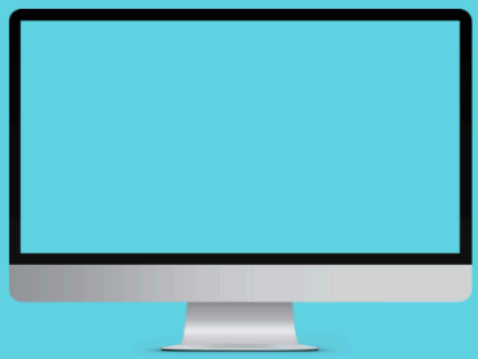
TMAS

TMAS Academy



ACE

AP Computer Science: Principles

2025

- 100+ Problems
- All Topics
- Detailed Solutions

Ipsaan Sedhai and Aviva Iyerkhan

Contents

§0.1 Credits/Acknowledgments

- I would like to acknowledge **Ritvik Rustagi** for giving me the opportunity to write this book on AP Computer Science Principles for TMAS Academy.
- I would also like to thank **The College Board** for outlining the curriculum which this book is built upon.
- Lastly, I would like to credit **Evan Chen**, a math Olympiad coach as well as a PhD MIT student for providing the template used to write this book.

§0.2 About the Author: Ipsaan Sedhai

My name is Ipsaan Sedhai. I'm a student at Sewanhaka High School who is interested in physics and enjoys programming/computer science.

Before writing this book, some of my computer science classmates would jokingly call me "The Professor" due to my knowledge of computer science allowing me to help them learn concepts in the class. Afterwards, I found out about TMAS Academy and decided to write a book on AP Computer Science Principles to help out others in the same way I was able to help out some of my classmates. My qualifications to write this book come from receiving a 5 on the AP Computer Science Principles Exam, as well as on AP Seminar and AP World History in the same year. I have also developed games with Lua in Roblox and I'm skilled in various programming languages such as C++, C, Python, JavaScript, and others.

I hope that with this book, more people can get a 5 on AP Computer Science Principles, especially if computer science is something new to them.

§0.3 About the Author: Aviva Iyerkhan

My name is Aviva Iyerkhan. I'm a student at Lynbrook High School and I worked on the end-of-chapter questions in this book. My qualifications include receiving a 5 on the AP CSP Exam and tutoring computer science (Python, Pygame, JavaScript, etc.) at and outside of school for the last 4 years. I hope this book is helpful as a learning and practice resource!!

§0.4 About TMAS Academy

TMAS Academy, previously known as Explore Math, was started by Ritvik Rustagi in 2020. TMAS Academy stands for The Math and Science Academy. TMAS Academy has previously published six books: *ACE The AMC 10/12, ACE AP Physics 1, ACE AP Calculus AB, ACE AP Calculus BC, ACE Physics C: Mechanics, AP Chemistry* and now the seventh, brought to you by Ipsaan Sedhai and Aviva Iyerkhan: *ACE AP Computer Science Principles*. For more information about TMAS Academy, check out the official website.
Website: https://www.tmasacademy.com/

§0.5 Benefits of Doing This Course

The technology of the world is evolving at a rapid rate. In the modern age it is impossible to escape from computer science. When you watch your phone, it is computer science. When you use a keyboard, there is computer science. Computers can be found throughout many other disciplines as well. Some computer programs are used to predict the shifting of the stock market, simulate physics models, help sort through DNA sequences to find diseases, attempt to disprove conjectures in mathematics, and much more. Thus, computer science has become vital knowledge and is only gaining popularity around the world. The possibilities are endless.

§0.6 Resources/Book Structure

The format of the code is based on the AP Computer Science Principles Reference sheet. You should familiarize yourself with this as it is the format for the official AP exam and for this book.

Also remember that the AP Exam will be **1 indexed**, meaning we start counting lists from index 1 and not 0. If you don't know what "1 indexed" means, then you might just be starting off with programming. If that is the case, the programming required for this course can surprisingly be learned in about two to three hours through YouTube.
There are numerous resources for programming online, but you can use:

1. **Code Camp YouTube Tutorials**
   a) JavaScript Tutorial
   b) C++ Tutorial
   c) Python Tutorial
   d) C Tutorial
   e) Check out Code Academy for unlisted languages
2. **Code.org**
   a) Code.org Website - You can create your own free account on code.org and follow the AP Computer Science Principles course that code.org offers. It covers exactly what you need to know in terms of programming, but for content perhaps books such as this and your teacher are your best bet. Code.org can also be used for the Create Task submitted to College Board.
3. **Khan Academy**
   a) Khan Academy AP CSP Course
4. **TMAS Academy Discord Server**
   TMAS Academy has a discord server dedicated for helping students, featuring a computer science channel. Feel free to ask questions!
   a) Discord Server
5. **Scratch**
   a) Scratch - For some people, programming can seem intimidating with its mysterious syntax. If that's the case for you, feel free to try code blocks in Scratch or Code.org. However, remember that a lot of questions on the exam

will be written in code format and not code block. However, code blocks can suffice to warm you up to programming.

6. **Runestone Academy**
    a) Runestone Academy - There are two CSP textbooks found on the website: one for line coding (StudentCSP) and one for block coding (mobileCSP). Runestone also has additional practice problems and activities to help learn how to use programming language for the create task.

7. **W3Schools**
    a) W3Schools- A completely free resource that offers help with CSP concepts and using programming languages for the create task.

I highly recommend learning some programming before beginning this book, or the course in general. That way, you'll be way ahead of the game and you will be able to do the questions in this book.

§0.7 Exam Format

The AP Computer Science Principles Exam has a multiple choice section, which is worth 70% of the exam grade and lasts 120 minutes. The other 30% is made up of a Performance Task, (i.e., a program that you create and submit to the College Board), and a Free Response Section.

The Performance Task is worth 10% and you **must** be given at least 9 hours to complete it in class by your teacher, though you can spend time outside of class completing it as well. The FRQs (Free Response Question) are worth 20% and are based on your Create Task. During the FRQ section, you will be given a Personalized Project Reference (a reference sheet) including some portions of your program code to help answer the FRQs. You will have 60 minutes for two free response questions, with question 2 being split into three parts (an a, b, and c).

In order to get points for the Performance Task and have your Personalized Project Reference, you must submit it to College Board through the official website before a specified date (usually somewhere in April). Make sure to adhere to all College Board specifications before submitting the Create Task or Personalized Project Reference. Certain things such as including comments and not having purely just source code in the Project Reference can lead to 0 points on that section.

This chart by College Board summarizes it:

Section	Question Type/Component	Number of Questions	Exam Weighting	Timing
I	Multiple-choice questions	70	70%	120 minutes End-of-course AP Exam
	Single-select	57		
	Single-select with reading passage about a computing innovation	5		
	Multi-select	8		
II	Create Performance Task	*See Below*	30%	*See Below*
	Program code, video, and Personalized Project Reference			At least 9 hours in class
	Written response questions related to the Create performance task	2		60 minutes End-of-course AP Exam

More info can be found here. I recommend reading it since it dives into the specifics.

1 Unit 1: Creative Development (10-13%)

§1.1 Benefits of Collaboration

The information about collaboration is short for AP Computer Science Principles. There are only a couple of main things that should be known:

Note 1.1.1

Why collaborate?
When a problem asks about the benefits of program collaboration, 90% of the time, the answer is that it reflects different viewpoints. This may not *always* be the answer, but these types of questions are meant to be simple, with obvious answers that could be benefits.

§1.2 Pair Programming

Pair programming is a system of collaboration for programming where one person is the "driver", that writes the code, and the other person is the "navigator", that reviews each line of code as it is written. The navigator serves as a second set of eyes to verify the validity of the code. Types of collaboration are unlikely to be asked on the exam, but here are the two that may show up:

Note 1.2.1

Iterative Development Process *rare
The iterative development process is a repetitive process in which the same steps are repeated several times to refine a program. This approach to program development is useful when the application is likely to have to undergo changes. For example, if a new technology is being used, as the new technology evolves, the program will have to as well.

Note 1.2.2

Incremental Development Process *rare
The incremental development process is when a program is broken into parts and developed independently before being put together to have a final product. This approach to program development is useful when the application is very large and when multiple people are working together. It also makes it easier to debug the program as each separate part is tested independently.

Sometimes on the AP exam, a question asks about the iterative AND incremental development process, which is a combination of the two.

§1.3 Input and Output

Input is something that is put into a program, and the output is the result after code functions on the input. It can be thought of like this:
Input → Program Code → Output

Note 1.3.1

Types of input
Types of input programs usually take in include
1. Clicking (Clicking a button for example)
2. Audio (Hey Siri)
3. Text (Entering username and password)

§1.4 Event Driven Programming

Event driven programming is a program that runs from a certain trigger called an "event". Most of these events/triggers are basically just the inputs that were listed above, and these are what make event driven programs run.

Example 1.4.1

Event Driven Example
An example of an event driven program is YouTube. When you click a video on YouTube, an event is triggered, firing program code to start displaying that video on your screen.

Example 1.4.2

Non-event Driven Example
An example of a non-event driven program would technically be a program that is just compiled and executed, without any user triggering it through input such as audio or clicking.

§1.5 Programs: How Do They Run?

Programs are written in a programming language that humans are able to easily comprehend, and we simply run the program, without really knowing what is going on under the hood.

Note 1.5.1

This is an example of an **abstraction**, a common concept in computer science where more complex details are hidden by a generalized simpler version. The specific low level complex instructions that are happening inside of the computer is abstracted into high level programming for our convenience.

The computer does not understand programming language, and instead it only understands **machine code**, or **binary**. Binary is all those 1s and 0s seen in stereotypical

film depictions of hackers, and they represent the code that computers can actually understand. Therefore, to get from the program code that humans use, to the binary code that computers use, we have something called a **compiler**. Typically, languages get **compiled** into machine code by the compiler, which makes the program executable by the computer.
For now, don't fret over binary. It will be explained in greater detail in Unit 2.

Note 1.5.2

Something to note is that computers can only use binary because the computer is divided into many tiny sections that can only represent "on" and "off". The same as the binary equivalent of a one or a zero.

On a related note, Alan Turing (a mathematician and computer scientist) is famous for the Turing theory. The theory states that if there was an infinite tape, divided into cells each with a zero or a one, a machine that could read and write on each cell would be theoretically capable of calculating anything. Putting this into perspective, it is practically the model of how modern computers work today.

§1.6 Program Errors

Programmers more often than they would like, run into errors. There are many different kinds of errors, which will be covered in this section.
First of all, is a **syntax error**. A syntax error occurs when someone typed the code incorrectly or misspelled lines of code, resulting in the program crashing.

Example 1.6.1

One example of a syntax error would be a variable declaration like:

```
fruit ← watermelon
```

In this case, there should be a quotation mark on both sides of the word watermelon to signify a string value. This incorrect syntax will result in the computer detecting an error and the program will not run.

A second important error is a **logic error**. This error means that nothing is inherently wrong with the code, all the syntax is followed, but the programmer made a mistake in writing it.

Example 1.6.2

For example, consider a program meant to give a student a grade of A if higher than 90:

```
grade ← 100
IF(grade > 50)
{
    RETURN("F")
}
ELSE IF (grade > 60)
{
    RETURN("D")
}
ELSE IF (grade > 70)
{
    RETURN("C")
}
ELSE IF (grade > 80)
{
    RETURN("B")
}
ELSE IF (grade > 90)
{
    RETURN("A")
}
```

Caught the error? If a person had a grade of 100 and deserved an A, they would still get an F because the program is designed so that it checks first if your grade is above 50, and if so, it returns an F.

The third kind of error that programs may encounter is a **run-time error**. A run-time error occurs when the syntax is correctly formatted, but the program tries to perform an illegal operation.

Example 1.6.3

For example:

```
favoriteItems ← ["Apple", "Banana", "Orange"];
IF(favoriteItems[0] > 42)
{
    DISPLAY("yay")
}
```

The program will compile successfully in this case, but after it runs, it will realize it cannot compare the word Apple to see if it is greater than the number 42.

Note 1.6.4

Another common run-time error includes dividing by zero. Just like in math, $\frac{x}{0} \rightarrow$ undefined. Thus, in most programming languages, the code will crash.

It is important to know the **difference** between a run-time error and a syntax error, as past AP Computer Science Principles FRQs have asked specifically for a run-time error. If a student used a syntax error in the response, such as mispelling a variable name, it would not count as a run-time error.

The third and fourth type of error are **integer overflow** and **round off** errors. A integer overflow error occurs when a number is too large to be represented. A round of error is when a number is too precise to be represented. These errors are also important, but are better explained in the context of the next chapter.

§1.7 End of Chapter Questions

Note: Topics in Big Idea 1 are generally tested in the free response section of the AP exam.

Problem 1.7.1 — Which is most likely a possible benefit of collaboration between programmers to create a program?

A. Collaboration can make the program take longer to code since several opinions need to be accounted for

B. Collaboration may make a program more complete and polished by using the varying viewpoints of the collaborators

C. The program may be confusing to work on together without proper documentation of the code

D. Constant communication is needed to properly develop the program

Solution:
Collaborating between multiple different people allows for numerous different perspectives to be incorporated into a program, which can help a program be better developed and be a benefit of collaboration. Thus, option B is correct. Option A is incorrect because taking more time is not a benefit. Option C is incorrect because confusion might arise through collaboration without proper documentation. Option D is also incorrect because communication constantly being needed makes development slower and more tedious.

Problem 1.7.2 — A programmer writes a program that is meant to return the factorial value of a number (For example, 2 factorial is $2 \cdot 1 = 2$. Similarly, 5 factorial is $5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$). The highest number that can be represented in the program is 31. When the program runs with an input of 3, the computer does not detect an error or crash, but the output is less than the correct value, 6. Which of the following errors has occurred?

A. Run-time error

B. Overflow error

C. Logic error

D. Syntax error

Solution:
There are a few things to note in this problem: the maximum value that can be represented (31) and the fact that the computer does not detect an error. Because the output value is less than the correct value which is also less than the maximum value, an overflow error has not occurred and Option B can be eliminated. Option A and D are also incorrect because the program's error is not recognized by the computer. Option C is correct because the output value is not the same as expected, but the program can run without problems.

Problem 1.7.3 — A student wants to manage complexity in their quiz app. Which of the following is a way abstraction can be used to do so?

A. Adding comments to their code so it is easier to understand.

B. Using lists to store variables such as questions, answers, or images.

C. Completely getting rid of the use of variables in the program.

D. Using many variables in the place of a list.

Solution:
As stated in the question, abstraction is used to manage complexity in a program. Option C and D can be eliminated because they INTRODUCE complexity to the program. Option C, getting rid of using variables, requires a new workaround to variables for creating a program. Option D, using many variables in the place of a list, can make it difficult to navigate through all the numerous variables. A list is more manageable and results in abstraction.

Option A is incorrect because program complexity is based on the actual code, and adding comments does not effect the actual program's complexity. Option B is correct because by organizing variables in lists, the code will be much less complex as indexes of items could be used instead of variables.

Problem 1.7.4 — Which of the following is NOT considered an event-driven program?

A. Saying "Hey Siri" to initiate Siri dialogue on your phone

B. Using WASD keys to move your character in a game

C. Executing a program that prints hello world

D. Entering an expression into a calculator

Solution:
An event-driven program relies on input, which triggers something to happen. The process of an input being detected is called an event. In this case, option A, which relies on your phone hearing your voice is triggered by an event. Option B, which relies on key inputs, is triggered by an event. Option D, which takes in as input numbers that are pressed on a keypad, is triggered by an event. The only option that doesn't rely on any input from the user is a program that simply executes and prints hello world. This makes Option C the correct answer.

Problem 1.7.5 — Which of the following would most likely result in a run-time error?

A. Creating a mistake in the logic of the program in which an integer that was supposed to increase by 2 increases by 4

B. Accessing an element that is at an index greater than the length of the list

C. Dividing an integer value by itself

D. Adding documentation to the program via comments

Solution:
Option A, while possibly causing a run-time error if the integer overflows, is less likely to cause a run-time error than Option B which accesses a list out of bounds. Since there is no element at an index that is greater than the index of the last element (aka trying to access an element after the last element), the program most likely crashes. Option C is a simple arithmetic process that will not result in the program crashing. Option D would have no effect on the program since comments don't interfere with program code. Therefore, option B is the answer.

Problem 1.7.6 — A group of programmers are working together on a project. Which of the following styles of collaboration between the programmers would result in the best communication?

A. Communicating on program development via email

B. Allowing each programmer to work on sections of the program by themselves and communicating their results every week

C. Working on the program together in real time

D. Each programmer leaving comments to explain their individual code

Solution:
The best choice of collaboration is option C, working on the program together in real time. By doing so, programmers can communicate thoughts, suggest ideas, and identify any errors that someone else might be making. They can better agree on how to develop the program in unison.

On the other hand, Option A is not a good choice since email responses often take long and don't allow for consistent and quick back and forth communication while developing a program. Option B is also incorrect because each programmer may develop the program in a unique way and thus not align the program with the other programmer's perspectives. This could make developing a program very difficult. Lastly, option D is incorrect because there is almost no communication besides comments and this could lead to conflicting code or confusion.

Problem 1.7.7 — In which of the following programs would it be MOST suitable to use iterative and incremental process development?

A. In a quiz game with 3 questions.

B. In a coin flip simulator.

C. In a complex mobile game with weekly updates.

D. In an online clicker game website.

Solution: Using the iterative and incremental process development allows programmers to collaborate to make a program in which each increment, or piece, can be made and released separately, and uses iterations to negate errors. It is generally used for more complicated programs that would require collaboration to produce in parts and would need refinement. Option B can be eliminated because it is a very simple program that would likely be made in one increment. Option A can be eliminated because a quiz game with only 3 questions is a simple program that, even if made in increments, would likely not require much iteration to refine. Option D is incorrect because an online clicker game is another simple program that would likely not require increments or iteration.

Option C is correct because the word "complex" implies that iteration may be needed to refine it and the weekly updates implies that parts will be made and released separately, signifying incremental development.

Problem 1.7.8 — Which of the following could possibly be seen as non-event driven programming?

A. A program used to translate sentences.

B. A program that does not take any input and is just used to display an image.

C. A program that records what a person says and plays it back.

D. A program that is a word game made with a block coding language.

Solution: Non-event driven programming is defined as being executed without needing any action from the user through inputs. Option A is incorrect because a program used to translate sentences needs an input sentence to work, in turn, requiring event driven programming. Option C is incorrect because the program requires an audio input. Option D is incorrect because a word game, or any game, has to have some type of input to work. The type of language used (block, line, etc.) also has no connection to the type of programming.

Option B is correct because the program does not take any input and just compiles and executes.

Problem 1.7.9 — Which of the following is FALSE about the relationship between a syntax and a run-time error?

A. If a syntax error occurs, a runtime error cannot.

B. Both errors can result in the program crashing or halting.

C. A syntax error can be considered a runtime error.

D. Both errors can occur when using any programming language.

Solution: A syntax error occurs when a program is compiling and a runtime error occurs when a program is running. Remember that in this question, we are looking for which statement is FALSE. Option A is incorrect because as a syntax error happens during the compiling of a program, a run time error cannot also occur because the program never runs, making it a true statement. Option B is also incorrect because both errors do result in the program being unable to run completely. Option D is incorrect because programming language has no effect on whether an error an occur or not.

Option C is correct because a syntax error cannot be considered a runtime error. Because a syntax error occurs when a program is compiling, it cannot happen when the program is running.

2 Unit 2: Data (17-22%)

§2.1 Binary Base 2 System

As previously explained, computers can only actually represent numbers in binary. Remember how computers are split into many little cells which can either be a zero or a one? These cells are called **bits**. 8 bits is called a **byte**.
To understand how the computer uses this system to represent numbers, we must first understand our own number system. The number system we use is called the base 10, or decimal system.

Note 2.1.1
It is called the base 10 because we can only count 10 numbers in one digits place. For example, in the ones place we can count 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. The binary system can only count 0 or 1, thus the name binary.

When we represent a number such as 642301, we are really representing it as
$(6 \cdot 10^5) + (4 \cdot 10^4) + (2 \cdot 10^3) + (3 \cdot 10^2) + (0 \cdot 10^1) + (1 \cdot 10^0)$

$$\Downarrow$$
$$600000 + 40000 + 2000 + 300 + 00 + 1$$
$$\Downarrow$$
$$642301$$

For the base 2, if we have a number such as 101, we are really representing it as
$(1 \cdot 2^2) + (0 \cdot 2^1) + (1 \cdot 2^0) \rightarrow 4 + 1 \rightarrow 5$

Problem 2.1.2 — Try converting the following from binary to decimal

1. $1100 \rightarrow$
2. $1001 \rightarrow$
3. $0111 \rightarrow$

Solutions:
For number one, we know that the most far left 1 is equivalent to 2^3, and second most far left 1 is equivalent to 2^2. The zeroes afterward can be accounted for in the calculation, but don't matter much since they will be $0 \cdot 2^x$ which will equal zero ultimately. Zeroes serve as placeholders, just as they do in the base 10 system. So adding $2^3 + 2^2 \rightarrow 8 + 4 \rightarrow 12$
For number two, the same process can be repeated. The far most left 1 is equivalent to 2^3, the far right 1 is equivalent to 2^0. $2^3 + 2^0 \rightarrow 1 + 8 \rightarrow 9$
Lastly, for number three, $2^2 + 2^1 + 2^0 \rightarrow 4 + 2 + 1 \rightarrow 7$

Note 2.1.3

It is important to know binary conversion for the AP Exam. It is almost guaranteed there will be some question on binary conversion. Number systems can also be represented by a subscript. For example, base 10 can be indicated by 120_{10}, and base 2 by 1010_2

Now that we understand how to convert binary to decimal, we must understand how to convert decimal to binary. Converting decimal to binary is an easy process once you know the steps. Let's take for example, the number 24.
First we find the greatest power of 2 just under or equal to 24. In this case it is 2^4 which equals 16. 2^5 would be 32 which is too great.
Since we know $2^4 = 16$, the answer in binary form must be a combination of 1s and 0s for 5 digits places

$$\Downarrow$$
$$\underline{1}\ \underline{x}\ \underline{x}\ \underline{x}\ \underline{x}$$

Note 2.1.4

5 digit places and not 4 because since 4 is the highest power, we would do some combination of $2^4, 2^3, 2^2, 2^1, 2^0$ to add up to 24. That means there are 5 places.

