Constraints on physical reality arising from a formalization of knowledge

David H. Wolpert
Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM, 87501
david.h.wolpert@gmail.com

There are (at least) four ways that an agent can acquire information concerning the state of the universe: via observation, control, prediction, via retrodiction, i.e., memory. Each of these four ways of acquiring information seems to rely on a different kind of physical device (resp., an observation device, a control device, etc.). However it turns out that certain mathematical structure is common to those four types of device. Any device that possesses a certain subset of that structure is known as an "inference device" (ID).

Here I review some of the properties of IDs, including their relation with Turing machines, and (more loosely) quantum mechanics. I also review the bounds of the joint abilities of any set of IDs to know facts about the physical universe that contains them. These bounds constrain the possible properties of any universe that contains agents who can acquire information concerning that universe.

I then extend this previous work on IDs, by adding to the definition of IDs some of the other mathematical structure that is common to the four ways of acquiring information about the universe but is not captured in the (minimal) definition of IDs. I discuss these extensions of IDs in the context of epistemic logic (especially possible worlds formalisms like Kripke structures and Aumann structures). In particular, I show that these extensions of IDs are not subject to the problem of logical omniscience that plagues many previously studied forms of epistemic logic.

INTRODUCTION

Ever since Wheeler discussed "it from bit" [37], there has been great interest in what constraints on the properties of the universe can be derived using some appropriate mathematical formulation of information. Some of this work relies on Shannon information theory [12, 19–21], and some of it on Fisher information theory [17]. There has also been work on this topic that focuses on the *processing* of information, i.e., that views the universe through the lens of Turing machine (TM) theory [13, 24, 26, 33, 43–47].

Here I adopt a different approach. I focus on the fact that information concerning the state of the universe typically *is held by some agent embedded in that universe*. For example, we cannot speak of Shannon information without specifying probability distributions — which reflect the uncertainty of some specific agent concerning the state of the universe that contains them (e.g., uncertainty of a scientist making a prediction).

Inference devices

There are (at least) four ways an agent can acquire some information concerning the universe in which it is embedded: via an observation device, via a control device, via a prediction device, and via a memory device, i.e., a "retrodiction" device. It turns out that there is some mathematical structure shared by all such information-acquiring devices. Devices with that structure are called "Inference Devices" (IDs) [9, 38–40].

In the first section of this paper I present two examples of how an agent can acquire information about the universe that contains them, illustrating that in both examples the agent has the mathematical structure of an ID. I then present some of the more elementary impossibility results concerning IDs. These results place strong constraints on what information about a universe can be jointly held by different IDs embedded in that universe. Importantly, these constraints arise only from the definition of IDs, without any assumptions about the laws of the universe containing the IDs; they hold in *any* universe that allows agents that have information about that universe. In particular, they would hold even in a classical, finite universe, with no chaotic processes. They would also hold even in a universe with agents who have super-Turing computation abilities, can transmit information at super-luminal rates, etc. It is worth noting as well that these impossibility theorems hold even though there is no sense in which IDs have the ability of self-reference.

After these preliminaries I present some of the connections between the theory of IDs and the theory of Turing Machines. In particular I analyze some of the properties of an ID version of universal Turing machines and of an ID version of Kolmogorov complexity [13, 24, 26, 33, 43–47]. I show that the ID versions of those quantities obey many of the familiar results of Turing machine theory (e.g., the invariance theorem of TM theory).

I then consider one way to extend the theory of IDs to the case where there is a probability distribution over the states of the universe, so that no information is ever 100% guaranteed to be true. In particular, I present a result concerning the products of probabilities of error of two separate IDs, a result which is formally similar to the Heisenberg uncertainty principle.

These results all concern subsets of an entire universe, e.g., one or two IDs embedded in a larger universe. However we can expand the scope to an entire universe. The idea is to *define* a "universe", with whatever associated laws of physics, to be a set of physical systems and IDs (e.g., a set of scientists), where the IDs can have information concerning those physical systems and / or one another. Adopting this approach, I use the theory of IDs to derive impossibility results concerning the nature of the entire universe.

Inference devices and epistemic logic

Most of the results presented to this point in the paper have appeared before, albeit in a more complicated, less transparent formalism than the one used here [9, 38–40]. In the last sections of the paper I present new results. These all involve an extension of IDs, one that includes some of the features that are shared by the four ways for an agent to acquire information (observation, control, prediction and memory) but that are not in the original definition of IDs.

I show that this strengthened version of IDs has a close relation to the various ways of formalizing "knowledge" that are considered in epistemic logic [3, 16, 30, 42]. However the ID-based theory of knowledge is not subject to what is perhaps the major problem of these earlier ways of formalizing knowledge, the problem of *logical omniscience*.

To explain that problem, it is easiest to work with the *event-based* formulation of epistemic logic pioneered by Aumann [5–7, 10, 18]. In this formulation we start with a space U of possible states of the entire universe across all time. (In the literature this is typically called a set of "possible worlds".) An *event* is defined as a subset of U. For example, let U be a set of all possible histories of the universe across all time and space. Then the event {there are no clouds in the sky in London during January 1, 2000} is all $u \in U$ such that {there are no clouds in the sky in London during January 1, 2000}. Belief by an agent concerning the universe is formalized in terms of a partition of U, which is called an *Aumann structure* and an associated *belief operator* $B : U \rightarrow \{R_i\}$. This is supposed to represent the intuition that in any universe u, the agent *believes* that an event $E \subset U$ holds iff $B(u) \subseteq E$. *Knowledge* is then defined as true belief, i.e., a belief operator K with the property that $u \in K(u)$ for all $u \in U$ [16]. Phrased differently, in the event-based framework we say that "agent i knows event E in world $u \in U$ if E holds for all worlds that agent i believes is possible when the actual world is u, one of which worlds is u itself".

Now suppose that some event E implies some event E', i.e., $E' \supseteq E$. This means that under the event-based definition of knowledge, if agent i knows event E in world u, E' is also true in world u — and agent i knows event E' in world u. In short, the agent cannot know a set of facts without knowing all logical implications of those facts. This is known as the property of **logical omniscience**. It says, for example, that if someone multiplies two huge prime numbers and then (honestly) tells that product to the agent — so that the agent knows that product — then the agent must know the two prime numbers. This of course is absurd.

This problem of logical omniscience plagues possible-worlds models of epistemic logic like those based on Aumann structures. Some extensions to possible-worlds models have been proposed to address this problem, e.g., bounds on the computational powers of the agent [3], assuming that the agent reasons illogically [16], and a set of related "impossible possible worlds" restrictions on the nature of the agent [3, 16, 30, 42]. However none of these has proven broadly convincing.

There are other difficulties with the event-based formalization of knowledge. In that framework, by simply defining the knowledge operator of a rock on the moon appropriately, we would say that the rock "knows" whether it is in sunlight or not (for example by having its knowledge operator pick sets of states of the universe based on the temperature of that rock). This pathology is due to the fact that the definition of knowledge operators in the event-based framework does not reflect the fact that knowledge is *held by a sentient agent*. Specifically, any sentient agent that knows something about the universe is able to correctly answer arbitrary questions about what they know, either implicitly or explicitly. (Note that a lunar rock cannot answer such questions.) However there is nothing in the formal structure of Aumann structures, Kripke structures, or the like that involves the ability of agents to correctly answer questions.

The ID framework is concerned precisely with such ability of an agent to answer questions about the information they have. As a result, in the extension of the ID framework into a full-fledged theory of knowledge, we cannot say that a rock on the moon "knows" whether it is in sunlight. Moreover, the ID-based theory of knowledge avoids the problem of logical omniscience. Specifically, in the ID-based formalization of knowledge, if an agent knows some fact A, and knows that A implies B, then B is true — but the agent may not know that.

None of the results below are difficult to prove; some of the proofs, especially those of the "Laplace demon theorems", are almost trivial. (The interest is the implications of the inference device axioms for metaphysics and epistemology, not the math needed to derive those implications.) Nonetheless, the interested reader can find all proofs that are not given below in [39].

INFERENCE DEVICES

In this section I review the elementary properties of inference devices, mathematical structures that are shared by the processes of observation, prediction, recall and control [9, 38–40]. The proofs of these results are, at root, simply variants of the Epimenides paradox. More sophisticated results, some of them new, are presented in the following section.

Observation, prediction, recall and control of the physical world

I begin with two examples that motivate the formal definition of inference devices. The first is an example of an agent making a correct observation about the current state of some physical variable.

Example 1. *Consider an agent who claims to be able to observe $S(t_2)$, the value of some physical variable at time t_2. If the agent's claim is correct, then for any question of the form "Does $S(t_2) = L$?", the agent is able to consider that question at some $t_1 < t_2$, observe $S(t_2)$, and then at some $t_3 > t_2$ provide the answer "yes" if $S(t_2) = L$, and the answer "no" otherwise. In other words, she can correctly pose any such binary question to herself at t_1, and correctly say what the answer is at t_3.*[1]

To formalize this, let U refer to a set of possible histories of an entire universe across all time, where each $u \in U$ has the following properties:

i) u is consistent with the laws of physics,

ii) In u, the agent is alive and of sound mind throughout the time interval $[t_1, t_3]$, and the system S exists at the time t_2,

iii) In u, at time t_1 the agent considers some L-indexed question q of the form "Does $S(t_2) = L$?",

iv) In u, the agent observes $S(t_2)$,

v) In u, at time t_3 the agent uses that observation to provide her (binary) answer to q, and believes that answer to be correct.[2]

The agent's claim is that for any question q of the form "Does $S(t_2) = L$?", the laws of physics imply that for all u in the subset of U where at t_1 the agent considers q, it must be that the agent provides the correct answer to q at t_3. Any prior knowledge concerning the history that the agent relies on to make this claim is embodied in the set U.

The value $S(t_2)$ is a function of the actual history of the entire universe, $u \in U$. Write that function as $\Gamma(u)$, with image $\Gamma(U)$. Similarly, the question the agent has in her brain at t_1, together with the time t_1 state of any observation apparatus she will use, is a function of u. Write that function as $X(u)$. Finally, the binary answer the agent provides at t_3 is a function of the state of her brain at t_3, and therefore it too is a function of u. Write that binary-valued function giving her answer as $Y(u)$.

Note that since U embodies the laws of physics, in particular it embodies all neurological processes in the agent (e.g., her asking and answering questions), all physical characteristics of S, etc.

So as far as this observation is concerned, the agent is just a pair of functions (X, Y), both with the domain U defined above, where Y has the range $\{-1, 1\}$. A necessary condition for us to say that the agent can "observe $S(t_2)$" is that for any $\gamma \in \Gamma(U)$, there is some associated X value x such that for all $u \in U$, so long as $X(u) = x$, it follows that $Y(u) = 1$ iff $\Gamma(u) = \gamma$.

I now present an example of an agent making a correct prediction about the future state of some physical variable.

Example 2. *Now consider an agent who claims to be able to predict $S(t_3)$, the value of some physical variable at time t_3. If the agent's claim is correct, then for any question of the form "Does $S(t_3) = L$?", the agent is able to consider that question at some time $t_1 < t_3$, and produce an answer at some time $t_2 \in (t_1, t_3)$, where the answer is "yes" if $S(t_3) = L$ and "no" otherwise. So loosely speaking, if the agent's claim is correct, then for any L, by their considering the appropriate question at t_1, they can generate the correct answer to any question of the form "Does $S(t_3) = L$?" at $t_2 < t_3$.*[3]

To formalize this, let U refer to a set of possible histories of an entire universe across all time, where each $u \in U$ has the following properties:

i) u is consistent with the laws of physics,

ii) In u, the agent exists throughout the interval $[t_1, t_2]$, and the system S exists at t_3,

iii) In u, at t_1 the agent considers some question q of the form "Does $S(t_3) = L$?",

iv) In u, at t_2 the agent provides his (binary) answer to q and believes that answer to be correct.[4]

The agent's claim is that for any question q of the form "Does $S(t_3) = L$?", the laws of physics imply that for all u in the restricted set U such that at t_1 the agent considers q, it must be that the agent provides the correct answer to q at t_2.

The value $S(t_3)$ is a function of the actual history of the entire universe, $u \in U$. Write that function as $\Gamma(u)$, with image $\Gamma(U)$. Similarly, the question the agent considers at t_1 is a function of the state of his brain at t_1, and therefore is also a function of u. Write that function as $X(u)$. Finally, the binary answer the agent provides at t_2 is a function of the state of his brain at t_2, and therefore it too is a function of of u. Write that function as $Y(u)$.

So as far as this prediction is concerned, the agent is just a pair of functions (X, Y), both with the domain U defined

[1] It may be that the agent has to use some appropriate observation apparatus to do this; in that case we can just expand the definition of the "agent" to include that apparatus. Similarly, it may be that the agent has to configure that apparatus appropriately at t_1. In this case, just expand our definition of the agent's "considering the appropriate question" to mean configuring the apparatus appropriately, in addition to the cognitive event of her considering that question.

[2] This means in particular that the agent does not lie, does not believe she was distracted from the question during $[t_1, t_3]$.

[3] It may be that the agent has to use some appropriate prediction computer to do this; in that case we can just expand the definition of the "agent" to include that computer. Similarly, it may be that the agent has to program that computer appropriately at t_1. In this case, just expand our definition of the agent's "considering the appropriate question" to mean programming the computer appropriately, in addition to the cognitive event of his considering that question.

[4] This means in particular that the agent does not believe he was distracted from the question during $[t_1, t_2]$.

above, where Y has the range $\{-1, 1\}$. The agent can indeed predict $S(t_3)$ if for the space defined above U, for any $\gamma \in \Gamma(U)$, there is some associated X value x such that, no matter what precise history $u \in U$ we are in, due to the laws of physics, if $X(u) = x$ then the associated $Y(u)$ equals 1 iff $\Gamma(u) = \gamma$.

Evidently there is a mathematical structure, in the form of functions X and Y, that is shared by agents who do observation and those who do prediction. As formalized below, I refer to any such pair (X, Y) as an "inference device". Say that for some function Γ, for any $\gamma \in \Gamma(U)$, there is some associated X value x such that, no matter what precise history $u \in U$ we are in, due to the laws of physics, if $X(u) = x$ then the associated $Y(u)$ equals 1 iff $\Gamma(u) = \gamma$. Then I will say that the device (X, Y) "infers" Γ.

See [39] for a more detailed elaboration of the processes of observation and prediction in terms of inference devices. Arguably to fully formalize each of these phenomena there should be additional structure beyond that defining inference devices. (See App. B. of [39].) Some such additional structure is investigated below, in the discussion of "physical knowledge".

It is also shown in [39] that a system that remembers the past is an inference device. (Intuitively, memory is just retrodiction, i.e., it is using current data to predict the state of non-current data, but rather than have the non-current data concern the future, in memory it concerns the past.) [39] also shows that a device that controls a physical variable is an inference device. All of this analysis holds even if what is observed / predicted / remembered / controlled is not the answer to a binary question of the form, "Does $S(t) = L$?", but instead an answer to question of the form, "is $S(t)$ more property A than it is property B?" or of the form, "is $S(t)$ more property$_A$ than $S'(t)$ is?"

In the sequel I will sometimes consider situations involving multiple inference devices, $(X_1, Y_1), (X_2, Y_2), \ldots$, with associated domains $U_1, U_2, \ldots$. For example, I will consider scenarios where agents try to observe one another. In such situations, when referring to "U", I implicitly mean $\cap_i U_i$, implicitly restrict the domain of all X_i, Y_i to U, and implicitly assume that the codomain of each such restricted Y_i is binary.

Notation and terminology

To formalize the preceding considerations, I first fix some notation. I will take the set of binary numbers $\mathbb{B}$ to equal $\{-1, 1\}$. For any function Γ with domain U, I will write the image of U under Γ as $\Gamma(U)$. I will also sometimes abuse this notation with a sort of "set-valued function" shorthand, and so for example write $\Gamma(V) = 1$ for some $V \subset U$ iff $\Gamma(u) = 1 \; \forall u \in V$. On the other hand, for the special case where the function over U is a measure, I use conventional shorthand from measure theory. For example, if P is a probability distribution over U and $V \subset U$, I write $P(V)$ as shorthand for $\sum_{u \in V} P(u)$.

For any function Γ with domain U that I will consider, I implicitly assume that the entire set $\Gamma(U)$ contains at least two distinct elements. For any (potentially infinite) set R, $|R|$ is the cardinality of R.

Given a function Γ with domain U, I write the partition of U given by Γ^{-1} as $\overline{\Gamma}$, i.e.,

$$\overline{\Gamma} \equiv \{\{u : \Gamma(u) = \gamma\} : \gamma \in \Gamma(U)\} \tag{1}$$

I say that two functions Γ_1 and Γ_2 with the same domain U are **(functionally) equivalent** iff the inverse functions Γ_1^{-1} and Γ_2^{-1} induce the same partitions of U, i.e., iff $\overline{\Gamma_1} = \overline{\Gamma_2}$.

