Algorithmic information and inference devices

Edward G. Huang

Santa Fe Institute

Department of Mathematics, University of California, Berkeley

eghuang@berkeley.edu

August 2018

There has been much interest surrounding what properties about the universe can be derived from applying a mathematical formalization of inference and knowledge. Previous work by Wolpert used the theory of "inference devices" (IDs) to demonstrate bounds on knowledge in any physical universe that allows agents to hold information concerning that universe. We extend previous work on the capacity and limitations of IDs to infer physical variables. Our results impose conditions on the inferrability of singular functions and sets of functions. We pursue analogues between IDs and their relation to Turing machine theory and algorithmic information theory. In particular, we show that any Turing machine can be strongly inferred and build upon that to demonstrate incompressibility of strong inference complexity. This incompressibility result has led to several analogues between Kolmogorov complexity and inference complexity that suggest further similarities between algorithmic information theory and the theory of inference devices.

1 Introduction

We begin with the observation that information concerning the universe is typically held by some agent embedded within that universe. Agents typically acquire information concerning the universe in at least four ways: prediction, observation, control, and memory. As discussed by Wolpert, there is a mathematical structure shared by all such information-acquiring avenues. Objects with that structure are called inference devices [1].

This work builds upon the relationship between inference devices, Turing machines, and Kolmogorov complexity. We introduce the mathematical structures of inference devices, weak inference, and strong inference and examine the conditions in which inference holds. Basic forms of Turing machines and their relationship to computability and partial recursive functions are reviewed. We then discuss the complexity of inference and its connections to Kolmogorov complexity. Finally we touch upon physical knowledge, a mathematical formalism of knowledge in a physical universe. Several properties of physical knowledge are extended and explored. These properties suggest that the mathematical definition of physical knowledge may require further examination and modification.

1.1 Notation and Definitions

This manuscript utilizes standard notation taken from set theory and vector algebra. We clarify notation specific to Turing machine theory and inference devices.

1.1.1 Turing Machine Notation

$\mathbb{B}^*$	The space of all finite bit strings, $\{\epsilon, 0, 1, 00, 01, \dots\}$
Λ	Symbol alphabet of a Turing Machine.
σ	A symbol on a Turing Machine tape.
Q	Set of finite states of a Turing Machine.
Δ	Transition function of a Turing Machine.
k	Number of tapes of a Turing Machine. The first tape is assumed to be read-only.
η	Non-halting state of a Turing Machine.
ϕ	The universal partial recursive function.

1.1.2 Inference Device Notation

U	Set of possible histories of the universe.
u	A history of the universe in U.
X	Setup function of an ID that maps $U \rightarrow X(U)$. A binary question concerning $\Gamma(u)$.
x	A binary question and a member of image $X(U)$.
Y	Single-valued conclusion function of an ID that maps $U \rightarrow \{-1, 1\}$. A binary answer of an ID for $X(u) = x$.
y	A single-valued answer, and member of image $Y(U) = \{0, 1\}$.
Γ	A function of the actual values of a physical variable over U, equivalent to $\Gamma(u) = S(t_i)(u)$.
γ	Possible value of a physical variable, a member of the image $\Gamma(U)$.
δ	Probe of any variable V parameterized by $v \in V$ such that :

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

$\wp$	Set of probes over $\Gamma(U)$.
$\mathcal{D}$	An inference device, consisting of functions (X, Y).
ξ	A function $\Gamma(U) \rightarrow \overline{X}$.
Γ^{-1}	Inverse. Given a function Γ over U, $\Gamma^{-1} = \bar{\Gamma} \equiv \{\{u : \Gamma(u) = \gamma\} : \gamma \in \Gamma(U)\}$.
$\overline{\Gamma}$	Given a function Γ over U, the partition of U given by Γ^{-1}.
$>$	Weak inference: a device $\mathcal{D}$ weakly infers Γ *iff* $\forall \gamma \in \Gamma(U), \exists x \in X(U)$ s.t. $\forall u \in U$, $X(u) = x \implies Y(u) = \delta_\gamma(\Gamma(u))$.
$\gg$	Strong inference: a device (X, Y) strongly infers a function (S, T) over U *iff* $\forall \delta \in \wp(T)$ and all $s \in S(U)$, $\exists x$ *such that* $X(u) = x \implies S(u) = s, Y(u) = \delta(T(u))$.
$\mathcal{C}_\mu(\Gamma; \mathcal{D})$	Inference complexity.
μ	Measure defined for $u \in U$.

