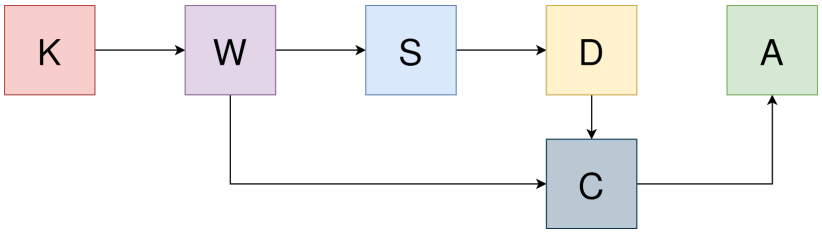
M20HSS316-ITP/Final-Assignment/20171047/CLD

Question 2

Section 2.1

We make the following assumptions about the two structures. In figure (a), the pull in the string causes the door to lift up, thereby activating the device. In figure (b), the tug in the string pulls down the door D which causes the door to get inside a cavity within the cubic table. When the pull disappears, the door comes up to its normal state of equilibrium blocking the path of light.

Part 1



All the variables in the above figure are binary in nature. We introduce three external variables: p, l, and Δ into the scope of the system Z (say). The external variables have the following semantics: p denotes the present time in seconds from absolute zero denoting some time sufficiently back in the past, l denotes the last time when the state of C was changed (on the same time scale as p), and Δ is simply (p - l). We update l as: if C changes state, then set l ← p. Now, we define system Z as Z(K, W, S, D, C, A, p, l, Δ) which is governed by the following structural equations:

1. $W = K$
2. $S = W$
3. $D = S$
4. $C = D \wedge W$ or $D \times W$
5. $A = C$ if $\Delta \geq t$, else $A = 0$

For the purpose of simplicity, we assume that there is no delay in propagation of activities (or that the delay is insignificant) and the switching of states based on equations happens instantaneously.

Part 2

In a case where the wind is blowing strongly as a whole but the alarm is not getting activated, that could mean that the threshold 't' is too high and within that time period, for at least a split second, there is a slow down in the wind speed which cuts off the laser. In order to troubleshoot purely based on the causal model, I would list out every possible legal assignment of values to the tuple (K, W, S, D, C, A, I(Δ))[1] and focus on the cases where $A = 0$. If I find a case of $A = 0$ where we'd expect it to be 1, I would revise the structural equations and update the model. Note that a causal model of

[1] I being an indicator function. $I(x) = 0$ if $x = 0$, else $I(x) = 1$

the above type is purely deterministic and it is an abstraction of the real world. Hence, simply leveraging the causal model in theoretic terms without a scrutiny of the real physical system, we cannot rule out incorrect implementations of the expected causal flow.

Thus, in order to pin-point any possible fault which could possibly occur in the physical system, we might need to look at the actual physical aspects and not just the causal model. For example, the string might be broken or the door might be jammed, or the control switch might be too sensitive. A causal model could be built in theory which would include all such events as nodes, but clearly, the size of such a model would be massive. Our present model sufficiently explains the major components and their interconnections as part of the system. It is not a model for troubleshooting, but one meant to be used as a concept blueprint.

Part 3

In going from implementation (a) to (b), the mechanism of state update of D in the physical system is all that changes. But for the causal model, whether the door is open or closed due to force acting on it via the string is all that matters. The open and closed positions of the door are abstracted out. Hence, the exact same causal model applies to case (b) and as far as troubleshooting based on the causal model is concerned, it is neither easier nor tougher, the problem is exactly the same.

In the physical world itself however, due to more moving parts such as one extra pulley, mechanism for the door to hide inside the table, the state of equilibrium for the door involving a higher gravitational potential state - physical troubleshooting might be more difficult due to the increased complexity.

