Exercise 1. *Counterfactuals in classical epistemic logic*

Counterfactual statements, or counterfactual modalities, concern what is not, but could have been. As an example, consider answers to the following questions: what would have happened if the pandemic didn't spread across the world in 2020? What if I were to take a holiday this week or the next?

Counterfactuals, as we have seen in the lecture, are also used in a variety of quantum thought experiments. Generally, these are simply statements expressing a particular logical inference, $a \Rightarrow b$: "from a it follows that b", "no pandemic $\Rightarrow$ classes in person". Counterfactual reasoning (in classical setting) is also used to solve logical puzzles. In this exercise, we will consider two of those.

(a) *Three logicians, Alice, Bob, and Charlie, walk into a bar. Alice asks, "Does everyone want a beer?" Bob answers, "I don't know." Charlie answers "I don't know" as well. Alice smiles, and orders beer for all three logicians. Explain her reasoning.*

Solution If Bob had *not* wanted a beer, he would have certainly answered negatively (since he belongs to "everyone"). Since this is not the case, Alice can conclude that he wants one. Analogously, she can deduce that Charlie wants a beer as well.

(b) *(Sleeping Beauty problem [1]) Sleeping Beauty volunteers to undergo the following experiment and is told all of the following details: On Sunday she will be put to sleep. Once or twice, during the experiment, Sleeping Beauty will be awakened, interviewed, and put back to sleep with an amnesia-inducing drug that makes her forget that awakening. A fair coin will be tossed to determine which experimental procedure to undertake:*

- *if the coin comes up heads, Sleeping Beauty will be awakened and interviewed on Monday only;*
- *if the coin comes up tails, she will be awakened and interviewed on Monday and Tuesday.*

Any time Sleeping Beauty is awakened and interviewed she will not be able to tell which day it is or whether she has been awakened before. During the interview Sleeping Beauty is asked: "What is your probability assignment now for the proposition that the coin landed heads?" You are the Sleeping Beauty; give your answer.

Solution The probability assignment carried out by the Beauty corresponds to her credence – a degree of her belief she puts in the event. Credences are numbers from 0 to 1 that represent how strongly we believe a claim to be true. A credence of 1 represents complete certainty that a claim is true, a credence of 0 represents complete certainty that a claim is false, and a credence of 0.5 represents complete neutrality.

Answers to the question asked to the Beauty can be divided into two camps: *thirders* and *halfers.*

1. **The thirder argument.** This point of view is based on the *principle of indifference.* It states that in the setting where we have n different possibilities and no additional information about whether one of these possibilities is more likely than the other, we should assign a uniform distribution over all these possibilities. Any possibility would occur with probability $\frac{1}{n}$. Consider an example of a die, which is claimed to be fair: we have no evidence that any one of the six identical die faces is preferred to the others in the die roll. Thus, we should assign a credence of 1/6 to the claim "the die will land 5 on the next roll."

For the Beauty, this uniformity is assigned over three options she has when she wakes up: the coin landed heads and it's Monday; the coin landed tails and it's Monday; and, finally, the coin landed tails and it's Tuesday. Hence, each of these event can be assigned the probability $\frac{1}{3}$, and the credence of the coin coming up heads is $\frac{1}{3}$.

2. **The halfer argument.** This point of view employs two assumptions: the *reflection principle* and the *principal principle*. The *reflection principle* states that, if I know that one will have credence c in a certain claim tomorrow (without learning any new information in the meantime), then one should have credence c in that claim today (and vice versa). The *principal principle* states that our degrees of belief ought to line up with real-world probabilities. When used together, these principles tell us that the Beauty should have credence $\frac{1}{2}$ in the event of the coin landing heads on Sunday night (per the principal principle), and, since she learns no new information when she wakes, she should have that same credence $\frac{1}{2}$ whenever she wakes.

Here's another argument for the unconvinced by the thirders' reasoning: imagine that, instead of waking up twice if the coin lands tails, Beauty instead wakes up a hundred times (and receives the memory drug after each waking). In this variant, it's easier to believe that Beauty gains some new, relevant information when she wakes: "I'm awake now," information she seems a lot more likely to receive if the coin landed tails instead of heads. Even if Beauty doesn't learn new information about the coin or the world outside of herself when she wakes, she does learn new information about her place in the world (that is, *self-locating information*) that justifies her change in credence. If this justification is correct, it applies in our original case too since, once again, Beauty should expect herself to wake up more frequently if the coin lands tails and so should take the new information that she is awake as evidence that the coin landed tails.

Debates over this problem are relevant to epistemology, the philosophy of science, the philosophy of probability, and more.

Exercise 2. *Measurements: from von Neumann to weak*

After carrying out a quantum measurement, we obtain a (classical) outcome. As a piece of information, this outcome is physical, and can be modelled as being contained in a physical system. In this exercise, we take this observation into account to arrive to von Neumann picture of measurement [2].