Recall that a partition A over a space U is a *refinement* of a partition B over U iff every $a \in A$ is a subset of some $b \in B$. If A is a refinement of B, then for every $b \in B$ there is an $a \in A$ that is a subset of b. Two partitions A and B are refinements of each other iff $A = B$. Say a partition A is finite and a refinement of a partition B. Then $|A| = |B|$ iff $A = B$. For any two functions A and B with domain U, I will say that "A refines B" if $\overline{A}$ is a refinement of $\overline{B}$. Similarly, for any $R \subset U$ and function A, I will say that "R refines A" (or "A is refined by R") if R is a subset of some element of $\overline{A}$.

I write the characteristic function of any set $R \subseteq U$ as

$$\mathcal{X}_R(u) = 1 \Leftrightarrow u \in R \tag{2}$$

As shorthand I will sometimes treat functions as equivalent to one of the values in their image. So for example expressions like "$\Gamma_1 = \Gamma_2 \Rightarrow \Gamma_3 = 1$" means "$\forall u \in U$ such that $\Gamma_1(u) = \Gamma_2(u), \Gamma_3(u) = 1$".

I define a **probe** of any variable V to be a function parametrized by a $v \in V$ of the form

$$\delta_v(v') = \begin{cases} 1 & \text{if } v = v' \\ -1 & \text{otherwise.} \end{cases} \tag{3}$$

$\forall v' \in V$. Given a function Γ with domain U I sometimes write $\delta_\gamma(\Gamma)$ as shorthand for the function $u \in U \rightarrow \delta_\gamma(\Gamma(u))$. When I don't want to specify the subscript γ of a probe, I sometimes generically write δ. I write $\mathcal{P}(\Gamma)$ to indicate the set of all probes over $\Gamma(U)$.

Weak inference

I now review some results that place severe restrictions on what a physical agent can predict and be guaranteed to be correct. To begin, I formalize the concept of an "inference device" introduced in the previous subsection.

Definition 1. *An* **(inference) device** *over a set U is a pair of functions (X, Y), both with domain U. Y is called the* **conclusion** *function of the device, and is surjective onto $\mathbb{B}$. X is called the* **setup** *function of the device.*

Given some function Γ with domain U and some $\gamma \in \Gamma(U)$, we are interested in setting up a device so that it is assured of correctly answering whether $\Gamma(u) = \gamma$ for the actual universe u. Motivated by the examples above, I will formalize this with the condition that $Y(u) = 1$ iff $\Gamma(u) = \gamma$ for all u that are consistent with some associated setup value x of the device, i.e., such that $X(u) = x$ for some x. If this condition holds, then setting up the device to have setup value x guarantees that the device will make the correct conclusion concerning whether $\Gamma(u) = \gamma$. (Hence the terms "setup function" and "conclusion function" in Def. 1.)

We can now formalize inference:

Definition 2. *Let Γ be a function over U such that $|\Gamma(U)| \geq 2$. A device $\mathcal{D}$* **(weakly) infers** *Γ iff $\forall \gamma \in \Gamma(U)$, $\exists x \in X(U)$ such that $\forall u \in U, X(u) = x \Rightarrow Y(u) = \delta_\gamma(\Gamma(u))$.*

If $\mathcal{D}$ infers Γ, I write $\mathcal{D} > \Gamma$. I say that a device $\mathcal{D}$ infers a set of functions if it infers every function in that set.

The following semi-formal example illustrates a scenario in which weak inference holds, and a related scenario in which it doesn't hold.

Example 3. *A scenario in which weak inference holds is illustrated in Fig. 1. In this example, for simplicity determinism is assumed. The full rectangle, including both colored rectangles, indicates the set of all possible histories of the universe across all time, U (i.e., the set of all "states of the world", in the language of epistemology). In this example the function Γ is whether the sky will (not) be cloudy at noon (at Greenwich, say). Since the ID is embedded in the universe, the precise question concerning the future state of the universe that it is instructed to answer picks out different subsets of the set of all possible histories of the universe across all time. There are two such sets indicated, corresponding to the ID being asked the question, "will the sky be cloudy at noon?" or being asked the question, "will the sky be clear at noon?". (Histories falling outside of both of those sets correspond to questions different from those two.) Again, since the ID is embedded in the universe, and since its answer can have two possible values, which answer it gives (say at 11am) is a partition across U. The separatrix between the two elements of that partition are indicated by the bold line. Finally, in all elements of U, the sky either will be clear at noon or will be cloudy. The two possibilities are indicated by the two colored rectangles.*

The ID weakly infers Γ, i.e., correctly predicts the state of the sky at noon, since whichever of the two possible questions it considers, it is guaranteed that its answer is correct.

A related scenario where weak inference does not hold is illustrated in Fig. 2. The only difference from the scenario depicted in Fig. 1 is that if the ID is asked the question, "will the sky be cloudy at noon?", and the sky in fact will be cloudy at noon, the ID will answer 'no' — which is incorrect.

Example 4. *While it is clearly grounded in a real-world scenario, Ex. 3 obscures the mathematical essence of weak inference. A fully abstract, stripped-down example of weak inference is given in the following table, which provides functions $X(u), Y(u)$ and $\Gamma(u)$ for all u in a space U. In this minimal example, U has only three elements:*

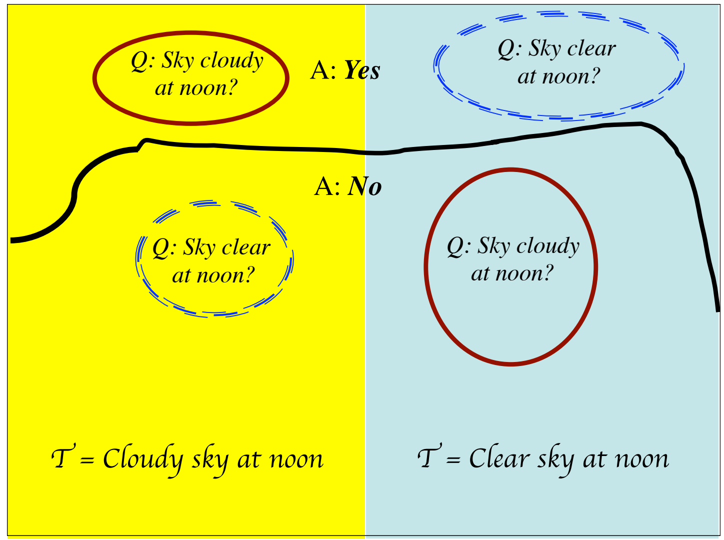

FIG. 1. An example of correct prediction as weak inference, where for simplicity determinism is assumed. The set U of all possible histories of the universe is the full rectangle, including both the yellow and blue subsets, which correspond to the two possible states of the sky at noon. Two of the possible questions of the ID are indicated: one of them is asked by the ID in all universes within the union of the two red ellipses, and the other question is asked in all universes within the union of the two blue ellipses. The ID weakly infers Γ, i.e., correctly predicts the state of the sky at noon, since whichever of the two possible questions it considers, it is guaranteed that its answer is correct.

u	$X(u)$	$Y(u)$	$\Gamma(u)$
a	1	1	1
b	2	-1	1
c	1	-1	2

In this example, $\Gamma(U) = \{1, 2\}$, so we are concerned with two probes, δ_1 and δ_2. Setting $X(u) = 2$ means that $u = b$, which in turn means that $\Gamma(u) = 1$ and $Y(u) = -1$. So setting $X(u) = 2$ guarantees that $\delta_2(\Gamma(u)) = Y(u)$ (which in this case equals -1, the answer 'no'). So the setup value $x = 2$ ensures that the ID correctly answers the binary question, "does $\Gamma(u) = 2$"? Similarly, setting $X(u) = 1$ guarantees that $\delta_1(\Gamma(u)) = Y(u)$, so that it ensures that the ID correctly answers the binary question, "does $\Gamma(u) = 1$"?

Ex. 4 shows that weak inference can hold even if $X(u) = x$ doesn't fix a unique value for $Y(u)$. Such non-uniqueness is typical when the device is being used for observation. Setting up a device to observe a variable outside of that device restricts the set of possible universes; only those u are allowed that are consistent with the observation device being set up that way to make the desired observation. But typically just setting up an observation device to observe what value a variable has doesn't uniquely fix the value of that variable. As discussed in App. B of [39], the definition of weak inference

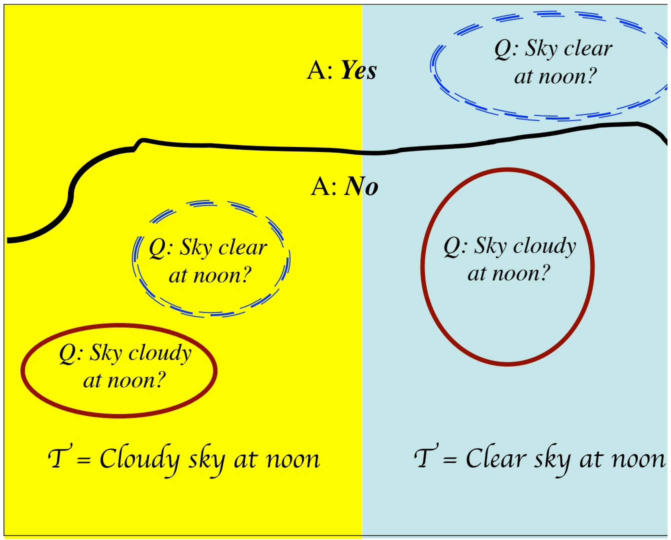

FIG. 2. An example where the prediction of an ID of the state of the sky at noon cannot be guaranteed of being correct, i.e., the ID does not weakly infer the function of u giving the state of the sky at noon. The scenario is identical to the one depicted in Fig. 1, except that if the ID is asked the question, "will the sky be cloudy at noon?", and the sky in fact will be cloudy at noon, the ID will answer 'no', which is incorrect.

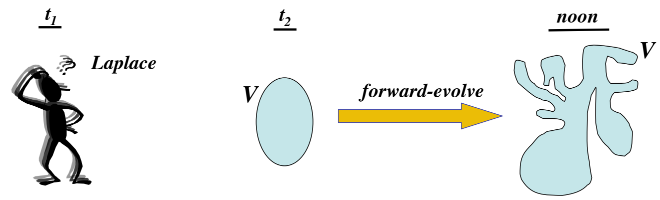

FIG. 3. The time t_1 is less than t_2, which in turn is less than noon. V is the set of all time-t_2 universes where Laplace is thinking the answer "yes" in response to the t_1 question Laplace heard — whatever that question was. V' is V evolved forward to noon. At t_1, we ask Laplace, "will the universe be outside V' at noon?" It is impossible for Laplace to answer correctly, no matter what his computational capabilities are, what the laws of the universe are, etc.

is very unrestrictive. For example, a device $\mathcal{D}$ is 'given credit' for correctly answering probe $\delta(\Gamma(u))$ if there is *any* $x \in X(U)$ such that $X(u) = x \Rightarrow Y(u) = \delta(\Gamma(u))$. In particular, $\mathcal{D}$ is given credit even if the binary question associated with x is not whether $\Gamma(u) = \gamma$, but some other question. In essence, the device receives credit even if it gets the right answer by accident.

Unless specified otherwise, a device written as "$\mathcal{D}_i$" for any integer i is implicitly presumed to have domain U, with setup function X_i and conclusion function Y_i (and similarly for no subscript). Similarly, unless specified otherwise, expressions like "$\min_{x_i}$" mean $\min_{x_i \in X_i(U)}$. I also say that a device $\mathcal{D}_1$ infers a device $\mathcal{D}_2$ iff $\mathcal{D}_1 > Y_2$, i.e., $\mathcal{D}_1$ infers $\mathcal{D}_2$ if it can infer what $\mathcal{D}_2$ will conclude. In general inference among devices is non-transitive (see [39] for an example).

The two Laplace's Demon theorems

"An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough ... nothing would be uncertain and the future just like the past would be present before its eyes."

— Pierre Simon Laplace, "A Philosophical Essay on Probabilities"

There are limitations on the ability of any device to weakly infer functions. Perhaps the most trivial is the following:

Proposition 1. *For any device $\mathcal{D}$, there is a function that $\mathcal{D}$ does not infer.*

Proof. Choose Γ to be the function Y, so that the device is trying to infer itself. Then choose the negation probe $\delta(y \in \mathbb{B}) = -y$ to see that such inference is impossible. (Also see [39].) □

It is interesting to consider the implications of Prop. 1 for the case where the inference is prediction, as in Ex. 2. Depending on how precisely one interprets Laplace, Prop. 1 means that he was wrong in his claim about the ability of an "intellect" to make accurate predictions: even if the universe were a giant clock, it could not contain an intellect that could reliably predict the universe's future state before it occurred.[5] More precisely, for all Γ as in Prop. 1, there could be an intellect $\mathcal{D}$ that can infer Γ. However Prop. 1 tells us that for any fixed intellect, there must exist a Γ that the intellect cannot infer. (See Fig. 3.) The "intellect" Laplace refers to is commonly called Laplace's "demon", so I sometimes refer to Prop. 1 as the "first (Laplace's) demon theorem".

One might think that Laplace could circumvent the first demon theorem by simply constructing a second demon, specifically designed to infer the Γ that thwarts his first demon.

[5] Similar conclusions have been reached previously [28, 31]. However in addition to being limited to the inference process of prediction, that earlier work is quite informal. It is no surprise than that some claims in that earlier work are refuted by well-established results in engineering. For example, the claim in [28] that "a prediction concerning the narrator's future ... cannot ... account for the effect of the narrator's learning that prediction" is refuted by adaptive control theory in general and by Bellman's equations in particular. Similarly, it is straightforward to see that statements (A3), (A4), and the notion of "structurally identical predictors" in [31] have no formal meaning.

Continuing in this way, one might think that Laplace could construct a set of demons that, among them, could infer any function Γ. Then he could construct an "overseer demon" that would choose among those demons, based on the function Γ that needs to be inferred. However this is not possible. To see this, simply redefine the device $\mathcal{D}$ in Prop. 1 to be the combination of Laplace with all of his demons.

This limitation on prediction hold even if the number of possible states of the universe is countable (or even finite), or if the inference device has super-Turning capabilities. It holds even if the current formulation of physics is wrong; it does not rely on chaotic dynamics, physical limitations like the speed of light, or quantum mechanical limitations.

Note as well that in Ex. 2's model of a prediction system the actual values of the times of the various events are not specified. So in particular the impossibility result of Prop. 1 still applies to that example even if $t_3 < t_2$ — in which case the time when the agent provides the prediction is *after* the event they are predicting. Moreover, consider the variant of Ex. 2 where the agent programs a computer to do the prediction, as discussed in Footnote 3 in that example. In this variant, the program that is input to the prediction computer could even contain the future value that the agent wants to predict. Prop. 1 would still mean that the conclusion that the agent using the computer comes to after reading the computer's output cannot be guaranteed to be correct.

Prop. 1 tells us that any inference device $\mathcal{D}$ can be "thwarted" by an associated function. However it does not forbid the possibility of some second device that can infer that function that thwarts $\mathcal{D}$. To analyze issues of this sort, and more generally to analyze the inference relationships within sets of multiple functions and multiple devices, we start with the following definition:

Definition 3. *Two devices* (X_1, Y_1) *and* (X_2, Y_2) *are* **(setup) distinguishable** *iff* $\forall x_1, x_2, \exists u \in U$ *such that* $X_1(u) = x_1, X_2(u) = x_2$.

No device is distinguishable from itself. Distinguishability is symmetric, but non-transitive in general (and obviously not reflexive).

Having two devices be distinguishable means that no matter how the first device is set up, it is always possible to set up the second one in an arbitrary fashion; the setting up of the first device does not preclude any options for setting up the second one. Intuitively, if two devices are not distinguishable, then the setup function of one of the devices is partially "controlled" by the setup function of the other one. In such a situation, they are not two fully separate, independent devices.

Proposition 2. *No two distinguishable devices* (X, Y) *and* (X', Y') *can weakly infer each other.*[6]

[6]In fact we can strengthen this result: If (X', Y') can weakly infer the distinguishable device (X, Y), then (X, Y) can infer neither of the two binary-valued functions equivalent to Y'.) I will call Prop. 2 the "second (Laplace's) demon theorem".

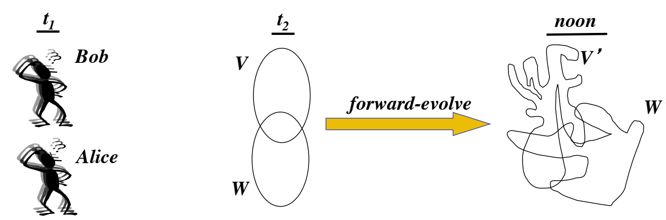

FIG. 4. The time t_1 is less than t_2, which in turn is less than noon. V is the set of all time-t_2 universes where Bob is thinking the answer "yes" in response to the t_1 question Bob heard — whatever that question was. W is the set of all time-t_2 universes where Alice is thinking the answer "yes" in response to the t_1 question Alice heard — whatever that question was. V' is V evolved forward to noon, and W' is W evolved forward to noon. At t_1, we ask Bob, "will the universe be in W' at noon?" (in other words, "Was Alice thinking 'yes' at t_2?"). At that time we also ask Alice, "will the universe be outside of V' at noon?" (in other words, "Was Bob *not* thinking 'yes' at t_2?"). It is impossible for both Bob and Alice to answer correctly, no matter what their computational capabilities are, what the laws of the universe are, etc.