1.2 Inference Devices

1.2.1 Types of Inference

1.3 Turing Machines

Church's Hypothesis

1.3.1 Deterministic Turing Machines

Arora and Barak denote a Turing Machine (TM) as $T = (\Lambda, Q, \Delta)$ containing:

1. An *alphabet* Λ of a finite set of symbols that T's tapes can contain. We assume that Λ contains a special blank symbol B, start symbol S, and the symbols 0 and 1.

2. A finite set Q of possible states that T's register can be in. We assume that Q contains a special start state q_s and a special halt state q_h.

3. A transition function $\Delta : Q \times \Lambda^k \to Q \times \Lambda^{k-1} \times \{L, \mathcal{S}, R\}^k$, where $k \geq 2$, describing the rules T use in performing each step. The set $\{L, \mathcal{S}, R\}$ denote the actions *Left, Stay,* and *Right*, respectively.

Suppose T is in state $q \in Q$ and $(\sigma_1, \sigma_2, \ldots, \sigma_k)$ are the symbols on the k tapes. Then $\Delta(q, (\sigma_1, \ldots, \sigma_k)) = (q', (\sigma_2^{'}, \ldots, \sigma_k^{'}), z)$ where $z \in \{L, \mathcal{S}, R\}^k$ and at the next step the σ symbols in the last $k-1$ tapes will be replaced by the σ' symbols, the machine will be in state q, and the k heads will move *Left, Right* or *Stay.* This is illustrated in the following figure.

FIGURE. The transition function Δ for a k-tape Turing Machine

$(q, (\sigma_1, \ldots, \sigma_k))$				$(q', (\sigma_2^{'}, \ldots, \sigma_k^{'}), z)$				
Input symbol	Work/output symbol read	$\cdots$	Current state	New work/output tape symbol	$\cdots$	Move work/output tape	$\cdots$	New state
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
σ_1	σ_i	$\ddots$	q	$\sigma_i^{'}$	$\ddots$	z_i	$\ddots$	$q^{'}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$

REMARK. Λ can be reduced to $\mathbb{B} = \{0, 1\}$ and k can be reduced to 1 without loss of computational power. Then, any Turing Machine can be expressed as a partial recursive function mapping $\mathbb{B}^* \to \mathbb{B}^* \cup \eta$, where η is the undefined non-halting output. Since $|\mathbb{B}^* \times \mathbb{B}^* \cup \eta| = |\mathbb{N} \times \mathbb{N}| = |\mathbb{N}|$, the set of all Turing machines is countably infinite.

1.3.2 Non-deterministic Turing Machines

Non-deterministic Turing Machines (NDTM) differ from deterministic Turing Machines by having two transition functions Δ_0, Δ_1 and a special state q_{accept}. From Arora and Barak:

> When a NDTM M computes a function, we envision that at each computational step M makes an arbitrary choice as to which of its two transition functions to apply. For every input x, we say that $M(x) = 1$ if there exists some sequence of these choices (which we call nondeterministic choices of M) that would make M reach q_{accept} on input x. Otherwise - if every sequence of choices makes M halt without reaching q_{accept} - then we say that $M(x) = 0$.

If $M(x) = 1$, we say that M accepts the input x. There are two ways to interpret the choice of update function to use in a NDTM. We can either assume that the NDTM chooses updates that will lead to an accepting state, or we can assume that the machine branches out into its choices such that it has a "computation tree" and if any of the branches reaches the accepting state then the machine accepts the input. From this second interpretation, the computational power of DTMs to NDTMs is analogous to the computational complexity of P to NP.

1.3.3 Universal Turing Machines

There exists a universal Turing machine that can simulate the execution of any other Turing machine M given the description of M as an input. It is a fundamental result that these machines can be explicitly constructed and that there are an infinite number of such. To imitate the behavior of M, a universal machine simulates the actions of M on a representation of the tape contents of M. The partial recursive function computed by any universal machine is called the universal partial recursive function.

2 Weak Inference

Definition 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 $\overline{\Gamma_1} = \overline{\Gamma_2}$.

The condition in this definition is equivalent to requiring the preimages of the two functions to be inequivalent. If two functions are equivalent and a device weakly infers one of the functions, then the device also infers the other. Now we address some conditions on functions that can be weakly inferred.