Now that we know $2^4 = 16$, we have 16 of the 24, and $24 - 16 = 8$, so we just repeat the process and find the greatest power of 2 under or equal to 8. In this case, it would be 2^3 which equals exactly to 8! This means we do not need to go any further. The first two numbers will be 1s, and the rest will be 0s.

$$\Downarrow$$
$$\underline{1}\ \underline{1}\ \underline{0}\ \underline{0}\ \underline{0}$$

Note 2.1.5

Note that with 8 bits, or a byte, the largest number we can represent is 255. If we added $2^7 + 2^6 + ... + 2^0$ we would get 255. However, the *amount* of numbers we can represent with a byte is **256** since we can represent everything from 0 to 255, and 0 to 255 inclusive means **256 numbers**.

The best way to become confident in binary and decimal conversion is practice. Create your own arrangements of 1s and 0s if you'd like, try converting them to decimal, and then use a binary decimal converter online.

Problem 2.1.6 — Here are some problems nonetheless to practice.

1. $101010_2 \rightarrow$
2. $240_{10} \rightarrow$
3. $10011_2 \rightarrow$
4. $160_{10} \rightarrow$

Solutions:
For number 1, we sum $2^5 + 2^3 + 2^1 \to \mathbf{42}$ As previously mentioned, the 0s are ignored since they are placeholder values. $0 \cdot 2^x$ always equals 0 regardless of the x value.
For number 2, $2^7 = 128$, and $240 - 128 = 112$. So currently we know it looks like this:

$$\Downarrow$$
$$\underline{1}\ \underline{0}\ \underline{0}\ \underline{0}\ \underline{0}\ \underline{0}\ \underline{0}\ \underline{0}$$

Then for 112, the largest power of 2 just under is $2^6 \to 64$. Then, $112 - 64 \to 48$

$$\Downarrow$$
$$\underline{1}\ \underline{1}\ \underline{0}\ \underline{0}\ \underline{0}\ \underline{0}\ \underline{0}\ \underline{0}$$

Next for 48, the largest power of 2 just under is $2^5 \to 32$. Subtracting, $48 - 32 = 16$.

$$\Downarrow$$
$$\underline{1}\ \underline{1}\ \underline{1}\ \underline{0}\ \underline{0}\ \underline{0}\ \underline{0}\ \underline{0}$$

Finally, the largest power of 2 for 16 is $2^4 \to 16$, which has an exact value of 16.

$$\Downarrow$$
$$\underline{1}\ \underline{1}\ \underline{1}\ \underline{1}\ \underline{0}\ \underline{0}\ \underline{0}\ \underline{0}$$

So our answer is **1111 0000**

Note 2.1.7

Note that binary is usually written in groups of 4. So instead of 10101111, we would write 1010 1111

For number 3, we sum $2^4 + 2^1 + 2^0 \to 16 + 2 + 1 \to \mathbf{19}$
For number 4, we list powers of 2 and realize that 2^7 is just below 160. $160 - 2^7 \to 160 - 128 \to 32$

$$\Downarrow$$
$$\underline{1}\ \underline{0}\ \underline{0}\ \underline{0}\ \underline{0}\ \underline{0}\ \underline{0}\ \underline{0}$$

Then 32 is simply exactly equivalent to 2^5.

$$\Downarrow$$
$$\underline{1}\ \underline{0}\ \underline{1}\ \underline{0}\ \underline{0}\ \underline{0}\ \underline{0}\ \underline{0}$$

§2.2 Integer Overflow

Now that we finally conquered binary conversions, we can truly understand why integer overflow errors occur. Imagine if you only had 3 bits but you had to represent 8. The way to get the maximum number value possible would be to do

$$\Downarrow$$
$$\underline{1}\ \underline{1}\ \underline{1}$$

However, this still is less than 8, $2^2 + 2^1 + 2^0 \to 4 + 2 + 1 \to 7$. You would need one extra bit to represent 8. This is when an overflow error occurs, the quantity is too high to be represented by the available amount of bits, resulting in an **overflow** from the number.

Note 2.2.1

A cool trick to find the maximum base 10 quantity that any number of bits, n, can represent, is evaluating $2^n - 1$. For example, recall that 8 bits can at maximum represent 255. To have summed this quickly, we could do $2^8 - 1 \to 256 - 1 \to 255$.

Overflow errors can cause a lot of problems for computers. Check out this link that shows the overflow error that will happen in 2038.

§2.3 Round-Off/Floating Point Imprecision Error

A **round-off error**, sometimes referred to by the AP with its fancy name *floating point imprecision error*, is when a decimal is too precise to be represented with a sequence of bits. Remember how $2^0 = 1$, $2^1 = 2$, $2^2 = 4$... and so on? In the same way, $2^{-1} = 0.5$, $2^{-2} = 0.25$... and so on. Thus, sometimes decimal point values are too precise to be exactly represented by bits, resulting in a round-off error.

Note 2.3.1

Floating point is a term that means a decimal value, not an integer. In languages like C++, a decimal variable is called a float. That's why floating point imprecision is just another way of saying round-off error.

Problem 2.3.2 — David is using a programming language that represents integers up to four bits. He creates a procedure that takes a list called numList as input. The purpose of the program is to sum the three integers provided by numList.

```
PROCEDURE summation(numList)
{
    sum ← 0
    i ← 1
    REPEAT UNTIL(i > LENGTH(numList)
    {
        sum ← sum + numList[i]
        i ← i + 1
    }
    RETURN(sum)
}
```

Which of the following arguments used as test cases would most likely result in a data type error?

A. [3, 10, 1]

B. [8, 4, 0]

C. [5, 5, 5]

D. [4, 7, 5]

Solution: According to the text, integers can be represented up to four bits. This means that the maximum number representable is $2^3 + 2^2 + 2^1 + 2^0$. Using the trick from a previous note, we can simply do $2^4 - 1 \to 16 - 1 \to 15$. Now all we need to do is see which one of these lists, where all the numbers are added together, exceeds 15. A $\to$ 14, B $\to$ 12, C $\to$ 15 on the dot, but D $\to$ 16. Therefore, the answer is **D. [4, 7, 5]**.

§2.4 American Standard Code for Information Interchange

How do computers know what letters are like? They can only represent numbers after all. Here comes **ASCII**, the American Standard Code for Information Interchange. It is basically a mapping of different letters to numbers.

Dec	Char	Dec	Char	Dec	Char	Dec	Char	Dec	Char	Dec	Char
...	...	48	0	64	@	80	P	96	`	112	p
33	!	49	1	65	A	81	Q	97	a	113	q
34	"	50	2	66	B	82	R	98	b	114	r
35	#	51	3	67	C	83	S	99	c	115	s
36	$	52	4	68	D	84	T	100	d	116	t
37	%	53	5	69	E	85	U	101	e	117	u
38	&	54	6	70	F	86	V	102	f	118	v
39	'	55	7	71	G	87	W	103	g	119	w
40	(	56	8	72	H	88	X	104	h	120	x
41	)	57	9	73	I	89	Y	105	i	121	y
42	*	58	:	74	J	90	Z	106	j	122	z
43	+	59	;	75	K	91	[	107	k	123	{
44	,	60	<	76	L	92	\	108	l	124	|
45	-	61	=	77	M	93	]	109	m	125	}
46	.	62	>	78	N	94	^	110	n	126	~
47	/	63	?	79	O	95	_	111	o	127	DEL

[1]

Note 2.4.1

More commonly used now is **Unicode** which is the same thing as ASCII but for a whole load more unique characters and languages besides the Latin alphabet.

§2.5 Red, Green, and Blue

We know how letters and numbers are represented now so how do we represent images and colors on computers? If you look very closely at your screen, it is divided into millions of tiny boxes. These tiny boxes each have a single solid color, and when put all together, form an image. The color of the boxes is determined by its RGB (Red, Green, Blue). RGB is an ordered triplet, and the numbers range from 0 to 255 (aka a byte) for each one.

[1] Chart from Maxim Aleksa

Note 2.5.1

So for example, the RGB for a pure red would be (255, 0, 0) since the first number is red, which is at its maximum value of 255, and green and blue are set to 0. You can test out more color configurations here.

These little boxes that the image is divided into is called **sampling**. For example, imagine you tried to draw something by drawing lines to make a paper into a grid with 100 boxes and coloring each box a specific color only. That would be an example of sampling, and is the exact same thing computers do. The exception is that computers make millions of boxes, not 100.

Note 2.5.2

Consider a pixel art image:
It looks pixelated since it does not have many squares. However, if we increased the sampling and created more and more squares, it would look increasingly realistic. That's what a monitor, or a phone screen basically is. A lot of sampling with millions of squares to make it look more realistic.

§2.6 Analog Data vs Digital Data

What we have been talking about is an example of digital data. Digital data is data that is divided into small intervals, and does not flow continuously. To make it easier to understand, consider a painting in real life vs online. In real life, the painting is not chopped up into pixels. It flows continuously, and computers would need an infinite amount of pixels to try to meet the same quality as a painting in real life. The online version of the painting would be considered **digital data** which uses sampling to break something into intervals. The real life version that flows continuously without any breaks or intervals is considered **analog**.

Note 2.6.1

Note that other forms of analog and digital data exist as well. Audio is a good example. Music listened to through headphone is actually a bunch of cut up pieces of audio data, sampled numerous times to make it higher quality. Music heard in real life by orchestras or bands is analogous, and is not chopped up at all.

§2.7 Files and Metadata

Files in computers are really just a bunch of binary code at their lowest level. File endings such as .mp4 or .png help the computer interpret how to open the file. Information about files or data about other data is called **metadata**. Metadata is like the context behind the data, which can be helpful in organizing it or gathering information.

Example 2.7.1

Examples of metadata include

1. Location
2. File creation date
3. File author
4. File size

§2.8 Lossy vs Lossless Compression

When file sizes are too big, people will often want to compress files to save storage. File compression is done in two main ways:

1. Lossy Compression - Compression where some data is lost
2. Lossless Compression - Compression where data is not lost

First, let's explain the general process of compression. When a file is compressed, there is some sort of **encoding algorithm** that compresses the data.

Note 2.8.1

For instance, let's take the string "giggling". An encoding algorithm may use a special symbol for "gg" such as @, substituting it for "gg", and thereby reducing the size of the file.

A **decoding algorithm** would be the inverse of the encoding algorithm and reconstruct the message. For lossless compression, it would **perfectly reconstruct** the data to its original form. For lossy, it may not be perfectly reconstructed. However, lossy compression often **yields greater saving for data storage**, in exchange for not perfectly reconstructed data.

Note 2.8.2

This means that lossless compression would be useful for intricate data such as programs, bank records, passwords, and texts.
Lossy compression would be better for images and videos since perfect reconstruction is not required, and sometimes pixel loss is not even perceptible by the human eye.

§2.9 Data Analysis

After computer scientists gather data, they typically analyze it through a visual. This includes **bar graphs**, **scatter plot graphs**, **line graphs**, and other forms of graphs. Though there are a few graph analysis questions on the AP Computer Science Principles exam, they will typically not ask too much about graphs themselves (except graphs representing packets, to be covered in the next unit). However, you should familiarize yourself with the listed graphs and generally know the following vocabulary:

1. **Datasets** are huge collections of data that can be used by computers for various tasks such as statistical analysis and even machine learning. As mentioned before, data is often visualized in graphs, charts, etc. so that we can find patterns and better analyze data.

2. **Big data** is a term that refers to datasets that are so large that they cannot be processed by a single computer. In these cases, parallel systems (multiple computers working at once) are used to analyze the data. A few sources of these huge datasets include scientific research, digital archive libraries, and records.

3. **Scalability** is the ability of something to accommodate for more information while not losing its computational capacity and performance (the internet, as an example).

4. **Data science** is a field of Computer Science which focuses on extracting, visualizing, and manipulating large datasets. It allows us to make observations that lead to discovering trends, connections, and finding possible solutions to problems.

5. **Machine learning** is when computers are able to learn and improve using large datasets which allows them to go beyond just what the user tells the computer to do. Knowledge is gained based on a dataset that is constantly being altered and the behavior is meant to mimic that of a human being's intelligence (artificial intelligence). A few examples of machine learning being used in everyday life is social media advertising and predicting the weather.

§2.10 End of Chapter Questions

Problem 2.10.1 — A programmer is using a data type that stores 4 bits. However, he finds that it would be more beneficial to use a data type storing 8 bits instead. How many times more storage will he have for values?

A. 3 times more

B. 10,000 times more

C. 16 times more

D. 8 times more

Solution:
4 bits can store 2^4 different values. Since one bit can either be 1 or 0, there are two different options for the first bit. Then there are another two possibilities for the second bit, which creates $2 \cdot 2$ or 4 combinations. This pattern repeats consistently. Therefore, four bits has 2^4 possible values.

By the same logic, eight bits has 2^8 values. To find how much times more possible combinations eight bits can store than four bits, we can do

$$2^4 \cdot x = 2^8$$
Dividing by 2^4 on both sides
$$x = 2^4 \text{ (which is 16)}$$

The answer is C. 16 times more. Also note that this is why the sum trick in this note works. Subtracting 1 from 2^n takes out the zero that is one of the possible values. For example, 8 bits, or 2^8 can represent 256 different values. The values it can represent starts from 0, 1, 2, 3 ... 255. Though 255 is the max number, the inclusion of 0 in the count causes the max possibilities to be 256. Therefore subtracting 1, or excluding the 0 from the count will give us the max again (255).

Problem 2.10.2 — A person wants to share a recording of his vacation with his friends. However, the size of the file is too big to send. He does not care about video or audio quality, and wants to send it as soon as possible. Which of the following actions should he take?

A. Use lossless compression to decrease the storage that the file takes up.

B. Use lossy compression to reduce the file size the most effectively.

C. Use lossless compression, though some parts of the video may be blurry or pixelated.

D. Use lossy compression so that the video remains in perfect quality while reducing storage.

Solution:
Lossless compression allows files to be restored perfectly while lossy leads to some loss in quality. Therefore, eliminate C and D. Though A is true, since the person wants to

reduce as much storage as possible in order to ensure a fast send time, while also not being concerned with quality, option B would be the answer.

Problem 2.10.3 — In a video game, a player uses an exploit to gain levels rapidly. Eventually his level becomes negative, while other normal players still have positive levels. Which of the following most likely explains this error?

A. The exploit used by the player led to levels being gained so rapidly that a calculation error using subtraction occurred.

B. The level achieved by the player was larger than the maximum level provided by the video game's integer data type.

C. The server that the game is running on was overloaded by too many players.

D. A rounding error occurred when the player's level was calculated.

Solution:
For a value to become negative, an overflow error must have occurred in which the number trying to be represented was larger than the maximum value that the data type could hold. This situation is what is described in Option B.

A rounding error adding numbers together would not make the level become negative, though it may result in the sum being rounded to the nearest integer or some other level of precision. Rapidly gaining levels would not lead to any "subtraction" error. Lastly, being overloaded by too many players may lead a server to crash but would not result in negative levels.

Problem 2.10.4 — In a photography class, students take pictures of different butterflies with a special camera that also records metadata about the exact date, time, and location of which the photo was taken. Which of the following cannot be determined by inspecting the metadata of the photos alone?

A. The day of the week in which the photo was taken.

B. How many photos were taken in a given place.

C. If the photo was taken in the morning photography class or the afternoon photography class.

D. The type of butterfly with the most pictures taken.

Solution:
Knowing that the metadata provided by the photo is the exact date, time, and location, we can first cross out Option A, since the exact date will also allow us to figure out the day of the week. Option B can be crossed out by looking at the location data of each photo to figure out how many were taken at a given place.

Option C can also be eliminated since we know the time the photo was taken. However, we cannot know the type of butterfly with the most pictures taken by the metadata alone, since the metadata does not indicate the type of butterfly. The photo itself would have to be seen. Thus, Option D is the answer.

Problem 2.10.5 — Someone goes to a high school music performance of the orchestra and band, and decides to record some of the music. Which best describes the difference in the music heard during the live performance, and in the recorded version of the music?

A. The recorded version of the music uses sampling techniques to record the analogous data in a digital form.

B. The recorded music kept as analog data on the device has lower sampling than the live performance music.

C. There is no difference or distinguishing feature between the music heard during the performance and the recorded version.

D. The music heard during the performance does not use sampling, just like the recorded version of the music.

Solution:
Music heard in real life is analog data, and recorded music is digital data since it is cut up into little intervals called samples that are stored as data. Thus, B is incorrect because recorded music is digital data, not analog, and analog does not have sampling. Analog is fully continuous.

Option C is also incorrect because the difference is within the analog and digital data representations of the music. Finally, Option D is wrong because while the music heard in the performance does not use sampling, the recorded version does use sampling. Therefore, A is the correct option since the recorded version does use sampling to transfer the analogous data into a digital form.

Problem 2.10.6 — In which of the following situations would using lossless compression be more appropriate than using lossy compression?

A. Attempting to maximize the number of photos that can be saved on a phone

B. Capturing thousands of audio files that need not be reconstructed perfectly

C. Compressing a program's source code

D. Saving the profile pictures of millions of users for a popular social media website

Solution:
Lossless compression has the advantage of perfectly reconstructing data, but has the disadvantage of not compressing data to as small file sizes as lossy compression can. Keeping this in mind, option A would be incorrect because if the intention is just to maximize the number of photos that can be saved, lossy compression would be more efficient as it typically reduces files to smaller file sizes than lossless. Option B would also be incorrect because the audio files do not need to be reconstructed perfectly and there are thousands of them. Thus, lossy compression is more optimal than lossless.

Option C is correct since using lossy compression to compress a program's source code means that some of the source code will be lost when reconstructing the program. Losing any source code could result in huge errors in the program and stop it from

functioning. Thus, lossless compression, which reconstructs perfectly, is better for compressing programs. Finally, option D is incorrect because saving profile pictures of millions of users takes up extreme amounts of storage, thus indicating lossy compression may be more efficient.

Problem 2.10.7 — Which of the following best describes how machine learning works?

A. Machine learning compiles line coding into machine language.

B. Machine learning provides an approximate solution to a problem which cannot be solved using an algorithm.

C. Machine learning processes big datasets to identify patterns in order to make predictions based off of it.

D. Machine learning writes documentation for a program.

Solution:
Machine learning is when computers are able to learn and improve using data from large datasets. This information allows a computer to predict a behavior through identifying patterns in the dataset. Therefore, Option B is correct.

Option A is incorrect because translating a programming language into "machine language" (aka binary) is done by a compiler or interpreter instead of a machine learning program. Option B is incorrect because the main purpose of a machine learning program is not to find a heuristic solution to an undecidable problem (covered in Big Idea 3). Option D is incorrect because the programmer is the one that writes documentation, such as comments, for a program.

Problem 2.10.8 — In a video game, a character uses the following abilities that are matched to corresponding binary values.

Ability	Binary Value
Dash	10101
Spell	11001
Parry	01101
Teleport	01111

What is the decimal value (base 10) of the difference between teleport and parry?

A. 2

B. 11

C. 8

D. 9

Solution:
To find the decimal value difference between teleport and parry, we must convert each into base 10 (unless you know how to do binary subtraction). To do so, we start with teleport, doing

$$01111 = 2^0 + 2^1 + 2^2 + 2^3$$
$$01111 = 15$$

(Also the sum of the powers of 2 for n bits is $2^n - 1 \to 2^4 - 1 \to 15$. Note we use 4 as the amount of bits since the 0 all the way to the left has no effect on the value of 01111 and we can thus treat the sequences of bits as 1111).

Next to find parry we do

$$01101 = 2^0 + (0 \cdot 2^1) + 2^2 + 2^3$$
$$01101 = 13$$

Therefore, $01111 - 01101 = 15 - 13$, meaning the answer is A. 2.

Problem 2.10.9 — In a video game, a player has an inventory that can hold up to a maximum of 200 items. Each item is represented with a unique binary sequence. If a player's inventory is full, how many bits will be required at minimum to represent each binary sequence?

A. 6

B. 7

C. 8

D. 9

Solution:
To represent 200 different unique binary sequences, the minimum amount of bits required is 8 since $2^8 = 256$, meaning that 8 bits can represent 256 unique binary sequences. Option A, 6 bits, means that only $2^6 = 64$ different sequences can be represented. Similarly, option B, 7 bits, represents $2^7 = 128$ different sequences. Both are less than 200. Option D, on the other hand, is $2^9 = 512$, which can also represent all 200 different sequences, but is greater than Option C. Thus, Option C is correct.

Problem 2.10.10 — A student uses a calculator for his science class. He attempts to perform calculations with a distance and time given by his teacher. After checking the answer key, he realizes that the decimals in his answers were calculated slightly imprecisely. Which of the following best describes what led to this imprecision?

A. An overflow error occurred due to the number type having a fixed number of bits.

B. A logic error occurred within the program due to a mistake in the program code.

C. A round-off error occurred due to the number type having a fixed number of bits.

D. A run-time error occurred, resulting in the imprecision.

Solution:
The fact that the decimals in his answer were imprecise but not totally off is most likely a consequence of the fact that computers only have a fixed number of bits with which they can represent numerical values, thus leading to some values being rounded because the exact precise decimal value is too specific to be contained in the fixed number of bits. This error is called a round-off error and fits the description of option C.

Option A is incorrect because that means the numbers used were extremely large and as a result the answer "overflowed" into the negatives or didn't compute. However, the problem states that the answer was computed with accuracy for the most part, except some slight imprecision. Option B is incorrect because a logic error implies that the code was written in an incorrect way such that it does not perform the design intent of the program. However, the code is functional and is doing what it is supposed to. The issue lies in the fixed bits.

Lastly, option D is wrong because a run-time error implies that upon execution the program crashed. This is not the case. The program executed and fully ran but the result was mathematically imprecise from the expected value.

Problem 2.10.11 — A data-storage system consisting of student data in a school stores the ID of each student matching with their name, as well as additional metadata. Which of the following pieces of metadata is the least likely to be in the dataset?

A. The student's grade point average

B. The student's schedule

C. The student's emergency contacts

D. The student's preferred lunch period

Solution:
Metadata is other additional data that is added along with a specific piece of data either because it is important additional information or adds context to the data. In the case of a data storage system that keeps data of students, a student's GPA, schedule, and emergency contacts would be highly important and are very likely to be kept as metadata. However, a student's preferred lunch period does not really matter as much since it is just a preference and isn't substantial enough to be kept as metadata. Therefore, Option D is the correct answer.