See Fig. 4 for an illustration of Prop. 2, for two IDs called "Bob" and "Alice".

The second demon theorem establishes that a whole class of functions cannot be inferred by $\mathcal{D}$ (namely the conclusion functions of devices that are distinguishable from $\mathcal{D}$ and also can infer $\mathcal{D}$). More generally, let $\mathcal{S}$ be a set of devices, all of which are distinguishable from one another. Then the second demon theorem says that there can be at most one device in $\mathcal{S}$ that can infer all other devices in $\mathcal{S}$. It is important to note that the distinguishability condition is crucial to the second demon theorem; mutual weak inference can occur between non-distinguishable devices.

In [8] Barrow speculated whether "only computable patterns are instantiated in physical reality". There "computable" is defined in the sense of Turing machine theory. However we can also consider the term as meaning "can be evaluated by a real world computer". If so, then his question is answered — in the negative — by the Laplace demon theorems.

By combining the two demon theorems it is possible to establish the following:

Corollary 3. *Consider a pair of devices* $\mathcal{D} = (X, Y)$ *and* $\mathcal{D}' = (X', Y')$ *that are distinguishable from one another and whose conclusion functions are inequivalent. Say that* $\mathcal{D}'$ *weakly infers* $\mathcal{D}$. *Then there are at least three inequivalent surjective binary functions* Γ *that* $\mathcal{D}$ *does not infer.*

In particular, Coroll. 3 means that if any device in a set of distinguishable devices with inequivalent conclusion functions is sufficiently powerful to infer all the others, then each of those others must fail to infer at least three inequivalent functions.

Strong inference — inference of entire functions

As considered in computer science theory, a computer is an entire map taking an arbitrary "input" physical variable $\Gamma_1(u)$ to an "output" physical variable $\Gamma_2(u)$ [22]. It is concerned with saying how the value of $\Gamma_2(u)$ would change if the value of $\Gamma_1(u)$ changed. In contrast, weak inference is only concerned with inferring the value of a single physical variable, $\Gamma(u)$, not the relationship between two variables.

So we cannot really say that a device "infers a computer" if we only use the weak inference concept analyzed above. In this subsection we extend the theory of inference devices to include inference of entire functions. In addition to allowing us to analyze inference of computers, this lays the groundwork for the analysis in the next section of the relation between inference and algorithmic information theory.

To begin, suppose we have a function f that arises in the physical universe, in the sense that there is some S that is a function of U, along with some T, and S refines T, so that for all $s \in S(u)$, $f(s) = T(S^{-1}(s))$ is single-valued. We want to define what it means for a device to be able to emulate the entire mapping taking any $s \in S(U)$ to the associated value $T(S^{-1}(s))$.

One way to do this is to strengthen the concept of weak inference, so that for any desired input value $s \in S(U)$, the ID in question can simultaneously infer the output value $f(s)$ *while also forcing the input to have the value* s. In other words, for any pair $(s \in S(U), t \in T(U))$, by appropriate choice of $x \in X(U)$ the ID (X, Y) simultaneously answers the probe δ_t correctly (as in the concept of weak inference) *and* forces $S(u) = s$. In this way, when the ID "answers δ_t correctly", it is answering whether $f(s) = t$ correctly, for the precise s that it is setting. By being able to do this for all $s \in S(U)$, the ID can emulate the function f.

Extending this concept from single-valued functions f to multifunctions results in the following definition:

Definition 4. *Let S and T be functions both defined over U. A device (X, Y)* **strongly infers** *(S, T) iff $\forall\ \delta \in \mathcal{P}(T)$ and all $s \in S(U)$, $\exists\ x$ such that $X(u) = x \Rightarrow \{S(u) = s, Y(u) = \delta(T(u))\}$.*

If (X, Y) strongly infers (S, T) we write $(X, Y) \gg (S, T)$.

By considering the special case where $T(U) = \mathbb{B}$, we can formalize what it means for one device to emulate another device:

Definition 5. *A device (X_1, Y_1)* **strongly infers** *a device (X_2, Y_2) iff $\forall\ \delta \in \mathcal{P}(Y_2)$ and all x_2, $\exists\ x_1$ such that $X_1 = x_1 \Rightarrow X_2 = x_2, Y_1 = \delta(Y_2)$.*

See App. B in [39] for a discussion of how unrestrictive Def. 5 is.

Example 5. *Suppose $\mathcal{D}_2$ is a device that (for example) can be used to make predictions about the future state of the weather. Let Γ be the set of future weather states that the device can predict, and let X_2 be the set of possible current meteorological conditions. So if this device can in fact infer the future state of the weather, then for any question δ_γ of whether the future weather will have value γ, there is some current condition x_2 such that if $\mathcal{D}_2$ is set up with that x_2, it correctly answers whether the associated future state of the weather will be γ. On the other hand, if $\mathcal{D}_2 \not> \Gamma$, then there is some such question of the form, "will the future weather be γ?" such that for* no *input to the device of the current meteorological conditions will the device necessarily produce an answer y_2 to the question that is correct.*

One way for us to be able to conclude that some device $\mathcal{D}' = (X', Y')$ can "emulate" this behavior of $\mathcal{D}_2$ is to set up $\mathcal{D}_2$ with an arbitrary value x_2, and confirm that $\mathcal{D}'$ can infer the associated value of Y_2. So we require that for all x_2, and all $\delta \in \mathcal{P}(Y_2)$, $\exists x'$ such that if $X_2 = x_2$ and $X' = x'$, then $Y = \delta(Y_2)$.

Now define a new device $\mathcal{D}_1$, with its setup function defined by $X_1(u) = (X'(u), X_2(u))$ and its conclusion function equal to Y'. Then our condition for confirming that $\mathcal{D}'$ can emulate $\mathcal{D}_2$ gets replaced by the condition that for all x_2, and all $\delta \in \mathcal{P}(Y_2)$, $\exists x_1$ such that if $X_1 = x_1$, then $X_2 = x_2$ and $Y = \delta(Y_2)$. This is precisely the definition of strong inference.

Say we have a Turing machine (TM) T_1 that can emulate another TM, T_2 (e.g., T_1 could be a universal Turing machine (UTM), able to emulate any other TM). Such "emulation" means that T_1 can perform any particular calculation that T_2 can. The analogous relationship holds for IDs, if we translate "emulate" to "strongly infer", and translate "perform a particular calculation" to "weakly infer". In addition, like UTM-style emulation (but unlike weak inference), strong inference is transitive. These results are formalized as follows:

Proposition 4. *Let $\mathcal{D}_1$, $\mathcal{D}_2$ and $\mathcal{D}_3$ be a set of inference devices over U and Γ a function over U. Then:*

i) *$\mathcal{D}_1 \gg \mathcal{D}_2$ and $\mathcal{D}_2 > \Gamma \Rightarrow \mathcal{D}_1 > \Gamma$.*

ii) *$\mathcal{D}_1 \gg \mathcal{D}_2$ and $\mathcal{D}_2 \gg \mathcal{D}_3 \Rightarrow \mathcal{D}_1 \gg \mathcal{D}_3$.*

In addition, strong inference implies weak inference, i.e., $\mathcal{D}_1 \gg \mathcal{D}_2 \Rightarrow \mathcal{D}_1 > \mathcal{D}_2$.

Most of the properties of weak inference have analogs for strong inference:

Proposition 5. *Let $\mathcal{D}_1$ be a device over U.*

i) *There is a device $\mathcal{D}_2$ such that $\mathcal{D}_1 \not\gg \mathcal{D}_2$.*

ii) *Say that $\forall\ x_1$, $|X_1^{-1}(x_1)| > 2$. Then there is a device $\mathcal{D}_2$ such that $\mathcal{D}_2 \gg \mathcal{D}_1$.*

Strong inference also obeys a restriction that is analogous to Prop. 2, except that there is no requirement of setup-distinguishability:

Proposition 6. *No two devices can strongly infer each other.*

Recall that there are entire functions that are not computable by any TM, in the sense that no TM can correctly compute the value of that function for every input to that function. On the other hand, trivially, any single output value of a function *can* be computed by some TM (just choose the TM that prints that value and then halts). The analogous distinction holds for inference devices:

Proposition 7. *Let U be any countable space with at least two elements.*

1. *For any function Γ over U such that $|\Gamma(U)| \geq 3$ there is a device $\mathcal{D}$ that weakly infers Γ;*
2. *There is a function (S, T) over U that is not strongly inferred by any device.*

Proof. Let $X(u)$ be the identity function (so that each $u \in U$ has its own, unique value x). Choose $Y(u)$ to equal 1 for exactly one u, $\bar{u}$. Then for the probe $\delta_{\Gamma(\bar{u})}$ we can choose $x = X(\bar{u})$, so that the device correctly answers 'yes' to the question of whether $\Gamma(u) = \Gamma(\bar{u})$. For any other probe δ_γ, note that since $|\Gamma(U)| \geq 3$, there must be a $u' \in U$ such that $\Gamma(u') \neq \gamma$. Moreover, by construction $Y(u') = -1$. So if we choose x to be $X(u')$, then the device correctly answers 'no' to the question of whether $\Gamma(u') = \gamma$. This proves the first claim.

To prove the second claim, choose both $S(u) = u$ and $T(u) = u$ for all u, so that $|S(U)| = |T(U)| = |U|$. So by the first requirement for some device (X, Y) to strongly infer (S, T), it must be that for any s, there is a value of X, $x(s)$, such that $X(u) = x(s) \Rightarrow S(u) = s$. This means that $x(s)$ must be a single-valued function, for each s choosing a unique (x which in turn choose a unique) u. This means that $Y(X^{-1}(x(s))$ must equal 1, in order for the device to correctly answer 'yes' to the probe of whether $T(u) = \delta_{T(S^{-1}(s))}$. However since this is true for all $s \in S(U)$, it is true for all $u \in U$. So $Y(U)$ is a singleton, contradicting the requirement that the conclusion function of any device be binary-valued. □

INFERENCE IN STOCHASTIC UNIVERSES

Stochastic inference

There are several ways to extend the analysis above to incorporate a probability measure P over U, so that inference is not exact, but only holds under some probability. In this subsection we present some of the elementary properties of one such measure of stochastic inference.

Note that once there is a distribution over U, all functions like X, Y and Γ become random variables. Also recall that $\delta_\gamma(\Gamma)$ is shorthand for the function $u \in U \to \delta_\gamma(\Gamma(u))$ — and so is a random variable. The measure of stochastic inference we will consider here is defined as follows:

Definition 6. *Let $P(u \in U)$ be a probability measure and Γ a function with domain U and finite range. Then we say that a device (X, Y) (weakly) infers Γ* **with (covariance) accuracy**

$$cov(\mathcal{D}, \Gamma) := \frac{\sum_{\delta \in \mathcal{P}(\Gamma)} \max_x \left[\mathbb{E}_P(Y\delta(\Gamma) \mid x)\right]}{|\Gamma(U)|}$$

Writing it out explicitly, for countable U, the numerator in Def. 6 is

$$\sum_{\gamma \in \Gamma(U)} \max_{x \in X(U)} \Big[\sum_u Y(u)\delta_\gamma(\Gamma(u))P(u \mid x)\Big] \tag{4}$$

Intuitively, this is a probe-averaged, best-case (over $x \in X(U)$) probability of answering the probe correctly.

Covariance accuracy is a way to quantify the degree to which $\mathcal{D} > \Gamma$ when the inference is subject to uncertainty. Clearly, $cov(\mathcal{D}, \Gamma) \leq 1.0$, and if P is nowhere 0, then $cov(\mathcal{D}, \Gamma) = 1.0$ iff $\mathcal{D} > \Gamma$.[7] Covariance accuracy obeys the following bound:

Proposition 8. *Let P be a probability measure over U, $\mathcal{D} = (X, Y)$ a device, and Γ a function over U with finite $|\Gamma(U)|$. Then*

$$cov(\mathcal{D}, \Gamma) \geq \frac{(2 - |\Gamma(U)|)\max_x \left[\mathbb{E}_P(Y \mid x)\right]}{|\Gamma(U)|}$$

Proof. For any probe δ_γ of $\gamma \in \Gamma(U)$, let $M_\gamma = \max_x \left[\mathbb{E}_P(Y\delta_\gamma(\Gamma) \mid x)\right]$. Define $x_m := \mathrm{argmax_x}[\mathrm{\mathbb{E}_P(Y \mid x)}]$. Then $M_\gamma \geq \mathbb{E}_P(Y\delta_\gamma(\Gamma) \mid x_m)$ and

$$\begin{aligned}
cov(\mathcal{D}, \Gamma) &= \frac{\sum_{\gamma \in \Gamma(U)} M_\gamma}{|\Gamma(U)|} \geq \frac{\sum_{\gamma \in \Gamma(U)} \mathbb{E}_P(Y\delta_\gamma(\Gamma) \mid x_m)}{|\Gamma(U)|} \\
&= \frac{\sum_u P(u \mid x_m) \sum_\gamma Y(u)\delta_\gamma(\Gamma(u))}{|\Gamma(U)|} \\
&= \frac{\sum_u P(u \mid x_m)(2 - |\Gamma(U)|)Y(u)}{|\Gamma(U)|} \\
&= \frac{(2 - |\Gamma(U)|)\mathbb{E}_P(Y \mid x_m)}{|\Gamma(U)|} \\
&= \frac{(2 - |\Gamma(U)|)\max_x \left[\mathbb{E}_P(Y \mid x)\right]}{|\Gamma(U)|}.
\end{aligned}$$

□

This bound is sharp, as can be seen from the following example.

Example 6. *Given any device $\mathcal{D}$ and $|\Gamma(U)| < \infty$, divide each cell of the partition $X \times Y$ into $|\Gamma(U)|$ parts of equal probability and map them to $1, \ldots, |\Gamma(U)|$, so $\Gamma(U) = \{1, \ldots, |\Gamma(U)|\}$. For any given $x \in X$, let $a_x = P(Y = 1 \mid x), b_x = P(Y = -1 \mid x)$. For any $x \in X(U), \gamma \in \Gamma(U), \delta_\gamma$ the corresponding probe,*

$$\begin{aligned}
\mathbb{E}_P(Y\delta_\gamma(\Gamma) \mid x) &= \frac{a_x + (|\Gamma(U)| - 1)b_x - (|\Gamma(U)| - 1)a_x + b_x}{|\Gamma(U)|} \\
&= \frac{(2 - |\Gamma(U)|)(a_x - b_x)}{|\Gamma(U)|} = \frac{(2 - |\Gamma(U)|)\mathbb{E}_P(Y \mid x)}{|\Gamma(U)|}.
\end{aligned}$$

Then

$$\begin{aligned}
M_m &= \max_x \left[\mathbb{E}_P(Y\delta_\gamma(\Gamma) \mid x)\right] \\
&= \frac{(2 - |\Gamma(U)|)\ \max_x \left[\mathbb{E}_P(Y \mid x)\right]}{|\Gamma(U)|}
\end{aligned}$$

---

[7] A subtlety with the definition of an inference devices arises in this stochastic setting: we can either require that Y be surjective, as in Def. 1, or instead require that Y be "stochastically surjective" in the sense that $\forall y \in \mathbb{B}$, $\exists u$ with non-zero probability such that $Y(u) = y$. The distinction between requiring surjectivity and stochastic surjectivity of Y will not arise here.

	(X,Y)	(X',Y')	Γ	P(.)
A	(1,1)	(1,1)	1	(1 - p)/8
B	(1.-1)	(1.-1)	1	"
C	(1,1)	(2,-1)	1	"
D	(1,-1)	(2,1)	1	"
E	(2,1)	(3,1)	-1	"
F	(2,-1)	(3,-1)	-1	"
G	(2,1)	(4,-1)	-1	"
H	(2,-1)	(4,1)	-1	"
I	(1,1)	(5,1)	1	p/2
J	(2,-1)	(5,-1)	1	"

FIG. 5. Specification of a scenario in which the stochastic version of Prop. 4(i), concerning "distributivity" of weak inference through strong inference, fails drastically.

and

$$cov(\mathcal{D}, \Gamma) = \frac{(2 - |\Gamma(U)|)\ \max_x [\mathbb{E}_P(Y \mid x)]}{|\Gamma(U)|}$$

The term $\frac{2-|\Gamma(U)|}{|\Gamma(U)|}$ in Prop. 8 depends only on the size of the space $\Gamma(U)$.[8] The other term, $\max_x(\mathbb{E}_P(Y \mid x))$, can be viewed as a measure of the "inference power" of the device, by analogy with the power of a statistical test. It quantifies the device's ability to say 'yes'.

In the previous section some *a priori* restrictions on the capabilities of IDs were presented. These restrictions involved whether certain properties of IDs can(not) be guaranteed with complete certainty. When we have a probability distribution over U it is appropriate to replace consideration of "guaranteed" properties with consideration of properties that are likely but not necessarily guaranteed, e.g., as quantified with covariance accuracy. When we do that the restrictions of the previous section get modified, sometimes quite substantially. This is illustrated in the next two propositions.

