INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD

M20HSS316/Final-Assignment/21-23 November/Points-20

Please read the following instructions carefully

(1) You should submit your answers as a single PDF in Turnitin by 6:30 pm, 22 November 2020.

(2) Use the usual format for identifying your submission by writing *M20HSS316-ITP/Final-Assignment/[your ID]/[your program]* at the top of the first page.

(3) DO NOT include the question in your PDF when you answer a question. Put the question number and start answering.

(4) Your ID should be the name of your PDF. E.g. A student with ID 20151005 should name the PDF "20151005.PDF".

(5) There 4 questions and 2 sections. You have to **attempt any two questions choosing one from each section**.

(6) Each question carries 10 points questions. Points allotted for sub-questions are indicated in square brackets [ ] at the end of each sub-question.

(7) **Begin answering a question (not sub-questions) on a new page**.

(8) The word limit for each answer is 1000.

(9) Write simple sentences, avoid grammatical errors and use spell check.

(10) Answer the question directly. You will not be given points for an introduction or for writing something that you know about the topic but is irrelevant to what is being asked in the question.

(11) Engage with views rather than merely stating claims. It is your reasoning ability that is being evaluated. Take time to explain your answers and most importantly, you should be able to understand what you have written. Be clear and try not to use terms you don't understand.

(12) Your submission will be screened for plagiarism using Turnitin. **If your similarity index is greater than 15% your submission will not be evaluated and you will be assigned zero points**. This rule will be strictly implemented.

(13) **Any doubts about the questions should be posted in the class MS Teams public chat** and I will answer them there.

Section **A**: Causality

Question **1**

Question **2**

Section **B**: Mind

Question **3**

Question **4**

Question 1

1.1 The following is a passage from Aristotle's *Physics* (Book 2, chapter 3):

> Here as elsewhere, we must always seek the most precise cause. A man, for example, is building because he is a builder, and he is a builder insofar as he has the building craft; his building craft, then, is the prior cause, and the same is true in all cases. Further, we must seek genera as causes of genera, and particulars as causes of particulars; a sculptor, for instance, is the cause of a statue, but this sculptor of this statue. And we must seek a potentiality as the cause of a potential effect, and something actualizing a potentiality as the cause of an actual effect. This, then, is an adequate determination of the number of causes, and of the ways in which they are causes.

Consider the following features of a causal relationship: X→Y is stable, X→Y is proportional, and X→Y is specific. *Stability* has to do with whether a causal relationship continues to hold under changes in background conditions. You could think of this as the counterfactual relationship holding under a wide range of relevant conditions. *Proportionality* has to do with whether the changes in the state of the cause "line up" in the right way with the changes in the state of the effect so that no irrelevant details are included. Think about a volume dial and the decibel level – the degree to which the latter changes is proportional to the degree to which the dial is rotated. Finally, *specificity* has to do with the extent to which a causal relation approximates to the ideal of one cause–one effect. A cause is specific when, among a collection of alternative effects (for example, diseases) it contributes to the production of just one of them, and is necessary for that effect. The smallpox virus is the specific cause of the disease smallpox but not of fever or red spots on the skin.

1.1.1 Which of the above three features comes close to Aristotle's idea of 'most precise cause'? Or do you think Aristotle's idea can be cashed out in terms of one of more of these features? Or is it that none of these three features are relevant to what Aristotle is taking about? Discuss by taking examples that Aristotle considers along with any of your own examples. **[3]**

1.1.2 How would you justify the advice that we must seek genera as the causes of genera and particulars as the causes of particulars? Illustrate with examples. **[2]**

1.2 The following is a passage from Hume's *An Enquiry Concerning Human Understanding* (Section VII):

> The only immediate utility of all sciences, is to teach us, how to control and regulate future events by their causes. Our thoughts and enquiries are, therefore, every moment, employed about this relation: Yet so imperfect are the ideas which we form concerning it, that it is impossible to give any just definition of cause, except what is drawn from something extraneous and foreign to it. Similar objects are always conjoined with similar. Of this we have experience. Suitably to this experience, therefore, we may define a cause to be an object, followed by another, and where all the objects, similar to the first, are followed by objects similar to the second. Or in other words, where, if the first object had not been, the second never had existed.

