Addendum to 'Relevance of the ensemble and actual theory distinction in the sleeping beauty problem': uncertainty on theories as opposed to uncertainty on states

Minseong Kim*

(Dated: December 4, 2025 (ver 2))

Abstract

This is an addendum (memo) that goes beyond the sleeping beauty problem. I discuss why 1) the sleeping beauty problem (SBP) essentially has the same structure as the black hole information problem, 2) uncertainty on theories should be distinguished from uncertainty on states. Vagueness can then be understood as uncertainty over theories, with each theory having a clear boundary. Backward induction paradoxes can also be resolved in terms of uncertainty over theories. The actual versus ensemble theory distinction matters everywhere.

* mkimacad@gmail.com; ORCiD:0000-0003-2115-081X

I. SBP = BLACK HOLE (BH) INFORMATION PROBLEM?

Structurally we can think of SBP as a simplification of the (black hole) information problem.

Consider observers that are inside some black hole. From the point of the black hole exterior, their time do not correspond to the exterior point of time. Coordinate-wise, the exterior time coordinate becomes spatial inside the black hole, while the exterior spatial coordinate becomes temporal. Therefore, the observers inside the black hole are essentially facing temporal self-locating uncertainty when they are trying to cast their story in terms of an observer on the black hole exterior.

We need to do more than just self-locating uncertainty to justify SBP being a simplification of the information problem, which is to consider the black hole evaporation process (though this will be drastically simplified). Observer C enters the black hole at exterior time $t = 0$. At $t = 1$, it is known that regardless of outcomes, C will not escape the black hole. At $t = 2$, C may escape the black hole with the probability of $1/2$. This is essentially the SBP setup.

Now C asks that conditional on C being inside the black hole, what is the credence to be assigned for the probability of escaping a black hole at $t = 2$. If the conventional thirder solution is correct, then it is $1/3$. But this is a paradox - the information paradox.

If the actual versus ensemble distinction does correctly resolve SBP, then we could see how the information problem may conceptually be resolved. The answer is holography of information, which states that the information regarding the black hole interior is already on the exterior. There is no actual uncertainty regarding C's history represented on exterior time t, and the theory that evolves C s already determined. C only experiences illusions of uncertainty because C does not have access to exterior degrees of freedom. When C knows this, C can self-correct for this and produce a coherent exterior account of C that nevertheless reflects C's uncertainty.

C's Page curve is therefore actually trivial - zero all the time. However, reflecting ensemble uncertainty, we can produce a coherent probabilistic story that reproduces the conventional Page curve with corrections (in case of SBP, probability of head from $1/3$ to $1/2$, and for the information problem, replica wormholes) as well.

II. BRIEF DIGRESSION: NEURAL NETWORKS AS ENSEMBLE THEORIES

A large neural network (which includes most LLMs) is a giant ensemble theory that has no single understanding of the world. In that sense, the probability that the network assigns can be misleading, even when it is correct. More precisely, neural network training procedures currently have no mean to enforce a single coherent understanding of the world when creating or assigning probabilities.

This can be good for ordinary worlds, where people use concepts and words vaguely, and the switch from a single coherent understanding to this ensemble understanding should be one reason why neural networks have been so successful. Yet they will have limitations in places where there should be no vagueness.

III. VAGUENESS FROM THE 'ACTUAL VERSUS ENSEMBLE'

From the 'actual versus ensemble theory' point of view, vagueness - or at least most instances of vagueness can be stated in a very simple way. Vagueness arises when there are uncertainties over actual theories, not just over states. The sleeping beauty problem suggests why uncertainties over actual theories slipping through uncertainties over states can cause problems when the two types of uncertainities are not cleanly distinguished.

Consider the conventional sand heap sorites paradox. When asked whether my pile of sands constitute a (large) heap, it is known that for each candidate boundary that distinguishes a heap from a non-heap:

$$CB(x) \to CB(x-1), \quad N \gg 0$$

where $CB(x)$ refers to 'x sand particles in a heap can constitute a candidate boundary' and N refers to the maximum number of sand particles. When taken to full induction, all possible values of x can constitute candidate boundaries.

But this only reflects uncertainties over actual theories (or 'boundary') $B(x)$, and CB acts as an ensemble theory over B. Since $CB(x) = 1$ (true) for all possible x, it is best to consider probabilities over actual boundary (theory) $B(x)$ in CB instead.

As Timothy Williamson would argue, for each actual theory $B(x)$, there is a sharp boundary. So on actual uses of a vague term, one actual theory $B(x)$ gets realized, and everyone using the term knows how it is being used. However, because there is probability over

how $B(x)$ is to be realized at each instance, different $B(x)$ is used for new contexts, given probabilistically.

In the language similar to SBP, the average actual post-measurement uncertainty over a vague term is zero (at least in this idealized world), because whenever a vague term is actually invoked, everyone knows how it is used (this can be inferred). However, before such a measurement is carried out, the ensemble theory rules, and uncertainties over a vague term can be significant.

For example, when someone calls some vague color (with a clear visual reference, realization or instance) that may be called blue or green as 'green', everyone having a conversation may form their color boundary (between blue and green) and continue to have coherent talks and understanding.

Even the quantum formalism is not currently well-suited for describing probabilities within an ensemble as opposed to actual states, which has confused many when utilizing quantum mechanics intuitions for SBP. But concepts like quantum state collapse and entanglement remain very useful for the analysis.

The questions like the ship of Theseus and backward inductions paradoxes, such as the chainstore paradox, can be handled similarly. There, the issue is that there are different candidate 'actual theories' that must be considered probabilistically. For both cases, over time, credibility of some theories decreases significantly for defining some vague term. For the chainstore paradox, this would be the probability that a monopolist may commit to some strategy profile regardless of a competitor action that arrives before it decides. At early time, the monopolist is very likely to choose the commitment theory ('deterrence theory') over the discretion theory ('backward induction' theory). At late times, the monopolist almost surely chooses the discretion theory over the commitment theory. That there is no one theory that governs the monopolist and yet the monopolist can be ultra-optimal by newly selecting an actual theory probabilistically when facing each competitor is not actually paradoxical. The chainstore paradox only arises because we force agents to use a single theory of rationality whenever making decisions.

If we force agents to use the same theory of rationality at all times, then the monopolist in the chainstore paradox has to adopt the discretion ('backward induction') theory. But if the concept of rationality can be vague, then there may be different rationality theories that agents can select at each time. The ensemble understanding will be provide the meta-

rationality analysis of whether such selections over time can be considered rational - or more precisely, meta-rational. In case of the chainstore paradox, the monopolist going from full deterrence to full cooperation over time is perfectly meta-rational.

These ideas can similarly be applied to other backward induction paradoxes, such as the centipede game or the surprise examination paradox.