# *Mathematical Modeling of the Epistemology of Simple Physical Agents*

Ian George and Prof. Benjamin Schumacher
Kenyon College Summer Science 2016

## Abstract

**Physics has a strong understanding of microstates (the complete description of particles which make up a system) and macrostates (the set of all possible microstates which make up an observed property such as temperature). However an agent, be it a detector or human being, may have partial knowledge of both the macrostate and the microstate. The interplay between the record of an agent, as described in terms of information, and the universe, as described by physics, has not been looked into thoroughly, even though solutions to Maxwell's Demon by Landauer and Bennett as well as ways to describe quantum physics in terms of information suggest a deep connection.**

**To look into this we first looked at epistemic logic, a formalization of the rules of knowledge, mapping the propositions themselves onto a phase space, as well as adding modal operators for time. Then we explored the connection between probability as inductive logic and epistemic logic. This then led us to questions about probability and how agents should quantify it- revealing the bundle of issues between probability as measured by an agent and physics.**

## Overview of Epistemic Logic

Epistemic Logic is a mathematical symbolic formalization of how you and I talk about knowledge- what we know and what follows from our knowing.

The basic formula of epistemic logic is $K_bA$ which translates to "agent b knows A" where A is a proposition. Using this formula, people generate claims about knowledge as axioms.

| Axiom Name | Axiom |
|---|---|
| K | $K_b(A \to B) \to (K_bA \to K_bB)$ |
| T | $K_bA \to A$ |
| 4 | $K_bA \to K_bK_bA$ |
| 5 | $\neg K_bA \to K_b\neg K_bA$ |

**Figure 1: Table of Common Epistemic Axioms**

Note that there are other axioms argued for, primarily ones that are stronger than Axiom 4 (i.e. if this other axiom can be proven, so too can Axiom 4), but weaker than Axiom 5.

Our first objective was to determine which of these axioms is true for all systems that have knowledge in the simplest sense– the ability to record information about the universe by some physical process. Towards these ends, we started with a definition of the knowledge operator and mapped it onto a phase space.

We take the following possible worlds definition of the Epistemic operator:

*$K_bA$ = In all possible world states compatible with what agent b knows, it is the case that A*

Which we translate into the following phase space definition:

$$K_bA = \{(s,t) \in C : \forall t' \in E_s, (s,t') \in A\}$$

Where $s$ is the given agent state of agent $b$, $t'$ is all possible environment states that do not contradict the agent state, and $C$ is a subset of all ordered pairs of agent states and environment states governed by a set of undetermined rules of the relationship between an agent state and an environment state. $E_s$ is a set which finds all environment states that can be paired with $s$ given $C$. We refer to $C$ as the context. This is because the ordered pairs of agent states and environment states are the universe states, and all the possible universe states are selected by the undetermined rules, an assumed context, and contained in $C$.

## Epistemic Logic Diagram

We can then use proof by picture to see some of the properties we proved using our phase space logic.