Part 4

There are various musings on the idea of multiple realization starting from Putnam's original ideas from the 1960s[2] right upto works by modern philosophers. Here, I will consider the most generic definition by Thomas W. Polger: there is a uniformity in the description of two systems at a high level while at the lower level, differences become apparent.[3] In that, the two systems share the same perceptual state but are realized by different physical states. In terms of mathematical functions, a description of what the functions achieve might be the same but their actual operations might differ. For example, consider f_1 and f_2 as: $f_1(x) = x - x$ and $f_2(x) = x * 0$. Both these functions have the exact same mapping: $f_n : x \rightarrow 0 \ \forall \ x \in \mathbb{C}$ where $n \in \{1, 2\}$. Thus the behaviours of two functions or systems might be the same even though their internal physical workings might differ. Using this idea, setup (a) and (b) collectively is indeed a case of multiple realization. The high-level view here is of the unique causal graph while the low-level views involve looking at the physical systems themselves.

Now we analyze the objection. The first point states that the two setups are mere variations of the same thing. The idea of what is 'mere' and what is not isn't well-defined. 'Mere' is highly subjective - what is mere to X might be significant to Y. We'll go by principle of charity and understand that the

[2] Stanford Encyclopedia of Philosophy - Multiple Realizability
[3] What is multiple realization? By Thomas W. Polger

argument author feels the change in manner of door opening is insignificant as compared to the whole physical design. The author makes a distinction between physical states (which are different in kind) and physical ways (which are 'just variations'). Notational changes from σ_1 and σ_2 to σ_1 and σ_1' are used to showcase closeness between variations as opposed to differences between physical states but that's just simple aesthetic tactics. The validity of the overall objection relies on the truth of the following two strong conditions:

1. Minting and acceptance of a strong legal definition of multiple realization which only allows cases where the core idea behind physical contraptions differ at the lower level.
2. The degree to which the realizations must differ in terms of specific pointers and laid out quantifiable metrics is well defined, agreed upon by a standardizing body and acceptable enough to be considered in a court.

However, without all that in place, the argument writer is relying too heavily on the gullibility of readers. The question of 'where do you draw the line?', still remains. Note that the very same idea is exploited in legal proceedings on whether a piece of art, say a song, was original or not and these debates do get intense. Same for deciding what should be the percentage threshold beyond which an assignment submission is flagged for plagiarism. The objection must be framed better. Based on the technical grounds alone, it can be discarded. However, we can have two driving responses to arrest the objection.

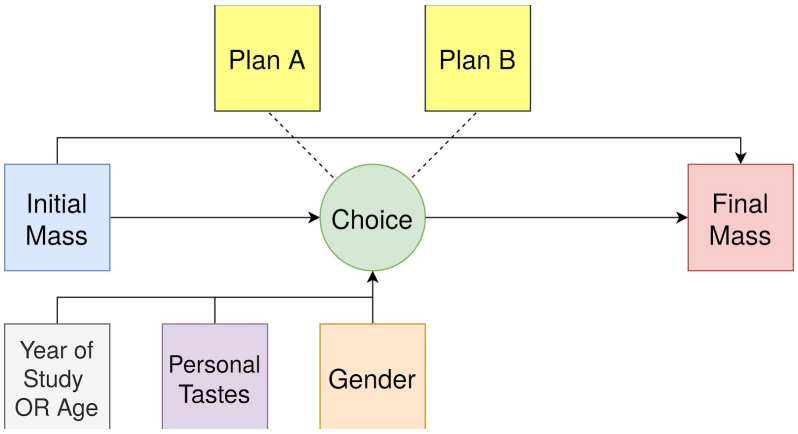
1. The update in door movement changes the manner in which light is allowed through. From the door moving up as open, we now have the door going down as the open state. We have a newer scheme where the door hides in a cavity within the table thereby changing a significant aspect of the realization. As opposed to the door dangling on top with the potential of swing, it is now being neatly tucked below which is a solid design change. We also have a new pulley as a component for the directionality of force application change. Even seemingly small changes within existing baseline systems which betters their performance in a few aspects often warrant a popular publication at a top computer science conference[4] and in that, stands as a new system. Thus, we have significant modifications solving key design issues and newer semantics for up and down door states.