Suppose we want to measure an observable $\hat{A}$ with the basis $\{|a_i\rangle\}_i$ on a system S. Consider another system (a test particle) M described by the canonical position $\hat{X}$ and conjugate momentum $\hat{P}$ which we couple the system S to via the interaction Hamiltonian

$$H_I = g(t)\hat{A}_S \otimes \hat{P}_M.$$

The time dependent coupling constant $g(t)$ describes the switching "on" and "off" of the interaction. For an impulsive measurement we need the coupling to be strong and short; we take $g(t)$ to be non-zero only for a short time around the moment of interest, t_0 and such that $\int g(t)dt = g > 0$. We assume that during the time of measurement the evolution is governed solemnly by the interaction term.

(a) *The initial state of the measured system and the measurement device (the pointer) is*

$$|\chi\rangle_{SM} = \sum_i \alpha_i |a_i\rangle_S \otimes \int dx\ \psi(x)|x\rangle_M.$$

Write down the final state of the system after the measurement.

Solution For simplicity, assume $g(t) = g = const$. Then the evolution of the global system during the interaction is described by the unitary

$$U(t) = e^{-iH_I t} = e^{-ig\hat{A}_S \otimes P_M t} = \sum_{n=0}^{\infty} \frac{(-itg)^n}{n!} A_S^n \otimes P_M^n.$$

Rewriting $A_S = \sum_j a_j |a_j\rangle\langle a_j|_S$, we obtain $A_S^n = \sum_j a_j^n |a_j\rangle\langle a_j|_S$, and arrive to

$$U(t) = \sum_j \sum_{n=0}^{\infty} \frac{(-itga_j)^n}{n!} |a_j\rangle\langle a_j|_S \otimes P_M^n = \sum_j |a_j\rangle\langle a_j|_S \otimes e^{-itga_j \hat{P}_M}.$$

Acting with the unitary on the initial state gives the final state

$$|\chi'\rangle_{SM} = U(t)|\chi\rangle_{SM} = \sum_i \alpha_i |a_i\rangle_S \otimes \int dx\ \psi(x - gta_i)|x\rangle_M.$$

(b) *Let us take a Gaussian as the initial state of the pointer $\psi(x) = \exp(-\frac{x^2}{\Delta^2})$. When does this measurement approach an ideal one?*

Hint: For an ideal measurement, the final state of the pointer (after tracing out the system S) is a density operator representing a series of peaks, each corresponding to a different eigenvalue a_i, with probability $|\langle a_i|\psi\rangle|^2$.

Solution When the uncertainty Δ in the initial position of the pointer is much smaller than the difference in the shifts of the pointer corresponding to different eigenvalues a_i, the measurement approaches an ideal measurement - the final state of the pointer (after tracing over the state of the measured system) is a density matrix representing a series of peaks, each corresponding to a different eigenvalue a_i, and having probability equal to $|\langle a_i|\psi\rangle|$.

(c) *We can reduce the disturbance caused by the measurement on the measured system by reducing the strength of the interaction g. What is the approximate final state of the system M in that case?*

Solution In this regime the measurement becomes less precise since the uncertainty Δ in the initial position of the pointer becomes larger than the difference in the shifts of the pointer ga_i, corresponding to the different eigenvalues. Nevertheless, even in the limit of very weak interaction the measurement can still yield valuable information – the final state of the measuring device is almost unentangled with the measured system and approaches a Gaussian centered around the average value $\bar{A}$, $\psi'(x) \approx e^{-\frac{(x-\bar{A})^2}{\Delta^2}}$.We need however to repeat the measurement many times to be able to locate the center.

(d) *Now let us add the pre- and post-selection on the system S to the picture. Write down the final state of the measuring device, given the initial state $|\Psi\rangle_S$ and the final state $|\Phi\rangle_S$ of the system S.*

Solution Given the final state calculated in (a), we can write the final state of the measurement device as (up to normalization)

$$|\xi\rangle_M = \sum_i \alpha_i \langle\Phi|a_i\rangle\langle a_i|\Psi\rangle \int dx\ \psi(x - gta_i)|x\rangle_M.$$

It can be shown [3] that the final state of the measuring device after post-selection is described by the Gaussian $\exp(-\frac{(x-g\,A_w)^2}{\Delta^2})$*, which corresponds to the pointer being shifted to a value* A_w*, called the weak value of the observable* $\hat{A}$*, and given by*

$$A_w = \frac{\langle\Phi|\hat{A}|\Psi\rangle}{\langle\Phi|\Psi\rangle}.$$

Note that in contrast to ordinary expectation values, weak values can lie outside the range of eigenvalues of $\hat{A}$ *and are generally complex!*

Most importantly, in the weak regime different measurements do not disturb each other so non-commuting variables $\hat{A}$ *and* $\hat{B}$ *can be measured simultaneously, and yield the same weak values* A_w *and* B_w *as when measured separately.*

References

[1] *for example,* https://en.wikipedia.org/wiki/Sleeping_Beauty_problem

[2] *John von Neumann. Mathematical foundations of quantum mechanics. Princeton University Press, Princeton, 1953.*

[3] *Yakir Aharonov, Lev Vaidman. Properties of a quantum system during the time interval between two measurements. Physical Review A, 41(1), 11–20, 1990.* DOI:10.1103/physreva.41.11