Problem 2.10.12 — An influencer makes many posts on social media platforms such as Threads. The data includes the content of the post and the metadata includes the time it was posted and the device used to post it. Which of the following can be observed using the data rather than the metadata of the influencer's posts.

A. The most common time of day the influencer posts.

B. The location of the influencer.

C. The most common topics that the influencer generally posts about.

D. The device the influencer uses to post most of the time.

Solution: As stated in the question, the data refers to what information directly contained in the post. Option A and Option D are incorrect because the most common time of the day the influencer posts and what device they generally use are both observations made using the metadata rather than the data. Option B is incorrect because location is not stored as data or metadata.

Option C is correct because the topic that the influencer generally discusses in posts is found in the information contained in the post itself which is the data, rather than the metadata.

Problem 2.10.13 — Many computer systems use RGB (Red, Green, Blue) to judge how to color a pixel. Higher values of red will result in a more red color, and the same applies to green and blue. The maximum number that a value of a color can be is 255. For example, (255, 0, 0) will return a pure red color. Which of the following is the binary code for the color purple assuming the numbers are listed as red, green, and blue respectively?

A. 11111111, 00000000, 11111111

B. 11111111, 11111111, 00000000

C. 11111111, 11111111, 11111111

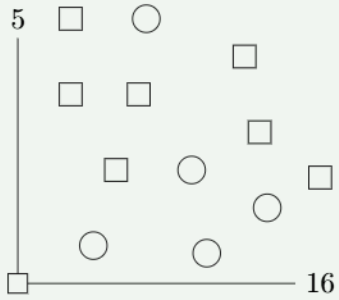
D. 00000000, 11111111, 11111111

Solution: Creating purple requires red and blue mixed together. Knowing that the first number is red and the last number is blue, we simply need to choose the option where red and blue have some value and green has no value. This matches Option A which is (255, 0, 255) = red + blue = purple.

Option B is red and green, which does not mix to purple. Option C is red, green, and blue (which is actually considered white for displays). Option D is yellow because its just green and blue.

Problem 2.10.14 — A student collected data to see the effect of hours spent doing non-academic activities based on the age of the student on standardized test scores. The data the student collected is represented in the graph below. A circle represents a student in at least the 80th percentile and a square represents those below the 80th percentile. Which of the following observations is MOST accurately reflected in the graph?

Hours spent doing non-academic activities



Age of student

A. Students who spent more time doing non-academic activities were more likely to be above the 80th percentile.

B. Students who spent less time doing non-academic activities were more likely to be above the 80th percentile.

C. Students who were younger were more likely to be above the 80th percentile.

D. Student who were older were more likely to be above the 80th percentile.

Solution: A simple way to approach this type of problem is by noticing a pattern in the shapes. In this case, there are more squares in the top half and more circles in the bottom half. Through this observation, we can see that the shape's condition is likely determined by the y-axis' condition, Hours spent doing non-academic activities. Option C and D are eliminated.

As stated in the question, a square represents a student above the 80th percentile and a circle represents a student below the 80th percentile. Since there are more squares in the top half of the graph, it means there are more squares, or students below the 80th percentile, for a higher y-value (more hours spent doing non-academic activities). Option A is incorrect as it states the opposite of this. Option B is correct.

Problem 2.10.15 — A binary number is represented as 11000. Which of the following represents the effect of removing the last three zeros in the binary number?

A. The new resultant number is 16 times greater

B. The new resultant number is 100 times less

C. The new resultant number is 6 times less

D. The new resultant number is 8 times less

Solution:
Removing the last three digits of 11000 is equivalent of dividing by 2^3 since each digit place of a binary sequence represents two possible values (1 or 0). This means that since since one digit place has 2 possible representations, the digit place after the first one, which also has 2 possible representations contributes to a combined total of 4 possible combinations when considering both digit places since $2 \cdot 2 = 4$. When adding a third digit place, it becomes $2 \cdot 2 \cdot 2 = 8$.

Therefore, removing three digit places is equivalent to dividing by 8. Alternatively, you could convert 11000 and 11 to base 10 and notice that 11000 is 24 while 11 is 3. Option D is the correct answer.

Problem 2.10.16 — Which of the following best explains why seeing a displayed image of a cat on a computer monitor may not look as high quality as seeing that cat in real life?

A. Due to the high sampling rate of the digital image, it looks more continuous than the real life cat

B. The displayed image uses digital data whereas in real life the cat is analog

C. The displayed image undergoes lossy compression when being displayed, resulting in a more pixelated image

D. Analog data is used to form the image of the cat on the monitor, while in real life the cat is displayed via digital data

Solution:
In real life, things that you see are continuous and infinitely smooth/high quality. This is analog data. On the other hand, images displayed digitally are cut up into intervals or samples called digital data. The image you see on a 1920x1080 monitor is cut up into $1920 \cdot 1080 = 2073600$ pixels that makes the images displayed look more realistic and less cut up or choppy. Each pixel is a solid color and those millions of solid colors form a realistic looking image.

Therefore, for the image to look as continuous as in real life, nearly an infinite amount of pixels is needed. This makes Option A incorrect since a digital image is not more continuous than the real life cat. Option B is correct because due to the limitations of digital data (mainly being broken into pixel intervals) the digital data is lower quality than the analog version. Option C is incorrect because lossy compression isn't related and images being displayed don't undergo lossy compression or compression in general.

Compression is used to store numerous files and not used for displaying the files. Lastly, Option D is incorrect because digital data is used for the image of the cat on the monitor, and the real life cat is displayed with analog data.

Problem 2.10.17 — A student is trying to predict the number of students that will receive an A in their class. They made a computer model that uses data from the previous 5 years that the class was offered as a course at the school, including the number of students taking the class and the number of students that received an A. In the model, it was predicted that 30 students would receive an A, but in reality only 23 students did. Which of the following would MOST likely result in the model becoming more accurate?

A. Using data from only the previous year's class.

B. Removing refinements from the model so that it will perform faster.

C. Using data from all of the years that the class was offered at the school.

D. Only using data of students that have at least a 3.8 GPA

Solution: As stated in the question, we are looking for a change that will make the model more accurate than before. Option A and Option D are incorrect because both result in reducing the sample size greatly and may lead to results that introduce bias. Option B is also incorrect because removing refinements in order for the model to preform quicker may result in a loss of accuracy as well.

Option C is correct because by taking into account all of the years that the class has been offered at the school increases the sample size which could help the student find more accurate patterns used to make predictions in the dataset.

3 Unit 3: Algorithms and Programming (30-35%)

This is a **major** unit. You need to know this unit inside out, like the back of your hand. Not only is this unit worth 30% of the exam at minimum, it can be considered to be worth even more since most questions from other units are a mix of that unit and this one at the same time.

Note 3.0.1

Make sure you are very good at interpreting the result of loops. A lot of the AP Computer Science Principles exam tests your algorithmic thinking skills. We will go over questions involving loops as well as the course content, but make sure analyzing code is your forte.

§3.1 Data Types

Data types represent the type of information a variable stores. The most essential datatypes to know are:

1. Integer - A whole number, negative or positive
2. String - A word, like "dog" or "cat", enclosed with quotations
3. Boolean - A variable that is either equal to true or false
4. List - A group of data types, enclosed in brackets. (For example: ["dog", 2, 1.2, cat])
5. Float - A decimal value, like 2.33
6. Char - Char means character, so for example 'a', or 'b'.

Each element of a list has an index, or a number that explains what position the element is on the list. Conventionally, the first element has a 0 index, and the second element has an index of 1, and so on.

However, for the AP Computer Science Principles exam, **lists are 1 indexed, meaning that the first element has an index of one**. This is very important as it could change your answer.

Note 3.1.1

The secret truth about strings:
Strings like "cat" and "dog" are actually lists of characters. Strings are just another example of an **abstraction**.

§3.2 Boolean Values

As previously mentioned, a **boolean value** is a variable that is either equal to true or false.
When you have set a variable equal to the result of a conditional expression that checks if one value is true AND another value is true, if both values are true, the outcome will be true. For example:

Example 3.2.1

```
b ← true
c ← true
a ← b AND c
DISPLAY(a)

Output: True
```

Think about it intuitively. Let's say I need water and food to survive.

```
survive ← water AND food
```

If I have water (e.g., water is true), and if I have food (e.g., food is true), then survive will also equal true. If I don't have water, and I only have food, survive will be false because I need water AND food. For OR boolean expressions, it works similarly. Let's say I need cash or credit to pay.

```
pay ← cash OR credit
```

If I have cash (e.g., cash is true) but not credit, it doesn't matter because I don't need both, I just need cash OR credit, so pay will be equal to true. If I don't have cash, and I don't have credit, pay will be equal to false because I have neither cash or credit.
Lastly, the AP Exam may use the keyword NOT, which basically inverses the value of a boolean. Take for example, a variable called cap equal to true.

```
cap ← true
```

If someone added a line of code such as

```
cap ← true
cap ← NOT(cap)
```

Then cap would be false. No cap.
The AP Exam will probably question you on boolean values, but in a verbose way. Consider the following code question.

Problem 3.2.2 — After compiling and executing the program, the output of the following boolean values would be:

```
a ← true
b ← false
c ← false
a ← b AND c
c ← NOT(a OR b)
DISPLAY(a)
DISPLAY(b)
DISPLAY(c)
```

A. False, false, false

B. False, false, true

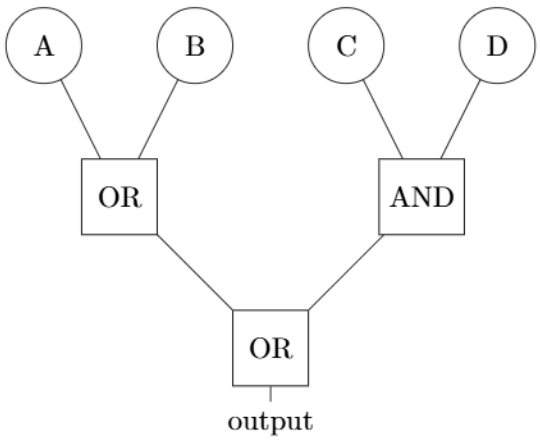
C. False, true, false

D. True, true, true

Solution:
The variable a is set to true at first, but since $a \leftarrow (b \text{ AND } c)$, and both b and c are equal to false, a will be equal to false as well (Even if one of b and c were equal to true, as long as the other one is false, they cannot evaluate to true in an AND statement). Then, $c = \text{NOT}(a \text{ or } b)$. Since a is equal to false, and b is equal to false, the inside of the parentheses evaluates to false. But remember the NOT? It inverses the value of the parentheses from false to true. So c is assigned the value of true. Therefore, a and b is false, while c is true. This makes option **B. False, false, true, the correct answer**.

§3.3 Logic Gates

Booleans can also be visualized using logic gates. The two types of logic gates include AND gates and OR gates. Here is an example:



This is the code equivalent of

```
e ← (a OR b) OR (c AND d)
DISPLAY(e)
```

First the (a OR b) grouping will be evaluated to be true or false, then the (c AND d) grouping will be evaluated to be true or false. Then the code will evaluate the resultant boolean values of both nested boolean expressions in the outer OR boolean expression. For example, suppose (a OR B) was false, and (c AND D) was true. Then the code would lastly evaluate the outer expression of (false OR true).

§3.4 Strings

Some useful string functions that you might want to know, which can appear in the MCQ are:

1. `len(string) or LENGTH(string)` - returns the number of characters in a string
2. `concat(string1, string2)` - combines two strings (for example, concat("comp", "uter") returns "computer"
3. `substring(string, start, length)` - substring returns a smaller string that is a part of the given string in the argument. The string will start at a certain character, whose position is provided as a number in the argument `start`, and will go `length` characters and then stop. For example: substring("hello", 2, 4) returns "ello", since it starts at e (whose index/position is 2), and goes to o which is the fourth letter if we start counting from e.

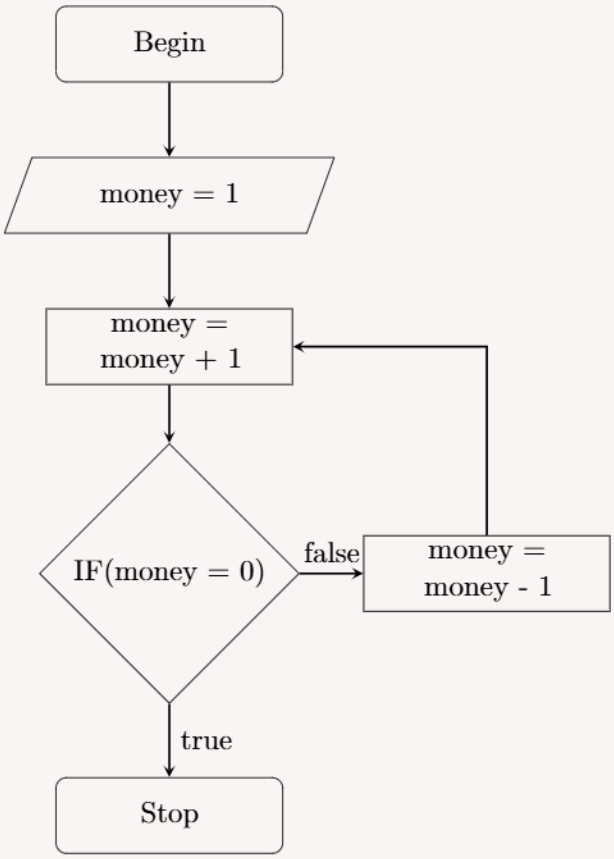
Don't worry about remembering these because they will most likely be provided as context for the question. The point was just to familiarize you with what you can be asked in advance, in case one of the concepts are confusing.

§3.5 Evaluating Conditionals

Another part of the AP Computer Science Principles exam that is very important is understanding how to evaluate conditions. Conditionals are just the if(statement1) expressions that you should already know. If not, sometimes conditionals will be represented in a flow-chart manner, like this:

Example 3.5.1

In the following example, there will be an infinite loop as money will always equal to two dollars, go back to one dollar, and then add again to 2. The program will most likely crash from constantly repeating the processes.



However, most of the time the conditionals will be represented in an if(statement) format. A problem would look something like this:

Problem 3.5.2 — In the following code segments, a programmer is attempting to program a video game, in which the character can only level up if he meets the following requirements:

Variable	Requirements
`Health`	greater than or equal to 90
Time	less than or equal to 30

Failing to meet these requirements should only grant 10 points, but not a level up, and succeeding should only level up the character. Which of the following procedures correctly executes the program's intended purpose?

A.
```
PROCEDURE canLevel(health, time, points)
{
    IF(health ≥ 90 OR time ≤ 30)
    {
        RETURN true
    }
    ELSE
    {
        points ← points + 10
    }
}
```

B.
```
PROCEDURE canLevel(health, time, points)
{
    points ← points + 10
    IF(health ≥ 90 AND time ≤ 30)
    {
        RETURN true
    }
}
```

C.
```
PROCEDURE canLevel(health, time, points)
{
    IF NOT(health < 90)
    {
        IF NOT(time > 30)
        {
            RETURN true
        }
    }
    points ← points + 10
}
```

Problem 3.5.3 — (part 2)

D.
```
PROCEDURE canLevel(health, time, points)
{
    IF NOT(health < 90 AND time > 30)
    {
        RETURN true
    }
    points ← points + 10
}
```

Solution:
Option A is not correct because it checks if health >= 90 **OR** time <= 30. However, according to the program requirements, both requirements must be met. The procedure would be correct if the OR was replaced with an AND.

Option B fixes the mistake of option A, using an AND operator. However, it gives 10 points immediately. According to the text, 10 points should only be granted if the character cannot level up. However, the procedure in option B would give 10 points even if the character levels up.

Option C is correct because first it checks if health is not less than 90. If the health is not less than 90, it continues to the second conditional. Checking if the health is not less than 90 is equivalent to checking if the health is equal to 90 or greater. The second conditional checks if the time is not greater than 30, and if so, continues. This is equivalent to checking whether the time is less than or equal to 30. Then it returns true if all conditionals are met. Otherwise, it grants 10 points.

Option D is not correct because instead of checking if NOT(health $<$ 90) and NOT(time $>$ 30) separately, it combines it into one conditional as NOT(health $<$ 90 AND time $>$ 30). This is different from checking separately and nesting the conditionals. It works in the case that all the requirements are met (e.g., health is greater than 90, and time is less than 30, which would lead to FALSE and FALSE, which would lead to the overall conditional inside the parentheses being false, but then inverted by the NOT to become true.

However, if only one requirement is met, such as health being greater than 90, while the other requirement is not met, such as time being greater than 30 when its supposed to be less than 30, the conditional inside the parentheses will be a FALSE and TRUE, which will evaluate to false because there is an AND operator. After, the NOT will invert the overall conditional inside the parentheses that evaluated to false into a true, and the program will end returning true, even though only one of the requirements were met and not both.

Note 3.5.4

If you could solve or understand this problem, this is probably the hardest the AP Computer Science Principles exam will get in terms of difficulty for conditional questions. Iteration questions might mix in conditionals, but will do so mostly straight forwardly.

§3.6 Iteration

Iteration is the process of repeating something multiple times. When we go through a list or an array in a programming language, we also call it iterating through the list.

Note 3.6.1

Iteration in the AP exam will not use a `for(i = 1; i <= variable; i++)` format. Instead it will either be `FOR EACH item in LIST`, or `REPEAT UNTIL (conditional)`.

If you don't understand iteration in the context of programming, you may want to look at the resources detailed here. An iteration problem would look something like this:

Problem 3.6.2 — A programmer is developing an algorithm to reverse all items in the list. Assume the list `aList` contains at least two elements.

```
1   PROCEDURE reverse(aList)
2   {
3       count ← 1
4       listLength ← LENGTH(aList)
5       REPEAT UNTIL(count > listLength/2)
6       {
7           temp ← aList[count]
8           aList[count] ← temp
9           aList[listLength - count + 1] ← temp
10      }
11  }
```

Which of the following changes is needed for the program to work as intended?

A. Assigning aList[count] to aList[listLength - count + 1] in line 8

B. Assigning temp to aList[listLength - count + 1] in line 7

C. Switching lines 3 and 9

D. Switching lines 8 and 9

Solution:
The answer to this problem is not C, because in option C, the variable count is declared inside of the REPEAT block which would not only not fix anything, but lead to an error as the variable count is being used before declared. The same thing would occur for line 9 since it would try to use the value of temp, and realize temp is not declared.

For option D, shuffling lines 8 and 9 would not change the program and both aList[count] and aList[listLength - count + 1] would be set to temp, which is aList[count].

Option B is also incorrect because if temp were set to aList[listLength - count + 1], then both aList[count] and aList[listLength - count + 1] would both be the value of aList[listLength - count + 1], and no values would be swapped.

Option A is the correct answer because to reverse the list, we need to swap the value of the element at index count, and the value of the element at index listLength-count + 1. By setting aList[count] to aList[listLength - count + 1], we have swapped the values, and the temp variable stores the previous value of aList[count], which is then set to aList[listLength - count + 1] to switch the values both ways.

Note 3.6.3

For these iteration questions with complex looking code, first try to figure out the general idea of the code. For the previous question, the idea is to reverse elements in a list. So how can someone reverse elements in a list? If the list had 8 elements, we would need to take the element at index 1, and the element at index 8, and swap them. Then we would do the same thing for 2 and 7, and 3 and 6, and 4 and 5. So in essence, we are swapping count, where count starts as 1, and the length of the list subtracted by the count plus 1.

§3.7 Modulo Operator

The modulo operator simply finds the remainder of two things divided. For example, 17 MOD 2 would equal 1, because 17 divided by 2 has a remainder of 1. 5 MOD 3 gives a remainder of 2, because 5 divided by 3 has a remainder of 2.

If the first number is less than the second number in a MOD operation, like 3 MOD 5, three is not divisible by 5, so you have 5 left over (aka the remainder), meaning the answer will be 3. In other words, a MOD b where a $<$ b, $=$ a.

One last thing you should know about the modulo operator is that it can be used to test for divisibility. For example, if I want to know if a number is divisible by 2, that number divided by 2 would have no remainder. That means the number MOD 2, should equal 0. This also means that if you want to check if a number is even, you can simply use MOD 2, and if the remainder is 0, the number must be even.

Problem 3.7.1 — Try the following problems:

37 MOD 7 →

10 MOD 2 →

4 MOD 8 →

The answers are: 2, 0, and 4 respectively. 37 divided by 5 goes in 7 times but leaves a remainder of 2. 10 divided by 2 goes in 5 times and leaves no remainder. 4 cannot be divided by 8, and thus 4 is left over.
A problem on the AP exam with the modulo operator may look like this.

Problem 3.7.2 — The program below iterates through list `numbersList`.

```
numbersList ← [24, 3, 12, 8, 9, 11]
count ← 1
FOR EACH number IN numbersList
{
    IF(number MOD 2 = 0)
    {
        count ← count + 1
    }
}
```

Which of the following would be the value of `count` at the end of this program?

A. 4

B. 5

C. 3

D. 1

Solution:
The answer to this problem is **C. 3**. Since the program is checking if an element MOD 2 is 0, it is the same as counting the amount of even numbers. In this case, there are three even numbers: 24, 12, and 8.

§3.8 Binary Search

Binary search is an algorithm used in a list of **sorted** items to find a specific item. The process of binary search is best explained with an example: Say we have a sorted list of 8 elements.

3	5	8	11	12	15	17	21

We want to find the number 5. We would first go to the middle, which is $8/2 = 4$th position.

3	5	8	11	12	15	17	21

We check if the number in the 4th position, which is 11, is less than, greater than, or equal to the number we want to find. In this case, 11 is greater than 5, so we know it can't be anything to the right of 11 or 11 itself, since all those values are greater. Therefore, we can eliminate all of those numbers, and we can repeat the process by going to the middle of our leftover section.