2. We should ideally follow the idea of having a theory which is the most economical. Defining explicit degrees of variation in physical models with their related thresholds and guidelines for interpretation is extremely cumbersome. Finding common ground: at the high-level of abstraction H, the encapsulated physical states are identical. As opposed to going by the objectioners' ideas, we can extend Polger's thoughts to include: a witnessed difference while considering a particular lower level L where we see a modification in the physical implementations entails a case of multiple realization. This is far more economical and much clearer than what we would end up with in going in line with the objectioner's claims.

Section 2.2

[4] See early papers attempting to solve neural question answering.

Part 1



The only factor mentioned in the text which can affect the student's decision is their initial mass. This is because the student might look at the plan details and decide which one would be better for them given how much they weigh at the moment. Apart from this, we can consider a few external factors such as the student's year of study OR age as a measure of maturity, the gender which the student identifies with, and their personal tastes. The choice can take two forms: Plan A or Plan B while the final mass of the student depends on their choice and initial mass. We leave the structural equations unspecified since there is no agreement on the domain of values for each factor. We only consider an edge from X to Y if the value of X directly affects that of Y.

Part 2

Yes. Purely based on average weight gains in plans A and B, one cannot conclude anything about how the plans affect the weights. Just from the causal graph, we see that the final weight is a function of the initial weight and plan choice, say $W_E = f(W_B, \text{Choice})$. Therefore, the change in weight, say $\Delta_W = W_E - W_B = f(W_B, \text{Choice}) - W_B$, is also functionally dependent on W_B. The average value of $\Delta_W = 0$ in both cases might simply be because of the grouping of students into the two classes.

For example, let P: set of students with weights = 80 kg, and Q: set of students with weights = 60kg. Suppose $P \cup Q$ is the entire population and $|P| = |Q|$. If Plan A is a weight loss programme and Plan B is a weight gain programme, and all students in P choose Plan A and all in Q choose Plan B, and at the end, let's say, everyone reaches a weight of 70kg, then the average Δ_W in P is 10kg while that in Q is -10kg. This shows that the individual plans are working with each having required losses and gains. We can abstract out one bit from here, Plan A causes a 10kg weight gain while Plan B causes a 10kg weight loss.

Now suppose we change the groups and have G_1 and G_2 each of which have students from P and Q in a 1:1 ratio. Now the mean change of weights in G_1 and G_2 are both equal to 0. Maybe the professor analyzed a case where the students did not consciously choose the meal plan based on its attributes and their initial mass. In a case of random grouping leading to the same distribution of students in

each group, the average weight changes might each be 0. But this does not say anything about the working of the programmes.

This case is very similar to the classical example of Berkson's paradox. The lesson here is simple, only looking at general trends and measures of central tendency to hypothesize is effectively misuse of statistics. One must look at the root cause factor or the conditioning variable which can explain the why behind the witnessed phenomena.

Question 4

Section 4.1

Part 1

The notion of knowledge is used in this response. Purified data embodies information and distilled information entails knowledge. Thus, the notion that new knowledge is not acquired by Mary can be entertained but with a pinch of salt since the boundaries between data, information and knowledge are subjective and Mary clearly has new data once she leaves the room. The example on Oedipus used to illustrate the claims is, however, wrong. Consider the two facts:

F_1: X is married to Jocasta.
F_2: X is married to his mother.

Now the truth of F_1 and F_2 does not necessarily imply that Jocasta is X's mother as one can be married to multiple people if they are polygamous. We need the extra bit that X is married to only one person to infer that Jocasta is X's mother. Now consider the fact F_3 which is clearly not inferable only from F_1 and F_2. F_3: Jocasta is X's mother. The entire response kind of contradicts itself by saying that at time t, X 'discovers' that Jocasta is his mother. Clearly at a time t' before t, X had no knowledge of this and it is only at t, that he learns about Jocasta being his mother. This in itself is a significant discovery. Getting to know who your mother is qualifies as gained knowledge. Thus, X gains a new fact F_3 and does learn something completely new about the world even if we totally ignore F_2.

Now, from F_1 and F_3, F_2 is inferrable. Here, we can say that F_2 is a new take on the facts already in X's possession. Whether, F_2 qualifies as a new piece of knowledge given X already knew F_1 and F_3 is subjective. But the response still fails because at time t, X acquires F_3 which qualifies as new knowledge thereby changing his mental view of the world.