Lemma *A function Γ can be weakly inferred by a device $\mathcal{D}$ if $|\Gamma^{-1}(\gamma)| \geq 2$ for any $\gamma \in \Gamma(U)$.*
Proof Let $U := \mathbb{N}$. Enumerate $\gamma \in \Gamma$ as $1, 2, \ldots, i, \ldots, n$ and define $V^i = \{u : \Gamma^{-1}(i) = u\}$. Then enumerate each element in V^i as $i_1, i_2, \ldots, i_j, \ldots, i_m$. Continue to define X and Y as

$$X(i_j) = \begin{cases} a_i & \text{if } j = 2 \\ b_i & \text{otherwise} \end{cases} \qquad Y(i_j) = \begin{cases} -1 & \text{if } j = 2 \\ 1 & \text{otherwise.} \end{cases}$$

Then for each $i \in \Gamma(U)$ choose $x = b_i$ to force $Y(X^{-1}(b_i)) = \delta_i(\Gamma(X^{-1}(b_i)) = i) = 1$. $\square$

Lemma *A function Γ can be weakly inferred by a device $\mathcal{D}$ if $|\Gamma(U)| \geq 3$.*
Proof Let $U := \mathbb{N}$. Enumerate $\gamma \in \Gamma(U)$ as $1, 2, \ldots, i, \ldots, n$ and define $V^i = \{u : \Gamma^{-1}(i) = u\}$. Then enumerate each element in V^i as $i_1, i_2, \ldots, i_j, \ldots, i_m$. Continue to define X and Y as

$$X(i_j) = \begin{cases} a & \text{if } i = 1 \text{ and } j = 1 \\ b & \text{if } i = 2 \text{ and } j = 1 \\ c & \text{otherwise.} \end{cases} \qquad Y(i_j) = \begin{cases} 1 & \text{if } i = 1 \text{ and } j = 1 \\ -1 & \text{otherwise.} \end{cases}$$

Then for each $i \in \Gamma(U)$ to force $Y(X^{-1}(x)) = \delta_i(\Gamma(X^{-1}(x)) = i)$ choose $x = a$ if $i = 1$, $x = c$ if $i = 2$, or else choose $x = b$. □

COROLLARY *A function* Γ *cannot be weakly inferred by ay device if* $|\Gamma{-}1(\gamma)| < 2$ *and* $|\Gamma(U)| < 3$.

PROOF Fix any such Γ and set $U = \{1, 2\}$. Let $\Gamma(1) = \gamma_1$, $\Gamma(2) = \gamma_2$. Then either $Y(1) = 1, Y(2) = -1$, or vice versa. For the first case, consider the probe δ_{γ_2} to see that there is no $x \in X(U)$ that forces $Y(u) = \delta_{\gamma_2}(\Gamma(u))$. Consider δ_{γ_1} to see that weak inference does not hold for the remaining case. □

These results are enough to fully determine whether a single function is weakly inferable. We continue to examine the inferability of sets of functions by a singular device.

THEOREM *A countable set of inequivalent functions* A^* *can be weakly inferred by a device if each function* $A_i \in A^*$ *is independently inferable.*

PROOF Let $U := \mathbb{N}$ and fix any A^*. Let a_i, a_j, a_k represent any distinct three elements in $A_i(U)$. Write $V = \{1, 2, 3\} \subset U$. The following table represents all possible combinations of a_i, a_j, and a_k over V.

u	$X(u)$	$Y(u)$	$A_1(u)$	$A_2(u)$	$A_3(u)$	$A_4(u)$	$A_5(u)$	$\ldots$
1	1	-1	a_i	a_i	a_i	a_i	a_i	$\ldots$
2	2	-1	a_i	a_j	a_j	a_j	a_i	$\ldots$
3	3	1	a_i	a_k	a_j	a_i	a_j	$\ldots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

Note that setting $X(u) = u$, $Y(1) = -1$, $Y(2) = -1$, and $Y(3) = 1$ immediately satisfies weak inference for any function A_i that is functionally equivalent to A_1, A_2, A_3 and A_4 over V. This holds regardless of the values of $Y(u)$ and $A_{i\in\{1,2,3,4\}}(u)$ that follow for $u > 3$. This is shown by selecting x for the following cases:

Case 1: $\overline{A_i(V)} = \overline{A_1(V)}$.
Choose $x = 3$ for $a = a_i$ or otherwise choose $x = 2$.

Cases 2, 3, 4: $\overline{A_i(V)} = \overline{A_2(V)}$ or $\overline{A_3(V)}$ or $\overline{A_4(V)}$.
Choose $x = 2$ for $a = a_i$ or otherwise choose $x = 1$.