3	5	8	X	X	X	X	X

We now check if the middle number, 5, is less than, greater than, or equal to our number. We find that 5 is in fact our number, and the process ends. This process took 2 individual steps:

1. Finding the middle number (11) and checking the side the desired value is on
2. Finding the middle number 5, and checking if it is equal to our desired number

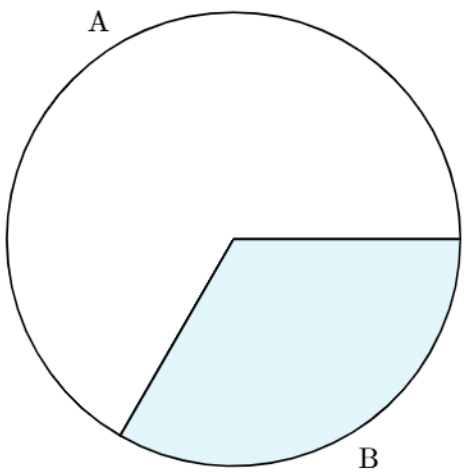
The maximum number of steps that could have been taken is four, had our number been 21. The maximum number of steps that binary search can take for an input size n, is $\lfloor \log_2 n \rfloor + 1$. The weird $\lfloor x \rfloor$ symbol around the x represents floor, meaning to round down no matter what. Sometimes, though not often, the AP exam will ask you the maximum number of steps that could be taken to search using binary. An intuitive way to think about it is that we are constantly dividing our list by 2 per step. $8/2 = 4$, which is 1 step, $4/2 = 2$, which is a second step, $2/2 = 1$, which is a third step. Then for the fourth step, we check if our remaining 1 element is the element we desire.

Note 3.8.1

Make sure you realize that binary search only works for **sorted** lists!! The lists do not need to be from least to greatest. They can be from greatest to least. They can even be alphabetically ordered instead of numbers. The main part that matters is that they are sorted and ordered. Without being ordered, binary search cannot perform.

§3.9 Randomizing/Probability

An important function to know for the AP exam is the random function. Random takes two numbers, such as RANDOM(1, 5), and it returns a number that is from 1 to 5, inclusive. This means the number 1 can also be returned, the number 1 is not excluded from the range of numbers. A random question takes two forms usually: code, or spinners.
For example you may see an image such as the one below, in which they tell you that the shaded region is one third of the area of the overall circle.



Then they will ask which code segment represents the probability of landing on the shaded region of the circle. For example:

Problem 3.9.1 — Referencing the image above, Jeremy wants to create a program that best simulates the outcome of the spinner. Landing on the shaded region, B, of the circle counts as a win. Landing on the non-shaded region, A, counts as a loss. The shaded region represents one third of the area of the overall circle.

A.
```
randomNumber ← RANDOM(1, 9)
IF(randomNumber > 3)
{
    DISPLAY("Victory")
}
ELSE
{
    DISPLAY("Loss")
}
```

B.
```
randomNumber ← RANDOM(1,3)
IF(randomNumber ≥ 2)
{
    DISPLAY("Loss")
}
ELSE
{
    DISPLAY("Victory")
}
```

C.
```
randomNumber ← RANDOM(1,3)
IF(randomNumber > 2)
{
    DISPLAY("Loss")
}
ELSE
{
    DISPLAY("Victory")
}
```

D.
```
randomNumber ← RANDOM(1,9)
IF(randomNumber > 6)
{
    DISPLAY("Loss")
}
ELSE
{
    DISPLAY("Victory")
}
```

Which of the code segments best represents the intention of Jeremy's program?

Solution: The correct answer is **option B**. Since the shaded region is $\frac{1}{3}$ of the area, there is a $\frac{1}{3}$ chance of landing on the shaded circle and "winning".

In option B, the random number can be 1, 2, or 3. If it is 2 or greater (aka 2 or 3), then it is a loss. This means that losing has a $\frac{2}{3}$ chance of happening, and the only way to win, rolling a 1, has a $\frac{1}{3}$ chance of happening. This accurately simulates the program.

The other options do not work because they simulate losing as a $\frac{1}{3}$ chance and winning as $\frac{2}{3}$ chance. Another type of question including the random function would appear as follows:

Problem 3.9.2 — Daniel is creating a program that simulates the outcome of 100 unbiased coin flips. He expects to get a roughly equal number of heads and tails, but instead gets 74 heads and 26 tails.

```
1       count ← 1
2       heads ← 0
3       tails ← 0
4       REPEAT UNTIL (count > 100)
5       {
6           random ← (1, 4)
7           IF(random ≥ 2)
8           {
9               heads ← heads + 1
10          }
11          ELSE
12          {
13              tails ← tails + 1
14          }
15      }
```

Which of the following changes could make the program run as intended?

A. Swapping lines 9 and 13

B. Assigning random to (1, 3)

C. Changing the first conditional to IF(random > 2)

D. Incrementing tails by 2 instead of 1

Solution:
The solution to this problem is **option C.** The general error of this program is that heads is increased if random is equal to or greater than two. This means heads is increased if the random value is one, two, or three. This gives heads a $\frac{3}{4}$ probability.

In option A, swapping lines 9 and 13 would do nothing but result in tails having a $\frac{3}{4}$ probability, while heads has a $\frac{1}{4}$.

In option B, heads would now have a $\frac{2}{3}$ probability, while tails has a $\frac{1}{3}$ probability. The aim of the program is to have $\frac{1}{2}$ probability of head, and $\frac{1}{2}$ probability of tails.

In option D, incrementing tails by 2 may make the final result around 50 instead of approximately 25. This would not fix the issue or give the correct output.

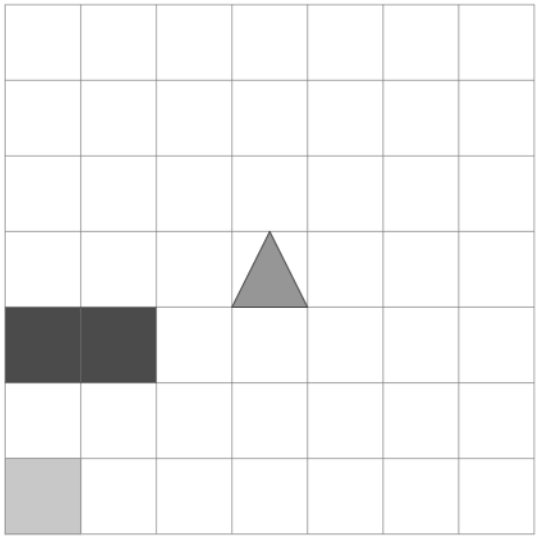
Finally, in option C, changing the conditional to random is **greater than** 2 instead of greater than or equal to results in head being incremented if random is 3 or 4, while tail is incremented if random is 1 or 2. This gives a $\frac{2}{4}$ chance for both heads and tails, which is the 50% probability the program is intended to have.

§3.10 Robot Questions

A practically guaranteed question on the AP exam will be a robot question. There will be a "robot" on a grid, which will be a black triangle. The triangle will point toward the direction it is facing. There will be a couple of functions for the "robot" such as:

1. `MOVE_FORWARD()` - Moving the robot one square forward in the direction it is facing.
2. `ROTATE_RIGHT()` - Rotating the robot 90° clockwise (to the right)
3. `ROTATE_LEFT()` - Rotating the robot 90° counterclockwise (to the left)
4. `CAN_MOVE(direction)` - Returns true if robot can move in a direction. The direction could be forward, left, right, or backward.

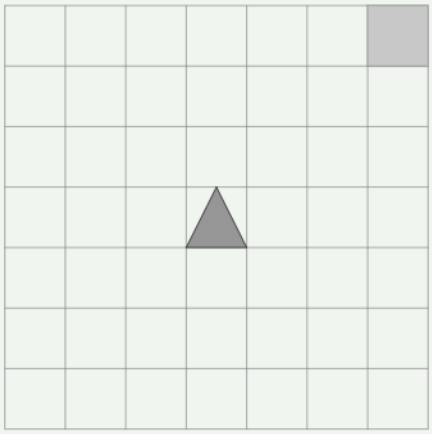
Such a grid would look like this:



The gray square often indicates the final square the robot wants to reach. The black squares indicate squares that the robot cannot enter. This has been the format consistently, but if a different format is used, the question will explain it as context.

Sometimes the question involves lots of iteration, but remember that the AP exam provides adequate time. You can draw out the path of the robot on the diagram to make sure.

Problem 3.10.1 — A robot is positioned at the center of a grid, facing forward. The function `GOAL_REACHED()` returns true when the correct position is reached.



Which of the following code segments would correctly move the robot to the gray square in its final position?

A.
```
count ← 1
REPEAT UNTIL(count > 4)
{
    MOVE_FORWARD()
    ROTATE_RIGHT()
    count ← count + 1
}
```

B.
```
REPEAT UNTIL(GOAL_REACHED())
{
    MOVE_FORWARD()
    IF NOT(CAN_MOVE(forward))
    {
        ROTATE_RIGHT()
    }
}
```

C.
```
n ← 3
REPEAT 2 TIMES
{
    REPEAT n TIMES
    {
        MOVE_FORWARD()
    }
    ROTATE_RIGHT()
    n ← n - 1
}
```

Problem 3.10.2 — (part 2)

D.
```
count ← 1
REPEAT UNTIL(count > 4)
{
    REPEAT count TIMES
    {
        MOVE_FORWARD()
    }
    ROTATE_RIGHT()
    count ← count + 1
}
```

Solution: The answer is option B. In option B, the loop repeats until the robot is in the correct square. In the loop, the robot moves forward, and when it cannot move forward any further, it turns to the right. This means, as soon as the robot collides with the upper wall, the conditional evaluates to true since it cannot move forward, and the robot turns right. Then it continues forward into the gray square.

Option A, moving forward and rotating right four times, would lead the robot in the same position as before.

Option C would make the robot move forward n squares, which would be three at first, then rotate right. Afterward, it moves forward two squares, being adjacent to the goal, but not there.

Option D would move forward one step at first, initially heading in the correct direction, but then rotate right and go forward two steps, and then rotate right and go forward three steps, and so on, until it is not headed the correct way.

§3.11 Algorithmic Efficiency

Algorithm efficiency describes how quickly the algorithm can solve a task of n steps, in terms of n. In order of efficiency, algorithms are often measured as:

1. Logarithmic (binary sort which is $\log_2(x)$)
2. Linear (mx + b form, such as $2n + 3$, or $4n$)
3. Polynomial (like n^2, $n^4 + 3n + 3$)
4. Exponential (like 3^n, 5^n, as long as n is the exponent)

Note 3.11.1

Exponential algorithms are considered **unreasonable** time to solve a task. You may be asked a question showing different algorithm solving times and asking for the unreasonable one.

Problem 3.11.2 — The following chart shows the number of steps it takes four algorithms to solve a task as input size increases.

Algorithm 1	Algorithm 2	Algorithm 3	Algorithm 4
10	5	2	0
20	9	4	1
30	17	8	2
40	33	16	2

Which of the following programs run in a time that can be considered to be unreasonable?

A. Algorithm 1

B. Algorithm 2

C. Algorithm 3

D. Algorithm 4

Solution: The answer to the problem is **option C**, algorithm 3. If you noticed, algorithm 3 is constantly doubling, meaning it an exponential function. More specifically, its 2^x

Algorithm 1 is linear since it is constantly adding tens. Algorithm 2 isn't really multiplying by anything (though roughly doubling), and doesn't have any clear pattern, indicating some polynomial time, maybe linear.

Algorithm 4's steps is barely increasing, possibly indicating a logarithmic time or something very short and efficient.

Sometimes, when looking at problems for algorithms to solve, the time that would be taken is too high (exponential). As such, **heuristics** or good solutions that approximate the best solution, but are not the best solution, are used.

Note 3.11.3

An analogy of a heuristic tactic can be seen when solving a Rubik's cube. Most people use a method to solve one such as the Beginner's method or CFOP. However, theoretically, there is *probably* a faster way to solve it through direct steps without using the aforementioned methods. Figuring out the most optimal way would require too much effort for a human though, and as such, the heuristics of CFOP/beginner's method are used.

Some problems in computer science are determined to be **undecidable problems**, meaning that no algorithm can solve them. A famous example is the "Halting problem", which asks if there is a program that can analyze the code of another program and determine whether it will loop forever or halt. Alan Turing famously used proof by contradiction to demonstrate that no such program exists.

What you need to know for AP Computer Science Principles is that **there are some problems called undecidable problems which cannot be solved by algorithms**. Though not necessary, for the proof of the halting problem, I recommend reading this.

§3.12 Simulations

Simulations are computer programs meant to copy events and their results. For example, someone may create a simulation that rolls a coin for heads or tails 100 times and keeps track of how many times heads/tails have been rolled.

In computer science, simulations are mainly useful when the real scenario:

1. Needs to be done many times
2. Takes a lot of money
3. Hazardous/dangerous

Simulations may be ineffective if the data needs to be analog, and not digital.

§3.13 End of Chapter Questions

Problem 3.13.1 — In which of the following programs would it be BEST to use the modulo operator?

A. A feature that checks if a password is shorter or longer than 10 characters.

B. A program that checks if a number is odd or even.

C. A simple drawing application.

D. A bubble sort algorithm.

Solution: The modulo operator is used to give the remainder when two numbers are divided (eg. 35 MOD 4 is 3). Option A is incorrect because checking if a password is either shorter than 10 characters or longer than 10 characters would not require the use of the modulo operator as it could be checked with a simple conditional (password < 10 or password > 10). Option C is incorrect because a simple drawing application would likely only require very basic mathematical operators such as addition or subtraction. Option D is incorrect because bubble sort (the simplest sorting algorithm) that sorts elements by comparing them to the ones adjacent to each other to eventually result in an ordered list which does not require the mod operator to do. (article)

Option B is correct because the modulo operator can be effectively utilized by doing x MOD 2 to see if there is a remainder (meaning x is an odd number) or if there is none (meaning x is an even number).

Problem 3.13.2 — A statistics teacher is building a model to demonstrate a graphing topic. In which of the following datasets would it make the MOST sense to use float values?

A. Number of pets in each household

B. Student GPAs

C. Number of people on different floors of a building.

D. Favorite class of each student

Solution: A float value is the same as a decimal value so we are looking for a dataset that has values that can be represented using decimals. Option D is the most incorrect as the favorite class of each student is a categorical variable which cannot be represented as a number. Option A is also incorrect because a number of pets has to be an integer value as there is no such thing as a partial pet. Option C is also incorrect because it is very unlikely that the number of people on one floor of a building will be represented using a decimal (may be used for representing a number in thousands or millions; e.g. 1.2 million).

Option B is correct because GPA is generally rounded to one or two decimal places which is represented as a float value.

Problem 3.13.3 — Consider the following procedure `calculate`

```
PROCEDURE calculate(x)
{
    answer ← 0
    i ← 1
    REPEAT UNTIL(i > x)
    {
        IF NOT(i MOD 2 = 0)
        {
            answer ← answer + i
        }
        ELSE
        {
            answer ← answer + 1
        }
        i ← i + 1
    }
    RETURN(answer)
}
```

Which of the following values would be returned when `calculate(x)` is run with a parameter of 8?

A. 4

B. 20

C. 28

D. 8

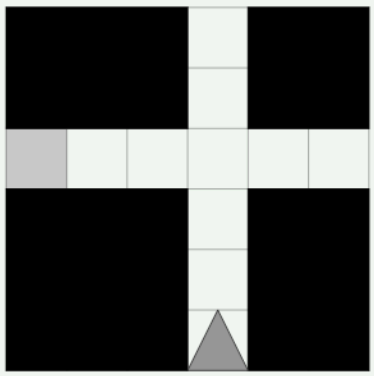
Solution:
The general idea of the program is to iterate through numbers 1 through 8, inclusive. This can be seen from the REPEAT UNTIL(i > x), where x is 8. While we iterate through these numbers, we want to check if i MOD 2 is *not* equal to zero, meaning that we want to check if i is an odd number. If i MOD 2 was zero, it would be an even number, and if i MOD 2 is not zero, it must be an odd number.

Therefore, we check if i is an odd number through checking if i MOD 2 is not equal to

zero, and if that is the case, we add i to the answer. If that is not the case, meaning if i is an even number, we simply add one to the answer. So overall, we are going through 1 to 8 inclusive, and when we encounter an odd number, (1, 3, 5, 7) we add them to the answer variable, and when we encounter an even number (2, 4, 6, 8), we only add one.

Thus the sum would look like $1+3+5+7+1+1+1+1$, where the $1+3+5+7$ are the odd numbers from 1 through 8 that are being added to `answer` and the four ones in $1+1+1+1$ represent the four even numbers of 2, 4, 6, 8, found in 1 through 8. Thus, the sum is 20, matching Option B.

Problem 3.13.4 — In the robot diagram below, the shaded gray square is the intended end goal of the robot. GOAL_REACHED returns true if the robot is on the shaded gray square.



Which of the following code segments would move the robot to the gray square?

A.
```
PROCEDURE move()
{
    REPEAT UNTIL(GOAL_REACHED())
    {
        IF(CAN_MOVE(forward))
        {
            MOVE_FORWARD()
        }
        ELSE
        {
            ROTATE_RIGHT()
        }
    }

}
```

Problem 3.13.5 — (part 2)

B.
```
PROCEDURE move()
{
    REPEAT UNTIL(GOAL_REACHED())
    {
        IF(CAN_MOVE(right) OR NOT(CAN_MOVE(forward))
        {
            ROTATE_RIGHT()
        }
        MOVE_FORWARD()
    }
}
```

C.
```
PROCEDURE move()
{
    n ← 3
    REPEAT n TIMES
    {
        MOVE_FORWARD()
        REPEAT (n + 1) TIMES
        {
            ROTATE_RIGHT()
        }
    }

}
```

D.
```
PROCEDURE move()
{
    REPEAT UNTIL(GOAL_REACHED())
    {
        IF(CAN_MOVE(forward))
        {
            MOVE_FORWARD()
        }
        ELSE
        {
            ROTATE_LEFT()
        }
    }
}
```

Solution:
Option A moves the robot straight forward, until it reaches the other end, and then it rotates right twice and repeats the pattern.

Option B checks if the robot can move right or if it cannot move forward. Since the robot will be able to move to the right in the middle, once the robot reaches the middle

intersection, it'll rotate to the right and go down the right side horizontal pathway. Then once the robot hits the wall of the pathway, the robot will rotate right twice, and repeat the pattern. Eventually the robot reaches the gray square, making **B the correct choice**.

In Option C, the robot moves forward, and then rotates n+1 times, meaning 4 times since n is 3. Rotating four times is a full 360, equivalent to not rotating at all. The robot does this two more times, and thus overall only reaches two squares forward from the starting position.

Lastly, Option D is identical to A, except for turning to the right, the robot turns to the left. The same problem with A would occur with option D.

Problem 3.13.6 — A programmer wants to find a certain number in a list of 1,000 sorted integers. Which of the following represents the maximum number of steps it could take to sort through this list?

A. 256 steps

B. 500 steps

C. 20 steps

D. 10 steps

Solution:
The solution to this problem is Option D, 10 steps. 1000 could be divided by 2 nine times before reaching a value less than 1. We see how many times 1000 can be divided by two because in binary search, we split the list of items in half and discard one half. Repeating this process is essentially the same as dividing the number of items by 2. After, +1 is added to the amount of steps since the last step is checking the final remaining number. $9 + 1 = 10$, making option D the correct answer. Option B and A are far too great, and either through repeated division or the previously mentioned formula, the answer could be found as Option D.

Problem 3.13.7 — A group of scientists want to study the spreading of poison through the body. Which of the following is least likely why the scientists would use a computer program to simulate the scenario?

A. Real life studies of poison can be hazardous

B. Poison studies can be expensive to replicate

C. Doing real experiments instead of computer simulations can provide more accurate data.

D. Real life experiments could be time consuming and take many trials.

Solution:
Out of all the options, C is the least likely reason to use a computer simulation instead of doing the actual experiment. A has the benefit of reduced risk, B has the benefit of less cost, and D makes the process quicker. However C indicates that simulations may not be as accurate as real experiments, providing reason to not use a computer simulation.

Problem 3.13.8 — In the following code segment, x represents an integer.

```
IF(x MOD 2 = 0)
{
    IF(x < 5)
    {
        DISPLAY("Perfect")
    }
    ELSE
    {
        DISPLAY("Banana")
    }
}
ELSE
{
    IF(x MOD 4 = 0)
    {
        DISPLAY("Apple")
    }
    ELSE
    {
        DISPLAY("Orange")
    }
}
```

Which of the following is not a possible value that can be displayed by the program?

A. Apple

B. Banana

C. Perfect

D. Orange

Solution:
The program first checks if x is divisible by 2. If not, it goes into the else statement, where it first checks if x is divisible by 4. However, if something is not divisible by 2, it could not be divisible by 4.

This is the case because to be divisible by 4 means to be divisible by $2 \cdot 2$, meaning that the number is divisible by 2. Try thinking of a number not divisible by 2, but divisible by 4. Thus, the answer is A. Apple.

In addition, through process of elimination, a number can be divisible by 2 and be less than 5, (e.g., 2 and 4). Therefore "Perfect" is possible. "Banana" is also possible because there are numbers divisible by 2 and greater than 5 (6, 8, 10, every even number ¿ 5).

Finally, orange is definitely possible because if x is not divisible by 2, it checks if x is divisible by 4, and it was previously established that a number couldn't be not divisible

by 2 yet divisible by 4, so it would go to the else statement automatically and display "Orange".

Note 3.13.9

If trying to conceptualize the answer to a problem is difficult sometimes, test out trial numbers and see how they interact within the program to better understand how the program works. You may even get an answer this way, though number bashing is not recommended.

Problem 3.13.10 — The following code segment with boolean values is executed.

```
yellow ← true
blue ← false
green ← yellow AND blue
blue ← NOT(green)
DISPLAY(yellow AND blue)
DISPLAY(blue)
```

Which of the following will be displayed in the output after the program is ran?

A. True, true

B. False, true

C. True, false

D. False, false

Solution: Blue is false when green is set to yellow and blue, assigning green a value of false. Then when blue is set to NOT(green) blue inverses its false boolean value to true. Thus, since yellow is already true, displaying yellow AND blue displays the result of true AND true which is true. Then blue is displayed, which is previously established to be true, resulting in the answer being option A.

Problem 3.13.11 — A programmer wants to create a code segment that moves an element in the middle of a list to the end of the list. For example, if `fruits` has [“apples”, “bananas”, “blue berries”, “peach”, “orange”], the new arrangement in the list should be [“apples”, “bananas”, “peach”, “orange”, “blue berries”]. You can assume that the length of the list is an odd number greater than or equal to three. The function `middle(list)` returns the index of the element positioned in the middle of the list.