Part 2

I agree with all statements except the final one. The usage of 'could have' helps clarify the author's stance. Based on a thorough analysis, study, and understanding of colours, when the world you live in is black and white (literally), Mary 'could have' possibly defined a mapping from possible colours to black and white counterparts (I cannot think of a way of doing it myself without having seen and analyzed true colours beforehand but for arguments sake, let us suppose that Mary figured out something of this kind). Then in theory, she could reason out what is red and what is green given she had some information about colour names as well. Also, the propositions: Mary knows the shape of an apple from multiple b&w images, Mary knows that an apple is red, Mary sees an object with the known shape of an apple having the colour C and Mary infers that C is red - forms a valid argument.

Even with all that, the actual visualization of colours would be new for Mary. Hence, the very last statement can be a point of debate where I'll take a stand against it. With all the information from inside the room, with careful study of properties of objects, of different colours, of mappings from

RGB[5] colours to black and white and back, and of colour names; the physical experience 'could' add new information or modify the previous understanding of red which Mary possessed. Imagine the quality of scents and perfumes. You can read the description very well, find out what are the mid notes, what are the base notes, is it fruity and citrousy or is it oriental, what is the strength, etc. And based on this knowledge you can carefully categorize what you are actually smelling into these buckets. However, the experience of inhaling the scent with the actual composition might differ in some respects with the scent you thought of or perceived in your head. Similarly for colours, it is possible to categorize and pinpoint, however, the first-hand experience might be very different. Consider someone color blind who has previously learnt about colours from suitable text-books and inside their inadequate ability to tell apart different colours. Observing such someone put on a pair of color blind lenses shall make this point more believable. From the video at this URL[6], it is clear that even simply expanding the spectrum of colours visible to a person from their previously restricted color-deficient range thereby inducing further granularity, causes them to understand colours in a much more vivid manner. Even though the boy can identify the colours he's seeing, it is still a new experience for him. Now, changing the start point from color-deficient to only sees black and white, the experience of seeing color for the first time would be much more breathtaking even after all the thorough logical reasoning and study.

Section 4.2

With pages and pages of text, the experimenter could potentially encode the pixel information and send it to mary. The resolution of the image would be provided, say 100 x 300. And for the induced pixel matrix, the intensity for every cell which is a value within a specified range, say between 0 and 255 would be provided to Mary. With this, Mary can essentially draw out the image in the same manner as a computer would. Given time, the experimenter can increase the resolution, further increasing the clarity of the image. If we are to assume that Mary has access to a computer (the question only states that there is no internet), she could easily teach herself to program such a text-to-image generator, given time. In the absence of a computer, she could sketch the image on a paper or on the wall. The lesser the amount of resources, the less vivid and realistic the image would get. Last but not the least, she could try to visualize each pixel and draw out the image in her mind (may not be practically possible assuming Mary is human).

If the text does not involve such an encoding and simply makes natural language statements about the image such as 'It appears to be a man with a beard and a moustache', a lot of the visualization is left to Mary's imagination. Again, like a sketch artist, Mary could try to draw but there is no feedback loop involved. The experimenter cannot see what she's drawing and comment on how it needs to change. It is also unclear if Mary can question the experimenter from inside the room. In such a case where an encoding which creates a one-to-one mapping from text input to image is not utilized, the image formed in Mary's head (or physically if she can manage to draw it out) will deviate from what the experimenter intended. This is theoretically sound since the mapping from textual description to the domain of person faces is many-to-many. Thus, in such a case, Mary will learn something new upon seeing the image. The probability of her actually nailing the

[5] Red Green Blue

[6] `https://www.youtube.com/watch?v=j6AqVzmbU-E`

representation in her mind is close to zero. Note that if Mary has seen the image before and the experimenter leverages that as, 'Recall the image of Avro Pärt which you saw in Page X of Book Y', things might be easier for Mary but even then, imagining the same image with cent percent precision will be extremely difficult.