1.2.1 Hume seems to propose the regularity account of causation (analyzing causation in terms of invariable patterns of succession) as well as the counterfactual account of causation (analyzing causation in terms of counterfactual conditionals of the form "If C had not occurred, E would not have occurred"). Do you think either of these two accounts would help science "to teach us, how to control and regulate future events by their causes"? If yes, explain how. If not, explain why and which account is best suited for this purpose. **[2]**

1.2.2 One problem with a regularity account is that we may assign causality in cases of spurious correlation. How would you defend Hume's account in such cases? **[1]**

1.2.3 Another problem with a regularity account is that most causes do not regularly follow their effects. For example, even though we know that smoking causes lung cancer, some smokers do not end up with lung cancer. Do you think Hume should have gone for a probability-based account and said something like 'C is a cause of E if and only if P(E|C) is greater than P(E)' or 'C is a cause of E if and only if P(E|C) is greater than P(E|~C)'? Explain. **[2]**

Question 2

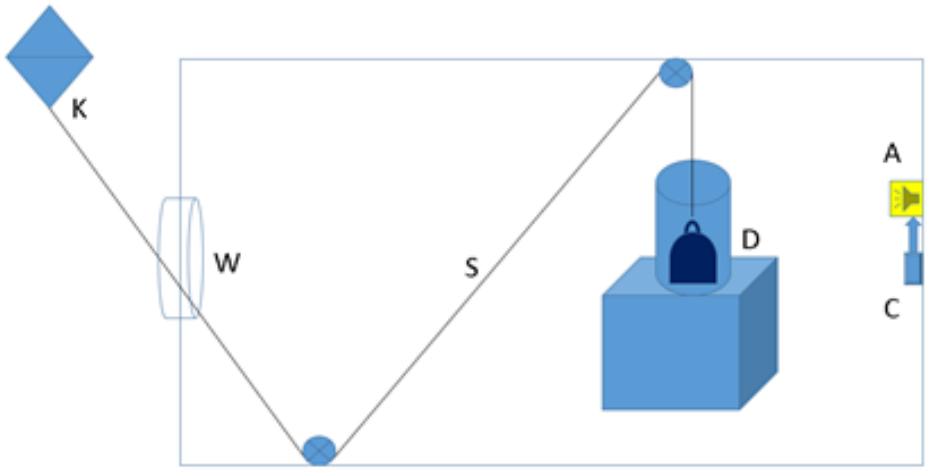


Fig a

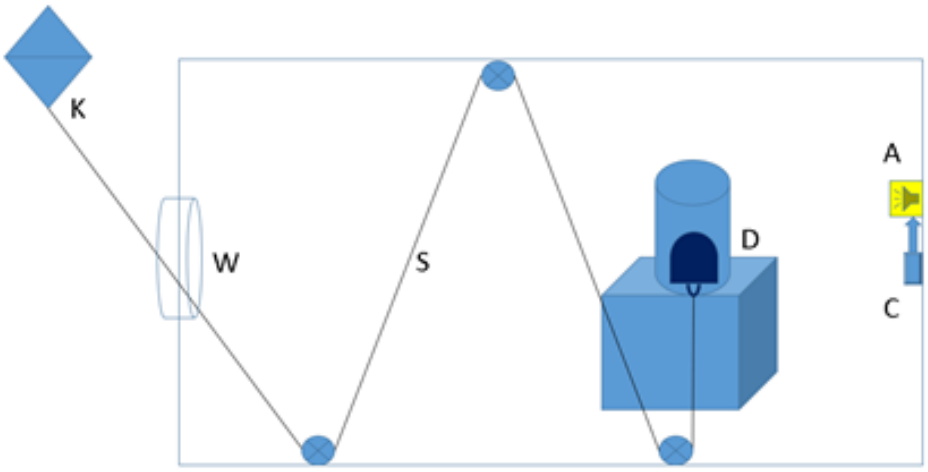


Fig b

2.1 Consider the elaborate wind alarm set-up shown in Fig a. The variables shown are binary and indicate events like 'kite flies (K=1), 'window opens' (W=1), 'door closed (D=0), etc. The set-up is designed to sound an alarm when there is a steady gust of wind in the following way: Flying the kite causes the window to open which allows string S to be pulled which opens door D of the device kept on the table. This device is activated by light from the window and fires a small laser to control switch C which activates the alarm A after a time t which is set by laser intensity. For short winds, the alarm must not sound. When the window is opened by the kite for a certain time t, the laser intensity becomes sufficient to activate C and sound the alarm.