A.
```
len ← LENGTH(fruits)
temp ← fruits[middle(fruits)]
REMOVE(fruits, len)
INSERT(fruits, middle(fruits), temp)
```

B.
```
len ← LENGTH(fruits)
temp ← fruits[middle(fruits)]
REMOVE(fruits, middle(fruits))
INSERT(fruits, len, fruits[middle(fruits)])
```

C.
```
len ← LENGTH(fruits)
temp ← fruits[len]
REMOVE(fruits, middle(fruits))
INSERT(fruits, middle(fruits), temp)
```

D.
```
len ← LENGTH(fruits)
temp ← fruits[len]
REMOVE(fruits, len)
INSERT(fruits, len, temp)
```

Solution: The general idea of the code should be to set a temporary variable to the element in the middle of the list, so that when it is removed the element is not completely gone, and then insert the element at the last index of the list provided by len which is the length of the list.

Option A correctly assigns len and the temporary variable, but it removes the last element in the list `fruits` instead of the middle element. In addition, it also inserts the temp variable into the middle instead of the end of the list.

Option B correctly assigns the temp variable and the length of the list. It also removes the middle element, and then inserts it at the end, thus making option B the correct answer. Option C assigns the temp variable as the last element in `fruits`, and then replaces the middle element with the last element which is nearly the opposite of the intended purpose. Option D sets the last element to variable temp, removes the last element, then inserts it again, effectively resulting in no change to the list at all.

Problem 3.13.12 — The following procedure searches for a number in a list of numbers. It returns true if the number is found, and false otherwise.

```
PROCEDURE find_number(number, numberList)
{
    FOR EACH element IN numberList
    {
        IF(element = number)
        {
            RETURN(true)
        }
    }
    RETURN(false)
}
```

Which of the following statements are true about the procedure `find_number()`?

I. The procedure uses a heuristic to find the specified number

II. The procedure runs in n steps where n represents the length of numberList.

III. The procedure runs a linear search through numberList.

A. Only statement I

B. Only statement II

C. Statement I and II

D. Statement II and III

Solution:
The program iterates through each element in the list `numberList` and checks once if the element is equal to the specified number. Thus, it linearly searches through the entire list. Additionally, since it checks every element once, the amount of steps it will take is equal to the amount of elements in the list. A linear search also has a runtime complexity that is equal to the size n of the list. Therefore, statement II and III are both true, making D the correct answer.

Statement I, the program uses a heuristic, is false because the definition of a heuristic is a "good-enough" solution when the actual answer to the solution takes too much time to compute or is impossible. However, finding an number in a list of numbers is not a situation that falls into that category.

Problem 3.13.13 — A programmer wants to create a program that can take as input another program's code and determine whether that program will be stuck in an infinite loop, or eventually terminate. The problem is deemed an undecidable problem. Which of the following statements is true?

A. Creating such a program is possible for all sets of possible inputs to the program.

B. Due to the resource strain of such a problem, advanced computer hardware that is not available in the present day is needed.

C. Creating such a program is not possible for all sets of inputs.

D. Creating a program such as the one the programmer intended is possible, but requires careful memory manipulation in assembly or other low level programming languages.

Solution:
The problem described in the question is known as the halting problem. The halting problem is proof that there are some undecidable problems that cannot be solved by an algorithm. Thus, A and D are incorrect. B is also incorrect because it is not a matter of the specs of a computer. It is logically impossible (check out the proof by clicking the reference link → proof link).

The only logical answer left is C. Creating such a program is not possible for all sets of inputs.

Problem 3.13.14 — A programmer wants to create a program that takes a list of random numbers in binary (base 2) as input, and returns a list of sorted numbers in decimal (base 10) with three times their original value. The programmer knows that these are the generic steps that must be followed:

Step	Explanation
Convert	Convert the list of binary numbers to base 10 number form.
Sort	Sorts the list from least to greatest
Multiply	Multiply each element to be three times its original value

Which of the following option describes the possible order of steps that will make the program function as intended?

I. Convert, multiply, sort

II. Multiply, convert, sort

III. Sort, convert, multiply

A. III only

B. I, II, and III

C. II and III

D. II only

Solution:
Multiplying then sorting or sorting then multiplying are commutative and result in the same end list. Consider a list of integers [2, 5, 1, 3, 4], multiplying each by 3 would result in $[2 \cdot 3, 5 \cdot 3, 1 \cdot 3, 3 \cdot 3, 4 \cdot 3]$. Since they are all multiplied by the same factor of 3, sorting them would be the same as if they were sorted and then multiplied by three. Converting can also happen at any point in the process, we could do all the operations in binary and then convert at the end, or convert in the beginning and then do all the operations. We could also convert in the middle, and either sort or multiply before. This makes option B the correct answer.

Problem 3.13.15 — In order to make the principal's list, a student must have a grade between 80 and 93 inclusive (any higher than 93 would be honor roll), and less than 10 absences. Which of the following boolean expressions evaluate to true if a student meets the requirements for principal's list?

A. `((grade ≤ 80) AND (grade ≥ 93)) AND (absences < 10)`

B. `((grade ≥ 80) AND (grade ≤ 93)) AND (absences < 10)`

C. `((grade ≥ 80) AND (grade ≤ 93)) OR (absences < 10)`

D. `((grade ≤ 80) OR (grade ≥ 93)) AND (absences < 10)`

Solution:
Knowing that a student must have a grade between 80 and 93 inclusive, a conditional expression must exist that checks if the student's grade is greater than or equal to 80 and less than or equal to 93. Therefore, option A and D are incorrect because they both check if a student's grade is less than or equal to 80 and greater than or equal to 93 (which is impossible as well).

Option C is incorrect because despite correctly checking if a grade is within the range of 80 to 93 inclusive, the boolean expression says OR absences < 10, indicating that either the grade could be in the required range OR absences could be less than 10 to be in principal's list. However, both requirements must be met. Thus, Option B is the correct answer since it checks the range (80-93 inclusive) AND checks if absences are less than 10.

Problem 3.13.16 — A program is intended to display the following output:
`apple apple apple banana apple apple apple banana tomato`
Which of the following code segments will display the intended output?

A.
```
REPEAT 2 TIMES
{
    REPEAT 3 TIMES
    {
        DISPLAY("apple")
    }
    DISPLAY("banana")
    DISPLAY("tomato")
}
```

B.
```
REPEAT 3 TIMES
{
    DISPLAY("apple")
    REPEAT 2 TIMES
    {
        DISPLAY("banana")
    }
    DISPLAY("tomato")
}
```

C.
```
REPEAT 2 TIMES
{
    REPEAT 3 TIMES
    {
        DISPLAY("apple")
    }
    DISPLAY("banana")
}
DISPLAY("tomato")
```

D.
```
REPEAT 3 TIMES
{
    DISPLAY("apple")
    REPEAT 2 TIMES
    {
        DISPLAY("banana")
    }
}
DISPLAY("tomato")
```

Solution: Looking at the output that is supposed to be displayed, apple repeats three times, followed by banana, and then that entire process is repeated twice. After, a tomato is added at the end.

$$\overbrace{\text{apple apple apple banana}}^{\text{repeats once}}\ \overbrace{\text{apple apple apple banana}}^{\text{repeats twice}}\ \overbrace{\text{tomato}}^{+\text{tomato}}$$

Thus, we know there must be an overarching REPEAT 2 TIMES loop. Tomato is probably outside of all the loops and added at the very end before the program is finished executing since it is not displayed multiple times and appears only once as the last displayed string. Automatically, this allows us to remove option A and option B since they both have tomato within the REPEAT 3 times loop, meaning tomato would be displayed three times.

Technically, option D can be removed as well considering that the outermost loop is supposed to be a REPEAT 2 times loop. However, even not knowing that, the program displays `apple`, then `banana banana` from the nested REPEAT 2 times loop and then repeats the overall process three times. This is totally different from the intended output.

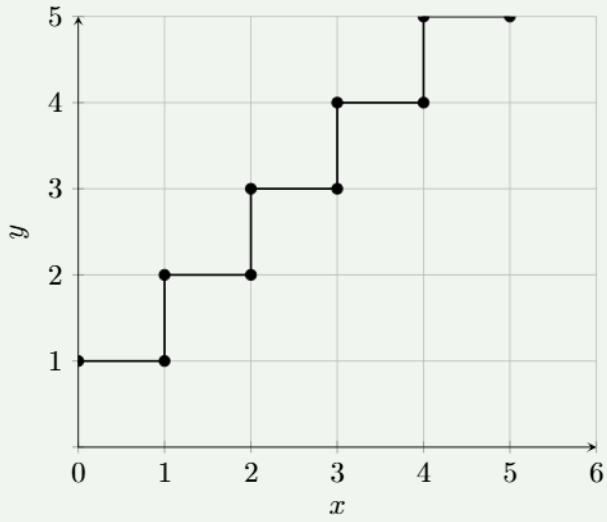
Additionally, by noticing

$$\underbrace{\text{apple apple apple}}_{\text{apple was repeated three times}}\ \text{banana}$$

It can be reasoned apple is in a REPEAT 3 TIMES loop that is nested in an overarching REPEAT 2 times loop, with tomato outside all of the loops. Therefore, the answer is **option C**. Of course bashing through the code by working it all out also works, but sometimes a higher level overview of the code can save some time and get the right answer quicker.

Problem 3.13.17 — Consider the following procedure and graph.

Procedure	Explanation
draw(x1, y1, x2, y2)	Draws a line segment from coordinates (x1, y1) to (x2, y2)



Which of the following code segments can be used to draw the figure?
Select <u>two</u> answers.

A.
```
xCord ← 0
yCord ← 1
len ← 1
REPEAT 9 TIMES
{
    IF NOT(xCord = yCord)
    {
        draw(xCord, yCord, xCord + len, yCord)
        xCord ← xCord + len
    }
    ELSE
    {
        draw(xCord, yCord, xCord, yCord + len)
        yCord ← yCord + len
    }
}
```

B.
```
xCord ← 5
yCord ← 5
len ← 1
REPEAT 9 TIMES
{
    IF NOT(xCord = yCord)
    {
        draw(xCord, yCord, xCord, yCord - len)
        yCord ← yCord - len
    }
    ELSE
    {
        draw(xCord, yCord, xCord - len, yCord)
        xCord ← xCord - len
    }
}
```

Problem 3.13.18 — (Part 2)

C.
```
xCord ← 0
yCord ← 1
len ← 1
REPEAT 9 TIMES
{
    IF NOT(xCord = yCord)
    {
        draw(xCord, yCord, xCord, yCord + len)
        yCord ← yCord + len
    }
    ELSE
    {
        draw(xCord, yCord, xCord + len, yCord)
        xCord ← xCord + len
    }
}
```

D.
```
xCord ← 5
yCord ← 5
len ← 1
REPEAT 9 TIMES
{
    IF NOT(xCord = yCord)
    {
        draw(xCord, yCord, xCord - len, yCord)
        xCord ← xCord - len
    }
    ELSE
    {
        draw(xCord, yCord, xCord, yCord - len)
        yCord ← yCord - len
    }
}
```

Solution:
There are two ways to draw the graph shown. Draw the line segments starting from $(0, 1)$ to $(5, 5)$, or start drawing from $(5, 5)$ to $(0, 1)$. If we begin drawing the line segments from the first method, we notice that when xCord = yCord, that means we need to draw a line segment that is vertical (in the draw function, xCord stays the same, and yCord to yCord + 1), and when xCord $\neq$ yCord, we need to draw a line segment that is horizontal (in the draw function, yCord stays the same, and we draw from xCord to xCord + 1).

Option A matches the first method since it checks IF NOT(xCord = yCord), which is the same thing as saying if xCord $\neq$ yCord. In the case that xCord $\neq$ yCord, it draws a horizontal line by drawing a line segment from (xCord, yCord,) to (xCord + len, yCord), in which we know len = 1. Thus, a horizontal line is drawn because the y remains the same but the x is increased by 1. In the case that xCord does equal yCord (the else statement of option A), a vertical line upward needs to be drawn. Thus, using the function draw, the program draws a line segment from coordinates (xCord, yCord) to

(xCord, yCord + len), resulting in a horizontal upward line segment.

Option C is the same thing as option A, but it reverses when the horizontal line segment should be drawn vs the vertical line segment, since it swaps the code for the horizontal line with the vertical line.

Method 2 of drawing the graph, starting from the top right coordinate $(5, 5)$ and working our way down, is done correctly in option B. Contrary to method 1 of drawing the graph, since we are going down from coordinate $(5, 5)$, when xCord $\neq$ yCord, we don't want to create a horizontal line segment, but rather a vertical line segment. To create this vertical line segment, we input the endpoints of the line segment into draw() as follows: draw(xCord, yCord, xCord, yCord - len). In the other case where xCord = yCord, we want a a horizontal line segment to be drawn to the left. Thus we draw from (xCord, yCord) to (xCord - len, yCord).

Option D is the same thing as Option B, but it again swaps when the horizontal line segment should be drawn vs the vertical line segment. Therefore, **Option A and B** are the answers.

Problem 3.13.19 — A list of numbers called list has `len` elements. A programmer decides to create pseudocode steps for an algorithm that swaps the first and last element in the list, the second and second last, and so on to effectively reverse the order of the elements in the list. For example, if the list had elements [2, 4, 6, 8, 10], after the algorithm it should have [10, 8, 6, 4, 2]. Steps 5 and 6 are missing.

Step 1: Declare a variable called count and set it equal to one

Step 2: Declare a variable called temp and set it equal to the element at index `len`

Step 3: Set the element at the index `len` equal to the element at index count

Step 4: Set the element at the index `count` equal to the temp variable

Step 5: Missing

Step 6: Missing

Which of the following choices could be used as steps 5 and 6 to make the program run as intended?

A. Replace **step 5** with increase count by 1 and decrease len by 1. Replace **step 6** with repeat steps 2-5 until len $\leq$ count.

B. Replace **step 5** with decrease count by 1 and increase len by 1. Replace **step 6** with repeat steps 2-5 until count $-$ len $= 0$.

C. Replace **step 5** with increase count by 1 and decrease len by 1. Replace **step 6** with repeat steps 2-5 until len $\geq$ count.

D. Replace **step 5** with decrease count by 1 and increase len by 1. Replace **step 6** with repeats step 2-5 until len $-$ count $= 0$.

Solution:
Essentially, the idea of the program is to store the last element of the list in a variable called temp (short for temporary), and then change the last element of the list to the value of the first element, and finally use the temporary variable to change the first element of the list to temporary. This swaps the values of the first and last element in the list. Next we want to swap the values of the second and second last element in the list, and so on.

The second element's index is really just the first element's index + 1. Thus the second elements index is 2 (because $1 + 1 = 2$). The second last element's index is the last element's index - 1. The last element always has an index of the length of the list (when working with 1-indexed lists). This indicates that the second last element's index is the length of the list - 1. We need to repeat this procedure until a certain point to swap all values in the list. Automatically, knowing this information, we can remove Option B and Option D since they decrease count by 1 (which is the same as finding the index of the second element by doing the first element index - 1) instead of increase count by 1. Also they increase len by 1 (which is the same as finding the index of the second last element by looking for the length of the list + 1).

In fact, trying to find an element of an index that is the length of the list + 1 doesn't even exist since an index of the length of the list is the last item, so an index of the length + 1 is out of bounds. Now, the difference between the two remaining options A and C is the step 6 of the program. Step 6 is supposed to repeat the prior steps 2-5 until a certain condition is satisfied. Knowing that len represents the length of the list, len will always be greater than or equal to count (since count = 1 and a list that has elements has to have at least one element minimum). Therefore, option C doesn't make sense because the repeating steps 2-5 part will immediately stop since len is already $\geq$ count normally. **Option A is the answer.**

Alternatively, to remove Option B and Option D from possible choices, the step 6 of both choices is not correct. If a list has an even amount of elements, len - count or count - len will never equal 0. Only if a list has an odd number of elements will len and count have the same value and thus subtract to equal 0. Thus the stopping condition provided by Option B and Option D only works for lists that have an odd number of elements.

Problem 3.13.20 — In a restaurant, people above 65 get a 10 dollar discount, people below 18 get a 7 dollar discount, and people in between those ages get no discount. A programmer creates the following procedure called calculate with parameters age and price in order to fulfill the purpose of calculating and returning the total a customer must pay. The parameter price is a number representing the cost before a discount is applied.

```
1       PROCEDURE calculate(age, price)
2       {
3           total ← 0
4           price ← 0
5
6           IF(age < 18)
7           {
8               total ← price - 7
9               RETURN(total)
10          }
11          ELSE IF(age ≤ 65)
12          {
13              total ← price - 10
14              RETURN(total)
15          }
16          ELSE
17          {
18              RETURN(price)
19          }
20      }
```

Which of the following options describes how to change the program to make it run as intended?

Select two answers

A. Switch lines 3 and 18

B. Remove line 4

C. Change line 10 to ELSE IF(age > 65)

D. Change line 5 to IF(age > 18)

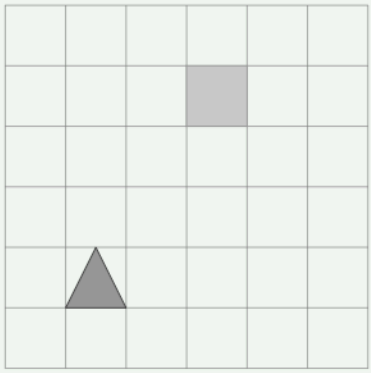
Solution:
The program is meant to check if someone is less than 18 in which case the price is decreased by 7, if someone is above 65, in which case the price is decreased by 10, and if neither condition is satisfied, the person must be between 18 and 65 inclusive meaning no discount.

Option A, switching lines 3 and 18, would result in the program crashing since the variable total would be used before it is defined. Thus, option A is incorrect. Option B, remove line 4, is correct since setting the price to 0 by default removes the price tag entirely. In fact, if the person is eligible for a discount, their total would be -7 or -10 dollars which is totally unreasonable.

Option C, changing line 10 to ELSE IF(age $>$ 65) is correct. The program first checks whether a person is below 18, in which case a discount of 7 dollars is applied, as intended. However, when they are not below 18, the next conditional ELSE IF(age $<$ 65) is checked. If this conditional is satisfied, it means that the person's age is 18 or greater, but less than or equal to 65. In such a case, no discount is supposed to be applied, rather than a discount of 10 dollars.

Finally Option D is incorrect because if a person's age is greater than 18 a discount of 7 dollars is not supposed to be applied. **Option B and Option C** are the answers.

Problem 3.13.21 — An image of a robot diagram as well as procedure is shown below. The gray square indicates the intended final location of the robot after the program is executed.



```
PROCEDURE walk(num)
{
    REPEAT num TIMES
    {
        MOVE_FOWARD()
        ROTATE_RIGHT()
        MOVE_FORWARD()
        ROTATE_LEFT()
    }
}
```

The parameter num is a positive integer. Which of the following code segments would result in the robot reaching the gray square as its final destination?

A.
```
ROTATE_RIGHT()
walk(4)
```

B.
```
walk(3)
ROTATE_LEFT()
MOVE_FORWARD()
```

C.
```
walk(4)
ROTATE_LEFT()
MOVE_FORWARD()
```

D.
```
walk(3)
MOVE_FORWARD()
ROTATE_RIGHT()
```

Solution: By going through all the iterations of the code, option B is the only answer that would work. Alternatively, the procedure walk with an argument of 1 inputted essentially moves the robot one square diagonally upward and to the right from its current

location by moving the robot up one square and then to the right one square. This pattern is useful as it simplifies the problem a lot.

For example, Option C can be eliminated because instead of iterative manually going through walk(4), we can just assume the robot ends up four squares diagonally upward to the right (the block that is on the top right of the grid). Then the robot turns left and moves forward once, finally positioning the robot diagonally adjacent to the shaded square but not on it.

In Option D, we know the robot will end up three squares diagonally upward and to the right of its initial location. The robot will therefore end up on the square that is directly to the right of the shaded square and touching it. The robot then moves forward and turns right which takes it further away from the shaded square. Therefore, option D can be eliminated.

In Option B, the robot ends up three squares upward and to the right of its initial location, just like Option D. This places the robot right next to the shaded square, on the empty square to the right. The robot then turns left and moves forward into the shaded square, thus making Option B the correct answer.

Problem 3.13.22 — A table is shown below displaying a boolean value called bool1, a logical operator, and another boolean value called bool2. The logical operator was used to put bool1 and bool2 together into a boolean expression. The result of the boolean expression is in the column called result.

bool1	**Operator**	**bool2**	**Result**
bool1	AND	bool2	FALSE
bool1	OR	bool2	TRUE
bool2	AND	bool2	FALSE

Which of the following represents the value of bool1 and bool2?

A. bool1 = true, bool2 = false

B. bool1 = false, bool2 = false

C. bool2 = false, bool2 = true

D. bool2 = true, bool2 = true

Solution:
Since (bool1 AND bool2) equal false, we know one or both must be false. A false and a false or a true and a false is the only way to get back a false, since one of the conditions are not satisfied. However, in the next row, evaluating (bool1 OR bool2) returns true, meaning that one of the boolean values was true. So currently we know either bool1 or bool2 is true, and the other is false.

Finally, the third row shows that bool2 is false since (bool2 AND bool2) evaluate to false and the only way to get a false is if one of the parts is a false or both. However, since (bool2 AND bool2) uses the same two variables, we know bool2 is equal to false. Alternatively, checking only the third row and second row suffices. Therefore, since bool1 is true and bool2 is false, option A is the correct answer.

Problem 3.13.23 — In the following program, if an integer `n` is divisible by 3, it should return "fizz". If the integer is divisible by 5, it should return "buzz". If the integer is divisible by both 3 and 5, it should return "fizzbuzz".

```
1   PROCEDURE fizz_buzz(n)
2   {
3       IF(n MOD 3 = 0)
4       {
5           RETURN("fizz")
6       }
7       ELSE IF(n MOD 5 = 0)
8       {
9           RETURN("buzz")
10      }
11      ELSE
12      {
13          RETURN("fizzbuzz")
14      }
15  }
```

The program does not work as intended. Which of the following choices could make the program function properly?

Select two answers.

A. Swap the boolean expressions of lines 3 and 7

B. Swap lines 5 and 9

C. Replace the boolean expression of line 3 to (`n MOD 3 AND n MOD 5`)

D. Swap lines 13 and 5

Solution: The main issue with this program is a logic error. The program should check whether a number is divisible 3 and 5 or not before checking whether it is divisible by 3, since a number divisible by 3 and 5 will also come up as a number divisible by 3 and thus the program might mistakenly return "fizz" when it is supposed to return "fizzbuzz". This also applies to mistakenly checking if the number is divisible by 5 before checking if it is divisible by 5 and 3, which would return "buzz" instead of "fizzbuzz".