Now we need to guarantee weak inference for A_i that are functionally equivalent to A_5 over V. Enumerate each $A_i \in A^* : \overline{A_i(V)} = \overline{A_5(V)}$ as $B_4, B_5, \ldots, B_i, \ldots, B_n$. To satisfy weak inference for any A_i and B_i, define

X, Y more explicitly as

$$X(u) = u, \qquad Y(u) = \begin{cases} -1 & \text{if } u = 1, 2 \text{ or if } B_i(u) = a_j, a_k : X(u) = i \\ 1 & \text{if } u = 3 \text{ or if } B_i(u) = a_i : X(u) = i \\ -1 & \text{otherwise.} \end{cases}$$

For each $a \in B_i(U)$ to force $Y(X^{-1}(x)) = \delta_a(B(X^{-1}(x)))$ choose $x = i$ if $a = a_i$ or otherwise choose $x = 1$.

$\square$

This result is an example of the weakness of the definition of weak inference. Note that a device only needs two values of u to infer any single function Γ, and that for a set, these u do not always need to be unique for each Γ. In fact, a device never needs to answer affirmatively that $\gamma = \Gamma(U)$. This flexibility allows for the weak inference of a countably infinite inequivalent set of functions using only two u. These properties of weak inference are crucial to later investigations concerning the magnitude of information required to infer sets of functions.

3 Strong Inference

In the next three examples we examine strong inference of integer-valued functions. The right table indicates x for each s, δ such that the definition of strong inference is satisfied.

Example. Let $T(U) = \{0, 1\}$ and $S(U) = \{0, 1, 2\}$. We construct (X, Y) in the table at the left such that it strongly infers (S, T).

u	$X(u)$	$Y(u)$	$S(u)$	$T(u)$
1	1	1	0	0
2	2	-1	0	0
3	3	1	1	0
4	4	-1	1	0
5	5	1	2	1
6	6	-1	2	1

$s \setminus \delta$	δ_0	δ_1
0	1	2
1	3	4
2	6	5

$\diamond$

Example. Let $T(U) = \{1, 2, 3\}$ and $S(U) = \{1, 2, 3, 4, 5\}$.

u	$X(u)$	$Y(u)$	$S(u)$	$T(u)$
1	1	1	1	1
2	2	-1	1	1
3	3	1	2	1
4	4	-1	2	1
5	5	1	3	2
6	6	-1	3	2
7	7	1	4	2
8	8	-1	4	2
9	9	1	5	3
10	10	-1	5	3

$s \backslash \delta$	δ_1	δ_2	δ_3
1	1	2	2
2	3	4	4
3	6	5	6
4	8	7	8
5	10	10	9

◇

Example. Let $T(U) = \{1, 2, 3\}$ and $S(U) = \{1, 2\}$. In this example, the inferred function $f : S \to T, f(s) = T(S^{-1}(s))$ is not single-valued.

u	$X(u)$	$Y(u)$	$S(u)$	$T(u)$
1	1	-1	1	1
2	2	-1	1	2
3	3	1	1	3
4	4	-1	2	1
5	5	-1	2	2

$s \backslash \delta$	δ_1	δ_2	δ_3
1	2	1	1
2	5	4	4

◇

4 Inference of Turing Machines

Theorem *A deterministic Turing machine (Λ, Q, Δ) can be strongly inferred by a device iff*

$$\forall s \in S(U),\ |S^{-1}(s)| \geq 2.$$

This holds for both the representation of a Turing machine as a partial recursive function and the representation as an update function.

Proof First examine the partial function case. Let f be the partial recursive function that describes the given Turing machine tuple. Let $U := \mathbb{N}$. Choose any convenient single valued surjective function $S : U \to \mathbb{B}^*$ and define $T : U \to \mathbb{B}^* \cup \eta$ by $T(u) = f(S(u))$ as the single-valued function mapping U to the halting and non-halting outputs of f. Then f can be written as the single-valued mapping $S \to T$ by $f(s) = T(S^{-1}(s))$.

Enumerate the elements of $S(U)$ as $1, 2, \ldots, s, \ldots$. Let $V^s = \{u : S^{-1}(s) = u\}$ for $s \in S(U)$. Similarly enumerate the elements of V^s as $s_1, s_2, \ldots, s_{|V^s|}$. Then define X and Y as follows:

$$X(s_i) = \begin{cases} a_s & \text{if } i = 1 \\ b_s & \text{otherwise} \end{cases} \qquad Y(s_i) = \begin{cases} 1 & \text{if } i = 1 \\ -1 & \text{otherwise} \end{cases}$$

Note that the condition $|S^{-1}(s)| \geq 2$ is required to guarantee $Y(V^s) = \{1, -1\}$. For each pair $(s, \delta_{t \in T(U)})$, to force $S(u) = s$ and $Y(u) = \delta_t(T(u))$, choose $x = a_s$ if $t = T(s_1)$ or otherwise choose $x = b_s$. Since the choices of s and t were arbitrary, this holds for all (s, δ_t) pairs.