2.1.1 Using only the variables shown in the Fig a, represent the causal mechanism of this wind alarm by a directed acyclic graph. Also write down the structural equations. **[1]**

2.1.2 Suppose there is a fault in the set-up. The wind is blowing strongly but the alarm is not activated. If your causal model is all that is available to you how would you go about troubleshooting the issue? Is the causal model sufficient to pin point any fault that could possibly occur in the system? Explain. **[2]**

2.1.3 Now consider the modified version of the wind alarm set-up shown in Fig b. Would trouble shooting a fault in this version be easier or more difficult than the one in Fig a? Explain. **[1]**

2.1.4 Do you think the wind alarm set-ups shown in Fig a and Fig b are an example of multiple realization? Explain. How would you respond to the following objection: "The set-ups are not giving multiple realizations of the causal model but merely variations in the set-up design. There is a difference between a phenomenon ψ being multiply realized by physical states σ_1 and σ_2 and ψ being realized in multiple ways by σ_1 and σ_1'. σ_1 is different in kind from σ_2 whereas σ_1' is just a variation of σ_1." **[3]**

2.2 At the beginning of the academic year, a IIIT hostel offers its students a choice between two meal plans for the year: Plan A and Plan B. The students' weights are recorded at the beginning and the end of the year. A concerned professor decided to determine how each plan affects students' weight gain and calculated the difference between each student's weight at the end of the year (W_E) and at the beginning of the year (W_B) and found that the average weight gain in each plan was zero.

2.2.1 Draw a causal graph representing the situation. (You could have a variable to represent the students' meal plan choice and make a suitable assumption about what affects this choice in terms of any of the other variables mentioned) **[1]**

2.2.2 Do you think the professor could be wrong in concluding that there was no effect of diet on weight gain? Explain using your causal graph. **[2]**

Question 3

3.1 The cartoon strip below illustrates an objection to Searle's Chinese Room argument. Is it a version of any of the objections we discussed in class (systems reply, robot reply, brain simulator reply and combination reply)? If yes, explain why it captures the intuition of that objection. If No, explain what this new objection brings out that the other objections don't and compose a reply to this objection on behalf of Searle. **[3]**

3.2 To counter Searle, suppose the proponent of strong AI proposes the following theory of meaning (TOM): A subject S, at time t, means *M* by the expression w if and only if S is disposed at t to apply w to all and only *M*s, under suitable situations K. For example, when you use the word 'table' in your conversations with others, you mean *table* (the property of being a table) by 'table' just when you are disposed to apply 'table' to all and only tables (objects that have the property of being a table) under suitable conditions. Take another example: when saying '68 plus 57 is 125', you mean addition by the word 'plus' since you are disposed to assert '125', which is the sum of 68 and 57, when asked 'what is 68 plus 57?'. Now consider a well-defined arithmetic function, *quus*, which is coextensive with the addition function for argument values less than 57. That is, x *quus* y = x plus y when both x and y are less than 57. x plus y performs addition whereas x *quus* y performs 'quaddition'. When x and y are greater than 57, the *quus* function yields 5. According to TOM, for S to mean quaddition by '*quus*' is to be disposed to respond with the quasum of the argument values presented. Suppose a subject, say Mary, uses 'plus' with x_i and y_i values that are less than 57. She calculates z_i from x_i and y_i following a set of arithmetic procedures and asserts "x_i plus y_i is z_i" implying that she meant addition when using the expression 'plus'. But given these values of x_i and y_i, the *quus* function would also map them to z_i and so we have equally good reasons to claim that z_i is the quasum of x_i and y_i. So how do we know that Mary didn't mean quaddtion by 'plus' when asserting 'x_j plus y_j is z_i"? Mary's judgement that '68 plus 57 is 125' is consistent with her possible judgement that '68 plus 57 is 5' and therefore we have no principled way to claim that she is mistaken in the latter case.