Therefore, the first conditional that should be checked is if the number is divisible by both 3 and 5. This means Option C is definitely one of the correct choices. To arrive at this conclusion, it might be helpful to test some trial numbers such as 15, which is divisible by both 3 and 5 and see where the program is making the mistake. Following the code line by line after using an arbitrary trial number can sometimes solve the problem by itself, as in this case.

Furthermore, though Option C correctly puts the conditional checking if a number is divisble by 3 and 5 first, an error it makes is that it would return "fizz" instead of "fizzbuzz". However, to fix this, lines 13 and 5 can be swapped. This matches option D. **Option C and Option D are the answers.**

We know Option C must definitely be in the answer, but other combinations with Option C wouldn't work because they would return the wrong output. For example, Option A and Option C would result in returning buzz if a number is divisible by 3 instead of fizz. Option B and Option C would result in buzz being displayed when fizzbuzz is supposed to be displayed.

Problem 3.13.24 — A teacher wants to assign her students into four separate groups. To do so, she walks to each student, giving the first student a number of one, the second student a number of two, the third a number of three, the fourth student a number of four. She repeats this cycle until all the students are accounted for (for example, the fifth student would receive a number of 1, the sixth a number of 2, etc.) To make it easier, she wants to create a program where students enter what position they are as a number and the program outputs the group they belong to.

Which of the following code segments correctly performs the intended purpose?

A.
```
PROCEDURE group(position)
{
    RETURN(position MOD 4)
}
```

B.
```
PROCEDURE group(position)
{
    IF(position MOD 2 = 0)
    {
        RETURN(4)
    }
    ELSE
    {
        RETURN(position MOD 4)
    }
}
```

C.
```
PROCEDURE group(position)
{
    IF(position MOD 4 = 0)
    {
        RETURN(4)
    }
    ELSE
    {
        RETURN(position MOD 4)
    }
}
```

D.
```
PROCEDURE group(position)
{
    RETURN(position MOD 2)
}
```

Solution:
Initially Option A may seem to be the correct answer since you could take a student position (for example 9), and find the remainder when divided by 4 (aka 9 MOD 4),

which in this case would be 1. This is indeed correct and finds the accurate group for a student of position 9. However, when the student is in position 4, 4 MOD 4 returns 0. However, a student in the fourth position is supposed to be in group number four, and furthermore a group zero does not exist. Therefore, the special case in which a student's position is a multiple of 4 causes the code segment in Option A to not return the correct values.

To fix this issue, we can simply check for the case where a student's position MOD 4 is 0 and return 4 since a student whose position MOD 4 is 0 is a multiple of four (4th, 8th, 12th, etc.) and thus should be in group 4. This matches option C since it checks if a students position MOD 4 is 0, returning 4 in the case that it is, and otherwise returning position MOD 4 to handle the other cases that work perfectly fine.

Option B is incorrect because if a student has a position of 2 (or a position that is a multiple of 2), it'll return group four when the student is supposed to be in group 2. This also applies to option D.

Problem 3.13.25 — A programmer wants to create a procedure that returns the minimum number in a list. Which of the following code segments correctly performs the intention of the program?

A.
```
PROCEDURE minList(list)
{
    minNumber ← list[1]
    len ← LENGTH(list)
    index ← 1
    REPEAT len TIMES
    {
        IF(minNumber < list[index])
        {
            minNumber ← list[index]
        }
        index ← index + 1
    }
    RETURN(minNumber)
}
```

B.
```
PROCEDURE minList(list)
{
    minNumber ← list[1]
    len ← LENGTH(list)
    index ← 1

    REPEAT len TIMES
    {
        IF(minNumber > list[index]])
        {
            minNumber ← list[index]
            RETURN(minNumber)
        }
        index ← index + 1
    }
}
```

C.
```
PROCEDURE minList(list)
{
    minNumber ← list[1]
    len ← LENGTH(list)
    index ← 1
    REPEAT len TIMES
    {
        IF(list[index] < minNumber)
        {
            minNumber ← list[index]
            index ← index + 1
        }
    }
    RETURN(minNumber)
```

Problem 3.13.26 — (part 2)

D.
```
PROCEDURE minList(list)
{
    minNumber ← list[1]
    len ← LENGTH(list)
    index ← 1
    REPEAT len TIMES
    {
        IF(minNumber > list[index])
        {
            minNumber ← list[index]
            RETURN(minNumber)
            index ← index + 1
        }
    }
}
```

Solution:
This program attempts to return the minimum element in a list by setting a variable minNumber to the first element of the list, then comparing the value of minNumber to each value in the list by incrementing index by one and using list[index]. If the list[index] is less than minNumber, then minNumber is set to list[index] since it has a lower value, and the process repeats. Overall the entire process should only take an amount of steps equal to the size of the list.

Option A is incorrect because it checks if the minimum number is **less than** the current number, and if so sets the minimum number to a new value of list[index]. This means it's finding the greatest value in the list.

Option B is incorrect because as soon as a new number less than the initial value of minNumber is found, it automatically returns minNumber without going through the entire list. So for example, for a list of [2, 4, 1, 8, 3, 0], minNumber would start of as list[1] which is 2, and be compared to 4, which it is less than, and then 1 which it is greater than. After 2 is compared to 1, the number 1 is returned instantly, and the program doesn't go through the entire list to realize that there is a value of 0 that is lower than 1.

Option C is correct because it follows the design intent of the program correctly. However, Option D is incorrect because just like option B, it automatically instantly returns the minNumber value without comparing to every element in the list first.

Problem 3.13.27 — In which of the following lists would binary search be most effective? Select two answers

A. [0, 3, 2, 8, 9]

B. ["apple", "banana", "carrot", "diamond" "egg"]

C. [1, 2, 1, 3, 9, 4]

D. [10, 8, 7, 5, 3]

Solution:
Binary search only works for lists that are ordered. This automatically means that Option C and Option A are incorrect since both are unordered lists. Binary search actually works for strings if they are ordered alphabetically (since, for example, you could compare ASCII values). This means option B is correct. Option D is also correct because a list doesn't need to be ordered from least to greatest for binary search. It can also be ordered from greatest to least. As long as the list is sorted, binary search works. Therefore, both option B and D are the correct answers.

Problem 3.13.28 — A program contains the following procedures explained in the chart. A programmer created the following code with the intention of combining the first three characters of each string in a list of strings called list. Each string can be assumed to have at least three characters. The code does not run as intended due to a run-time error and a logic error.

Procedure	Explanation
`concat(str1, str2)`	Combines two strings together and returns it as a single string. For example, concact("wa", "ter") will return "water"
`substr(str, start, length)`	Using string `str`, and starting at a certain letter of the string specified by the index `start`, the function returns a substring from the character at the `start` position to the character at the `length` position.

```
PROCEDURE combine(list)
{
    i ← 1
    finalStr ← ""
    REPEAT UNTIL (i > LENGTH(list))
    {
        finalStr ← concat(finalStr, substr(list[i], 1, 2)
    }
    RETURN(finalStr)
}
```

Which change needs to be made for the code to work properly?

A. Change `i = 1` to `i = 2`

B. Change `substr(list[i], 1, 2)` to `substr(list[i], 1, 3)`

C. Move `RETURN(finalStr)` within the `REPEAT UNTIL` loop

D. Set `finalStr = list[1]` instead of `""`

Solution:
To refresh on substrings and concatenation, read this. In order to combine the first three letters of each string in the list, we need to iterate through each element of the list (aka each string) and then take the substring of that string from the 1st to the 3rd letter and combine it with a final string variable.

Knowing this, we realize that we need the substring function to be `substr(list[i], 1, 3)` and not `substr(list[i], 1, 2)` because the latter only returns the first two characters of a given string and not the first three. Thus, option B is the correct answer.

Opotion A is wrong because setting `i = 2` instead of 1 means skipping the first element in the list. Option D is wrong because setting `finalStr = list[1]` means that when the loop runs, the first three characters of the first element in the list will appear twice in the final string. Option C is also wrong because if the `RETURN(finalStr)` was moved

within the REPEAT loop, then the finalStr would be returned before the list was even fully iterated through and each element was checked.

Problem 3.13.29 — Which of the following best explains what change must be made to Program B in order for its output to be the same as Program A?

Program A:

```
1       a ← 10
2       REPEAT UNTIL a < 5
3       {
4           DISPLAY(a)
5           a ← a - 1
6       }
```

Program B:

```
1       a ← 10
2       REPEAT UNTIL a < 5
3       {
4            a ← a - 1
5           DISPLAY(a)
6       }
```

A. Change a = 10 to a = 11

B. Change DISPLAY(a) to DISPLAY(b)

C. Change REPEAT UNTIL a ¡ 5 to REPEAT UNTIL a ¿ 5

D. None of the above

Solution:
Option B is incorrect because there is no variable b to be displayed. Option C is incorrect because variable a starts off as a value greater than 10 which means the loop will never run.

- The output for Program A: 10, 9, 8, 7, 6
- The output for Program B: 9, 8, 7, 6, 5

As you can see, all of the output values for Program B are one less than the output values for Program A. Option A adds one to the starting value of variable a, which may seem like the correct answer, but the output for Program B will actually have an additional output 5 at the end, and so the outputs are not the same and Option A is incorrect. Option D is correct because none of the listed changes will resulted in the same output. As you can see, all values in the output of Program B are one less than the values in Program A.

Problem 3.13.30 — A teacher has a test curving policy in which she takes the highest grade, and puts all the other students' grade out of the highest grade. For example, if the highest grade was an 80%, and someone scored a 40%, they would now score a 50%. Which of the following would correctly calculate a student's score, if `maxGrade` is the highest grade obtained by a student and `currentGrade` is the student's grade before the adjustment by the curve?

A. `newGrade ← (currentGrade + maxGrade)/2`

B. `newGrade ← currentGrade/maxGrade`

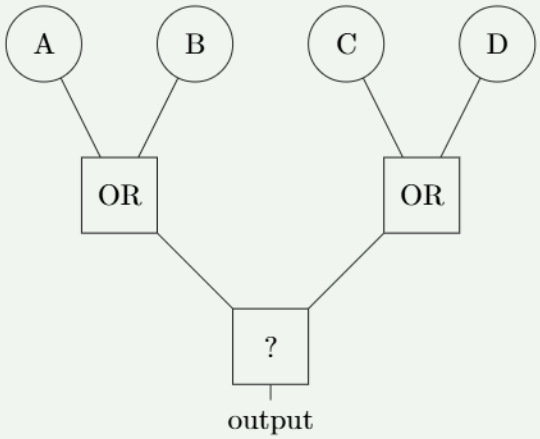
C. `newGrade ← (currentGrade - maxGrade)/2`

D. `newGrade ← (maxGrade - currentGrade)/maxGrade`

Solution:
If all the other student's grade's are out of the maximum grade, that means they are a percentage of the maximum grade. Therefore, the new grade would be calculated as $\frac{\text{currentGrade}}{\text{maximumGrade}}$. This matches Option B.

Option A would be finding the midpoint between your grade and the maximum grade. Option C would make your grade negative unless it was equal to the maxGrade, in which case your grade would become 0. This is because subtracting a grade less than the max grade from maxGrade will result in a negative, or at best a zero. Option D would calculate how much percent you were away from the maxGrade, then divide that by the maximum grade. This curve also (for the most part) hurts grades and can give a zero if a student's grade is equal to the maximum grade.

Problem 3.13.31 — The diagram below shows an algorithm with three logic gates. Each gate accepts two inputs and generates one output. If when the A = FALSE, B = FALSE, C = TRUE, and D = FALSE, the output is FALSE then which statement must be true?



A. It must be an AND gate because both OR gates have an output of TRUE.

B. It must be an AND gate because only one of the OR gates have an output of TRUE.

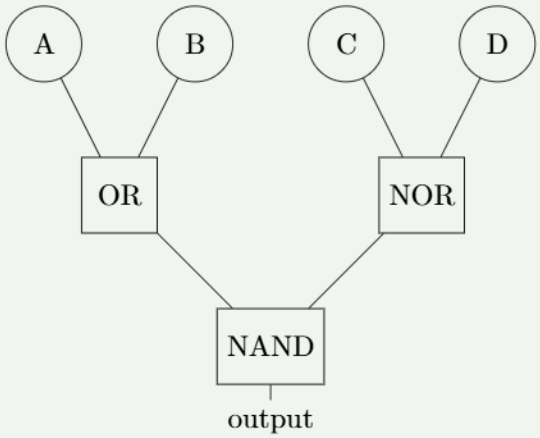
C. It must be an OR gate because both OR gates have an output of TRUE.

D. It must be an OR gate because only one of the OR gates have an output of TRUE.

Solution: An OR gate will have an output of TRUE if one of the two inputs is equal to TRUE. An AND gate will have an output of TRUE only if both inputs are equal to TRUE. In this case, since both A and B are equal to FALSE, one of the OR gates has an output of FALSE. Option A and Option C can be eliminated for this reason.

As stated before, an OR gate will have an output of TRUE if one of the two inputs is equal to TRUE and an AND gate will have an output of TRUE only if both inputs are equal to TRUE. Since one of the OR gates has an output of FALSE, and the overall output is FALSE as well, it must be an AND gate. Option D can be eliminated and Option B is correct.

Problem 3.13.32 — A NOR gate is another type of logic gate that has an output of FALSE when one or both of inputs are TRUE and an output of TRUE in all other cases. A NAND gate is also another type of logic gate that has an output of FALSE when both inputs are TRUE and an output of TRUE in all other cases. Which of the following best explains the output if A = TRUE, B = FALSE, C = TRUE, D = FALSE?



A. The output is TRUE because the both of the gates above have an output of FALSE.

B. The output is TRUE because one of the gates above has an output of FALSE.

C. The output is FALSE because both of the gates above have an output of TRUE.

D. The output is FALSE because one of the gates above has an output of FALSE.

Solution: Before looking closely at the logic gates, Option D can be eliminated as it incorrectly describes how a NAND gate works (If one of the outputs is FALSE, then the NAND gate has an output of TRUE). Then we can go through the logic gates to see that the OR gate has an output of TRUE (One of the inputs is TRUE - A) and the NOR gate has an output of FALSE (One of the inputs is TRUE - C). Option A and Option C can be eliminated since both describe the outputs of the two above gates as the same (either both TRUE or both FALSE).

Since the output of the OR gate is TRUE and the output of the NOR gate is false, the inputs of the NAND gate are TRUE and FALSE meaning that it has an output of TRUE. Option B is the correct answer.

Problem 3.13.33 — A teacher creates two programs to show the percentage of students that scored higher than an 85 on a test. The list `gradeList` is a numerical list of all student grades.

Program I

```
PROCEDURE percentAbove(gradeList)
{
    i ← 1
    count ← 0
    percent ← 0
    listLength ← LENGTH(gradeList)
    REPEAT UNTIL(i > listLength)
    {
        IF(gradeList[i] > 85)
        {
            count ← count + 1
        }
        percent ← (count / listLength) * 100
    }
    RETURN(percent)
}
```

Program II

```
PROCEDURE percentAbove(gradeList)
{
    i ← 1
    count ← 0
    percent ← 0
    listLength ← LENGTH(gradeList)
    REPEAT UNTIL(i > listLength)
    {
        IF(gradeList[i] > 85)
        {
            count ← count + 1
        }
    }
    percent ← (count / listLength) * 100
    RETURN(percent)
}
```

Which of the following most accurately compares program I and II?

A. Program I returns an incorrect value of percent, while program II returns the correct percent

B. Program I continuously updates the value of percent per each count, while program II waits till the entire list is iterated through

C. Program II returns the correct value of percent, while program I returns an incorrect percent

D. Program I runs in an unreasonable amount of time, while program II runs in a reasonable amount of time

Solution:
Both programs correctly return the percent of students that received a score above 85. While program I does update it for each count (thus initially not considering all the students) it only returns percent at the end once the entire list is checked and count is updated for all the students. Therefore, Option A and Option C are incorrect.

Option D is also incorrect because both programs run in a linear amount of time. If the time it took was exponential, then the run time would exponential. Option B is correct because percent is calculated within the REPEAT loop in program I, resulting in percent being re-updated as count is updated each time. However, in program II, percent is only updated after the REPEAT loop is complete, and all the students who got above an 85 have been counted.

Problem 3.13.34 — The code below was written to rearrange the three values in the opposite order using another variable called temp, but does not work as intended. The order is currently: A, B, C.

```
1       temp ← A
2       A ← B
3       C ← temp
4       print(A)
5       print(B)
6       print(C)
```

Which of the following changes to line 2 would make the program run as intended?

A. `A ← D`

B. `temp ← B`

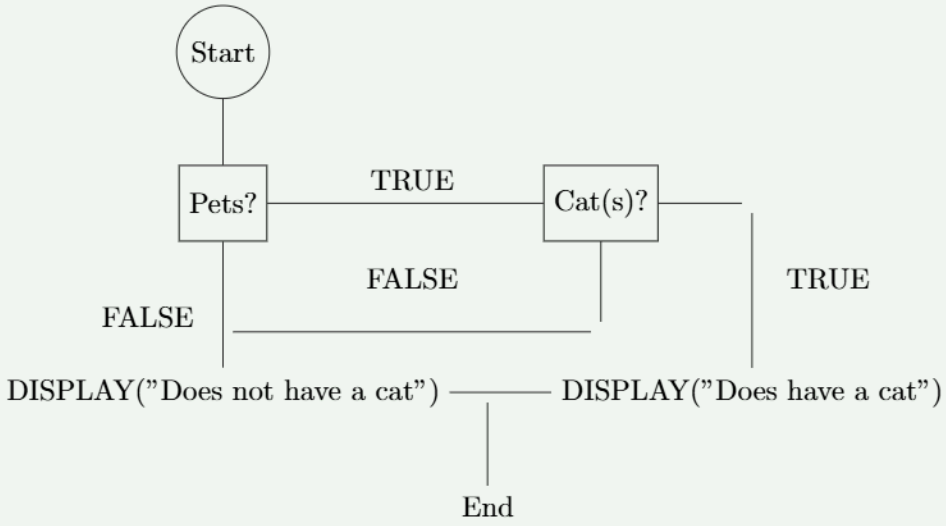
C. `A ← C`

D. `B ← temp`

Solution: As there are 3 values being reversed in order, the center value, B, should remain the same. Option B and Option D are incorrect as they change the value of B. Option A is also incorrect because there is no value D.

Option C is the correct answer because A and C need to swap values. By setting the value of A to C, the 3rd value (C) become the 1st value (A) and in the last line of code the 3rd value (C) is assigned to the first value (A) through the variable temp (which equals A).

Problem 3.13.35 — An algorithm is represented using a flowchart below. A circle represents the start of the algorithm and a square represents a conditional. Which of the following programs could the algorithm be used most effectively for?



A. A program used to find how many pets are in a household.

B. A program used to find how many cats are in a household.

C. A program to find if there are pets in a household.

D. A program to find if there are cats in a household.

Solution: By observing the flowchart above closely, we can see that the algorithm checks if both the conditionals "Pets" AND "Cat(s)?" equal to TRUE to display "Does have a cat" and otherwise displays "Does not have a cat". Since there is no numerical quantity determined by the program, Option A and B can be eliminated (The question how many refers to a quantitative value).

Although the program does check if there are pets, it specifically checks if there is are cats and if there is no cat, the program will not display whether there are other pets or not. Therefore, Option C is incorrect and Option D is correct.

4 Unit 4: The Internet (11-15%)

§4.1 Packets

To send data from one device to another, data is chopped up into sections called packets. Packets are then sent to another device. Packets can be sent and arrive in **any order** and they may **not follow the same path**. Those two pieces of information are crucial to know about packets. Since packets are numbered, they are generally reassembled when all packets arrive at their destination.

The maximum amount of data that can be sent over some time interval through the internet is called the **bandwidth** of that network. Decreasing bandwidth means sending/receiving less data per second, meaning that data will take longer to send or receive.

§4.2 The World Wide Web

The World Wide Web was first created at CERN, in Geneva, Switzerland. It's the reason there's a www. in your URL, and it allows devices on the internet to connect with HTTP (Hyper Text Transfer Protocol).

HTTP, as in the name, is a protocol that standardizes how computers send web pages to each other. Remember the distinction between the World Wide Web and the Internet: the World Wide Web is devices linked to send web pages to each other, and the Internet is the overall, larger system of devices that are connected together and sent each other data.

Note 4.2.1

Remember that standardized protocols are extremely important on the internet and the world wide web. Without them, computers would not know how to read or send data to other computers.

For some insight on how the internet works, each device and website has an IP address, IP standing for Internet Protocol. The reason that the internet is built this way is so that it can **scale as more devices are added** to the internet. In the World Wide Web, when a URL such as youtube.com is typed, the device asks a DNS (Domain Name System) server for the IP address of that website URL. Receiving the IP address allows the device to know where to send data or request data from.

In essence, a DNS server is just a list of website URLs and their corresponding IP addresses. Numerous DNS servers exist, so that even if one collapses, others can respond. After the DNS server sends an IP address, the device knows which other devices to communicate with to receive the web page information.

§4.3 Fault Tolerance

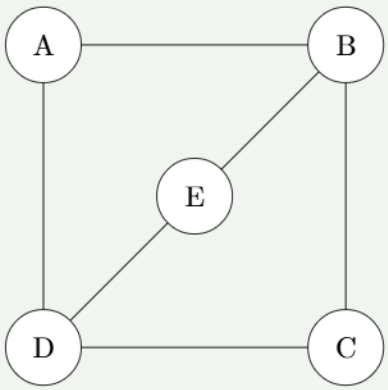
Previously, it was explained that even if one DNS server collapsed, there are still others that can translate domain names (the beginning part of a URL such as website.com/...)

into IP addresses. This example of **redundancy** through multiple DNS servers creates something called fault tolerance.

Fault tolerance is when, even if one component of a system fails, the system can still run as a whole. This applies to the internet as well. Just because one device fails in the internet, doesn't mean all of them will. One DNS server shutting down won't cause the entirety of the internet to fail.

Sometimes, fault tolerance problems show graphs with lines connected between circles to represent pathways between devices.

Problem 4.3.1 — The image below shows lines that are connections between devices labeled A through E.



Removing which of the following devices would result in B no longer being able to communicate with D?

A. Removing devices A and E.

B. Removing devices A and C.

C. Removing devices C and E.

D. Removing devices A, C, and E.

Solution: The answer to the problem is D. Removing devices A, C, and E. These problems may have curved lines instead of straight or possibly many devices, but they should be more time-consuming rather than difficult.