First, by Prop. 4(i), if for devices $\mathcal{D}_1$, $\mathcal{D}_2$ and function Γ, $\mathcal{D}_1 \gg \mathcal{D}_2$ and $\mathcal{D}_2 > \Gamma$, then $\mathcal{D}_1 > \Gamma$. In covariance terms, this says that if $\mathcal{D}_1 \gg \mathcal{D}_2$ and $cov(\mathcal{D}_2, \Gamma) = 1.0$, then $cov(\mathcal{D}_1, \Gamma) = 1.0$. What happens to $cov(\mathcal{D}_1, \Gamma)$ if $cov(\mathcal{D}_2, \Gamma) < 1.0$? A partial answer is given by the following result:

Proposition 9. *There are devices $\mathcal{D}$, $\mathcal{D}'$, probability distribution P defined over U, and function Γ, such that $\mathcal{D}' \gg \mathcal{D}$ and $cov(\mathcal{D}, \Gamma)$ is arbitrarily close to 1.0 while $cov(\mathcal{D}', \Gamma) = 0$.*

Proof. The proof is by example.

Let U have ten states, labeled A, . . . , J and suppose that the functions $P, \Gamma, \mathcal{D} = (X, Y)$ and $\mathcal{D}' = (X', Y')$ are as in Fig. 5, with $0 \leq p \leq 1$.

1. To verify that $\mathcal{D}' \gg \mathcal{D}$, for the 1-probe, for $x = 1, 2$, choose $x' = 1, 3$, respectively. For the -1-probe, for $x = 1, 2$, choose $x' = 2, 4$, respectively.

2. $cov(\mathcal{D}, \Gamma) = p$. To see this, for the 1-probe, evaluate $\max_x \mathbb{E}_P(Y\delta_1(\Gamma) \mid x) = p$, the maximum occurring for $x = 1$. Similarly, for the -1-probe, evaluate $\max_x \mathbb{E}_P(Y\delta_{-1}(\Gamma) \mid x) = p$, the maximum occurring for $x = 2$.

3. $cov(\mathcal{D}', \Gamma) = 0$. To see this for both probes, note that $\mathbb{E}_P(Y'\delta(\Gamma) \mid x') = 0$ for each x'.

The proof is completed by taking $p \to 1$. □

To understand Prop. 9, recall that the definition of $\mathcal{D}' \gg \mathcal{D}$ requires that for any $x \in X(U)$ and for any probe $\delta_\gamma \in \mathcal{P}(\Gamma)$, there be *some* x' and associated $X'^{-1}(x') \subseteq U$ for which $\mathcal{D}'$ successfully emulates $\mathcal{D}$'s behavior at inferring δ_γ. If the inference $\mathcal{D} > \Gamma$ is perfect, then $\mathcal{D}'$ also infers Γ. However, if the inference $\mathcal{D} > \Gamma$ is only partially correct, then that value x' and associated subset of U, under which $\mathcal{D}' \gg \mathcal{D}$ may be precisely those u for which $\mathcal{D}$ performs badly at inferring δ_γ. Thus, $\mathcal{D}$ may do an excellent, though imperfect, job overall of inferring Γ while $\mathcal{D}'$ fails completely.

The second example of how the restrictions of the previous section get modified by introducing a probability distribution is that this makes the second Laplace's impossibility theorem become "barely true":

Proposition 10. *There are devices $\mathcal{D}$ and $\mathcal{D}'$ with X and X' setup-distinguishable and a distribution P where both $cov(\mathcal{D}, \mathcal{D}')$ and $cov(\mathcal{D}', \mathcal{D})$ are arbitrarily close to 1.*

Proof. The proof is by example.

Let U have sixteen states, labeled A, . . ., P and suppose that the functions $P, \Gamma, \mathcal{D} = (X, Y)$ and $\mathcal{D}' = (X', Y')$ are as in Fig. 6, with arbitrary $0 < b < 1/6$, and $a = (1 - 6b)/2$.

By inspection, X and X' are setup distinguishable. Next, plugging in yields $cov(\mathcal{D}, \mathcal{D}') = cov(\mathcal{D}, Y') = a/(a + b)$. Moreover $cov(\mathcal{D}', \mathcal{D}) = cov(\mathcal{D}, \mathcal{D}')$ by symmetry of the columns in Fig. 6. ($\mathbb{E}_P(YY' \mid X = 1) = (a - b)/(a + b)$ and $\mathbb{E}_P(YY' \mid X = -1) = -1$.)

So by taking b arbitrarily close to 0, both of the covariances can be made arbitrarily close to 1. □

Prop. 10 shows that in a certain sense, as soon as any stochasticity is introduced into the universe, having two devices be setup-distinguishable no longer restricts their ability to simultaneously infer each other. However if we replace setup-distinguishability with the property that the setup functions of the two devices are statistically independent, then we recover strong restrictions on simultaneous inference.

To illustrate this, let M be the four-dimensional hypercube $\{0, 1\}^4$. Define the following three functions over $\vec{z} \in M$:

1. $k(\vec{z}) = z_1 + z_4 - z_2 - z_3$;

2. $m(\vec{z}) = (z_2 - z_4)$;

[8] Note that this term $[2 - |\Gamma(U)|]\ /\ |\Gamma(U)|$ can be negative for $|\Gamma(U)| > 2$. This reflects our use of expected values and the convention that $\mathbb{B} = \{-1, 1\}$.

	(X,Y)	(X',Y')	P()
A	(1,1)	(1,1)	a
B	(1,1)	(1,-1)	0
C	(1,-1)	(1,1)	0
D	(1,-1)	(1,-1)	a
E	(1,1)	(-1,1)	0
F	(1,1)	(-1,-1)	b
G	(1,-1)	(-1,1)	b
H	(1,-1)	(-1,-1)	0
I	(-1,1)	(1,1)	0
J	(-1,1)	(1,-1)	b
K	(-1,-1)	(1,1)	b
L	(-1,-1)	(1,-1)	0
M	(-1,1)	(-1,1)	0
N	(-1,1)	(-1,-1)	b
O	(-1,-1)	(-1,1)	b
P	(-1,-1)	(-1,1)	0

FIG. 6. Specification of a scenario in which the stochastic version of Prop. 2, concerning simultaneous inference of two setup-distinguishable IDs, fails drastically.

3. $n(\vec{z}) = (z_3 - z_4)$.

Proposition 11. *Let P be a probability measure over U, and $\mathcal{D}_1$ and $\mathcal{D}_2$ two devices where $X_1(U) = X_2(U) = \mathbb{B}$, and those variables are statistically independent under P. Define $P(X_1 = -1) \equiv \alpha$ and $P(X_2 = -1) \equiv \beta$. Say that $\mathcal{D}_1$ infers $\mathcal{D}_2$ with accuracy ϵ_1, while $\mathcal{D}_2$ infers $\mathcal{D}_2$ with accuracy ϵ_2. Then*

$$\epsilon_1\epsilon_2 \;\leq\; max_{\vec{z}\in M}\Big|\alpha\beta[k(\vec{z})]^2 + \alpha k(\vec{z})m(\vec{z}) + \beta k(\vec{z})n(\vec{z}) + m(\vec{z})n(\vec{z})\Big|.$$

In particular, if $\alpha = \beta = 1/2$, then

$$\begin{aligned} \epsilon_1\epsilon_2 &\leq \frac{max_{\vec{z}\in M} \mid (z_1 - z_4)^2 - (z_2 - z_3)^2 \mid}{4} \\ &= 1/4. \end{aligned}$$

The maximum for $\alpha = \beta = 1/2$ can occur in several ways. One is when $z_1 = 1$, and z_2, z_3, z_4 all equal 0. At these values, both devices have an inference accuracy of 1/2 at inferring each other. Each device achieves that accuracy by perfectly inferring one probe of the other device, while performing randomly for the remaining probe.

The ID framework as developed to date has no function measuring distance, nor one measuring time. So at present, one cannot even formulate an ID-analog of Heisenberg's uncertainty principle, never mind try to derive it. It is intriguing that despite this, Prop. 11 is a bound on the product of uncertainties, exactly like Heisenberg's uncertainty principle. This suggests it may be worth exploring extensions of the ID framework that do involve distance and time, to see what *a priori* constraints there might be on the product of uncertainties of two IDs that are measuring different aspects of the same system. (This idea is returned to in the last section below.)

Finally, it should be noted that there are other ways to quantify the degree of weak inference when there is intrinsic uncertainty, in addition to covariance accuracy. For example, we could change Def. 6 by replacing the sum over all probes δ and associated division by $|\Gamma(U)|$ with a minimum over all probes δ. (This amounts to replacing an average-best-case expression with a worst-case expression.)

The complexity of inference

Constraints on what can be computed by a physical device can be derived from the laws of physics [25]. There have also been attempts to go the other way, and derive constraints on the laws of physics from computation theory, in particular from algorithmic information theory (AIT) [13, 23, 43–46]. These often implicitly involve uncertainty about the state of the universe. For example, the use of Kolmogorov complexity to model physical reality is often intimately related to the use of algorithmic probability [23, 33, 47]. (Indeed, the very first line in [33] is "The probability distribution P from which the history of our universe is sampled represents a theory of everything".) One way to justify consideration of such a probability distribution in the first place is to identify it with uncertainty of some agent (e.g., a scientist) concerning the state of the universe.

This importance of an agent in attempts to analyze physics using AIT suggests we extend the inference device framework to include structures similar to those considered in AIT. There are several ways to extend the ID framework this way. In this subsection I sketch the starting point for one of them.

Given a TM T, the *Kolmogorov complexity* of an output string s is defined as the size of the smallest input string s' that when input to T produces s as output. To construct our inference device analog of this, we need to define the "size" of an input region of an inference device $\mathcal{D}$. To do this, we assume we are given a measure $d\mu$ over U, and for simplicity restrict attention to functions Γ over U with countable range. Then we define the **size** of $\gamma \in \Gamma(U)$ as -ln$[\int_{\Gamma^{-1}(\gamma)} d\mu(u)\ 1]$, i.e., the negative logarithm of the measure of all $u \in U$ such that $\Gamma(u) = \gamma$.[9] We write this size as $\mathcal{M}_{\mu;\Gamma}(\gamma)$, or just $\mathcal{M}(\gamma)$ for short.[10]

We define inference complexity in terms of such a size function using the shorthand introduced just below Eq. (2):

Definition 7. *Let $\mathcal{D}$ be a device and Γ a function over U where $X(U)$ and $\Gamma(U)$ are countable and $\mathcal{D} > \Gamma$. The* **inference complexity** *of Γ with respect to $\mathcal{D}$ and measure μ is defined*

[9] As usual, if U is countable, μ is a point measure, and the integral is a sum.

[10] If $\int d\mu(u)\ 1 = \infty$, then we instead work with differences in logarithms of volumes, evaluated under an appropriate limit of $d\mu$ that takes $\int d\mu(u)\ 1 \to \infty$. For example, we might work with such differences when U is taken to be a box whose size goes to infinity.

as

$$C_\mu(\Gamma;\mathcal{D}) \;\triangleq\; \sum_{\delta\in\mathcal{P}(\Gamma)} min_{x:X=x\Rightarrow Y=\delta(\Gamma)}[\mathcal{M}_{\mu,X}(x)].$$

In the sequel I will often have the measure implicit, and (for example) simply write C rather than C_μ. I will also mostly restrict attention to the case where μ is either a distribution or a semi-measure.

As an example, for the case where inference models the process of prediction, Γ corresponds to a potential future state of some system S external to $\mathcal{D}$. In this case $C(\Gamma;\mathcal{D})$ is a measure of how difficult it currently is for $\mathcal{D}$ to predict that future state of S. Loosely speaking, the more sensitively that future state depends on current conditions, the greater the inference complexity of predicting that future state.

Inference complexity of any function Γ with respect to a device (X, Y) is bounded by the Shannon entropy of $\mu(X)$:

Proposition 12. *For any ID $\mathcal{D}$, probability distribution μ, and function Γ with a countable image such that $\mathcal{D} > \Gamma$,*

$$C_\mu(\Gamma;\mathcal{D}) \leq |\Gamma| \times H_\mu(X)$$

where $H_\mu(X)$ is the Shannon entropy of $\mu(X)$.

Proof. Expand

$$\begin{aligned}\sum_{\delta\in\mathcal{P}(\Gamma)} \min_{x:X=x\Rightarrow Y=f(\Gamma)}[\mathcal{M}_{\mu,X}(x)] &\leq \sum_{x\in X(U)} \mathcal{M}_{\mu,X}(x)\\ &\leq -|\Gamma| \sum_{x\in X(U)} \frac{\log_2 \mu(x)}{|\Gamma|}\\ &\leq |\Gamma| H_\mu(X)\end{aligned}$$

□

Kolmogorov complexity concerns TMs computing a single output, rather than TMs emulating an entire function from inputs to outputs. The field of algorithmic information theory then analyzes the relation between Kolmogorov complexity and UTMs, i.e., TMs that emulate entire functions from inputs to outputs. Analogously, inference complexity concerns inferring a single value of a variable, i.e., it is defined in terms of *weak* inference. So to investigate the inference device analog of algorithmic information theory means investigating the relation between inference complexity and IDs that emulate entire functions.

To begin, recall perhaps the most fundamental result in AIT, the *invariance theorem*. This theorem gives an upper bound on the difference between the Kolmogorov complexity of a string using a particular UTM T_1 and its complexity if using a different UTM, T_2. This bound is independent of the computation to be performed, and can be viewed as the Kolmogorov complexity of T_1 emulating T_2. Similarly, we can bound how much greater the inference complexity of a function can be for a device $\mathcal{D}_1$ than it is for a different device $\mathcal{D}_2$ if $\mathcal{D}_1$ can strongly infer $\mathcal{D}_2$:

Proposition 13. *Let $\mathcal{D}_1$ and $\mathcal{D}_2$ be two devices and Γ a function over U where $\Gamma(U)$ is finite, $\mathcal{D}_1 \gg \mathcal{D}_2$, and $\mathcal{D}_2 > \Gamma$. Then for any distribution μ,*

$$\begin{aligned}C_\mu(\Gamma;\mathcal{D}_1) - C_\mu(\Gamma;\mathcal{D}_2) \;&\leq\; |\Gamma(U)| \;\times\\ &\max_{x_2}\Big(\min_{x_1:\{X_1=x_1\Rightarrow X_2=x_2, Y_1=Y_2\}}[\mathcal{M}_{\mu,X_1}(x_1) - \mathcal{M}_{\mu,X_2}(x_2)]\Big).\end{aligned}$$

Note that since $\mathcal{M}_{\mu,X_1}(x_1) - \mathcal{M}_{\mu,X_2}(x_2) = \ln\Big[\frac{\mu(X_2^{-1}(x_2))}{\mu(X_1^{-1}(x_1))}\Big]$, the bound in Prop. 13 is independent of the units with which one measures volume in U. (Cf. footnote 10.) Furthermore, note that that $X_1 = x_1 \Rightarrow X_2 = x_2, Y_1 = Y_2$ iff $X_1^{-1}(x_1) \subseteq X_2^{-1}(x_2) \cap (Y_1 Y_2)^{-1}(1)$. Accordingly, for all (x_1, x_2) pairs arising in the bound in Prop. 13, $\frac{\mu(X_2^{-1}(x_2))}{\mu(X_1^{-1}(x_1))} \geq 1$. So the upper bound in Prop. 13 is always non-negative.

The max-min expression on the RHS is independent of Γ. Accordingly, the bound in Prop. 13 is independent of all aspects of Γ except the cardinality of $\Gamma(U)$. Intuitively, the bound is $|\Gamma(U)|$ times the worst-case amount of "computational work" that $\mathcal{D}_1$ has to do to "emulate" $\mathcal{D}_2$'s behavior for some particular value of x_2.

Suppose that it takes a lot of computational work for $\mathcal{D}_2$ to infer Γ, and so it also takes a lot of computational work for $\mathcal{D}_1$ to infer Γ by emulating $\mathcal{D}_2$. However, it might take very little work for $\mathcal{D}_1$ to infer Γ directly. In fact, it may even be that $C(\Gamma;\mathcal{D}_1) < C(\Gamma;\mathcal{D}_2)$:

Proposition 14. *There are devices $\mathcal{D}$, $\mathcal{D}'$, probability distribution P defined over U, and function Γ, such that $\mathcal{D} > \Gamma$, $\mathcal{D}' \gg \mathcal{D}$, and $C_P(\Gamma;\mathcal{D})$ is arbitrarily large, while $C_P(\Gamma;\mathcal{D}')$ is arbitrarily close to the minimum value of $\Big|\Gamma\Big| \times \ln(|\Gamma(U)|)$.*

Proof. The proof is by example.

Let U have twelve states, labeled A, . . . , L and suppose that the functions $P, \Gamma, \mathcal{D}' = (X, Y)$ and $\mathcal{D} = (X, Y)$ are as in Fig. 7, with $1/4 < p < 1$.