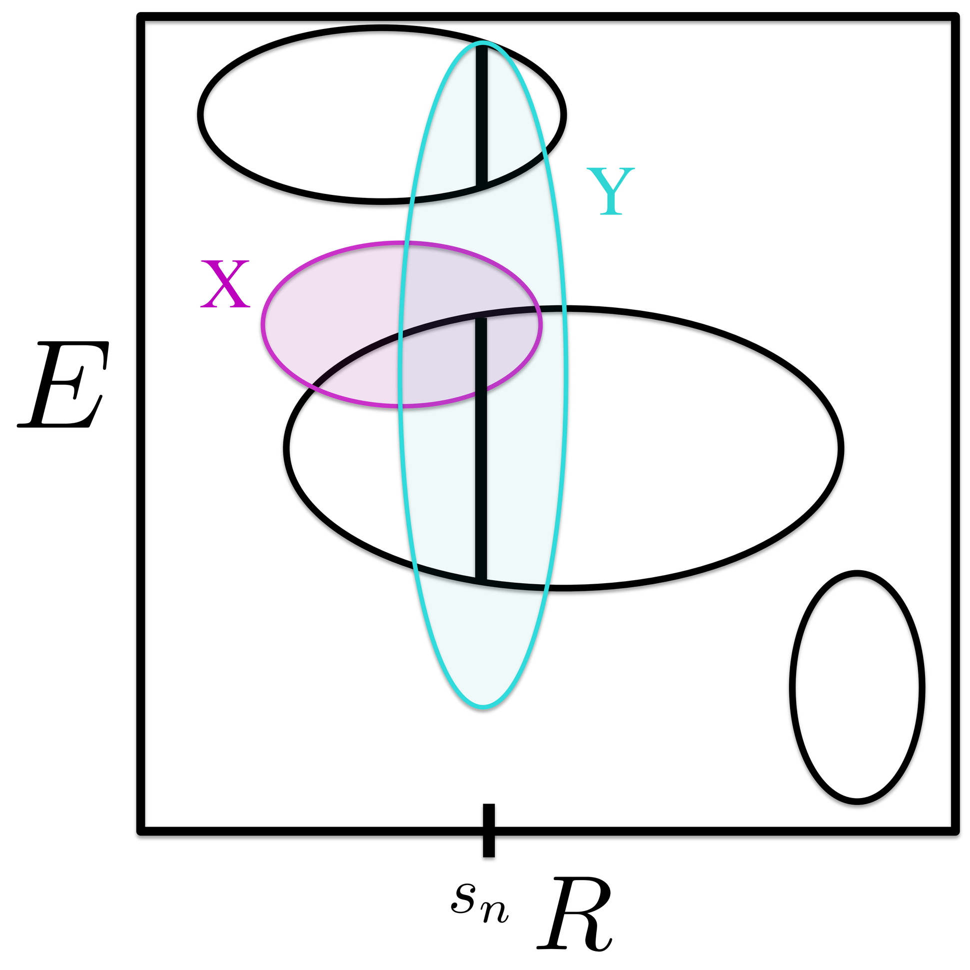



**Figure 2: A Possible Universes Phase Space**

The axis labeled $E$ is all possible environment states. The axis labeled $R$ is all possible agent states. Every point on this graph is an ordered pair, a universe state, $(s,t)$. Those universe states within the black ellipses are universe states within the context $C$. As they are in the context, they are possible universe states. The black line projected from agent state $s_n$ is only drawn within the context- it is $E_s$ for $s_n$.

We can see that all environment states paired with agent state $s_n$ are contained in the set of universe states in which proposition $Y$ is true. Thus we can conclude the agent $(b)$ in state $s_n$ knows $Y$ is true. Symbolically stated, $K_bY$. By the same reasoning we can conclude $\neg K_bX$.

We can also see that for any given state in $E_s$ for $s_n$, all other universe states in $E_s$ are also in $K_bY$. Therefore $K_bK_bY$, so Axiom 4 holds. By a similar argument we proved that Axiom 5 is true. Therefore with the simplest reasonable definition of knowledge, we have shown even the simplest machine knows its state perfectly.

## Time Evolution

How do we handle time evolution given our phase space system then?

- Classical mechanics suggests the world evolves deterministically and reversibly.
- We take the context $C$ to describe the world for all times, and therefore the context is time invariant.

Therefore, we can describe time evolution with a bijective function $f$ where $f(C) = C$.