§4.4 Parallel Processing/Distributed Computing

Parallel processing and distributed computing are similar, but with slight differences. Parallel processing refers to using multiple processors of an individual computer to break down a task into smaller tasks to be solved by each processor. Distributed computing refers to breaking down a problem for multiple different computers to solve.

Note 4.4.1

Eventually, distributed computing and parallel processing are not as useful the more processors or computers are used. The benefits in speed up time begin to slowly decrease, somewhat like a logarithmic graph.

To calculate the speed up of parallel processing/distributed computing, we can take the final time, and divide by the original time. For example, assume before parallel processing was used, a task took 60 seconds. Afterward, the task only takes 20 seconds. The speed up is therefore 3, $\frac{\text{final time}}{\text{original time}} \to \frac{60}{20} \to 3$.

Problem 4.4.2 — Using distributed computing, a programmer is attempting to decrease the amount of steps it takes to compute a program. The table below shows processes that are sub-tasks of the programs. The programmer has two computers, computer 1 and computer 2, available.

Process	Time Complexity
Process A	n^2
Process B	n
Process C	2^n
Process D	1

Which two processes should be assigned to each computer in order to minimize the amount of steps it takes for all four of the listed processes to run?

A. Processes B and C on computer 1, and processes A and D on computer 2.

B. Processes B and D on computer 1, and processes A and C on computer 2.

C. Processes A and D on computer 1, and processes B and C on computer 2.

D. Processes A and B on computer 1, and processes C and D on computer 2.

Solution:
This question combines algorithm efficiency and distributed computing. To minimize the amount of steps that it takes for the processes to run, each computer should be given roughly an equal workload. Process D occurs in 1 step, so it should be paired with process C which is exponential and runs in an unreasonable amount of time. Consequently, process A should be paired with B because both run in roughly polynomial time. Though it may not be exactly equivalent to the workload of computer 2 running an exponential time complexity, it is the most balanced distribution. This makes option D the answer.

§4.5 Crowd Sourcing/Citizen Science

Crowd-sourcing is defined as gathering large amounts of data via the internet. An example may include a YouTuber posting a poll asking which thumbnail is better, or a game that has a bug submission forum to fix any issues that the game has.

Citizen science is pretty similar to crowd-sourcing, except the focus is on the general public volunteering to submit data that may help scientists in research. For example,

public volunteers submit photos of butterflies in the local area so that scientists can categorize the species of butterflies living in that locality.

§4.6 End of Chapter Questions

Problem 4.6.1 — In which of the following situations is using parallel processing likely most disadvantageous?

A. When the task has many trivial sub-tasks that need to be completed

B. When a task of the program requires information from the previous sub-task

C. When the program takes an unreasonable time to run on a single processor

D. When a program has been developed with another creator

Solution:
Parallel processing would be the most disadvantageous when a program requires information from a previous step because all the processors running the program are acting independently and cannot obtain the data needed, thus not allowing them to complete the task. This makes Option B the correct choice. Option A would be an optimal opportunity for parallel processing since there are numerous sub-tasks to complete, and they are trivial to complete. Option C is also a good time since a program that takes too long to run on a single processor can have its workload divided to other processors. Option D has no effect on whether parallel processing will be advantageous or disadvantageous.

Problem 4.6.2 — Which of the following best describes how data is sent throughout the internet?

A. Data is sent in a digital form, and is transmitted all at once instantaneously

B. Data is cut into segments called packets that do not always arrive in the order they are sent

C. Data streams are sent to a central computer of the internet which then redirects the data by checking the IP address

D. Data is transmitted through the shortest path possible while making sure that lossy compression does not occur.

Solution:
Data is sent through the internet as packets, which do not arrive in the order that they are sent, and may not even take the same path. Option A is initially right, in the fact that data is sent digitally, but it is not transmitted all at once instantaneously. Option C is incorrect since there are no "central computers" of the internet. Option D is also wrong because data does not always follow the shortest path possible, as packets may go through inefficient, redundant paths.

Problem 4.6.3 — A person goes to their friend's house and finds that their friend's network has a much higher bandwidth than his own. Which of the following is most likely the effect of going to his friend's house?

A. Less packets need to be sent at the friends house to send messages

B. The amount of time it takes for a game to download will increase

C. The higher bandwidth may lead to internet security issues

D. Watching videos on the internet will be less likely to buffer and appear pixelated.

Solution:
Increasing bandwidth means that more data can be sent and received per second through the internet, since bandwidth measures the max amount of data capable of being sent through the internet over some time frame. Thus, watching videos on the internet will not pause to buffer as often and since more data can be sent, the video will be higher quality and likely appear less pixelated. Therefore, Option D is the correct choice.

Option A isn't right because the same amount of packets will still be sent, they will just be sent quicker. Option B is incorrect because the amount of time to download a game will decrease since more data can be received quicker. Option C is incorrect because there is no direct correlation between higher bandwidth and internet security issues.

Problem 4.6.4 — Which of the following best describes the design intent of the internet?

A. The internet uses standard protocols such as IP addresses in order to ensure that the internet is scalable.

B. The internet is intentionally designed to lack fault tolerance to minimize the route packets must take to arrive at their destination.

C. In the internet, packets arrive exactly by the order they were sent to make sure they can be reassembled.

D. The internet has no redundancy such as multiple DNS servers.

Solution:
The internet was designed to be scalable, meaning that more devices could be added easily. This is one of the main reasons why IP addresses and standard protocols, even on the World Wide Web such as HTTP, exist. Thus, A is the correct answer.

The internet is not designed to lack fault tolerance as that means that if one of the devices on the internet shut down, every device would, which is not the case. Thus, Option B is incorrect. Option C is also incorrect as packets can arrive at any order or go through any path and be reordered upon arrival. Option D is incorrect because the internet has a lot of redundancy such as numerous paths that packets can be sent to, and definitely has multiple DNS servers.

Problem 4.6.5 — Which of the following best describes a use of large scale crowd-sourcing?

A. The redundancy of DNS servers that are interconnected with other devices throughout the internet.

B. An application that has user ratings on shows as well as user written reviews.

C. Using parallel processing so that multiple processors of a computer can compute a complex problem faster.

D. Packets traveling through sometimes non-optimal paths.

Solution:
Crowd-sourcing is the use of gathering large amounts of information from the internet. In this case, Option B, an application that has user ratings/reviews on shows, would be correct since numerous users are submitting data (ratings and written reviews) via the internet. Option A is incorrect because the redundancy of DNS servers has nothing to do with crowd-sourcing. Redundancy is related to fault tolerance.

Option C is incorrect because parallel processing relies on processors of a single computer to split up a task to work on it faster. It doesn't take data from the internet and uses its own local processors. Distributed computing, on the other hand, uses multiple computers, which can thus lead to using the internet and be a form of crowd-sourcing.

Finally, Option D is incorrect because packets traveling through non-optimal paths isn't related to crowd-sourcing.

Problem 4.6.6 — Which of the following is false about the Internet?

A. An Internet protocol (IP) address is assigned to every device connected to the internet.

B. The Internet is an application that runs on the World Wide Web.

C. The Internet was made with the intent of being scalable.

D. Because of the digital divide, internet is not accessible to everyone around the world.

Solution:
Option A is true because every device connected to the internet must have an IP Address. Option C is true because the internet was made to be able to accommodate for more devices, meaning it is scalable. Option D is true because the digital divide is defined by the concept that some populations do not have access to computer technologies.

Option B is correct because the statement is false. The World Wide Web is an application that runs on the Internet, not the other way around.

Problem 4.6.7 — Which of the following is the best situation to use citizen science?

A. When a complex program could have its computation time decreased considerably by having additional processors work on the sub tasks of the program

B. When specific advanced scientific data needs to be collected to solve a sophisticated issue

C. When water-quality scientists needs data on whether or not public water fountains are functional in a specific area.

D. When a program is considered an undecidable problem.

Solution:
The best situation to use citizen science is when a task requires a large amount of data that can be gathered by the contributions of regular everyday citizens. Following this definition, option B can be eliminated since most citizens often are not capable of producing advanced scientific data to be sent to scientists. Such a task is not openly accessible to everybody.

Option A can be eliminated because in the scenario, parallel processing is needed, and not citizen science. Citizen science has no effect on a computer that requires additional processors to decrease computation time. Option D can also be eliminated since citizen science has no correlation to undecidable problems, and no amount of citizen science can solve an undecidable problem. Option C on the other hand, is a simple check by citizens to see if a water fountain is working, allowing easy contributions that result in large amounts of helpful data for scientists concerned with water quality. Therefore, Option C is the correct answer.

Problem 4.6.8 — A researcher wants to create a dataset that gives the cell viability of 100 different cell lines when exposed specific concentrations of a drug. Why should citizen science not be used in this instance?

A. It is unlikely that the public would be interested in partaking in the project.

B. The data cannot be accurately determined by the general public due to requiring advanced lab protocols.

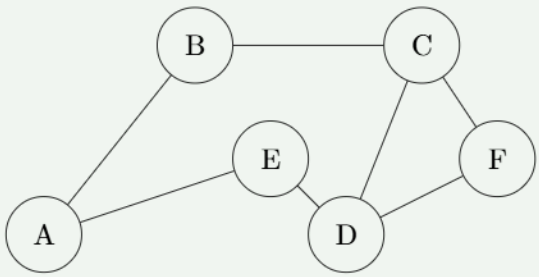
C. It should not be used because all the cell lines must be tested within the same week.

D. Citizen science should actually be used because will make testing 100 different cell lines easier and faster.

Solution: Citizen science is generally used when data needed for a project can be easily obtained by regular people who volunteer to take part in it. Generally, it is very unlikely for a project of such a small sample size (only requires 100 people in this case) to face a shortage of volunteers and so Option A is incorrect. Option C also is incorrect because the question never states that all of the lines have to be tested in a week.

Option B is correct because calculating cell viability requires lab procedures that are not accessible to the public or realistic for untrained individuals to perform. Since citizen science is unsuitable for this project, Option D is incorrect.

Problem 4.6.9 — Which of the following connections have to be cut off to prevent devices E and F from communicating?



A. The connection from E to D and D to C

B. The connection from A to E and B to C

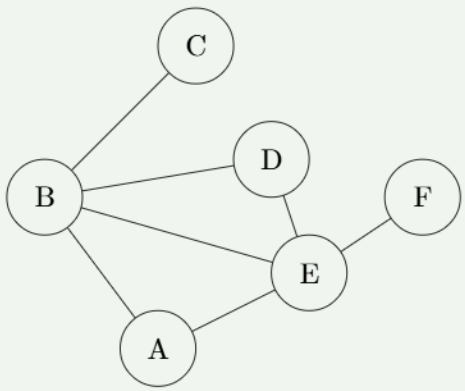
C. The connection from C to F and B to A

D. The connection from C to F and D to E

Solution:
Option A is incorrect because removing E to D and D to C still allows E to go through A, B, C, and then to F. Option B is incorrect because removing A to E and B to C still allows the device E to communicate from E to D and then D to F.

Option C is incorrect because removing the connection from C to F and B to A will still allow the device to communicate from E to D and then D to F. Lastly, option D is correct because there are no longer any connections to F, so it is impossible for device E to reach device F.

Problem 4.6.10 — What is the minimum number of connections that must be removed to prevent communication between Devices A and D?



A. 1

B. 2

C. 3

D. 4

Solution: Option A is incorrect because if one connection is removed, there is still one other path that can be taken to connect Device A and Device D. Option C and Option D are incorrect because, although they do prevent communication between the devices, they do not represent the minimum number of connections that must be removed.

Option C is correct because removing the connection between Device A and Device B and the connection between Device A and Device E are the only two connections that must be removed in order to prevent communication between Devices A and D.

Problem 4.6.11 — Which of the following best emphasizes the fault tolerant design of the internet?

A. The tendency of packets to take routes that may not be optimal throughout the internet

B. Encryption techniques such as public key encryption used to protect against malicious users

C. The bandwidth limit of devices on the internet

D. The centralized design of the internet focusing on a few main devices that regulate traffic

Solution:
Fault tolerance means that even if one component of the internet stops working, the internet as a whole will work correctly. This fits Option A since even if some routes are blocked for packets, they can still use other non-optimal routes through the internet. Option B is incorrect because encryption might make parts of the internet more secure but doesn't have anything to do with fault tolerance.

Option C is incorrect because bandwidth only has to do with the maximum amount of data that can be sent within a time frame, and not with fault tolerance. Option D is incorrect because it is the opposite of fault tolerance. If the internet was centralized and a few devices controlled internet traffic, those devices malfunctioning would result in the entire internet collapsing.

Problem 4.6.12 — Which of the following best describes the relation between the World Wide Web and the Internet?

A. The World Wide Web and the Internet are two completely separate entities.

B. The World Wide Web has fault tolerance built into it via redundancy while the Internet is not fault tolerant

C. The Internet is the larger over-arching connection between millions of devices, under which the World Wide Web operates

D. The Internet makes usage of packets whereas the World Wide Web does not

Solution:
The Internet describes the global connection of devices and the World Wide Web describes the system of linked URLs and pages translated into IP addresses by DNS servers. This

means that the World Wide Web is merely a part of the overall internet, making option C correct. Option A is incorrect because since the World Wide Web is part of the internet, they are not separate.

Option B is incorrect because both have fault tolerance, and if the World Wide Web has fault tolerance, consequently the Internet does as well. Option D is incorrect because both use packets to send data.

Problem 4.6.13 — Which of the following best explains a possible limitation of parallel processing?

A. Parallel processing may speed up the computation of a task that has trivial parts

B. Using too many processors may eventually result in diminishing returns in computation time of a task

C. Parallel processing requires high tech computers that are not readily commerically available

D. Parallel processing can decrease the stress on a single processor for computation

Solution:
Option A, speeding up the computation of a task, is a benefit of parallel processing and not a limitation. Option A is incorrect. Option B, using many processors may result in diminishing returns is correct because the more you split up a task into tinier and tinier tasks, the increase in speedup given by parallel processing becomes exponentially less.

5 Unit 5: Impacts of Computing (21-26%)

§5.1 Digital Divide

The digital divide frequently shows up on the less programming related questions. The digital divide is the concept that in the world, some groups and communities do not have access to the internet, technology, or technical training. This puts them at major disadvantage compared to those who do have those luxuries.

The AP exam may ask you what can worsen the digital divide as a question, like this:

Problem 5.1.1 — The local government of a city is planning on changing the voting process. Which of the following changes may widen the digital divide?

A. Mailing homes physical forms to fill out for elections

B. Requiring that all voters fill out details online

C. Offering access to spaces with computers for online voting

D. Allowing only black ink to be used to fill out ballots

Solution:
The answer is option B. Requiring all voters to fill out details online means that those who don't have access to the internet will be greatly hindered in attempting to vote.

These problems are not bad too often, as long as you think from the perspective of someone who does not have technology.

§5.2 Algorithmic Bias

Some programs take in data input, and learn from analyzing that data. Some programs that do this, which you may be familiar with, are ChatGPT and Google Translate.

These programs that continuously learn from data are considered machine learning programs. For such programs, a diverse range of training data is needed, which the program learns from to more accurately perform its intended purpose. In fact, this video shows this process works with Google Translate.

Going back to the facial recognition idea, some face detection programs trained with a very specific group of faces only recognize that ethnicity/race, but not others. This could end up with a racial bias formed by the algorithm due to the limited training dataset.

§5.3 Legal Considerations

When creating programs, some people may not want the program to be used or distributed to others. To do such a thing, the creator would use an **all rights reserved copyright** license.

However, if a creator wants their work to be distributed or used in a certain way, they can modify what permissions over their work they want to be given in more detail using a **creative commons license**. It's like a some rights reserved, instead of an all rights reserved.

Lastly, **fair use** is using a portion of a copyrighted work for educational or non-profit goals.

Two other terms used on the AP exam also include **open source** and **open access**. Open source refers to software/code that is available to anyone on the internet to read and review. Open access refers to online research, data, etc. that is available to anyone on the internet to use.

§5.4 Cyber Security

Typically, there are one to two questions about cyber security on the AP exam. The main things you need to know for cyber security are:

1. Keylogging - A malicious program that tracks the keystrokes inputted on a computer
2. Phising Attack - Pretending to be another person or authority in order to ask for private information, usually via email/messaging
3. Rogue Access Point - A wireless access point used to steal information from a network
4. DDoS (Distributed Denial of Service) attack - an attack in which a server is interrupted through flooding of false internet traffic preventing user access.

These attacks are the main ones that will be asked about in cyber security questions on the exam.

§5.5 Cryptography/Encryption

Cryptography is the method of encrypting data so that it cannot be read without knowing how to decrypt it. A famous way of disguising data is a Caesar cipher. In a Caesar cipher, each letter of a message is shifted some spaces in its alphabetical order to become another letter. For example, a Caesar cipher of the word "hello", with a key of 2, replaces each letter of "hello" with the letter that is 2 after it in alphabetical order. So h would j, e would become g, and so forth to create "jgnnq".

Encryption methods often have a **key** that either deciphers or encrypts the message. In a Caesar cipher, it was a number that indicated how many places to the right the letters should be shifted. Using that same key, but subtracting the number (going to the left) would decrypt the message. This is an example of **symmetric encryption.** Encryption and decryption use the same key, making it symmetric.

Another type of encryption is **public key encryption**. Public key encryption has a public key that everyone knows and can use to encrypt a message, but a private key to decrypt the message only known by the decrypter. For example, consider a mail box that has a lock on it so that nobody can access the contents, but a thin slit for papers to pass through so that the mailman or others can deliver letters. This analogy is similar to public key encryption since everyone has the public key (the ability to put papers through the thin slit), but only one person has the private key to decrypt the message (the key to the lock to open the mailbox and see the letters inside).

Note 5.5.1

Public key encryption is also known as asymmetric key encryption. A common example, that rarely shows up on the exam, is the Diffie-Hellman key exchange in which the modulo operator is used.

§5.6 Authentication

Though this is mostly known already, the AP also tests knowledge of two factor authentication. Two factor authentication, also called 2FA for short, provides another additional way of assuring that a person logging into an account is the correct person. It can do this by emailing you a code, messaging you a link, and more.

§5.7 End of Chapter Questions

Problem 5.7.1 — Which of the following most closely matches the description of a phishing attack?

A. A message asking for login info from malicious users claiming to be Windows administrators.

B. Exploiting a wireless access point to engage with data flowing through a network.

C. Software running in the background that tracks all keys pressed by the computer user.

D. Decrypting data containing private information such as passwords or other login info.

Solution:
A phising attack is when an email or message is somehow sent to a user, in which hackers pretend to be an authority in order to claim that they need the login info of your account for some purpose. This matches Option A, as malicious users claim to be Windows administrators to ask for login info. Option B is the description of a rogue access point, and Option C is an example of keylogging. Option D does not have to do with requesting login info or pretending to be some other person in order to hack an account.

Problem 5.7.2 — Daniel has a key for encrypting messages that he sends to his classmates. Despite all of his classmates sharing the key to encrypt the message, only Daniel has the required key to decrypt the message. Which of the following encryption types best match this situation?

A. Symmetric encryption

B. Public key encryption

C. Caesar cipher

D. Prime factorization encryption

Solution:
A key that is publicly available to a selected group for encrypting messages, but a private key that can only decrypt messages is Option B, public key encryption. Symmetric key encryption is when both the encryption key and decryption key are the same, and knowing one means knowing the other. Caesar cipher has to do with shifting letters a certain amount of spaces on the alphabet. Prime factorization is not relevant to the provided description.

Problem 5.7.3 — A creator would like to share a song that he made for others to use, listen to, and edit in any way they'd like as long as he gets credited for creating the song. Which of the following copyright licenses best fits his needs?

A. All rights reserved

B. No rights reserved

C. Creative commons license

D. Enabling the work to be public domain for anybody to use

Solution:
In order to modify the copyright restrictions in such a way that the creator's song can be shared and edited as long as credit is given matches Option C, a creative commons license. Option A, all rights reserved means that others could not use his song in any form. Option B, no rights reserved, means that other potential users of the song are not obligated to give him credit. Finally, Option D, making the work public domain means that anybody can use it, and there are no copyright restrictions such as crediting the creator.

Problem 5.7.4 — A programmer wants to create a program that can identify flowers based on an image file. Which of the following actions should be taken to ensure that algorithmic bias can be reduced in the training model?

A. Implementing proper documentation on the program and flower recognition process

B. Gathering a specific subset of flowers to help the program first be accustomed to a general shape of a flower

C. Using a diverse set of flower images from different regions

D. Fixing bugs in the program with other collaborators

Solution:
Algorithmic bias is created from a biased data training set. Thus, Option A is incorrect because implementing proper documentation may help other programmers understand how the flower recognition program works, but won't directly reduce algorithmic bias. Option B would likely increase algorithmic bias, rather than reduce it, since the training model would only be used to a specific set of flowers, and thus only recognize those flowers.

Option C, using a diverse set of flower images, is the best option to reduce algorithmic bias. By incorporating a wide range of possible shapes and colors of flowers, the program is less likely to be biased to one specific type of flower, thus reducing algorithmic bias. Option D is incorrect because fixing bugs, even with other collaborators, has no effect on algorithmic bias of a training model. Instead, the data used to train the model must be changed.

Problem 5.7.5 — Which of the following best explain a potential issue with using a commonly guessed password?

A. Malicious individuals may be able to undo encryption used to protect the password with a decryption algorithm, allowing them to steal private information from the user.

B. Malicious individuals may use a program that has a large dataset of potential passwords in order to brute force access to the user's account through login attempts

C. Malicious individuals may send phishing attacks to the user, which results in the leaking of private information.

D. Malicious individuals could possibly use a rogue access point to find the password.

Solution:
The problem with a commonly used password is that others hacking an account may attempt that password and gain access to the account. This fits Option B, as it describes malicious individuals using a large collection of passwords and spamming login attempts with different passwords to gain access to the account. With a secure, random password, such an issue is much less likely to occur.

Option A is incorrect because undoing encryption with a decryption algorithm can be done to any password, and is not the result of using a commonly guessed password. The same logic applies to Option C and Option D. In Option C, a phishing attack could be sent to any user, regardless of whether the password is common or not. In Option D, a rogue access point attack to find a password is also an issue regardless of password strength or rarity.

Problem 5.7.6 — A bank wants to store information securely and decides to use symmetric encryption to encrypt private information of each customer. Would this be a secure system?

A. Symmetric encryption does not allow for a secure system because it allows people to guess the key to decode the credit card information.