1. To verify that $\mathcal{D} > \Gamma$, for the 1-probe, choose $x = 1$. For the -1 probe, choose $x = 2$.
2. To verify that $\mathcal{D}' \gg \mathcal{D}$, first, for the 1-probe, for $x = 1, 2, 3$, choose $x' = 1, 3, 5$, respecitively. Then for the -1-probe, for $x = 1, 2, 3$, choose $x' = 2, 4, 6$, respectively.
3. To verify that $C(\Gamma;\mathcal{D})$ can be arbitrarily large, first expand it as $-2\ln((1-p)/2) = 2\ln(2) - 2\ln(1-p)$. (For the 1-probe, $x = 1$ and $\mathcal{M}_{P,X}(x) = -\ln((1-p)/2)$ and similarly for the -1-probe and $x = 2$.)
4. To verify that $C(\Gamma;\mathcal{D}')$ can be arbitrarily close to its minimal value, write it as $-2\ln(p/2) = 2\ln(2) - 2\ln(p)$. (For the 1-probe, $x' = 5$ and $\mathcal{M}_{P,X}(x) = -\ln(p/2)$ and similarly for the -1-probe and $x' = 6$.)

Finally, by taking p arbitrarily close to 1, $C(\Gamma;\mathcal{D})$ becomes arbitrarily large while $C(\Gamma;\mathcal{D}')$ becomes arbitrarily close to the minimum of $2\ln(2)$. □

	(X,Y)	(X',Y')	Γ	$P(.)$
A	(1,1)	(1,1)	1	$(1-p)/8$
B	(1, -1)	(1, -1)	-1	”
C	(1, 1)	(2, -1)	1	”
D	(1 -1)	(2, 1)	-1	”
E	(2, 1)	(3, 1)	-1	”
F	(2, -1)	(3, -1)	1	”
G	(2, 1)	(4, -1)	-1	”
H	(2, -1)	(4, 1)	1	”
I	(3, 1)	(5, 1)	1	$p/4$
J	(3, -1)	(5, -1)	-1	”
K	(3, 1)	(6, -1)	-1	”
L	(3, -1)	(6, 1)	1	”

FIG. 7. Scenario illustrating discrepancies of complexities of two IDs where one strongly infers the other.

Although there is not space to analyze them here, it is worth noting that there are several ways to translate some of the mathematical structures of algorithmic information theory into the inference device framework. For example, just as a given Turing machine may fail to produce an output for some specific input, so an inference device may fail to reach a conclusion for some specific setup. This motivates the following definition:

Definition 8. *A device (X, Y) **halts** for setup value x iff $X = x \Rightarrow Y = y$ for some single value y.*

We say that x is a "halting setup" if (X, Y) halts for x. As usual, we say that an ID is **total**, or **recursive** iff it halts for all $x \in X(U)$. So an ID (X, Y) is recursive iff X refines Y.

Given this definition of what it means for a device to halt on a given input, we can define the inference analog of a "prefix-free Turing machine":

Definition 9. *Given a semi-measure μ, a device (X, Y) is **prefix(-free)** iff*

$$\sum_{x:\mathcal{D}\ halts\ on\ x} 2^{-\mathcal{M}_{\mu,X}(x)} \leq 1$$

By Kraft's inequality, if $\mathcal{D}$ is prefix-free for semi-measure μ, then there is a prefix-free code for the halting $x \in X(U)$. Therefore we can identify those x with semi-infinite bit strings, or equivalently with the natural numbers [23].

As a final example, note that the min over x's in Def. 7 is a direct analog of the min in the definition of Kolmogorov complexity (there the min is over those strings that when input to a particular UTM result in the desired output string). A natural modification to Def. 7 is to remove the min by considering all x's that cause $Y = \delta(\Gamma)$, not just of one of them:

$$\begin{aligned}\hat{C}(\Gamma;\mathcal{D}) &\triangleq \sum_{\delta\in\mathcal{P}(\Gamma)} -\ln\left[\,\mu\left(\cup_{x:X=x\Rightarrow Y=\delta(\Gamma)} X^{-1}(x)\right)\,\right] \\ &= \sum_{\delta\in\mathcal{P}(\Gamma)} -\ln\left[\sum_{x:X=x\Rightarrow Y=\delta(\Gamma)} e^{-\mathcal{M}(x)}\right],\end{aligned}$$

where the equality follows from the fact that for any $x, x' \neq x$, $X^{-1}(x) \cap X^{-1}(x') = \varnothing$. The argument of the ln(.) in this modified version of inference complexity has a direct analog in TM theory: The sum, over all input strings s to a UTM that generates a desired output string s', of $2^{-n(s)}$, where $n(s)$ is the bit size of s. (This is sometimes known as the "algorithmic" or "Solomonoff" probability of s' [23].)

MODELING THE PHYSICAL UNIVERSE IN TERMS OF INFERENCE DEVICES

I now expand the scope of the discussion to allow sets of many inference devices and / or many functions to be inferred. Some of the philosophical implications of the ensuing results are then discussed in the next subsection.

Formalization of physical reality involving Inference Devices

Define a **reality** as a pair $(U; \{F_\phi\})$ where the space U is the **domain** of the reality, and $\{F_\phi\}$ is a (perhaps uncountable) non-empty set of functions all having domain U. We are particularly interested in **device realities** in which some of the functions are binary-valued, and we wish to pair each of those functions uniquely with some of the other functions. Such realities can be written as the triple $(U; \{(X_\alpha, Y_\alpha)\}; \{\Gamma_\beta\}) \equiv (U; \{\mathcal{D}_\alpha\}; \{\Gamma_\beta\})$ where $\{\mathcal{D}_\alpha\}$ is a set of devices over U and $\{\Gamma_\beta\}$ a set of functions over U.

Define a **universal device** as any device in a reality that can strongly infer all other devices and weakly infer all functions in that reality. Prop. 6 means that no reality can contain more than one universal device. So in particular, if a reality contains a universal device and there is a given distribution over U, then the reality has a unique natural choice for an inference complexity measure, namely the inference complexity with respect to its (unique) universal device. (This contrasts with Kolmogorov complexity, which depends on the arbitrary choice of what UTM to use.)

For simplicity, assume the set of ϕ is countable, indexed $\phi_1, \phi_2, \ldots$. It is useful to define the **reduced form** of a reality $(U; \{F_\phi\})$ as the image of the function $u \rightarrow (F_{\phi_1}(u), F_{\phi_2}(u), \ldots)$. In particular, the reduced form of a device reality is the set of all tuples $([x_1, y_1], [x_2, y_2], \ldots; \gamma_1, \gamma_2, \ldots)$ for which $\exists\ u \in U$ such that simultaneously $X_1(u) = x_1, Y_1(u) = y_1, X_2(u) = x_2, Y_2(u) = y_2, \ldots; \Gamma_1(u) = \gamma_1, \Gamma_2(u) = \gamma_2, \ldots$. By working with reduced forms of realities, we dispense with the need to explicitly discuss U entirely.[11]

Example 7. *Take U to be the set of all possible histories of a universe across all time that are consistent with the laws of physics. So each u is a specification of a trajectory of the state of the entire universe through all time. The laws of physics are then embodied in restrictions on U. For example, if one wants to consider a universe in which the laws of physics are time-reversible and deterministic, then we require that no two distinct members of U can intersect. Similarly, properties like time-translation invariance can be imposed on U, as can more elaborate laws involving physical constants.*

Next, have $\{\Gamma_\beta\}$ be a set of physical characteristics of the universe, each characteristic perhaps defined in terms of the values of one or more physical variables at multiple locations and/or multiple times. Finally, have $\{\mathcal{D}_\alpha\}$ be all prediction / observation systems concerning the universe that all scientists might ever be involved in.

In this example the laws of physics are embodied in U. The implications of those laws for the relationships among the agent devices $\{\mathcal{D}_\alpha\}$ and the other characteristics of the universe $\{\Gamma_\beta\}$ is embodied in the reduced form of the reality. Viewing the universe this way, it is the $u \in U$, specifying the universe's state for all time, that has "physical meaning". The reduced form instead is a logical implication of the laws of the universe. In particular, our universe's u picks out the tuple given by the Cartesian product $[\bigtimes_\alpha \mathcal{D}_\alpha(u)] \times [\bigtimes_\beta \Gamma_\beta(u)]$ from the reduced form of the reality.

As an alternative we can view the reduced form of the reality as encapsulating the "physical meaning" of the universe. In this alternative u does not have any physical meaning. It is only the relationships among the inferences about u that one might want to make and the devices with which to try to make those inferences that has physical meaning. One could completely change the space U and the functions defined over it, but if the associated reduced form of the reality does not change, then there is no way that the devices in that reality, when considering the functions in that reality, can tell that they are now defined over a different U. In this view, the laws of physics i.e., a choice for the set U, are simply a calculational shortcut for encapsulating patterns in the reduced form of the reality. It is a particular instantiation of those patterns that has physical meaning, not some particular element $u \in U$.

See [36] for another perspective on the relationship between physical reality and mathematical structures.

[11] This means that all of the non-stochastic analysis of the previous sections can be reduced to satisfiability statements concerning sets of categorial variables. For example, the fact that a device cannot weakly infer itself is equivalent to the statement that there is no countable space X with at least two elements and associated set of pairs $\mathcal{U} = \{(x_i, y_i)\}$ where all $y_i \in \mathbb{B}$, such that for both probes δ of y_i, there is some value $x' \in X$ such that in all pairs $(x', y) \in \mathcal{U}$, $y = \delta(y)$.

Given a reality $(U; \{(X_1, Y_1), (X_2, Y_2), \ldots\})$, we say that a pair of devices in it are **pairwise (setup) distinguishable** if they are distinguishable. We say that the reality as a whole is **mutually (setup) distinguishable** iff $\forall\ x_1 \in X_1(U), x_2 \in X_2(U), \ldots\ \exists\, u \in U$ s.t. $X_1(u) = x_1, X_2(u) = x_2, \ldots$.

Proposition 15. **i)** *There exist realities $(U; \mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3)$ where each pair of devices is pairwise setup distinguishable and $\mathcal{D}_1 > \mathcal{D}_2 > \mathcal{D}_3 > \mathcal{D}_1$.*

ii) *There exists no reality $(U; \{\mathcal{D}_i : i \in \mathscr{N} \subseteq \mathbb{N}\})$ where the devices are mutually distinguishable and for some integer n, $\mathcal{D}_1 > \mathcal{D}_2 > \ldots > \mathcal{D}_n > \mathcal{D}_1$.*

iii) *There exists no reality $(U; \{\mathcal{D}_i : i \in \mathscr{N} \subseteq \mathbb{N}\})$ where for some integer n, $\mathcal{D}_1 \gg \mathcal{D}_2 \gg \ldots \gg \mathcal{D}_n \gg \mathcal{D}_1$.*

There are many ways to view a reality with a countable set of devices $\{\mathcal{D}_i\}$ as a graph, for example by having each node be a device while the edges between the nodes concern distinguishability of the associated devices, or concern whether one weakly infers the other, etc. In particular, given a countable reality, define an associated directed graph by identifying each device with a separate node in the graph, and by identifying each relationship of the form $\mathcal{D}_i \gg \mathcal{D}_j$ with a directed edge going from node i to node j. We call this the **strong inference graph** of the reality.

Prop. 7(ii) means that no reality with $|U| > 3$ can have a universal device if the reality contains all functions defined over U. Suppose that this is not the case, so that the reality may contain a universal device. Prop. 6 means that such a universal device must be a root node of the strong inference graph of the reality and that there cannot be any other root node. In addition, by Prop. 4(ii), we know that every node in a reality's strong inference graph with successor nodes has edges that lead directly to every one of those successor nodes (whether or not there is a universal device in the reality). By Prop. 15(iii) we also know that a reality's strong inference graph is acyclic.

Note that even if a device $\mathcal{D}_1$ can strongly infer all other devices $\mathcal{D}_{i>1}$ in a reality, it may not be able to infer them *simultaneously* (strongly or weakly). For example, define $\Gamma : u \to (Y_2(u), Y_3(u), \ldots)$. Then the fact that $\mathcal{D}_1$ is a universal device does not mean that $\forall \delta \in \mathcal{P}(\Gamma)\ \exists\, x_1 : Y_1 = \delta(\Gamma)$. See the discussion in [38] on "omniscient devices" for more on this point.

We now define what it means for two devices to operate in an identical manner:

Definition 10. *Let U and $\hat{U}$ be two (perhaps identical) sets. Let $\mathcal{D}_1$ be a device in a reality with domain U. Let R_1 be the relation between X_1 and Y_1 specified by the reduced form of that reality, i.e., $x_1 R_1 y_1$ iff the pair (x_1, y_1) occurs in some tuple in the reduced form of the reality. Similarly let R_2 be the relation between X_2 and Y_2 for some separate device $\mathcal{D}_2$ in the reduced form of a reality having domain $\hat{U}$.*

Then we say that $\mathcal{D}_1$ **mimics** *$\mathcal{D}_2$ iff there is an injection, $\rho_X : X_2(\hat{U}) \to X_1(U)$ and a bijection $\rho_Y : Y_2(\hat{U}) \leftrightarrow Y_1(U)$, such that for $\forall x_2, y_2,\ x_2 R_2 y_2 \Leftrightarrow \rho_X(x_2) R_1 \rho_Y(y_2)$. If both $\mathcal{D}_1$*

mimics $\mathcal{D}_2$ and vice-versa, we say that $\mathcal{D}_1$ and $\mathcal{D}_2$ are **copies** *of each other.*

Intuitively, when expressed as devices, two physical systems are copies if they follow the same inference algorithm with ρ_X and ρ_Y translating between those systems. In particular, say a reality contains two separate physical computers that are inference devices, both being used for prediction. If those devices are copies of each other, then they form the same conclusion for the same value of their setup function, i.e., they perform the same computation for the same input.

The requirement in Def. 10 that ρ_Y be surjective simply reflects the fact that since we're considering devices, $Y_1(U) = Y_2(U) = \mathbb{B}$. Note that because ρ_X in Def. 10 need not be surjective, there can be a device in U that mimics multiple devices in $\hat{U}$. The relation of one device mimicing another is reflexive and transitive. The relation of two devices being copies is an equivalence relation.

Say that an inference device $\mathcal{D}_2$ is being used for observation and $\mathcal{D}_1$ mimics $\mathcal{D}_2$. The fact that $\mathcal{D}_1$ mimics $\mathcal{D}_2$ does not imply that $\mathcal{D}_1$ can emulate the observation that $\mathcal{D}_2$ makes of some outside function Γ. The mimicry property only relates $\mathcal{D}_1$ and $\mathcal{D}_2$, with no concern for third relationships with any third function. (This is why for one device to "emulate" another is defined in terms of strong inference rather than in terms of mimicry.)

Proposition 16. *Let $\mathcal{D}_1$ be a copy of $\mathcal{D}_2$ and both exist in the same reality.*

i) *It is possible that $\mathcal{D}_1$ and $\mathcal{D}_2$ are distinguishable and $\mathcal{D}_1 > \mathcal{D}_2$, even for finite $X_1(U), X_2(U)$.*

ii) *It is possible that $\mathcal{D}_1 \gg \mathcal{D}_2$, but only if $X_1(U)$ and $X_2(U)$ are both infinite.*

Philosophical implications

Return now to the case where U is a set of laws of physics (i.e., the set of all histories consistent with a set of such laws). The results above provide general restrictions that must relate any devices in such a universe, regardless of the detailed nature of the laws of that universe. In particular, these results would have to be obeyed by all universes in a multiverse [4, 11, 35].

Accordingly, it is interesting to consider these results from an informal philosophical perspective. Say we have a device $\mathcal{D}$ in a reality that is distinguishable from the set of all the other devices in the reality. Such a device can be viewed as having "free will", in limited sense that the way the other devices are set up does not restrict how $\mathcal{D}$ can be set up. Under this interpretation, Prop. 2 means that if two devices both have free will, then they cannot predict / recall / observe each other with guaranteed complete accuracy. A reality can have at most one of its devices that has free will and can predict / recall / observe / control the other devices in that reality with guaranteed complete accuracy.[12]

Prop. 6 then goes further and considers devices that can emulate each other. It shows that independent of concerns of free will, no two devices can unerringly emulate each other. (In other words, no reality can have more than one universal device.) Somewhat tongue in cheek, taken together, these results could be called a "monotheism theorem".

Prop. 16 tells us that if there is a universal device in some reality, then it must be infinite (have infinite $X(U)$) if there are other devices in the reality that are copies of it. Now the time-translation of a physical device is a copy of that device.[13] Therefore any physical device that is *ever* universal must be infinite. In addition, the impossibility of multiple universal devices in a reality means that if any physical device is universal, it can only be so at one moment in time. (Its time-translation cannot be universal.) Again somewhat tongue in cheek, taken together this second set of results could be called an "intelligent design theorem".

In addition to the questions addressed by the monotheism and intelligent design theorems, there are many other semi-philosophical questions one can ask of the form "Can there be a reality with the following properties?". Such questions can often be reduced to a constraint satisfaction problem, potentially involving infinite-dimensional spaces. In this sense, many of the questions that have animated philosophy can be formulated as constraint satisfaction problems.