3.1.1 Do you think the puzzle of "When uttering claims like '2 plus 2 is 4' does Mary mean addition or quaddition?" demonstrates the intuition that Searle has with his Chinese Room argument or do you think it is attacking strong AI's TOM from another perspective? Explain. **[1]**

3.1.2 Is it possible to modify TOM in any way to take care of the addition-quaddition puzzle? Here are some things to consider which may or may not be helpful in coming up with a better version of TOM: (i) The procedures followed by Mary when performing addition are different from those involved when she performs quaddition. These procedures could be her using a mental look-up table for directly reading off the sum of the queried numbers or her step-wise algorithm for calculating large numbers. (ii) Mary might make a false judgement like '69 plus 37 is 116' and yet know what 'plus' means. She has done numerous additions in the past and her aptitude in basic mathematics is quite robust. Occasional instances of error is not a good reason to say that she doesn't know what 'plus' means. (iii) The way Mary applies 'plus' is complex since there are many things going on: she reliably perceives the stimulus presented to her in the form of the question, uses her working memory, recalls if necessary, certain steps from long term memory, uses her inferential skills, and so on. **[2]**

3.3 In a 2014 article, John Searle makes the following observation**:**

> "Except for the cases of computations carried out by conscious human beings, computation, as defined by Alan Turing and as implemented in actual pieces of machinery, is observer relative. The brute physical state transitions in a piece of electronic machinery are only computations relative to some actual or possible consciousness that can interpret the processes computationally. For example, in the Chinese room, I interpret the input symbols according to instructions I have been given, and I give back an output containing the Chinese answers. Observer-relative computation, though containing an element of ontological subjectivity, can nonetheless be subject to claims that are epistemically objective[1]. It is an epistemically objective fact that I am writing this on a Word program, but a Word program, though implemented electronically, is not an electrical phenomenon; it exists only relative to an observer."

Consider a physical system S that has two input channels and only one output channel. S is put to work as a voltage gate in the following way. If V_{i1} and V_{i2} are the input voltage variables and V_o the output voltage: V_o= 5–10 when V_{i1}, V_{i2} > 5, V_o= 0–2.5 when V_{i1}, V_{i2} < 2.5 and V_o= 2.5–5 otherwise. This state of affairs is expressed in Table 1. Now, if we take the 0–5 V range as a single unit and assign '0' to it and '1' to the 5–10 V range, we get Table 2 which describes the AND gate. There are lesser number of rows in Table 3 because the redundant rows have been removed. On another interpretation, when we assign the 0–2.5 V range as '0' and '1' to the 2.5–10 V range, we get Table 3 which describes the OR gate (after removing redundant rows). Do you think this example illustrates what Searle means by observer relativity of computation? How would you respond to Searle's objection that if our minds are computational systems, they are not observer relative? **[4]**

Input channel 1 V_{i1} (volts)	Input channel 2 V_{i2} (volts)	Output channel V_o (volts)
5–10	5–10	5–10
5–10	2.5–5	2.5–5
5–10	0–2.5	2.5–5
2.5–5	5–10	2.5–5
2.5–5	2.5–5	2.5–5
2.5–5	0–2.5	2.5–5
0–2.5	5–10	2.5–5
0–2.5	2.5–5	2.5–5
0–2.5	0–2.5	0–2.5

Table 1: physical state description of S

Input variable 1	Input variable 2	Output variable
1	1	1
1	0	0
0	1	0
0	0	0

Table 2: computation carried out by S under the assignment '0' = 0–5 V; '1' = 5–10 V

Input variable 1	Input variable 2	Output variable
1	0	1
1	1	1
0	0	0
0	1	1

Table 3: computation carried out by S under the assignment '0' = 0–2.5 V; '1' = 2.5–10 V

[1] Please read section 1 of the Searle's article this is linked to know what he means by ontological subjectivity and epistemological objectivity.

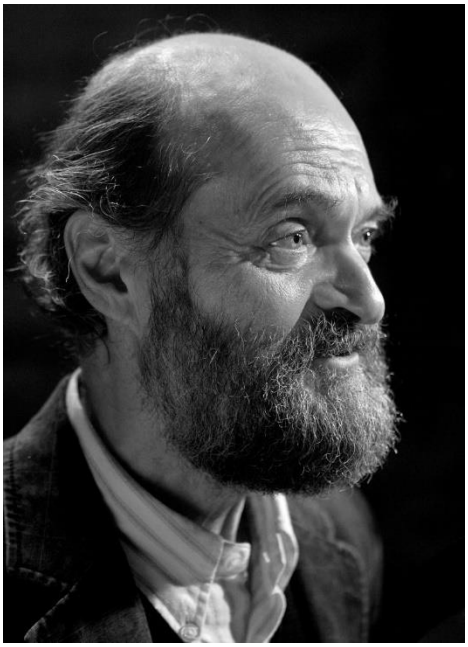
Question 4

4.1 Do you think the following are good responses by a physicalist in response to the Mary's Room thought experiment? Explain.