B. Symmetric encryption does not allow for a secure system because it can still be solved by a brute force attack.

C. Symmetric encryption allows for a secure system it is still very difficult to solve even by powerful computing devices.

D. Symmetric encryption allows for a secure system because the credit card information is stored in a data table.

Solution: Symmetric encryption is an encryption method in which the same key is used to both encrypt and decrypt the information. Option A and Option B are incorrect because it is very difficult to guess an encryption key or use brute force to find it. Option D is also incorrect because credit card information being stored in a data table does not affect its ability to be decrypted or increase security.

Option C is correct because symmetric encryption, in most cases, cannot be decrypted even by very powerful computing devices due to the length of the keys used in symmetric encryption.

Problem 5.7.7 — A teacher records data about exams, specifically:

- The exam date (day/month/year)
- The average class grade for an exam
- The exam subject

The exams are divided into two categories:

- Humanities (Social Studies, English, etc.)
- Sciences (Math, Chemistry, Physics, etc.)

Which of the following questions CANNOT be answered through only using the recorded data?

A. Has the average exam grade increased as the end of the school year approached?

B. Are class scores typically better on Mondays or Fridays?

C. Did students typically perform better in terms of average class grade by taking an exam in the morning versus the afternoon?

D. Do humanities or sciences typically have better average class grades?

Solution:
Option A, has the average exam grade increased as the end of the school year approached, is answerable through the recorded data. Regardless of when the end of the school year is, the teacher records the dates of all the exams including the day, month, and year, which thus allows the teacher to compare it to average exam grades to look for a connection. Option B is also answerable since the precise date is recorded, meaning that we know what exams were taken on Monday vs Friday and can thus deduce if scores are better in one or the other.

Option C, cannot be figured out with just the recorded data. Though the date is recorded, the time of the exam is not. Therefore an exam taken in the morning vs an exam taken in the afternoon cannot be distinguished with just the available data. Lastly, Option D is also answerable since the exams are divided into humanities vs sciences and average class grades are recorded.

Problem 5.7.8 — Which of the following would be MOST useful in reducing the digital divide

A. Providing free computing devices in schools.

B. Building more internet towers in urban areas.

C. Increasing the cost of computing devices to cater to high-income individuals.

D. Preventing phishing from occurring through emails.

Solution:
The digital divide can be defined as the idea that some populations do not have access to the internet, technology, or technical training which puts them at major disadvantage

compared to those who have access to those resources. Option B is incorrect because providing more resources to an already developed and highly populated region that likely already has access will not greatly combat the digital divide. Option C is incorrect because increasing the cost of resources to provide more for upper class citizens misses the demographics of populations that are less likely to have access, in turn, not greatly improving the digital divide. Option D is incorrect because preventing negative internet practices does not affect the digital divide.

Option A is correct because providing free computing devices to schools allows a larger population to have access to these computing resources which likely will help reduce the digital divide.

Problem 5.7.9 — Which of the following is most likely to make a person's social media account more secure?

A. Using the same password as other accounts to make the account password easier to remember, and therefore more secure

B. Enabling authentication measures such as receiving an email to verify a login attempt

C. Using high bandwidth internet to decrease the probability of any rogue access points

D. Using a password that is frequently used by others

Solution:
Using authentication makes sure that even if a password is guessed, an additional step of verification by the account owner must happen in order for a hacker to gain access to the account. This makes B the correct answer. Option A is incorrect because if another account with the same password is hacked, a malicious user may try that password on the person's social media account and luckily hack that account as well. Therefore Option A decreases the security of the person's social media account.

Option C is incorrect because higher bandwidth does not decrease the probability of rogue access points. They are two uncorrelated topics. Option D is also incorrect because if your password is frequently used by others, then a malicious user may use a dataset of commonly used passwords and brute force through them to gain access to your account. Thus, Option D does not make a person's social media account more secure and is incorrect.

Problem 5.7.10 — Which of the following describes an attack that would result in a website becoming unusable?

A. Phishing Attack

B. Rogue Access Point Attack

C. Keylogging Attack

D. DDoS (Distributed Denial of Service) Attack

Solution:
Option A is incorrect because phishing attacks are generally used to steal a user's information or download malware on a user's device, rather than preventing them from accessing a webpage. Option B is incorrect because rogue access point attacks also aim to download malware on a user's device through harming a network's security. Option C is incorrect because keylogging attacks are used to steal personal information such as passwords.

When a DDos (Distributed Denial of Service) attack occurs, fake internet traffic inundates the server causing users to lose access and disrupt other operations. Option D is correct users would likely be prevented from using the website as a result of the influx of false internet traffic created by the DDoS attack on the server.

Problem 5.7.11 — Using symmetric key cryptography, a message can be encrypted with a sender's public key. Which of the following is used in symmetric key cryptography to decrypt the message?

A. A private key only held by the receiver and no one else

B. The same key that the sender used

C. A high bandwidth computer that can decrypt the key

D. The use of a wireless access point

Solution:
In symmetric key encryption, the same key used to encrypt the message can be used to decrypt the message. Thus, Option B is correct. Option A is incorrect because in symmetric key cryptography there is no such thing as a private key only held by the receiver. This describes public key cryptography. Option C is incorrect because bandwidth has nothing to do with key decryption. The same applies to Option D.

Problem 5.7.12 — A school has two separate datasets: A general student dataset and an athletes dataset.
The general students dataset contains the following info:

- Student ID number
- First name
- Last name
- DOB
- Grade

The athletes dataset contains the following info:

- Student ID number
- First name
- Last name
- Current sport
- Level (JH/JV/Varsity)
- Medical history (y/n)

The school decides to create another dataset that combines the two former datasets together. Which of the following can be determined from the new combined dataset?

A. The last time a student played a certain sport

B. The amount of 11th graders in Varsity

C. The name of the captain of each sport team

D. The average absences of student athletes in JV

Solution:
The last time a student played a certain sport cannot be determined since the second dataset only contains the current sport a student is playing and not past sports. This makes option A wrong. Option B, the amount of 11th graders in Varsity, correct because the dataset could be filtered to all students in 11th grade, then the level at which they play (JH or JV or Varsity) can be checked to find any matches. Thus, Option B is correct.

Option C is incorrect because there is no way to distinguish if a student is a captain of a sports team. Option D is also incorrect because there is no data concerning absences of a student.

Problem 5.7.13 — A student signs into their school account using two factor authentication. Which of the following is one possibility on how this is done?

A. Using Face ID to sign into the account.

B. Being signed in on other websites after signing into the account through the usage of single sign-on authentication.

C. Entering a one time password sent by SMS after trying to sign in with a username and password.

D. Telling a friend to sign into the school account by sharing the username and passsword.

Solution: Two factor authentication is security measure in which an additional way of verifying a user is utilized. Option A is incorrect because using using Face ID only involves one way of verifying a user. Option B is incorrect for the same reason. Option D is also incorrect because after telling a friend one's log in information, they are again signing in with only one way of verification.

Option C is correct because both a user's username and password as well as an additional code are required to sign in which means there are 2 forms of verification.

Problem 5.7.14 — Which of the following MOST accurately describes a possible harmful effect of computing?

A. Computing innovations leading to advancements in various fields.

B. Forcing programmers to spend more time and effort to ensure that their work will be used for its intended purpose.

C. Algorithmic bias resulting in majority groups being discriminated.

D. Information gathered by targeted advertising being misused by third parties or hackers.

Solution: Option A is incorrect because it describes a beneficial effect of computing. Option B is incorrect because, although programmers must act more responsibly to prevent their computing innovations to be used immorally, it is not a direct harmful effect of computing. Option C is incorrect because algorithmic bias primarily affects minority groups, rather than majority groups.

Option D is correct because personal user information gathered in order to make targeted advertisements successful can be intercepted by hackers or third parties to be used for malicious reasons.

Problem 5.7.15 — A teacher wants to encrypt a message and have students attempt to decrypt it without using a key. Which of the following encryption methods would be the best to use for this scenario?

A. Caesar cipher

B. Symmetric key encryption

C. Public key encryption

D. Vigniére cipher

Solution: In this scenario, we are looking for an encryption method that can be decrypted relatively easily without a key and so the correct option should be the weakest/simplest encryption method. Option B and C can be eliminated because both require at least one private key which will not be given to the students (would have to be guessed) and so it is unrealistic for a student to solve it. The vignieré cipher is a simpler polyalphabetic cipher which would be easier to decode than a symmetric or public key encryption, but is still likely too difficult because the key length is not known so it is hard to guess. Consequently, Option D is incorrect.

Option A is correct. The caesar cipher is a simple substitution cipher in which letters are shifted a set number of places in alphabetical order to become a different letter. Because of the simplistic nature of the cipher, a frequency analysis can be done to find which letter is most common in the message. In the English language, the most commonly occurring letter is e and so the shift can be determined based off of how far away the most commonly occurring letter is from the letter e. Additionally, single letter words such as "I" or "a" could be used verify if the shift is accurate.

Problem 5.7.16 — A student is creating a presentation and wants to use images. Which of the following images are the MOST likely to result in a copyright violation if used?

A. Popular images that are found on many websites.

B. Screenshots from videos found on the internet.

C. Images published with an all rights reserved copyright license.

D. Drawing the images manually.

An image that is likely to result in a copyright violation has a license that prohibits its use and distribution. Option D can be easily eliminated because a student using their own drawn images does not involve a copyright license. Option A is also incorrect because images that are found on many different sites likely have a creative commons license which allows for specific distribution/use. Option B is incorrect because, although it is likely an unethical way of procuring images, the copyright license, if any, of the video may allow for the image.

Option C is correct because an all rights reserved copyright license does not allow for any usage or distribution of the image, which would mean that the usage of the image in the student's presentation is a copyright violation.

6 Extra - Passage/Multi-Select Questions

§6.1 Multi-Select Question Structure

Usually passage/multi-select questions appear towards the end of the AP exam. There are 8 multi-select questions on the exam, typically with two correct options, not more or less. Some multi-select can contain code that is complex and will take time to analyze, so don't worry if you spend more time on these types of questions more than your typical question. Multi-select questions usually have a note saying "Select two answers" to remind you that it is multi-select. There are 4 multi-select questions in the Chapter 3 End of Chapter Questions section to try out: 1, 2, 3, 4.

Problem 6.1.1 — The following program is meant to count the number of perfect squares that are divisible by 2 in a list of numbers called `numbers`. The program uses procedure `isPerfect(n)` to check if a number is a perfect square. Procedure `isPerfect(n)` returns `true` if a number `n` is a perfect square, and false otherwise. After counting the number of perfect squares divisible by 2, the program is meant to display the total.

The program does not work correctly.

```
1       FOR EACH n IN numbers
2       {
3           count ← 1
4           IF(n MOD 2 = 0 OR isPerfect(n))
5           {
6               count ← count + 1
7           }
8       }
9       DISPLAY(count)
```

Which two lines of code should be changed so that the program will operate as intended? Select two answers.

A. Move line 9 within the FOR EACH loop

B. Move line 3 outside of the FOR EACH loop, placing it directly before line 1

C. Change the OR boolean operator in line 4 to AND

D. Remove line 6

Solution:
In order to make the program run as intended, the count variable has to be declared before the for each loop otherwise the count will be set to 0 on each iteration of the loop. In addition, to check if a number is a perfect square and divisible by 2, the boolean expression must use an AND operator instead of OR. Thus option B and option C are correct.

Option A doesn't help fix the program errors and instead the count will be displayed on every iteration of the for loop. Option D would result in the count not being incremented by 1, which would make a counter that doesn't work.

§6.2 Passage Questions

There are 5 questions that are single select about some reading passage, often related to a new app or innovation.

Structurally, passage questions contain long detailed paragraphs describing a made up app or innovation. Fortunately, they are more often analyzing the paragraph than having to logically deduce code. The answer to the question is practically always in the text, so don't rush to complete the question and make sure to read thoroughly. Paragraph form questions are meant to be more time consuming.

Problem 6.2.1 — -

PathTracker is a GPS (global positioning system) app that allows cars to navigate to a location by generating suggested routes. After downloading PathTracker, users can create a profile, username, and friends list. The application generates a list of possible friends to send a friend request to by taking the phone number contacts list of the user's device and checking if any of the phone numbers in that list are linked to an account in PathTracker.

When creating a path, users input the end location that they want to reach. The application uses the GPS to find the speed, distance, duration of the journey, and best paths. The application then suggests several routes, and the user picks a preferred suggested route. The app also displays other users of PathTracker on the GPS map.

Downloading PathTracker is free, but premium accounts can be bought for a monthly fee. On non-premium accounts advertisements are randomly displayed with no personalization. Purchasing a premium account will remove any advertisements.

Which of the following is needed in order to find a path for the user?

A. A premium account that pays the monthly fee

B. Username

C. Geographic position of user/destination

D. The most commonly run path by the user.

Problem 6.2.2 — - Which of the following is most likely to be a data privacy concern for a PathTracker user?

A. The application displays the speed of the vehicle

B. The application takes phone number contact information from a user's device

C. The application requires that you take the device the GPS is running on with you during the travel

D. The application has a premium version costing a monthly fee, otherwise displaying advertisements.

Which of the following is most likely to be a benefit of the application for users?

A. Displaying other PathTracker users on the map can help the user identify other users in their location

B. Allowing users to more easily navigate to desired locations

C. Advertisements will frequently display products that the user is interested in

D. A premium version of PathTracker is available

Which of the following groups of users may be hindered the most in using the application?

A. Citizens in rural areas with a lack of technology due to the digital divide

B. Non-premium users of the application

C. Users of the app who intend to navigate to a place without a car

D. Users who want to find the time it will take to get somewhere via car

If a user wants to start at some location and go somewhere, which of the following pieces of data are required by app but not taken from the user's device?

A. The user's geographic position

B. The destination desired by the user

C. The location of other PathTracker users

D. The path generated by the program through using an algorithm

Solution:
Q1. To find a path for the user, the location of the user must be known otherwise the program will not know the starting area of the journey. Therefore, Option C is the correct choice. Option A is incorrect because a premium account only removes advertisements and has no effect on finding a path for the user. Option B is incorrect because a username is not required in the process to find a path for the user. Only the location of the user and the location of the destination is required. Option D is incorrect because the most

commonly run path by the user doesn't give any information that can be used by the program to generate a path for where the user wants to go.

Q2. In terms of data privacy, the largest concern would arise in the fact that the application takes the phone number contact info from a device immediately and searches to see if any other users have the same contact info. Thus, Option B would be correct since the device intrusively searches a device for contact information. Option A is incorrect because the application displaying the speed of a vehicle is not a huge privacy concern and it merely displays the speed but has no mention of storing it. Option C is incorrect because taking the device with you isn't revealing any personal data (besides location, but using a GPS necessitates that the user is willingly giving location data, as it is needed to navigate to their destination). Option D would only be correct if the advertisements displayed were personalized, indicating that the app may be tracking the user across websites or other applications. However, the paragraph states that the advertisements are random.

Q3. The most likely benefit of the application, would be to help users figure out the path to a destination, as the purpose of the application is to act as a GPS. Thus, Option B would be correct. Option A is more of a privacy concern and not a benefit since a user's approximate location can be determined by strangers by looking at the PathTracker map. Option C is incorrect because displaying appropriate advertisements for users is unlikely considering that the advertisements are random and not personalized. Option D is not the most likely to be a benefit, especially considering that the premium version of the app may remove advertisements, but costs a monthly fee.

Q4. The group of users that may find the most difficulty in using the application is Option A, citizens in rural areas with a lack of technology due to the digital divide. Citizens in rural areas with a lack of technology may not even be able to download the app, due to not having devices. However, even if they do have devices, it may be difficult as a result of a lack of experience. Option B is incorrect because non-premium users of the app still have all features except not being able to remove advertisements. Option C is incorrect because even without a car, users of the app who intend to navigate to a place can still use the GPS to find an optimal path. Option D is also incorrect because the passage states that the GPS indicates the "duration of the journey", so users can still find the amount of time that will need to be taken.

Q5. The user's geographic position is required, but it is taken from the device, therefore making Option A incorrect. Option B is also wrong because the destination desired by the user is inputted, and hence taken from the device. Option C is not required to generate a path for the user. Option D is required since the program needs to calculate the best path, and it is not taken from the device since it is calculated by the application. Therefore, Option D is the correct answer.

7 Free Response Questions/Create Task (30% of score)

§7.1 Free Response Questions

The FRQ section of the AP exam is rapidly changing. For 2024 and after, the FRQ section will be part of the exam as an additional hour of test-taking. You will have a 10 minute break in between the multiple choice and the FRQ section. For the 2024 AP exam, there were two questions on the FRQ, with question 2 split into question 2a, 2b, and 2c. This format is likely to stay as the new format.

Problem 7.1.1 — (Set 2)

1. Identify the expected group of users of your program. Explain how your program addresses at least one concern or interest of the users you identified.

2. Refer to your Personalized Project Reference when answering this question.

   (a) Consider the first conditional statement included in the Procedure section of your Personalized Project Reference. Describe your conditional statement including its Boolean expression. Describe what the procedure does in general when the Boolean expression of this conditional statement evaluates to `false`.

   (b) Consider the procedure and procedure call identified in parts (i) and (ii) of the Procedure section of your Personalized Project Reference. Describe the outcome that your procedure call is intended to produce. Write a new procedure call with at least one different argument value that will produce the same outcome, if possible, and explain why this procedure call produces the same outcome. If it is not possible to write a new procedure call that produces the same outcome, explain why this is not possible

   (c) Consider the procedure identified in part (i) of the Procedure section of your Personalized Project Reference. Identify the parameter(s) used in this procedure. Explain how your identified parameter(s) use abstraction to manage complexity in your program.

Source: 2024 AP Computer Science Principles Exam, College Board

All the past FRQs can be accessed here, but be wary that the format changed for the 2024 exam, so anything before 2024 is not what you will have on the exam. However, some of the questions such as identifying an expected user of your program have been repeated.

In general, to prepare for the FRQ section, you should know your program and the procedures/lists used in your Project Reference very well. You should practice the past FRQs available (though there are only these two sets from 2024 currently). Make sure that the Create Task and Personal Project Reference follows all the College Board guidelines.

§7.2 Create Task

The Create Task has three parts:

1. Program Code
2. Video Displaying Program
3. Personalized Project Reference

First, the requirements for the Program Code are:

1. A single PDF file containing all the source code (try this website to make PDF)
    a) Text size should be at minimum 10 pt font
    b) Make sure PDF is not blurry
2. **Contains all comments**
    a) Comments must have instruction for program input
    b) Comments must explain program output instructions as well
    c) To acknowledge other authors for parts of the program, you can use comments
3. Contains at least one list or collection datatype
    a) **Must be relevant to the program's purpose** and not just redundantly added
    b) Helps to abstract complexity of program
    c) The abstraction must make the program easier to develop (meaning not using the list would introduce new difficulties), or easier to modify list values.
    d) Other data collection types include arrays, dictionaries, lists, databases, sets, etc.
4. A procedure/function
    a) **Must be relevant to the program's purpose** and not just redundantly added
    b) Defined procedure name, and return type (return type if applicable)
    c) **At least one** parameter
    d) The function must be called in the code (otherwise it wouldn't be used)
5. Sequencing, selection, and iteration
    a) Sequencing just means the program code runs lines of code (as long as your program works, you're fine)
    b) Selection - An if statement is needed, meaning some boolean expression must be evaluated
    c) Iteration - Some loop must be used to iterate through your data collection type (for loop, while loop, etc.)

All of these steps are absolutely crucial, as on the FRQ section, your answers depend on if you completed these steps. The FRQ may ask how your list helps your program or manages complexity, so if you missed the list or it wasn't relevant, then you may lose point on the FRQ. If you didn't include selection or iteration, then you also may lose points on the FRQ. On top of the FRQ points you would lose points just solely on the program code section of the Create Task.

The requirements for the second part of the Create Task, the video displaying your program, is much shorter:

1. Shows program input
2. Shows program functionality
3. Shows program output
4. DO **NOT** put:
    a) Voice narration
    b) Anything revealing your identity/personal info
5. File must be formated as:
    a) .webm, .mp4, .wmv, .avi, or .mov
    b) Must be less than 60 seconds
    c) Must be less than 30 MB of file size
6. According to the College Board, captions are encouraged but not necessary
7. Collaboration with other students is **not** allowed when making this video

Finally, the Personalized Project Reference is the third part of the create task. You will have this reference with you on exam day.

1. Make sure you include **NO comments** on the reference. This is different from the source code of your program that you upload as a PDF in part 1, in which comments are required. In addition, **no student collaboration** is allowed

(i) Procedure Part i

a) Capture a code segment of your entire procedure, defining procedure name and return type (if applicable to programming language)

b) Procedure must have one or more parameters that have an effect on the program (e.g., are not redundantly added)

c) The procedure must include sequencing, selection (if statements), and iteration (looping through data collection types)

(ii) Procedure Part ii

a) Shows where procedure is called in program

(i) List Part i

a) Shows how data has been stored in a list (or other data collection type)

(ii) List Part ii

a) Capture a code segment showing the list being used in code

b) Code segment must show either new data being created from list, accessing multiple elements in the list, or fulfilling program functionality in some way.

2. Code capture format
    a) Can be block code
    b) If text, the capture should be at least 10 point font

For any component of the program not made by yourself, make **100% sure** to credit the proper authors. Crediting authors can be done through comments. This includes media such as images, sounds, as well as program code templates, APIs, databases, etc.

The College Board gives the following examples:

- If the program code has been made available for your use by your teacher, add a comment that states: This code was provided as starter code by my teacher.
- If the program code has been made available through an API or open- source code, add a comment that states: This code was made freely available by [source of code].
- If the program code has been co-created with the assistance of a generative AI tool, add a comment that states: This code was generated using [Generative AI Tool Name]

Giving proper credit is crucial for ethical and academic reasons. The College Board states "A student who commits plagiarism will receive a **score of 0** on the Create performance task, including their responses to the written response prompts on the end-of-course AP Exam." For more information, and the source of this quote, check out page 8 of this PDF. It also contains an explanation of the Create Task.

Here are a few examples of programs that work well for the create task:

- Word games (e.g. wordle, anagrams, etc.) - Using substrings or a large dataset (list) of words usually in a CSV file
- Quiz games - Parallel lists with a global index
- Encryption program (e.g. caesar cipher) - Procedure depends on how text is encrypted. Caesar cipher example is found in Runestone mobileCSP book.