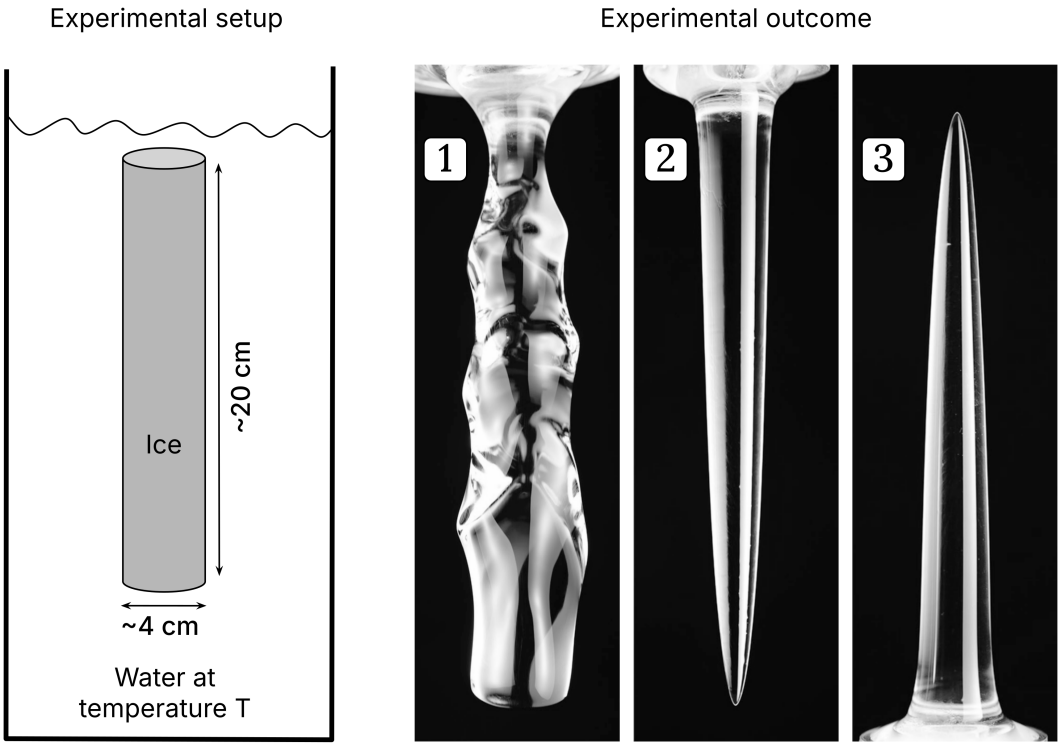


Figure 1: Left: Diagram of the experiment. Right: Photographs of the melted ice taken out after a few minutes in three different experiments that varied the initial water temperature.

Web Engineering
Nikita Korobkov

Skills Questionnaire: https://forms.gle/qP8Dr1BDmrhzGeiv6

You are given a 6-row by 7-column **Connect 4** board state as input. Your task is to write code to determine whether the board configuration is **valid** based on the rules of the game. If the board is valid and one player has won, print the winning player.

Validation Rules

A valid board state must satisfy the following conditions:

1. The board contains only three types of symbols:
   - `X` for Player 1's pieces
   - `O` for Player 2's pieces
   - `.` for empty spaces
2. Pieces must be stacked correctly:
   - A piece **cannot** float above an empty space (every `X` or `O` must be supported by another piece or the bottom of the board).
3. The number of pieces must be balanced:
   - The number of `X` pieces must be **equal to or exactly one more** than the number of `O` pieces.
   - If the board contains a winning sequence then number of pieces should be exactly equal if `O` won, and differ by 1 if `X` won.
4. There should be **at most one winning player** (with 4 in a row horizontally, vertically, or diagonally).
   - If both players have a winning sequence, the board is **invalid**.
5. A hidden condition that you can figure out on your own for an extra point. (not required)

Input Format

The input consists of a list of **six** strings, each containing exactly **seven** characters (`X`, `O`, or `.`), representing the board from **top to bottom**.

Output Format

- Print `VALID` if the board state is correct and no player has won.
- Print `VALID X` if **Player 1 (X)** has won.
- Print `VALID O` if **Player 2 (O)** has won.
- Print `INVALID` if the board violates any of the rules.