Now consider the update function that describes the given Turing machine. Recall that the update function is written as $\Delta : Q \times \Lambda^k \to Q \times \Lambda^{k-1} \times \{L, \mathcal{S}, R\}^k$, $k \geq 2$. Consider a convenient single-valued surjective function $S : U \to Q \times \Lambda^k$ representing the possible inputs for a Turing Machine and a corresponding single-valued $T : U \to Q \times \Lambda^{k-1} \times \{L, \mathcal{S}, R\}^k$ as $T(u) = \Delta(S(u))$. Observe that Δ can be written as the single-valued function $\Delta(s) = T(S^{-1}(s))$. Then define V^s, X, Y, and choose x for each (s, δ_t) as described in the preceding portion of the proof. Hence, the claim holds for the update function of a Turing machine.

To show that the condition is necessary for either representation, suppose that $|V^s| < 2$ for some s. If $V^s = \varnothing$ then there exists no x that can force $S = s$. If $|V| = 1$, then we can assign $Y(s_i) = y \in \{-1, 1\}$. However, whichever value is assigned, there exists a t such that $\delta_t(T(s_i)) \neq Y(s_i)$ since $|T(U)| \geq 2$. $\square$

Remark. Conventionally all functions over U must have a range of at least two elements. This implies that Turing machines that never halt cannot be strongly inferred.

5 Inference Complexity

Definition Let $\mathcal{D}$ be an inference device and Γ be a function over U where $X(U)$ and $\Gamma(U)$ are countable and $\mathcal{D} > \Gamma$. Let the **size** of $\gamma \in \Gamma(U)$ be written as $\mathcal{M}_{\mu:\Gamma(\gamma)} = -\ln[\int_{\Gamma^{-1}(\gamma)} d\mu(u)1]$ such that $d\mu$ denotes a measure over U. Then the **inference complexity** of Γ with respect to $\mathcal{D}$ and measure μ is defined as

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

The **strong inference complexity** for any single $\gamma \in \Gamma$ is

$$\mathcal{C}_\mu(\gamma; \mathcal{D}) \triangleq \min_{x : X(u) = x \implies S(u) = s, Y(u) = \delta_\gamma(\Gamma(u))} [\mathcal{M}_{\mu, X}(x)].$$

Claim *Let c be a positive integer. There exists a countable set of inferable functions A^* and device $\mathcal{D}$ that infers all functions $a \in A^*$ such that every finite subset $A \subseteq A^*$ with cardinality m has at least $m(1 - 2^{-c-1}) + 1$ elements $a \in A$ such that $\mathcal{C}_\mu(a; \mathcal{D}) \geq \log m - c$.*

Proof Choose A^* such that $\overline{a(V)}$ is equivalent to $\overline{A_5(V)}$ for $a \in A^*$. Then choose any subset $A \subseteq A^*$ with cardinality m. Let $U := \mathbb{B}^*$. Construct $\mathcal{D}$ such that it weakly infers $a \in A$. Take $d\mu(u) = \ell(u)$ where $\ell(b)$ is the length of a bit string b. Recall that the inference complexity of $a \in A$ with respect to $\mathcal{D}$ is

$$\mathcal{C}_\mu(a;\mathcal{D}) = \sum_{\delta\in\wp(a)} \min_{x:X=x \implies Y=\delta(a)} [\mathcal{M}_{\mu,X}(x)] : \mathcal{M}_{\mu,X}(x) = -\ln \sum_{X^{-1}(x)} \ell(u)$$

By the construction of A^* and $\mathcal{D}$, it requires at least two values of u to weakly infer each a, and at least one of these u must be unique between all a. Then the inference complexity of a with respect to $\mathcal{D}$ is

$$\mathcal{C}_\mu(a;\mathcal{D}) = \sum_{\delta\in\wp(a)} \min_{x:X=x \implies Y=\delta(a_i)} \sum_{X^{-1}(x)} \ell(a^{-1}(a_i))$$

$$= \sum_{\delta\in\wp(a)} \min_{x:X=x \implies Y=\delta(a_i)} na^{-1}(a_i) = 1 + (a^{-1}(a_i)) = \ell(u) + 1.$$

Since a unique u is used by $\mathcal{D}$ to infer every $a \in A$, we can apply the pigeonhole principle. The number of bit strings of length less than $m - c - 1$ is

$$\sum_{i=0}^{m-c-2} 2^i = 2^{\log m - c - 1} - 1.$$

By the pigeonhole principle, there are at least $m - m2^{-c-1} + 1$ functions $a \in A$ which have inference complexity greater than $\log m - c$. $\square$