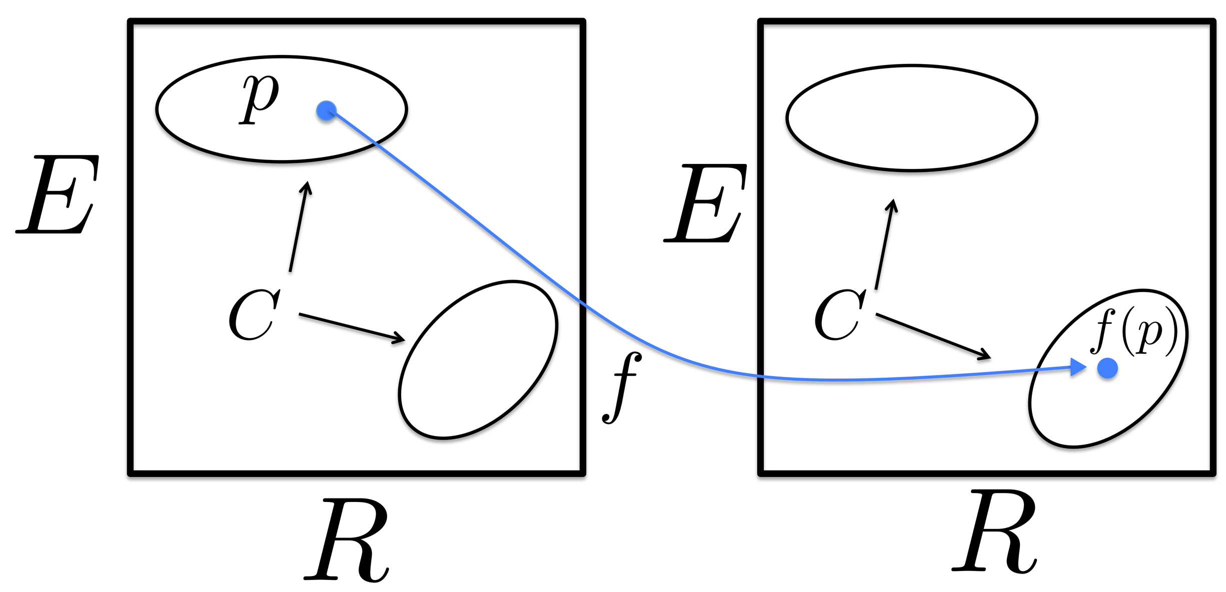



**Figure 3: Time Evolution**

Using our function $f$ we can create phase space definitions for linear temporal operators- modal logic operators for formalizing claims about the world throughout time. Let $u$ be a universe state $(s,t)$ in $C$.

| Modal Operator | Phase Space Definition |
|---|---|
| $\bigcirc A$ | $f^{n+1}(u) \in A$ |
| $\Diamond A$ | $\exists n \in \mathbb{N} : f^n(u) \in A$ |
| $\Box A$ | $\forall n \in \mathbb{N}, f^n(u) \in A$ |

## Extension to Probability

We now want to generalize our model for uncertainty by taking probability theory as an extension of epistemic logic.

As the context is time invariant, it is reasonable to assume a uniform probability distribution over all universe states in the context. With this assumption, we defined prior probabilities- which are in fact conditional on any information the agent has- to be calculated as follows:

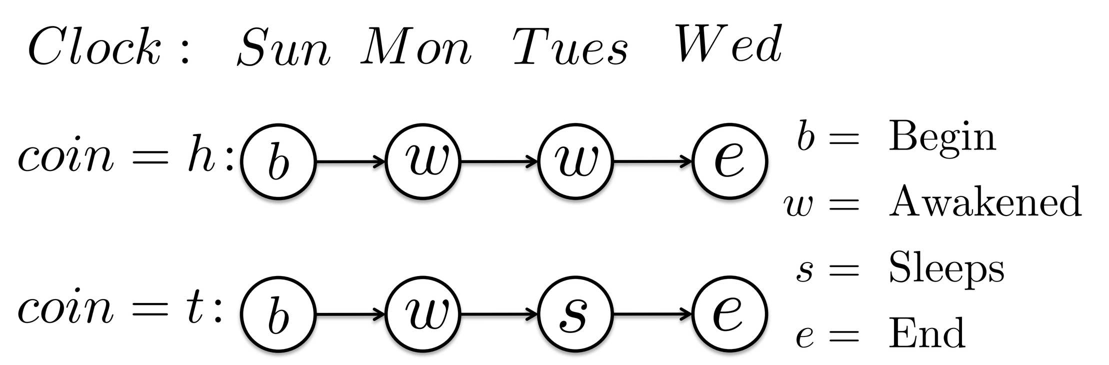
$$P(X|s) = \frac{\#((s,t) \in C : (s,t) \in X)}{\#((s,t) \in C)}$$

## Sleeping Beauty Problem

Consider the following scenario:

Sleeping Beauty undertakes the following experiment.
(1) On Sunday, she goes to sleep. A fair coin is flipped, but she does not know its result.
(2) If the coin is "heads", she is awakened on both Monday and Tuesday. If "tails", she is awakened only on Monday, but remains asleep on Tuesday.
(3) Each time she is awakened in (2), she is given a drug to erase her memory of being awakened.
(4) On Wednesday, the experiment ends.
Question: Sleeping Beauty is awakened. According to her, what is P(heads)?

One might have a gut reaction for a correct answer, but there is a real issue in this problem. The coin is fair, so Sleeping Beauty might say the probability of heads is 1/2, but Sleeping Beauty also knows she should be woken twice as often if the coin is heads, so perhaps Sleeping Beauty should answer the probability of heads is 2/3. Let us try to solve this.




The agent state of Sleeping beauty is written in the circle and the environment state (or at least the part we are concerned with), is written to the left. The two universe states in which coin = $h$ and Sleeping Beauty = $t$ are distinct because of the clock state which is also part of the environment state.

Therefore, given our definition of prior probability, Sleeping Beauty's only rational evaluation of the probability of the coin being heads is 2/3!

## Key References

[1] Hendricks, Vincent and Symons, John, "Epistemic Logic", The Stanford Encyclopedia of Philosophy (Fall 2015 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2015/entries/logic-epistemic/>.

## Acknowledgments

I would like to thank my advisor Prof. Schumacher for his constant patience, support, and insight; Prof. Westmoreland at Denison University for his help; and, the rest of my research group.

This work was funded by the Kenyon College Summer Science Scholars program and FXQI grant FXQi-RFP-1517.