4.1.1 Mary does not acquire any new knowledge when she leaves her room. *What* Mary thinks once she is out of the room is something that she is already knows. What is new is the *way* she is thinking what she is thinking. If what Mary knows before and after her release is the same, she does not make a discovery in an epistemically significant sense. To illustrate this, consider Oedipus from Greek Mythology who marries his mother Jocasta. Prior to time *t*, Oedipus is not aware that he married his mother but knows that he is married to Jocasta. Only at *t* does he *discover* that Jocasta is his mother. Since the fact that he is married to his mother is *identical* to the fact that he is married to Jocasta, Oedipus does not learn any new fact about the world. What Oedipus acquires is, as it were, a new take on a fact already in his possession. **[2]**

4.1.2 What Mary comes to know when she is outside the room and sees a red object is something that she could have logically inferred from what she learned inside the room. If a subject S knows p and S also knows that p implies q, then it is not necessary that S knows that q (here p and q are propositions). Only logically omniscient beings can figure out the logical consequences of all the knowledge claims they possess. Mary just did not infer something from what was in her possession. What she comes to possess once she is outside the room and sees the red object is something that she *could* have possessed inside the room had she diligently worked through the logical implications of what was available to her inside the room. **[2]**

4.2 It has been many years since Mary walked out of her colorless room and experienced seeing a red object for the first time. She has been living a normal life like you. Mary is now recruited for a new experiment. She is kept inside a room without being able to access the internet or search any information in any book (there aren't any useful resources in the room). The experimenter outside the room has the picture shown here. Mary has not seen the person in the picture or any picture of him. The experimenter describes the picture to Mary in language and sends her pages and pages of text. You can assume that experimenter does the best job possible, perhaps with the aid of some technology that translates images to text. After having learned everything that is given to her, Mary is let out of the room and is shown the picture. Would you think Mary would exclaim "Oh, so that's what Arvo Pärt looks like"? That is, does Mary learn anything new when she comes out of the room and sees the picture that has been described to her? If No, explain why not. If yes, explain what she learns and discuss whether this knowledge is different from the kind of knowledge she learns when she saw red for the first time. **[3]**



4.3 Frank is a Qualia freak. He believes in qualia. His friend, Phyllis, does not. Frank tells Phyllis that once you attend to your (say, visual) experience you can't help but notice that there are intrinsic features of one's experience that are not about anything in the environment. Qualia are such intrinsic properties of experience that one can be directly aware of via introspection. Phyllis rejects this and says that when we have a visual experience of an object, we are not aware of any features of our experience. We are not aware of the experience itself and cannot attend to its features. The only properties of which we are aware of and to which we can attend are the properties of the object in the environment that we are perceiving. "But what about cases of illusion?" asks Frank. Phyllis replies that in such cases, the properties of which we are aware of and to which we can attend are properties experienced as belonging to the object in the environment. For example, when misperceiving a red pen as being blue, we think that the *pen* is blue. Frank finally asks "What if it is case of hallucination?" Phyllis responds by saying that when we hallucinate, the properties of which we are aware of and to which we can attend are properties that are not predicated of anything in the immediate environment but that if they belong to anything or could belong to anything, they belong to external objects and are not features of the visual experience itself. Explain Phyllis' skepticism about qualia. You should be charitable to Phyllis and bring out her intuition against the existence of qualia in a forceful way. **[3]**

Bonus Question

I once met a 'magician' performing tricks with cups and balls. To dispute his title, I asserted "If you are a magician, you can fly". When he agreed, I said "But you cannot fly". He agreed to that as well. "Then you are not a magician!" I retorted, adding, "If you agree with the first two claims, then you have to agree with the inference that follows, else you are being irrational". To show that he was not, he proceeded with the following demonstration. He picked a ball at random from a non-transparent bag containing 100 balls and placed it under a cup. The balls in the bag vary in size (small, big) as well as color (black, white) and the mix was as follows: big black balls – 10, big white balls – 30, small black balls – 50, and small white balls – 10. "Given these numbers," he said "the following claims can be made about the ball under the cup: (1) If the ball is big, then it is probably white, and (2) The ball is not probably white. Do you agree with (1) and (2)"? I agreed, since (1) is true because there are 30 big white balls and only 10 big black balls and (2) is true because there are 60 black balls and only 40 white balls. With a smirk, he asked "Well, aren't you being irrational if you infer from (1) and (2) that the ball is not big"? Should I stop believing in the validity of modus tollens? Explain. **[5]**