THEOREM *There exist A^* and $\mathcal{D}$ such that for all functions of any subset $A \subseteq A^*$, $\mathcal{C}_\mu(A_i;\mathcal{D}) = 0$.*

PROOF Recall the functions and spaces defined in the Countable Inference Theorem and take $\mu(u) = \ell(u)$. Then choose $A^* : \overline{A_i(V)} \in A^* = \overline{A_{j\in\{2,3,4\}}(V)}$. Then the inference complexity for any function $A_i \in A$ is

$$\mathcal{C}_\mu(A_i;\mathcal{D}) = \sum_{\delta\in\wp(A_i)} \min_{x:X=x \implies Y=\delta(A_i)} -2\ln \sum_{X^{-1}(x)} \ell(u) = \sum_{\delta\in\wp(f)} \min_{x:X=x \implies Y=\delta(A_i)} -\ln 1 = -2\ln 1 = 0.$$

$\square$

6 Incompressibility

INCOMPRESSIBILITY THEOREM *Let c be a positive integer. For each fixed y, every finite set A of cardinality m has at least $m(1 - 2^{-c}) + 1$ elements x with $\mathcal{C}(x|y) \geq \log m - c$. This result is taken from Li and Vitanyi (1993).*

PROOF The number of programs of length less than $\log m - c$ is

$$\sum_{i=0}^{\log m - c - 1} 2^i = 2^{\log m - c} - 1.$$

Hence, there are at least $m - m2^{-c} + 1$ elements in A which have no program of length less than $\log m - c$. $\square$

Note. Let $\mathcal{M}_{\mu:\Gamma(\gamma)} = [\int_{\Gamma^{-1}(\gamma)} d\mu(u)1]$ for the remainder of this section. The ln term is not needed because U will be taken over the space of bit strings instead of the natural numbers.

THEOREM *There is a constant c, such that $\mathcal{C}_\mu(t;\mathcal{D}) \leq 2\ell(t) + c$ for all t.*

PROOF Let $U := \mathbb{B}^* \times \{0,1\}$. Fix a Turing machine ϕ over U such that $S : U \to \mathbb{B}^*$ is written as $S((u,b)) = u$. Then define $T : U \to \mathbb{B}^*$ as $T(u) = S(u)$. Construct a device $\mathcal{D}$ that strongly infers (S,T). Then for every $t \in \mathbb{B}^*$, $\mathcal{C}_\mu(t;\mathcal{D}) = 2\ell(t) + 2$. □

THEOREM *Let c be a positive integer. For a fixed $\mathcal{D}$, every finite set $T(U)$ of cardinality m has at least $m(1 - 2^{-c-1}) + \frac{1}{2}$ elements t with $\mathcal{C}_\mu(t;\mathcal{D}) \geq \log m - c$.*

PROOF Set $U := \mathbb{B}^*$. Fix any universal partial recursive function ϕ that can be strongly inferred. Let $\mu(u) = \ell(u)$. Construct a device $\mathcal{D}$ and functions (S,T) over ϕ such that $\mathcal{D} \gg (S,T)$. By the construction of $\mathcal{D}$, it takes at least two unique values of u to force strong inference for each value t. The number of bit strings of length less than $\log m - c$ is

$$\sum_{i=0}^{\log m - c - 1} 2^i = 2^{\log m - c} - 1.$$

Then the greatest number of unique pairs of bit strings with combined length less or equal to $\log m - c$ is $m2^{-c-1} - \frac{1}{2}$. Hence, there are at least $m - m2^{-c-1} + \frac{1}{2}$ values t with $\mathcal{C}_\mu(t;\mathcal{D}) \geq \log m - c$. □

THEOREM *Consider Γ with countably infinite image over U. (i) The function $\mathcal{C}(\gamma;\mathcal{D})$ is unbounded. (ii) Define a function m by $m(\gamma) = \min\{\mathcal{C}(y;\mathcal{D}) : y \geq \gamma\}$. That is, m is the greatest monotonic increasing function bounding $\mathcal{C}$ from below. The function $m(\gamma)$ is unbounded.*

PROOF (i) This follows immediately from (ii). (ii). For each i there is a least γ_i such that for all $\gamma > \gamma_i$ the smallest set of $u \in U$ strongly inferring γ has a total length greater or equal to i. This follows immediately from the fact that there are only a finite number of u of each length i. Clearly, for all i we have $\gamma_{i+1} \geq \gamma_i$. Now observe that the function m has the property that $m(\gamma) = i + 1$ for $\gamma_i < \gamma \leq \gamma_{i+1}$. This proves (ii). □

7 Physical Knowledge

DEFINITION 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 $\exists \xi : \Gamma(U) \to \overline{X}$ such that

i) $\forall \gamma' \in \Gamma(U), u \in \xi(\gamma') \implies \delta_{\gamma'}(\Gamma(u)) = Y(u)$
ii) $\varnothing \neq \xi(\gamma) \cap W \subseteq Y^{-1}(1)$
iii) For all $\gamma' \neq \gamma$, $\varnothing \neq \xi(\gamma') \cap W \subseteq Y^{-1}(-1)$.