Examples

Example 1: Valid Board (No Winner)

```
Input:
.......
.......
.......
...X...
...XO..
..XXOO.

Output:
VALID
```

Example 2: Valid Board (X Wins)

```
Input:
.......
.......
..X....
..XOX..
..XOO..
XXXOOO.

Output:
VALID X
```

Example 3: Invalid Board (Floating Piece)

```
Input:
.......
.......
...X...
.......
...XO..
..XXOO.

Output:
INVALID
```

Example 4: Invalid Board (Two Winners)

```
Input:
.......
.......
...XXO.
..XXXO.
..XOOO.
XXXOOO.

Output:
INVALID
```

Scientific Machine Learning for Science
Helmut Strey

Skills Questionnaire: https://forms.gle/J9UYc6wH8jAs3jcx7

Protein Folding Simulator: Basic Path Finding

You are studying a simple protein chain made up of 8 amino acids. Each amino acid can be either hydrophobic (H) or polar (P), represented as a string like 'HHPPHHPP'. The protein folds on a 2D grid where:

- Each amino acid occupies one grid point
- Adjacent amino acids in the chain must be orthogonally adjacent on the grid
- The protein cannot cross itself
- Hydrophobic amino acids (H) prefer to be adjacent to other H's

The goal is to find a valid folding pattern that maximizes H-H contacts.

Please write code for the following functions:

1. Write a function that checks if a folding pattern is valid.

   The function should accept a protein chain and a folding pattern and should print `VALID` or `INVALID`.

2. Implement a simple algorithm to find at least three valid folding patterns. Print those foldings using the output format together with the number of H-H contacts.

   The function should accept a protein chain and print several folding patterns.

3. (Bonus) Try to find the optimal folding that maximizes H-H contacts. Are there several folding patterns that are optimal?

   The function should accept a protein chain and print several folding patterns.

Protein chain format (function input): 8 character string from the set of {'`H`','`P`'} (e.g. '`HHPPHHPP`')

Folding pattern format (function output): String of six characters in set {'`R`', '`L`', '`S`'} (e.g. '`SRSRSL`') to indicate whether the segments of the 2nd to 3rd amino acid, 3rd to 4th, etc. turn right, turn left, or go straight from the previous segment.

Mathematical Modeling: Asking the Right Questions
Boris Barron

Skills Questionnaire: https://forms.gle/JR9yRbUVBxkyJ3VZA

Bob wants to estimate the maximum value of a concave-down quadratic function. The function is given in factored form as

$$f(x) = a(x - x_1)(x - x_2), \quad a < 0,$$

where the roots satisfy $x_2 > x_1$. The maximum of this function, M, occurs at

$$x_M = \frac{x_1 + x_2}{2} \quad \text{with} \quad M = f(x_M) = -\frac{a}{4}(x_2 - x_1)^2.$$

However, Bob's measurements of the roots are noisy. Instead of x_1 and x_2, he observes

$$r_1 = x_1 + e_1 \quad \text{and} \quad r_2 = x_2 + e_2,$$

where e_1 and e_2 are independent noise terms with zero mean ('independent' means that an observed value for e_1 does not affect the value of e_2 and vice versa, while 'zero mean' implies that averaging over many measurements of e_1 results in a value of 0, and similarly for e_2). Using a set of measured roots, Bob estimates the maximum as

$$M_{\text{est}} = -\frac{a}{4}(r_2 - r_1)^2.$$

Bob repeats this process many times, collecting multiple M_{est} values and then averaging them to form his final estimate.

Questions:

1. Bob's first estimate underestimated the actual maximum. As he continues collecting many more measurements, what should we expect his average estimate to do?
   (a) Underestimate M,
   (b) Equal M, or
   (c) Overestimate M?
2. Alice claims she can remove the discrepancy between Bob's average estimate and M by discarding some of Bob's measurements before averaging. She realizes that there are multiple ways she can accomplish this—*i.e.*, she can remove different combinations of measurements from the overall set of Bob's measurements—but notices something peculiar: regardless of which specific measurements she removes to resolve the discrepancy, the average value of the noise in the left root (e_1) is greater than the average value of the noise in the right root (e_2) in her resulting 'filtered' set of measurements.

   Is this evidence that the noise distributions of e_1 and e_2 were different in Bob's overall set of measurements or entirely a consequence of Alice's filtering procedure?

Note: This question is designed to be conceptually challenging. You are being evaluated on your ability to navigate conceptual challenges and your ability to justify conclusions in mathematical terms.