PHYSICAL KNOWLEDGE

Say that colloquially speaking you "know" the sky's color is blue, for u in some subset W of all histories. How can we formalize that? Well, one thing it means if you "know the sky's color is blue" for any $u \in W$ is that for such u's you can ask yourself "Is the sky green?" and answer 'no', ask yourself "Is the sky red?" and answer 'no', ask yourself "Is the sky blue" and answer 'yes', etc., and always be correct in your

[12] There are other ways to interpret the vague term "free will". For example, Lloyd has argued that humans have "free will" in the sense that under the assumption that they are computationally universal, then due to the Halting theorem they cannot predict their own future conclusions ahead of time [27]. The fact that an ID cannot even weakly infer itself has analogous implications that hold under a broader range of assumptions concerning human computational capability, e.g., under the assumption that humans are not even computationally universal, or at the opposite extreme, under the assumption that they have super-Turing reasoning capability.

[13] Formally, say that the states of some physical system S at a particular time t and shortly thereafter at $t+\delta$ are identified as the setup and conclusion values of a device $\mathcal{D}$. In other words, $\mathcal{D}$ is given by the functions $(X(u), Y(u)) \triangleq (S(u_t), S(u_{t+\delta}))$. In addition, let R_S be the relation between X and Y specified by the reduced form of the reality containing the system. Say that the time-translation of $\mathcal{D}$, given by the two functions $S(u_{t'})$ and $S(u_{t'+\delta})$, also obeys the relation R_S. Then the pair of functions $(X_2(u), Y_2(u)) \triangleq (S(u_{t'}), S(u_{t'+\delta}))$ is another device that is copy of $\mathcal{D}$. So for example, the same physical computer at two separate pairs of moments is two separate devices, devices that are copies of each other, assuming they have the same set of allowed computations.

answer. So to "know" something implies you can weakly infer it. Intuitively speaking, weak inference formalizes an aspect of the semantic content of "knowledge".

To properly formalize knowledge of the sky's color however, we need to use more structure than is just provided by weak inference of the sky's color. The problem is that it is possible that $(X, Y) > \Gamma$ even if for each $\gamma \in \Gamma(U)$, the associated x that causes $Y(u) = \delta(\gamma, \Gamma(u))$ always results in $Y(u) = -1$.[14] Loosely speaking, (X, Y) can infer the sky's color by always setting itself up so that it (correctly) answers that the sky does not have a given color c, so long as it can do that for any given color c.[15] So to say that (X, Y) knows $\Gamma = \gamma$ over (all u in) W, it makes sense not just to require that $(X, Y) > \Gamma$, but also that for all $\gamma \in \Gamma(U)$, there exists some $u \in W$ such that $X(u)$ is a setup value that arises for the question, "Does $\Gamma(u) = \gamma$?", and that $Y(u) = 1$, i.e., that the device answers 'yes'.

Similarly, it is problematic that (X, Y) can infer the sky's color by always setting itself up so that it (correctly) answers that the sky *does* have a given color c, so long as it can do that for any given color c. This suggests we want to add the requirement that for all $\gamma \in \Gamma(U)$, there exists some $u \in W$ such that $X(u)$ is a setup value that arises for the question, "Does $\Gamma(u) \neq \gamma$?", and that $Y(u) = -1$, i.e., that the device answers 'no'.

To model knowledge in this sense, not just inference, we need to guarantee that there is some color c such that whenever the history u is in some set W, for the question, "Is the sky's color c?", the inference device will answer 'yes', and be correct. In other words, you don't know the sky's color whenever $u \in W$ if you can only ever say what it is *not* whenever $u \in W$. For it to be the case that whenever $u \in W$ you know that the sky's color is c, at a minimum, it must be that you can correctly answer "yes, the sky's color is c", for some such $u \in W$.

Formal definition of physical knowledge

We can formalize this strengthened version of inference as follows:

Definition 11. *Consider an inference device (X, Y) defined over U, a function Γ defined over U, a $\gamma \in \Gamma(U)$, and a subset $W \subseteq U$. We say that "(X, Y)* ***(physically) knows*** *$\Gamma = \gamma$ over W" iff $\exists\, \xi : \Gamma(U) \rightarrow \overline{X}$ such that*

i) $\forall \gamma' \in \Gamma(U), u \in \xi(\gamma') \Rightarrow \delta_{\gamma'}(\Gamma(u)) = Y(u)$,

ii) $\varnothing \neq \xi(\gamma) \cap W \subseteq Y^{-1}(1)$.

iii) For all $\gamma' \neq \gamma$, $\varnothing \neq \xi(\gamma') \cap W \subseteq Y^{-1}(-1)$;

(Recall that $\overline{X}$ is the partition of U induced by X.) When I want to specify the precise function ξ used in Def. 11, I will say that "by using ξ, (X, Y) knows that $\Gamma = \gamma$ over W".

By Def. 11(i), if (X, Y) physically knows $\Gamma = \gamma$ over W, then (X, Y) weakly infers Γ. So (X, Y) is always correct in its inference — even if $u \notin W$. We impose this requirement because the agent using the device does not have any *a priori* reason to expect that $u \in W$ So it does them no good to be able to set up a device that will correctly say whether some function has a certain value — but only if $u \in W \subset U$.

Def. 11(ii) and Def. 11(iii) are the extra conditions, forcing the ID to answer 'yes' at least once, and to answer 'no' at least once. Neither of those conditions depend on the precise form of the function $\Gamma(u)$, only its image, $\Gamma(U)$. It's also worth noting that most of the analysis below does not invoke Def. 11(iii). The reason for including that condition is to make clear that physical knowledge can avoid logical omniscience even when that condition holds.

Note that the definition of physical knowledge does not require that $\xi(\gamma) \subseteq Y^{-1}(1)$, but only that $\xi(\gamma) \cap W \subseteq Y^{-1}(1)$ (and similarly for $Y^{-1}(-1)$). The simple fact that $\bar{x} \in \xi(\gamma)$ *and nothing more* does not imply that the device must answer 'yes' if $X(u) = \xi(\gamma)$. Furthermore, there may be more than one $\xi(.)$ with which the ID can "know $\Gamma = \gamma$ over W". There may even be some ξ that can be used to instead know $\Gamma = \gamma' \neq \gamma$ over W. This means that physical knowledge does not require that Γ have the same value over all of $\xi(\Gamma)$. So it includes knowledge that occurs by observation of the value of Γ, just like inference does.

The following properties are immediate:

Lemma 17. *Let (X, Y) be a device defined over U, Γ a function over U, and W a subset of U. Say that by using ξ, (X, Y) knows that $\Gamma = \gamma$ over W. It follows that:*

i) $\Gamma(u) = \gamma\ \forall u \in \xi(\gamma) \cap W$;

ii) If W refines Γ, then $\Gamma(W) = \gamma$.

Proof. To prove the first claim, note from Def. 11(ii) that for all $u \in \xi(\gamma) \cap W$, $Y(u) = 1$. By Def. 11(i), this means that at all such u, $\Gamma(u) = \gamma$, completing the proof. Given this, if in addition W refines Γ (so that $\Gamma(u)$ has the same value across all W), then it must be that $\Gamma(u) = \gamma$ for all $u \in W$. (Similar arguments for $Y(u) = -1$ follow by using Def. 11(iii).) This establishes the second claim. □

Note that we do not require that W refine Γ to have a device know that $\Gamma = \gamma$. This freedom allows the device to know that $\Gamma = \gamma$ over W even if the value of Γ depends on the value of X, the question the device is asking. In other words, it is possible that the device both knows that $\Gamma = \gamma$ over W and knows that $\Gamma = \gamma'$ over W for some $\gamma' \neq \gamma$. In this sense, the definition of physical knowledge is extremely non-restrictive.

[14] Note that there must be *some* x that allows $Y(u) = 1$, since $|Y(U)| = 2$. However it may be that none of those specific x's that are involved in the ID's inferring Γ have that property.

[15] This characteristic of weak inference is an example of how flexible and unrestrictive the definition of weak inference is, mentioned above. This particular flexibility is most reasonable for the inference process of control, where typically x directly influences the value of Γ, and to a somewhat lesser degree for the inference process of observation.

This lack of restriction means that physical knowledge allows for quantum-mechanical-style coupling of an observation device and the system being observed, and more generally, it allows (X, Y) to be a device that "controls" the property Γ of the system being observed. Typically though, when we are interested in knowledge in the sense of accurate prediction or observation that does not affect the system being predicted / observed, W will refine Γ.

The following example illustrates Def. 11 in more detail:

Example 8. *Say that the sky above Greenwich, UK at time t is* {*blue, cloud-free, with the sun less than 15 degrees above the horizon*}*. Furthermore, say that at some time t', Bob knows that the sky above Greenwich, UK at time t is blue. (It does not matter whether $t' = t$.) To formalize this knowledge in terms of Def. 11, let U be the set of all histories in which both Bob and Greenwich, UK exist, and where in addition the following conditions hold:*

i) There is a partition C of all possible distributions of the intensity of light in optical wavelengths. For example, one element of that partition is 'green', one is 'red', and one is the color 'blue';

ii) Bob asks himself at t', "Is c the color of the sky above Greenwich at t?", for some color $c \in C$;

iii) Bob answers that question at that time t' to the best of his abilities, with either a 'yes' or a 'no'.

Define $\Gamma(u)$ as the map taking each $u \in U$ to the associated element of C that characterizes the color of the sky above Greenwich at t. Define $X(u)$ as the map taking each $u \in U$ to the associated color c where at t' Bob is asking himself the question, "Does the sky's color at Greenwich at t equal c?". Assume that the image of X is all C. Next, let $Y(u)$ specify the binary answer in Bob's mind at t'. Finally, let $W \subseteq U$ be the set of all histories u such that the sky above Greenwich at time t is {*blue, cloud-free, with the sun less than 15 degrees above the horizon*}*, and assume $u \in W$.*

Given all this, "at t', Bob knows the sky is blue above Greenwich at time t, when the sky above Greenwich at time t is {*blue, cloud-free, with the sun less than 15 degrees above the horizon*}*", in the sense of Def. 11 if three conditions are met:*

i) There is a u for which "the sky above Greenwich at time t is {*blue, cloud-free, with the sun less than 15 degrees above the horizon*}*" and for which $X(u)$ specifies the question, "Is the sky blue above Greenwich at t?", i.e. for which Bob is asking himself that question at t';*

ii) For the u in (i), $Y(u)$ specifies that Bob answers 'yes' at t';

iii) If the sky above Greenwich at time t were still {*blue, cloud-free, with the sun less than 15 degrees above the horizon*}*, but instead Bob were able to ask himself at t' any question of the form "Does the sky's color at Greenwich at t equal c'?" where $c' \neq$ blue (i.e., if the history were some different $u' \in W$ where $X(u') \neq X(u)$) and did so, Bob would answer 'no' at t' (i.e., $Y(u)$ would equal -1).*

This is a very simple example. In particular, in some situations for Bob to know at t' that the sky is blue at Greenwich at t, Bob will need to configure an apparatus to have a particular state at a particular time (e.g., he may need to configure an automatic camera to photograph the sky above Greenwich at t). In such situations, X not only specifies the question that Bob asks himself at t' but also specifies how Bob configures the apparatus. See [39].

Note that the set U in this example might be a proper subset of the set of all histories that are consistent with the laws of physics. This is just an example of the fact that the definition of weak inference in general implicitly specifies a set U that is a subset of the set of all physically possible histories. Indeed, there might very well be histories that are consistent with the laws of physics in which Bob does not exist, or perhaps even Greenwich does not exist. Clearly we cannot speak of whether Bob does or does not "know the sky is blue" in any such histories.[16]

Finally, note that Bob could conceivably also know at time t that the sky above Greenwich, UK at time t is cloud-free, or that the sun in that sky is less than 15 degrees above the horizon. Any such alternative knowledge would require posing a question to Bob about a different subject. Therefore it would require a different value of $X(u)$, and therefore a different u. However W, the state of the sky above Greenwich at t, doesn't vary with the question that Bob asks concerning that sky. This means that W contains each of those different u's that result in different values of $X(u)$. Ultimately, it is to allow this possibility of multiple questions all concerning the same state of the sky that the definition of physical knowledge involves sets W.

Epistemic logic based on physical knowledge

Inference knowledge obeys many of the usual properties considered in theories of logic. Here I illustrate this by presenting some of those properties.

For the rest of this subsection assume that any space U we consider is countable. I will consider Boolean-valued functions, i.e., functions Γ such that $\Gamma(U) = \{-1, 1\}$. It will also be convenient to identify each such binary-valued function Γ over U with the associated set $\Gamma^{-1}(1)$. Using this identification, any so-called *concrete* Boolean algebra over subsets of

[16] This need to restrict the universe of discourse to a subset of all physically possible histories holds no matter how we formalize "knowledge". It has nothing to do with formalizing knowledge using inference devices.

U defines a Boolean algebra over an associated set of binary-valued functions, which we call the **function Boolean algebra**. Moreover, we can always express an arbitrary Boolean algebra as a concrete Boolean algebra [2]. Accordingly, given any Boolean algebra with propositions Φ, we can identify any specific proposition $\phi \in \Phi$ as a subset of U in the concrete Boolean algebra, and then identify that element of the concrete Boolean algebra with an element of the function Boolean algebra. So we can identify any proposition ϕ with a specific binary function, which we write as Γ_ϕ (with the set Φ usually implicit).

We now use the function Boolean algebra to define the standard shorthands of propositional logic for binary-valued functions. For example, for any two binary-valued functions Γ_1 and Γ_2, their logical AND is

$$\Gamma_1(u) \wedge \Gamma_2(u) = \{u \in U : \Gamma_1(u) = \Gamma_2(u) = 1\} \tag{5}$$

and the logical NOT is given by

$$\neg\Gamma_1(u) = \{u \in U : \Gamma_1(u) = -1\} \tag{6}$$

This allows us to apply the usual axioms of Boolean algebra to binary-valued functions. We can also adopt the usual abbreviations from Boolean algebra, e.g.,

$$\Gamma_1 \vee \Gamma_2 = \neg(\neg\Gamma_1 \wedge \neg\Gamma_2); \tag{7}$$
$$\Gamma_1 \equiv \Gamma_2 = (\Gamma_1 \wedge \Gamma_2) \vee (\neg\Gamma_1 \wedge \neg\Gamma_2); \tag{8}$$
$$\Gamma_1 \Rightarrow \Gamma_2 = \neg\Gamma_1 \vee \Gamma_2; \tag{9}$$
$$\Gamma_1 \Leftrightarrow \Gamma_2 = (\Gamma_1 \Rightarrow \Gamma_2) \wedge (\Gamma_2 \Rightarrow \Gamma_1) \tag{10}$$

I extend this terminology to cases where we are considering subsets $W \subseteq U$ in the obvious way, e.g., $\Gamma_1 \Rightarrow \Gamma_2$ is true over W iff $\neg\Gamma_1(u) \vee \Gamma_2(u)$ is true for all $u \in W$.

Similarly, though it is not used here, as is conventional we can take the function TRUE(u) to be an abbreviation for some fixed propositional tautology such as $(\Gamma_1 \vee \neg\Gamma_1)(u)$ (i.e., the function that equals 1 for all $u \in U$) and FALSE to be an abbreviation for the function $\neg$ TRUE.[17] In keeping with this, I will sometimes use the term 'true' to mean the value 1, and use the term 'false' to mean the value -1.

In many conventional types of epistemic logic, in particular Kripke structures, knowledge is defined in such a way that is impossible for an agent to know two contradictory things. However as discussed above, to allow physical knowledge to capture the case where the agent "knows a function has a specific value" because they control the value of that function, an agent can physically know contradictory things. This occurs when that function, Γ, takes on more than one value across the set under consideration, W, and by appropriate choice of (ξ and therefore) $\bar{x}$, the agent can cause different probes of $\Gamma(u)$ to have the value 1. (Note that in this case Γ is *not* refined by W.) Because of this, certain epistemic properties that are automatically satisfied in conventional types of epistemic logic must be carefully derived in analysis of physical knowledge.

The following proposition presents one of these properties. For pedagogical clarity, in the remainder of this subsection, "is true" is shorthand for "is true over W" and similarly "is false" is shorthand for "is false over W".

Proposition 18. *Let Γ_1, Γ_2 and Γ_3 be any binary-valued functions over U, $\mathcal{D} = (X, Y)$ any device over U, and W some (implicit) subset of U. Then $\mathcal{D}$ knows that Γ_1 is false iff $\mathcal{D}$ knows that $\neg\Gamma_1$ is true.*

Proof. I prove the forward direction; the inverse follows the same way. Let ξ_{Γ_1} be the operator establishing that "$\mathcal{D}$ knows that Γ_1 is false". So $\xi_{\Gamma_1}(-1) \cap W \subseteq Y^{-1}(1)$. Define $\xi_{\neg\Gamma_1}(\gamma) = \xi_{\Gamma_1}(-\gamma)$ for all $\gamma \in \mathbb{B}$ (i.e., for all γ in the codomains of both Γ_1 and $\neg\Gamma_1$). It follows that $\xi_{\neg\Gamma_1}(1) \cap W \subseteq Y^{-1}(1)$. This establishes that if the condition Def. 11(ii) hold for "$\mathcal{D}$ knows that Γ_1 is false (over W)", then it must also hold for "$\mathcal{D}$ knows that $\neg\Gamma_1$ is true". Property (iii) is established using an analogous argument. Finally if property (i) in Def. 11 holds for "$\mathcal{D}$ knows that Γ_1 is false" using ξ_{Γ_1}, it must also hold for "$\mathcal{D}$ knows that $\neg\Gamma_1$ is true" using $\xi_{\neg\Gamma_1}$. This establishes the (forward direction of the) claim in full.