A device $\mathcal{D}$ physically knows a function Γ over W iff $\forall \gamma \in \Gamma(U)$, $\exists \xi : \mathcal{D}$ knows γ over $W \subseteq U$ using ξ.

Definition The **knowledge complexity** of Γ given $\mathcal{D}$, ξ, and measure μ is defined as:

$$\mathcal{K}_\mu(\Gamma;\mathcal{D}) \triangleq \mathcal{M}_\mu[\bigcup_{\gamma\in\Gamma(U)} \min_{\overline{\{\xi_i\}}:\mathcal{D} \text{ knows } \Gamma=\gamma \text{ using } \xi_i\in\xi^*}]$$

$$\mathcal{M}_{\mu;\xi}(\gamma) = -\ln \int_{\xi(\gamma)} d\mu(u)1$$

Example. Let $U := \{1,2,3\}$ and take $W = U$. The following table and formulas demonstrate physical knowledge for $\gamma \in \Gamma : |\Gamma(U)| = 2$.

u	$X(u)$	$Y(u)$	$\Gamma(u)$
1	1	1	1
2	2	-1	2
3	3	1	2
4	4	-1	1

$$\xi_1(\gamma) = \begin{cases} \{1\} & \text{if } \gamma = 1 \\ \{4\} & \text{if } \gamma = 2. \end{cases}$$

$$\xi_2(\gamma) = \begin{cases} \{2\} & \text{if } \gamma = 1 \\ \{3\} & \text{if } \gamma = 2. \end{cases}$$

$\diamond$

Lemma *For every countably-ranged function Γ there exists a device $\mathcal{D}$ such that $\mathcal{D}$ physically knows all $\gamma \in \Gamma(U)$ over $W = U$ if $|\Gamma^{-1}(\gamma)| \geq 2$ for all $\gamma \in \Gamma$.*

Proof Take $U := \mathbb{N}$ and $W = U$. Choose any Γ as described in the claim. Enumerate $\gamma \in \Gamma(U)$ as $1, 2, \ldots, i, \ldots$. Define $V^i = \{u : \Gamma^{-1}(\gamma) = u\}$ for each $i \in \Gamma(U)$. Similarly enumerate each $u \in V^i$ as $i_1, i_2, \ldots, i_j, \ldots$. Now define the device (X, Y) as

$$X(i_j) = \begin{cases} a_i & \text{if } j = 1 \\ b_i & \text{if } j = 2 \\ c_i & \text{otherwise} \end{cases} \qquad Y(i_j) = \begin{cases} 1 & \text{if } j = 1 \\ -1 & \text{otherwise.} \end{cases}$$

Define $\xi_\gamma : \Gamma(U) \to \overline{X}$ for each $i \in \Gamma(U)$ as

$$\xi_\gamma(i) = \begin{cases} \overline{a_i} & \text{if } i = \gamma \\ \overline{b_i} & \text{otherwise.} \end{cases}$$

Then for all $\gamma \in \Gamma(U)$ and all $u \in \xi_\gamma(\gamma')$, $\delta_\gamma(\Gamma(u)) = Y(u)$. Furthermore, for each ξ_γ, $\varnothing \neq \{(\xi_\gamma(\gamma' = \gamma) \cap W) \subseteq Y^{-1}(1)\}$ and $\varnothing \neq \{(\xi_\gamma(\gamma' \neq \gamma) \cap W) \subseteq Y^{-1}(-1)\}$. Hence, $\mathcal{D}$ physically knows $\gamma \in \Gamma(U)$. $\square$

Theorem *For any set of independently physically knowable characteristic functions A^* there is a device $\mathcal{D}$ and set of functions $\xi*$ such that $\mathcal{D}$ knows $A \in A^*$ over a given subset $W \subseteq U$ using ξ^*.*

Proof
Let $U := \mathbb{N}$ and take $W = U$. Fix A^* and construct $\mathcal{D}$ such that $\mathcal{D} > A \in A^*$. For all $A \in A^*$, or all $\gamma \in A(U)$, define ξ_γ as specified in the preceding theorem. Then for all $\gamma \in A(U)$ and all $u \in \xi_\gamma(\gamma')$, $\delta_\gamma(\Gamma(u)) = Y(u)$. Furthermore, for each ξ_γ, $\varnothing \neq \{(\xi_\gamma(\gamma' = \gamma) \cap W) \subseteq Y^{-1}(1)\}$ and $\varnothing \neq \{(\xi_\gamma(\gamma' \neq \gamma) \cap W) \subseteq Y^{-1}(-1)\}$.