This acquired knowledge upon viewing the actual picture is very different from experiencing colours such as red for the first time. Firstly, seeing colours involve activations of cones in the retina and usage of particular aspects of the visual cortex. In the previous case, a particular set of cells is activated for the very first time in your life which is a new experience at the raw physical level. Secondly, the activity here is more like a guessing game: based on language input, you are trying to draw out a face. No new physical faculties are getting created, trained or evolved due to this. It simply requires usage of existing intellect and world knowledge to try to analyze and visualize the text as nicely as possible.

Section 4.3

Qualia, in essence, is the way in which things seem to us. A rejector of qualia such as Phyllis believes that the focus should be upon the object that we are perceiving rather than mental image formed due to the observation. We have a two-level distinction here. Suppose there is an object O in the real world which we observe to have a mental representation M of O in our minds. A proponent of qualia believes that M has a set of intrinsic features which are specific to M and not derivable purely from O. In that, if we were to infer something about O or answer a question about O, we would refer to M with its intrinsic features. On the other hand, someone like Phyllis believes that when we observe O, we abstract out certain features or properties of O itself which compose our mental model M'. Thus, we do not have any intrinsic properties associated with the created mental representation.

To answer Frank's counters, Phyllis sticks to showing how each feature or attribute of any object in our mind is actually based on a real world feature. An illusion is a case of misrepresentation of an objective feature in the real world or perhaps false categorization and/or mistaken identification of a specific aspect. When we see flashing dots in images, it is a physical phenomena due to our vision apparatus which makes us falsely classify the input in making us see the flashing dots. This is in stark contrast to Frank's view where we first view the image and form a mental representation in which we add the existence of dots flashing.

In the case of hallucination, the properties of the object which we are hallucinating are attributed to objects which may not be in our immediate environment but do in fact belong to external objects. The features of objects which we hallucinate are a result of witnessed external objects and not independent features of the mental realization process. To make it more clear, it is not the activation of rods and cones, passage of signals to visual cortex and further processing of these signals which produce the features of objects which we hallucinate. The features themselves belong to external objects, which are not in the present field of view.

To illustrate this further, can see examples of the strong form of the Sapir-Whorf hypothesis. A person living in the Sahara desert with no contact with the outside world will not have a

hallucination about snow. Similarly, an eskimo won't be hallucinating about sand, the Sentinelese won’t be hallucinating the Burj Khalifa and so on. Thus, it is our world view and shared knowledge which is based on pre-observed objects that compose the sets of things we can and cannot think of. Clearly, there are faculties in our bodies (eyes and brain) which allow us to perceive and form mental representations of the objects we see, but these do not induce intrinsic features into the mental objects themselves. The features are extrinsic and belong to the actual objects.

Bonus

The logical rule of modus tollens is unchallenged in this scenario. The magician's dialogue, although snappy and seemingly logical, is not an illustration of modus tollens. It is mostly the choice of words and the form of sentencing which creates the illusion that it is one particular ball which is being considered throughout but that is not the case in actuality.

In (2), the statement: 'The ball is probably not white' tries to assign a probability value of being a particular colour to a definitely identifiable ball. The statement is rhetorical and what it actually says is: 'A ball selected randomly from the well mixed collection of balls within the non-transparent bag has a lower probability of being white than black.' When we fixate on a particular ball, we know it is of a single colour, i.e either white or black. We may not know which but the statement: 'This particular ball is white w.p 0.4 and black w.p 0.6' is nonsensical[7] (a ball is not an electron which can exist in multiple quantum states with varying probability values).

Same kind of clarification is required for (1), the statement can be expressed as: 'If a randomly selected ball from the well mixed collection within the non-transparent bag is big, then it has a higher probability of being white than black'.

Looking at the revised statements, we see that modus tollens is not directly applicable as at no point of time, do they talk about the colour of any particular ball. The statements do not state simple propositions such as 'This well-identified ruler is 15 centimetres long'. The magician oversimplified the argument rhetoric ('a ball selected randomly from …' to 'the ball') which caused a subtle change in how readers infer the meaning on the very first read. On deeper scrutiny, we see that 'the ball' in question in statements (1) and (2) are not even the same ball. It is simply an abstraction to the idea of a ball in the bag with all of the bag's associated properties and the available statistics on the different kinds of balls.

[7] w.p: with probability