□

Note that applying Prop. 18(i) with a new function $\Gamma'_1 := -\Gamma_1$ establishes that a device $\mathcal{D}$ knows that an arbitrary function Γ'_1 is true iff $\mathcal{D}$ knows that $\neg\Gamma'_1$ is false.

In contrast to these results, if the function under consideration is refined by W, then the agent cannot know contradictory things concerning the value of that function. We start our discussion of this kind of situation with an immediate corollary of Lemma 17(ii), which intuitively says that a device cannot physically know something if that thing is false:

Corollary 19. *Suppose that W refines Γ. Then if $\mathcal{D}$ knows that Γ is true (over W), Γ is true.*

(Similarly, if W refines Γ and $\mathcal{D}$ knows that Γ is false, Γ_1 is false.) Note that Coroll. 19 tells us that if W refines Γ, then it is possible that $\mathcal{D}$ knows that Γ is true, or that $\mathcal{D}$ knows that Γ is false — but not both. So if W refines Γ, we have the usual property that $\mathcal{D}$ cannot know two contradictory things.

Since binary-valued functions obey the rules of propositional logic, Coroll. 19 means that if W refines Γ as well as Γ', and $\mathcal{D}$ both knows that Γ is true and that Γ' is true, it follows that $\Gamma \wedge \Gamma'$ is true. This immediately establishes many properties of physical knowledge, in particular if Γ' involves the logical $\Rightarrow$ operator, including the following:

Corollary 20. *Let Γ_1, Γ_2 and Γ_3 be any binary-valued functions over U, $\mathcal{D} = (X, Y)$ any device over U.*

[17] Note that these two function each violate the requirement of the usual formulation of inference devices that every function take on at least two values. The possibility of such a violation does not affect the elementary results below however, i.e., we can relax the requirement that a function take on two values for the purposes of those results. Such relaxation means in particular that the definition of weak inference has to be modified to allow a single value in the image of functions like TRUE.

i) *Say that W refines Γ_1. Then if $\mathcal{D}$ knows that Γ_1 is true, and $\Gamma_1 \Rightarrow \Gamma_2$ is true, it follows that Γ_2 is true.*

ii) *Say that W refines Γ_1 as well as $\Gamma_1 \Rightarrow \Gamma_2$. Then if $\mathcal{D}$ knows that Γ_1 is true, and knows that $\Gamma_1 \Rightarrow \Gamma_2$ is true, it follows that Γ_2 is true.*

iii) *Say that W refines $\Gamma_1 \Rightarrow \Gamma_2$ as well as $\Gamma_2 \Rightarrow \Gamma_3$. Then if $\mathcal{D}$ both knows that $\Gamma_1 \Rightarrow \Gamma_2$ is true and knows that $\Gamma_2 \Rightarrow \Gamma_3$ is true, it follows that $\Gamma_1 \Rightarrow \Gamma_3$ is true.*

iv) *Say that we have a set of binary-valued functions $\{\Gamma_i : i = 1, \ldots N\}$ and that W refines Γ_1 as well as $\Gamma_i \Rightarrow \Gamma_{i+1}$ for all $i \in \{1, \ldots, N-1\}$. Then if $\mathcal{D}$ knows that Γ_1 is true and knows that $\Gamma_i \Rightarrow \Gamma_{i+1}$ is true for all $i \in \{1, \ldots, N-1\}$, it follows that Γ_i is true for all $i \in \{1, \ldots, N\}$.*

(In interpreting these results, the reader should remember to insert "over W" after every statement about whether a function is true or false.)

In general, the properties described in Coroll. 20 do not hold without the conditions that certain functions are refined by W. So for the most part, they need not hold for the case where an agent knows the value of a function because they control its value. Note though that Coroll. 20(ii) does not require that W refine Γ_2. Similarly Coroll. 20(iii) does not require that W refine either $\Gamma_1, \Gamma_2, \Gamma_3$, or $\Gamma_1 \Rightarrow \Gamma_3$.

We can weaken the last two claims in Coroll. 20:

Corollary 21. *Let Γ_1 and Γ_2 be any binary-valued functions over U, $\mathcal{D}$ any device over U, and W any (implicit) subset of U.*

i) *Say that W refines Γ_1 and refines $\Gamma_1 \Rightarrow \Gamma_2$. Then if either $\mathcal{D}$ knows that Γ_1 is true and $\Gamma_1 \Rightarrow \Gamma_2$ is true, or Γ_1 is true and $\mathcal{D}$ knows that $\Gamma_1 \Rightarrow \Gamma_2$ is true, it follows that Γ_2 is true.*

ii) *Say that W refines $\Gamma_1 \Rightarrow \Gamma_2$ and refines $\Gamma_2 \Rightarrow \Gamma_3$. Then if either $\mathcal{D}$ both knows that $\Gamma_1 \Rightarrow \Gamma_2$ is true and $\Gamma_2 \Rightarrow \Gamma_3$ is true, or $\Gamma_1 \Rightarrow \Gamma_2$ is true and $\mathcal{D}$ knows that $\Gamma_1 \Rightarrow \Gamma_3$ is true it follows that $\Gamma_2 \Rightarrow \Gamma_3$ is true.*

Impossibility results concerning physical knowledge

There are major restrictions on physical knowledge. The first such restriction follows from the first demon theorem.

Corollary 22. *For any device $\mathcal{D}$, there exists a function Γ over U such that for no $W \subseteq U, \gamma \in \Gamma(U)$ does $\mathcal{D}$ know $\Gamma = \gamma$ over W.*

The second major restriction follows from the second demon theorem.

Corollary 23. *Let (X_1, Y_1) and (X_{-1}, Y_{-1}) be two distinguishable devices. Then for at least one of the two devices $i \in \{-1, 1\}$, there is no pair $(W \subseteq U, y_{-i} \in Y_{-i}(U))$ such that (X_i, Y_i) knows that $Y_{-i} = y_{-i}$ over W.*

Similarly, Coroll. 3 provides another restriction on physical knowledge:

Corollary 24. *Consider a pair of devices $\mathcal{D} = (X, Y)$ and $\mathcal{D}' = (X', Y')$ that are both distinguishable from one another and whose conclusion functions are inequivalent. Say that there is a $W \subseteq U$ that refines Y such that $\mathcal{D}'$ knows that $Y = Y(W)$ over W. Then there are at least three inequivalent surjective binary functions Γ such that there is no W' with the following two properties: W' refines Γ, and $\mathcal{D}$ know that $\Gamma = \Gamma(W')$ when W'.*

In other words, if $\mathcal{D}'$ knows the value of $\mathcal{D}$'s conclusion function over W, then there are at least three separate functions that $\mathcal{D}$ never knows, no matter what the subset of U we are in.

Physical knowledge and the first three rules of *S5*

S5 is a set of five rules obeyed by many epistemic logics, including Kripke structures. The "knowledge axiom" of *S5* says that if an agent knows a Boolean proposition ϕ, then ϕ must be true. We can use the map from propositions to binary functions (discussed just before Eq. (5)) to formulate a physical knowledge version of this axiom. Coroll. 19 confirms that the physical knowledge analog of the knowledge axiom holds (assuming we only consider sets W that refine Γ_ϕ).

In addition to the knowledge axiom, *S5* also includes the "knowledge generalization rule", which says that if proposition ϕ is true in all possible states of the world (i.e., if ϕ is necessarily true rather than contingently true), then the agent knows ϕ. An analogous rule in terms of physical knowledge might be that if W refines Γ_1, and Γ_1 is true over W, then the agent physically knows Γ_1 is true over W. However this rule need not hold; an agent (X, Y) may not be able to weakly infer Γ_1, whether or not Γ_1 is true over W. (And even if they can infer Γ_1, it may be that $Y^{-1}(1) \cap W = \varnothing$.)

The "distribution axiom" of *S5* says that if an agent both knows proposition ϕ_1 and knows $\phi_1 \Rightarrow \phi_2$, then ϕ_2 is true. In contrast, Coroll. 20(ii) only establishes that if an agent both physically knows that Γ_1 is true over W and that $\Gamma_1 \Rightarrow \Gamma_2$ is true over W, then Γ_2 is also true over W. However it is easy to construct examples where the conditions for Coroll. 20(ii) hold but the agent does not know Γ_2 is true. Ultimately, this is due to the requirement of physical knowledge of Γ that the device weakly infer Γ — a requirement that involves considering counterfactual scenarios, something not done in conventional epistemic logics.

For what are ultimately the same reasons, the conditions for Coroll. 20(i) may hold even if the agent does not physically know that Γ_2 is true. This is illustrated in the following example:

Example 9. *For purposes of this example, fix some particular location and time. Suppose that the (binary-valued) function $\Gamma_1(u)$ is defined by whether the temperature in u is / isn't* *ten degrees celsius at that location and time. Have $\Gamma_2(u)$ be whether the temperature in u is / isn't above freezing. Presume U is large enough that there are $u \in U$ that satisfy each of the three possible values of $(\Gamma_1(u), \Gamma_2(u))$. Finally, have W be the set of all u at which the temperature is ten degrees. Note that $(\Gamma_1 \Rightarrow \Gamma_2)(u)$ is always true. (This means it is a "valid" statement, in the language of epistemic logic.) So in particular it is true for all $u \in W$.*

To ground thinking suppose that the ID is a thermometer that outputs a 1 or -1, depending on the temperature. x is the value of the temperature that the thermometer is checking. By Def. 11(i), if the agent physically knows $\Gamma_1 = 1$ over W, then there is some setup function ξ they can use to configure their thermometer to correctly give a 1 if the temperature is ten degrees, and to correctly give a -1 otherwise. In other words, such physical knowledge requires that when answering the question, "Is the temperature ten degrees?" (i.e., does $\Gamma_1(u) = 1$), the agent will set X to be in the state $\xi(1)$, and they will be guaranteed that the associated conclusion $Y(u)$ will equal $\delta_1(u)$ for any $u \in \xi(1)$ whether or not $u \in W$.

In general, depending on whether that ξ obeys $\xi(1) \subseteq W$, this phenomenon may mean that the ID gives the correct answer for some u in which the temperature is not ten degrees. This is key: the ID is set up with the same x value regardless of whether $u \in W$. After the thermometer is set up this way, then the agent learns whether the temperature equals ten degrees. If (after having been set up with the x value corresponding to ten degrees) the ID tells the agent that the temperature is indeed ten degrees (since $u \in W$), at that point the agent has physical knowledge that the temperature is ten degrees. But not before.

In particular, consider the situation where $\xi(1)$ is big enough to contain both a u where the temperature is five degrees, and one where it is negative five degrees (in addition to containing one where it is ten degrees). Since we require that $Y(u) = \delta_1(\Gamma_1(u))$ for all $u \in \xi(1)$, this means that $Y(u) = -1$ for both of those u's. Note though that $\Gamma_2(u)$ has different values for those two temperatures. This means that the agent cannot use this same ξ that allows them to weakly infer Γ_1 to also weakly infer Γ_2. (This is how the intuitive notion of regular implication would work too; if all I know is that the temperature is not 10 degrees, then I don't know whether it is -5 or 5.) Moreover, in general, there may not be any alternative to this ξ that the agent can use with that thermometer to weakly infer Γ_2, i.e., weakly infer whether the temperature is above freezing or not. In this situation, the agent cannot weakly infer Γ_2. (Recall again the key point that in the definition of physical knowledge we require that weak inference holds for all $u \in U$, not just the $u \in W$.)

So in this situation, the agent does not physically know that $\Gamma_2 = 1$ for $u \in W$. Loosely speaking, there are thermometers that can be used to always tell us correctly whether the temperature is (not) ten degrees – both when it is and when it isn't ten degrees. However some such thermometers cannot be used to always tell us correctly whether the temperature is (not) above freezing (both when it is and when it isn't above freezing).

Given such a thermometer, for the particular situation where the thermometer is set up to detect whether the temperature is ten degrees, and in addition it answers 'yes', you and I can use our reasoning ability to realize that the temperature must above above zero. But the ability to use such reasoning to come to a conclusion is not the same thing as physical knowledge of that conclusion.

In the rest of this example I establish this argument in a fully formal manner, making the simplifying assumption that W refines Γ_1, as in Coroll. 20(i). First, note that by Lemma 17(ii), since the ID physically know that $\Gamma_1 = 1$ when W, $\Gamma_1(u)$ is true throughout W. This in turn means that Γ_2 is true throughout W (since $\Gamma_1 \Rightarrow \Gamma_2$).

Let ξ be the function that establishes that condition Def. 11(ii) holds for Γ_1. We can use that same function to establish that condition Def. 11(ii) holds for $\Gamma_1 \Rightarrow \Gamma_2$ and for Γ_2. Similar arguments hold for condition Def. 11(iii). So $\mathcal{D}$ meets conditions (ii) and (iii) for having physical knowledge of both $\Gamma_1 \Rightarrow \Gamma_2$ and Γ_2. So to complete our analysis of whether $\mathcal{D}$ has physical knowledge that $\Gamma_2 = 1$, we must consider whether condition Def. 11(i) for Γ_2. We do this by considering two cases, one in which $\mathcal{D}$ does physically know $\Gamma_2 = 1$ when W, and one where it does not:

1. *First, assume $\xi(\gamma) \subseteq W$ both for $\gamma = 1$ and $\gamma = -1$. Since Def. 11(i) holds for Γ_1 for that ξ, and since $\Gamma_1(u) = \Gamma_2(u)$ throughout W, it is immediate that Def. 11(i) also holds for that ξ. So $\mathcal{D}$ knows that Γ_2 is true over W, under our assumption.*

   Next, plug the fact that $\Gamma_1(u) = \Gamma_2(u) = 1$ for all $u \in W$ into the definition of $\Gamma_1 \Rightarrow \Gamma_2$ to see that $\delta_1((\Gamma_1 \Rightarrow \Gamma_2)(u)) = 1$ for all $u \in \xi(1) \subset W$. So $\delta_1((\Gamma_1 \Rightarrow \Gamma_2)(u)) = Y(u)$ for all $u \in \xi(1)$. This establishes that Def. 11(i) holds for the function $\Gamma_1 \Rightarrow \Gamma_2$ for the case of $\gamma' = 1$. For the remaining case of $\gamma' = -1$, note that for all $u \in \xi(-1)$, again $\Gamma_1(u) = \Gamma_2(u) = 1$. So $\delta_{-1}((\Gamma_1 \Rightarrow \Gamma_2)(u)) = -1$. Since $Y(u) = -1$ throughout $\xi(-1)$, this establishes that Def. 11(i) also holds for the function $\Gamma_1 \Rightarrow \Gamma_2$ for the case of $\gamma' = -1$. Accordingly, under our assumption that the support of both $\xi(1)$ and of $\xi(-1)$ is restricted to W, $\mathcal{D}$ knows that $\Gamma_1 \Rightarrow \Gamma_2$ over W.

2. *In many situations however, even though W refines Γ_1, it will* not *be the case that the associated function ξ that establishes that Def. 11(i) and Def. 11(ii) both hold for Γ_1 always produce sets that are confined to W. Very often either $\xi(1) \not\subseteq W$ and / or $\xi(-1) \not\subseteq W$. An example of this is given just above, in the discussion involving thermometers. As mentioned in that discussion, for such a ξ, it may be that (for example) $\xi(1)$ contains points $u' \notin W$ such that $\Gamma_1(u') = -1$ but $\Gamma_2(u') = 1$. Now for any such u', it must be that $Y(u') = -1$ (since $\mathcal{D}$ weakly infers Γ_1). On the other hand, for any $u \in W \cap \xi(1)$, $Y(u) = 1$. Since the value of $\Gamma_2(u)$ does not change*

across $\xi(1)$, this means that Y and Γ_2 cannot have the same value across all of $\xi(1)$.

This means that that function ξ could not be used to establish that $\mathcal{D}$ weakly infers Γ_2 — and therefore all bets are off concerning whether $\mathcal{D}$ can physically know Γ_2 over W. This reflects the fact that while under the conditions of Coroll. 20(i) Γ_1 and Γ_2 must be identical for all $u \in W$, they will in general differ outside of W, and so a device that can say both whether Γ_1 is true or not may not be able to tell us whether Γ_2 is true or not.

Recall from the introduction that in many epistemic logics, if an agent knows a proposition ϕ is true (more generally, that a set of propositions are true), and $\phi \Rightarrow \phi'$, then not only is ϕ' true — but the agent knows that it is.

This "(full) logical omniscience" is a major problem with these logics, since logical omniscience implies for example that if someone knows the axioms of numbers theory, then they know all the theorems of number theory that are implied by those axioms. However as illustrated in Ex. 9, physical knowledge need not obey logical omniscience.[18]

A closely related point is that the definition of physical knowledge does not fully agree with the colloquial meaning of the term "knowledge". It should really be viewed more of a strengthened form of inference, capturing more of the common structure of real-world prediction, observation, memory and control, rather than an attempt to provide an accurate entry in an English language dictionary.