Hence $\mathcal{D}$ physically knows $\gamma \in A(U)$ for $A \in A^*$. $\square$

THEOREM *Let Γ be a knowable function with countable range and $\mathcal{D}$ be a device that physically knows all $\gamma \in \Gamma$ over U. Every such Γ with $|\Gamma(U)| = m$ has knowledge complexity of at most $-\ln \sum_{i=1}^{2m} i$.*

PROOF Set $U := \mathbb{N}$. Fix Γ with $|\Gamma(U)| = m$. Let $\mu(u) = \ell(u)$. Construct a device $\mathcal{D}$ and a set of functions ξ^* such that $\mathcal{D}$ physically knows $\gamma \in \Gamma$ over U. By the construction of $\mathcal{D}$, it takes at least $2m$ values of u to physically know each value γ over U. Then the knowledge complexity of Γ with $\mathcal{D}$ using $\xi_{\gamma\in\Gamma(U)}$ is

$$\mathcal{K}_\mu(\Gamma;\mathcal{D}) = \mathcal{M}_\mu(\bigcup_{\gamma\in\Gamma(U)} \min_{\overline{\{\xi_i\}}:\mathcal{D} \text{ knows } \Gamma=\gamma \text{ using } \xi_i\in\xi^*}) =$$

$$\mathcal{M}_\mu(\{\epsilon, \ldots, 2m\}) = -\ln \sum_{i=1}^{2m} i$$

$\square$

CLAIM *There exists a countable set of inequivalent characteristic functions A^* and a device $\mathcal{D}$ such that for all $A \in A^*$, $\mathcal{D}$ physically knows A over U and $\mathcal{K}_\mu(A;\mathcal{D}) = -\ln 10$.*

PROOF Set $U := \mathbb{N}$. Fix A^* such that for any $A \in A^*$, $A(1) = A(2) = 0$ and $A(2) = A(3) = 1$. Let $\mu(u) = \ell(u)$. Construct a device $\mathcal{D}$ and a set of functions ξ^* such that $\mathcal{D}$ physically knows $a \in A$ over U. By the construction of $\mathcal{D}$, it takes only the values $u = 1, 2, 3, 4$ to physically know each A over U. Then the knowledge complexity of all A with $\mathcal{D}$ using $\xi_{\gamma\in\Gamma(U)}$ is

$$\mathcal{K}_\mu(\Gamma;\mathcal{D}) = \mathcal{M}_\mu(\bigcup_{\gamma\in\Gamma(U)} \min_{\overline{\{\xi_i\}}:\mathcal{D} \text{ knows } \Gamma=\gamma \text{ using } \xi_i\in\xi^*}) = -\ln \sum_{i=1}^{4} i = -\ln 10$$

$\square$

8 Open Questions

Much of the results covered in this manuscript are preliminary and leaves much to be desired in our understanding of the theory of inference devices and its connections to algorithmic information theory. There are many immediate questions deserving attention. The question of whether non-deterministic Turing machines may be strongly inferred was left unattended. Our incompressibility results with strong inference of universal Turing machines may be extended by exploring possible analogues between Kolmogorov complexity and strong inference complexity. The numerous negative results we found concerning the definitions of weak inference complexity and physical knowledge suggest that a revisitation of these definitions may be appropriate or fruitful. Any revisal of these definitions should consider possible compatibility with the similar topics of Shannon information and prefix code.

9 Acknowledgements

I am grateful to David Wolpert for his mentorship and for many detailed discussions. I would like to thank the education office of the Santa Fe Institute (SFI) and the participants of the SFI Research Experience for Undergraduates (REU) program for providing valuable resources and assistance. This work was supported by the National Science Foundation, Grant No. 1757923.

---

[1] Wolpert DH. *Constraints on physical reality arising from a formalization of knowledge.* arXiv preprint arXiv:1711.03499v3 [physics.hist-ph] (2018).

[2] Li M & Vitnyi P. *An Introduction to Kolmogorov Complexity and Its Applications* (1st ed.). Springer-Verlag. (1993).

[3] Arora S & Barak B. *Computational Complexity: A Modern Approach.* Cambridge University Press. (2009).