For example, it is possible that a device knows that $\Gamma_1 \wedge \Gamma_2 = 1$ over W, but does not know that $\Gamma_1 = 1$ over W. In particular, physical knowledge by a device that $\Gamma_1 \wedge \Gamma_2 = 1$ over W provides no guarantees that the device weakly infers Γ_1; loosely speaking, the ID may not be able to correctly answer questions concerning the value of $\Gamma_1(u)$ for $u \notin W$.[19,20]

Nonetheless, it is worth noting that the definition of physical knowledge could be weakened to agree with this aspect of the colloquial meaning of "knowledge". One way to do that would be drop the requirement that the ID infer Γ in full, including for $u \notin W$. Under this modified definition of what it means for the ID to know that $\Gamma = \gamma$ over W, we would still require that for all $u \in \xi(\gamma)$, if $Y(u) = 1$, then $\Gamma(u) = \gamma$ (whether or not $u \in W$). So no matter what u is, we would require that if the device is answering the question, "does $\Gamma(u) = \gamma$?" and it answers 'yes', then it is correct. However for all $\gamma' \neq \gamma$, we only require that for all $u \in W \cap \xi(\gamma')$, if $Y(u) = -1$, then $\delta_\gamma(\Gamma(u)) = Y(u)$. Under this modification, we would allow there to be u outside of W, and $\gamma' \neq \gamma$, where the device is answering the question, "does $\Gamma(u) = \gamma'$?" and incorrectly answers 'no'.

[18] It is known that so long as both the distribution axiom and the knowledge generalization rule hold — which is the case in all so-called "normal modal logics" — then so does (full) logical omniscience. However neither of those need hold with physical knowledge.

[19] As an example, suppose that some of the subsets $\bar{x}$ that are images of ξ extend beyond W, and in particular include points u at which $\Gamma_1(u) \wedge \Gamma_2(u) = -1$ while $\Gamma_1(u) = 1$. $Y(u)$ must equal -1 for such a u, since the device uses ξ to weakly infer $\Gamma_1 \wedge \Gamma_2$. But this means that ξ does not weakly infer Γ_1, and so does not know $\Gamma_1 = 1$ over W.

[20] This should not be surprising; if logical omniscience held, then knowledge that $\Gamma_1 \wedge \Gamma_2 = 1$ over W *would* imply knowledge that $\Gamma_1 = 1$ over W.

Physical knowledge that you have physical knoweldge

The final two rules of *S5* are known as the *positive introspection rule* and the *negative introspection rule*. Intuitively, they stipulate that when an agent knows something, they know that they know it, and when they don't know something, they know that they don't know it (resp.).

Perhaps the simplest formalization of these rules occurs in the event-based framework based on Aumann structures [5–7, 10, 16, 18, 42]. As discussed in the introduction, in this framework, in this framework events are defined as subsets of U. We say "Alice knows event E" if $A(u) \subseteq E$, where A is Alice's knowledge operator. So the event, "Alice knows E" is just the union of all u such that Alice knows E for $U = u$, i.e., the union of all $u \in U$ such that $A(u) \subseteq E$. (Note that even if $u \in E$, $A(u)$ may include points $u' \notin E$ — no elements of such a set $A(u)$ are contained in the event "Alice knows E".) It is immediate that if Alice knows event E, then Alice knows {Alice knows E}. This is the (event-based approach version of the) positive introspection rule.

Physical knowledge is formulated in terms of functions and subsets W, not in terms of events, so we need to extend it to consider the introspection rules. We say that "$\mathcal{D}$ **(physically) knows** event $E \subseteq U$" if $\mathcal{D}$ knows $\mathcal{X}_E = 1$ over E for some ξ. Next, we must define what subset of U is represented by "the event that {$\mathcal{D}$ knows event E}", i.e., by "the event that {$\mathcal{D}$ knows $\mathcal{X}_E = 1$ over E for some ξ}". We adopt the interpretation that this set is the union of all sets $\bar{x}$ that might arise in the image of some ξ such that $\mathcal{D}$ knows $\mathcal{X}_E = 1$ over E for ξ. We write this set as

$$K(\mathcal{D}\ knows\ E) \quad := \bigcup_{\xi\,:\,\mathcal{D}\ knows\ \mathcal{X}_E = 1\ over\ E\ for\ \xi} \xi(1) \cup \xi(-1) \tag{11}$$

(Note that in general, $K(\mathcal{D}\ knows\ E)$ can include points u that lie outside of E.) This allows us to translate from the event-based framework to the physical knowledge framework: we say that "$\mathcal{D}$ obeys positive introspection" if for every event that $\mathcal{D}$ knows, $\mathcal{D}$ also knows the event $K(\mathcal{D}\ knows\ E)$.

Corollary 25. *For every event that a device $\mathcal{D}$ knows, $\mathcal{D}$ also knows the event $K(\mathcal{D}\ knows\ E)$*

Proof. Plugging in, "$\mathcal{D}$ knows the event $K(\mathcal{D}\ knows\ E)$" will be established if we can show that

$$\mathcal{D}\ knows\ \mathcal{X}_{K(\mathcal{D}\ knows\ E)} = 1\ over\ K(\mathcal{D}\ knows\ E)$$

for some ξ.

Now by hypothesis $\mathcal{D}$ knows event E. So there is at least one function ξ such that $\mathcal{D}$ knows $\mathcal{X}_E = 1$ over E for ξ. By Def. 11(i), this means that $\mathcal{D}$ weakly infers $\mathcal{X}_E$ using ξ. Therefore for both $\gamma \in \mathbb{B}$, $u \in \xi(\gamma) \Rightarrow \delta_\gamma(\mathcal{X}_E(u)) = Y(u)$. Moreover for both $\gamma \in \mathbb{B}$, $\mathcal{X}_E(u) = \mathcal{X}_{K(\mathcal{D}\ knows\ E)}(u)$ for all $u \in \xi(\gamma)$. Therefore $\mathcal{D}$ weakly infers $\mathcal{X}_{K(\mathcal{D}\ knows\ E)}$ using that same function ξ.

Next, note that $\xi(1) \cap K(\mathcal{D}\ knows\ E)$ equals $\xi(1) \cap E$. Moreover, since $\mathcal{D}$ knows $\mathcal{X}_E = 1$ over E for ξ, by Def. 11(ii), $\varnothing \neq \xi(1) \cap E \subseteq Y^{-1}(1)$. So $\varnothing \neq \xi(1) \cap K(\mathcal{D}\ knows\ E) \subseteq Y^{-1}(1)$. Therefore the condition in Def. 11(ii) holds for knowledge that $\mathcal{X}_{K(\mathcal{D}\ knows\ E)} = 1$ over $K(\mathcal{D}\ knows\ E)$ by using ξ .

Similarly, $\varnothing \neq \xi(-1) \cap K(\mathcal{D}\ knows\ E) \subseteq Y^{-1}(-1)$. So all three criteria in Def. 11 are met for physical knowledge that $\mathcal{X}_{K(\mathcal{D}\ knows\ E)} = 1$ over $K(\mathcal{D}\ knows\ E)$ by using ξ. □

In this sense, the positive introspection rule of *S5* holds for physical knowledge.

The negative introspection rule cannot hold for the event-based framework. This is because the event "Alice does not know E" cannot contain any u obeying $u \in A(u)$, and so Alice can never know that she does not know E. (As a result, investigations in the event-based approach focus on negative introspection of belief rather than negative introspection of knowledge.) Not surprisingly then, it is not clear how to formalize a physical knowledge version of the negative introspection rule, since that requires defining a function over U that captures the case that the device does *not* know that $u \in E$. (N.b., that is not the same as having the device know that $u \notin E$.)

FUTURE WORK

Much more work remains to complete our understanding of inference. Perhaps most obviously, a lot remains to be investigated concerning the relationship between structures like inference complexity (the ID version of Kolmogorov complexity) and all the results in algorithmic information theory, from Chaitin's "incompleteness theorem" to the Halting theorem to computational complexity theory.

There is also a lot of future work to be done concerning physical knowledge. To begin, it might be useful to extend the analysis of physical knowledge to include all the concepts introduced in the analysis of inference, e.g., strong inference, covariance accuracy, and inference complexity. Other future work on physical knowledge would be to develop Prop.18, Coroll. 19 and Coroll. 20 into a complete axiomatization of physical knowledge, i.e., a set of axioms that are logically equivalent to the definition of physical knowledge. The goal would be to parallel the kind of axiomatization which has been done for Kripke structures (See Chap. 3 of [16].) As a final example, it might be illuminating to construct and then investigate physical knowledge versions of common knowledge, of distributed knowledge, and associated results in conventional epistemic logic, e.g., Aumann's famous proof that "no two Bayesians can disagree" [7].

There are also many ways to extend the concept of physical knowledge to capture attributes of our physical world, like space and time. As an example, suppose we are given a **distance function** $D(\gamma, \gamma') : \Gamma(U) \times \Gamma(U) \to R^+$. We want to use that $D(.,.)$ to define the distance between what a given ID "says" that $\Gamma(u)$ is, and what $\Gamma(u)$ really is. One way to do this builds on the physical knowledge formalism. For simplicity, assume W refines Γ. We say that $\mathcal{D}$ **claims** γ over W if $\exists\ \xi : \Gamma(U) \to \bar{X}$ such that for all $u \in \xi(\gamma) \cap W$, $Y(u) = 1$, and for all $\gamma' \neq \gamma, u \in \xi(\gamma') \cap W$, $Y(u) = -1$. (Note that there may be more than one γ that $\mathcal{D}$ claims over W.) We define the **error** of $\mathcal{D}$ over W (for γ) as the smallest $\epsilon \in \mathbb{R}$ such that $\mathcal{D}$ claims γ over W and $D(\Gamma(W), \gamma) = \epsilon$.

By supposing a probability distribution over U as well as a distance function, we can analyze concepts like the expected error of the claim of a device, the variance of what a device claims, etc. In particular, it may be possible to use an error function to extend the analysis that led to Prop. 11, to investigate the possible relationship between inference and Heisenberg's uncertainty principle. (As part of such an investigation, it may be helpful to focus on the specific case where U is a Hilbert space.)

Other future work is to investigate the use of inference devices in general, and physical knowledge in particular, as a formalization of "semantic information", a concept that has been extensively debated by people ranging from the founders of information theory and cybernetics [34] to philosophers [1, 29] to people working in statistical physics [14, 15, 32, 41].

Finally, despite the relation of physical knowledge with epistemic logic, physical knowledge is designed only to capture properties of knowledge concerning physical reality. It is not designed to capture properties of knowledge concerning mathematical systems, e.g., predicate logic. However it may be worth investigating its application to such systems. For example, one could identify each $u \in U$ with a (perhaps infinite) string over some alphabet. U might then be defined as the set of all strings that are "true" under some encoding that translates a string into axioms and associated logical implications. Then an inference device would be a (perhaps fallible) theorem-proving algorithm, embodied within U itself. The results of this subsection would then concern the relation among such theorem-proving algorithms.

ACKNOWLEDGEMENTS

I would like to thank Philippe Binder, Artemy Kolchinsky, Brendan Tracey, Alexi Parizeau and especially Walt Read for many useful discussions. In particular, Prop. 8, Ex. 6, and Prop. 12 are due to Walt. This work was supported by the Santa Fe Institute, Grant No. FQXi-RFP-1622 from the FQXi foundation, Grant No. FQXi-RFP3-1349 from the FQXi foundation, and the Silicon Valley Foundation.

[1] *Stanford encyclopedia of philosophy entry on semantic information*, $https://plato.stanford.edu/entries/information-semantic/$, 2017.

[2] *Wikipedia 2017 entry on boolean algebra*, 2017, $https://en.wikipedia.org/wiki/Boolean\_algebra$.

[3] Scott Aaronson, *Why philosophers should care about computational complexity*, Computability: Turing, Gödel, Church, and Beyond (2013), 261–327.

[4] A. Aguirre and M. Tegmark, *Multiple universes, cosmic coincidences, and other dark matters*, hep-th/0409072, 2005.

[5] R. J. Aumann, *Interactive epistemology ii: Probability*, Int. J. Game Theory **28** (1999), 301–314.

[6] R. J. Aumann and A. Brandenburger, *Epistemic conditions for nash equilibrium*, Econometrica **63** (1995), no. 5, 1161–1180.

[7] Robert J Aumann, *Agreeing to disagree*, The annals of statistics **4** (1976), no. 6, 1236–1239.

[8] John D Barrow, *Godel and physics*, Kurt Gödel and the Foundations of Mathematics: Horizons of Truth (2011), 255.

[9] P. Binder, *Theories of almost everything*, Nature **455** (2008), 884–885.

[10] K. Binmore and A. Brandenburger, *Common knowledge and game theory*, ST/ICERD Discussion Paper 88/167, London School of Economics, 1988.

[11] B. Carr (ed.), *Universe or multiverse?*, Cambridge University Press, 2007.

[12] Ariel Caticha, *Entropic dynamics, time and quantum theory*, Journal of Physics A: Mathematical and Theoretical **44** (2011), no. 22, 225303.

[13] Gregory J Chaitin, *Algorithmic information theory*, vol. 1, Cambridge University Press, 2004.

[14] James P Crutchfield, *Knowledge and meaning... chaos and complexity*, Modeling complex phenomena (1992), 66.

[15] ———, *Semantics and thermodynamics*, Santa Fe Institute studies in the sciences of complexity - proceedings volume, vol. 12, Addison Wesley, 1992, pp. 317–317.

[16] Ronald Fagin, Joseph Y Halpern, Yoram Moses, and Moshe Vardi, *Reasoning about knowledge*, MIT press, 2004.

[17] B Roy Frieden, *Science from fisher information: A unification*, Cambridge University Press, 2004.

[18] D. Fudenberg and J. Tirole, *Game theory*, MIT Press, Cambridge, MA, 1991.

[19] Philip Goyal, *From information geometry to quantum theory*, New Journal of Physics **12** (2010), no. 2, 023012.

[20] ———, *Information physics—towards a new conception of physical reality*, Information **3** (2012), no. 4, 567–594.

[21] James B Hartle, *The physics of now*, American journal of physics **73** (2005), 101.

[22] John E Hopcroft, Rajeev Motwani, and Ullman Rotwani, *Jd: Introduction to automata theory, languages and computability*, 2000.

[23] M. Li and Vitanyi P., *An introduction to kolmogorov complexity and its applications*, Springer, 2008.

[24] Seth Lloyd, *Valuable information*, Complexity, entropy and the physics of information (Wojciech H Zurek, ed.), 1990, pp. 193–197.

[25] ———, *Ultimate physical limits to computation*, Nature **406** (2000), no. 6799, 1047–1054.

[26] ———, *Programming the universe: a quantum computer scientist takes on the cosmos*, Vintage, 2006.

[27] ———, *A turing test for free will*, Phil. Trans. R. Soc. A **370** (2012), no. 1971, 3597–3610.

[28] Donald M. MacKay, *On the logical indeterminacy of a free choice*, Mind, New Series **69** (1960), no. 273, 31–40.

[29] ———, *Information, mechanism and meaning*, (1969).

[30] Rohit Parikh, *Knowledge and the problem of logical omniscience.*, ISMIS, vol. 87, 1987, pp. 432–439.

[31] K. Popper, *The impossibility of self-prediction*, The Open Universe: From the Postscript to the Logic of Scientific Discovery, Routledge, 1988, p. 68.

[32] Carlo Rovelli, *Meaning= information+ evolution*, arXiv preprint arXiv:1611.02420 (2016).

[33] Jürgen Schmidhuber, *Algorithmic theories of everything*, arXiv preprint quant-ph/0011122 (2000).

[34] Claude Elwood Shannon and Warren Weaver, *A mathematical theory of communication*, 1948.

[35] L. Smolin, *The life of the cosmos*, Weidenfeld and Nicolson, 2002.

[36] Max Tegmark, *The mathematical universe*, Foundations of Physics **38** (2008), no. 2, 101–150.

[37] J.H. Wheeler, *Information, physics, quantum: The search for links*, Complexity, Entropy, and the Physics of Information (W. H. Zurek, ed.), Addison-Wesley, 1990.

[38] D. H. Wolpert, *Computational capabilities of physical systems*, Physical Review E **65** (2001), 016128.

[39] D. H. Wolpert, *Physical limits of inference*, Physica D **237** (2008), 1257–1281, More recent version at http://arxiv.org/abs/0708.1362.

[40] ———, *Inference concerning physical systems*, CiE Proceedings on Programs, proofs, and processes, 2010, pp. 438–447.

[41] David H. Wolpert, *Presentation on statistical physics definition of semantic information*, $http://fqxi.org/data/documents/conferences/2016-talks/Wolpert.pdf$, 2016.

[42] Edward N Zalta et al., *Stanford encyclopedia of philosophy*, 2003.

[43] Hector Zenil, *A computable universe: Understanding and exploring nature as computation*, World Scientific Publishing Co., Inc., 2012.

[44] W. H. Zurek, *Algorithmic randomness and physical entropy*, Phys. Rev. A **40** (1989), 4731–4751.

[45] ———, *Thermodynamic cost of computation, algorithmic complexity and the information metric*, Nature **341** (1989), 119–124.

[46] Wojciech H Zurek (ed.), *Complexity, entropy and the physics of information*, Addison-Wesley, 1990.

[47] Konrad Zuse, *Rechnender raum (calculating space)*, (1969).