

# Intelligence via ultrafilters: structural properties of some intelligence comparators of deterministic Legg-Hutter agents

**Samuel Allen Alexander**

SAMUELALLENALEXANDER@GMAIL.COM

*Quantitative Research Analyst*

*The U.S. Securities and Exchange Commission*

*New York Regional Office*

**Editor:** Florin Popescu

## Abstract

Legg and Hutter, as well as subsequent authors, considered intelligent agents through the lens of interaction with reward-giving environments, attempting to assign numeric intelligence measures to such agents, with the guiding principle that a more intelligent agent should gain higher rewards from environments in some aggregate sense. In this paper, we consider a related question: rather than measure numeric intelligence of one Legg-Hutter agent, how can we compare the relative intelligence of two Legg-Hutter agents? We propose an elegant answer based on the following insight: we can view Legg-Hutter agents as candidates in an election, whose voters are environments, letting each environment vote (via its rewards) which agent (if either) is more intelligent. This leads to an abstract family of comparators simple enough that we can prove some structural theorems about them. It is an open question whether these structural theorems apply to more practical intelligence measures.

## 1. Introduction

Who is more intelligent,  $A$  or  $B$ ? This is a paper about how to formalize a definition of the relative intelligence of agents who interact deterministically with deterministic environments based on observations and rewards<sup>1</sup> (formalized below). Starting with the landmark paper of Legg and Hutter (2007), various attempts (Hernández-Orallo and Dowe, 2010; Hibbard, 2011; Legg and Veness, 2013; Gavane, 2013) have been made to formalize intelligence measures of such agents<sup>2</sup>. We will refer to such an agent as a *deterministic Legg-Hutter agent*, or *DLHA* for short, to distinguish it from other agent notions. DLHAs' intelligence can be defined in various ways using real numbers (for example, by means of infinite series involving Kolmogorov complexities). The underlying principle is that a DLHA with higher intelligence should earn higher rewards from environments, in some aggregate sense. Thus, we are talking about a type of intelligence one would consider if one were evaluating a general-purpose utility-maximizing AI for usage in many well-defined, rational, emotionless tasks.

- 
1. Essentially, the agents and environments of *general reinforcement learning*, but with a more universal point of view.
  2. Authors differ on whether or not to require determinism.

Aside from quantifying and comparing the intelligence of individual DLHAs, this field of research has the potential to shed light on (at least certain aspects of) intelligence itself, abstracted over many DLHAs. To that end, we intend to present a notion of the relative intelligence of DLHAs, which notion exhibits *structural properties*. What we mean by a “structural property” is roughly the following: if one knows something about the intelligence of an agent  $A$ , a “structural property” is a theorem which allows one to infer something about the intelligence of some agent  $A'$  which is a variation of  $A$  in some sense.

Here is an example to make the notion of a “structural property” more concrete. Suppose  $A, B, A', B'$  are agents, and we have some way of letting  $A$  and  $B$  (resp.  $A'$  and  $B'$ ) “team up” into a new agent  $C$  (resp.  $C'$ ). It might seem intuitively plausible that if  $A$  is more intelligent than  $A'$ , and if  $B$  is more intelligent than  $B'$ , then the team  $C$  obtained from  $A$  and  $B$  should be more intelligent than the team  $C'$  obtained from  $A'$  and  $B'$ . This property, “higher-intelligence team-members make higher-intelligence teams,” if it held, would be an example of a structural property of intelligence. Later in this paper, we will explore this property for two specific notions of DLHA teamwork.

In the future, we hope that the investigation of structural properties of an algebraic nature will help to reveal qualitative facts about intelligence. As a precedent, one of the structural properties we prove (Proposition 22) helped us realize that if environments can give negative rewards, then the relative intelligence of different DLHAs really ought to depend on how risk-averse we are; more on this below.

Rather than directly compute a real number intelligence measure for any particular DLHA, we instead focus on the simpler problem of how to determine whether one DLHA is more intelligent than another<sup>3</sup>.

We will introduce an abstract intelligence comparison based on the following insight: when considering whether DLHA  $A$  is more intelligent than DLHA  $B$ , we can imagine that different environments are *voters* voting in an election with three candidates<sup>4</sup>. Those three candidates are: “ $A$  is more intelligent”, “ $B$  is more intelligent”, and “ $A$  and  $B$  are equally intelligent”. Arrow’s impossibility theorem (Morreau, 2016) would crush all hope of easily<sup>5</sup> obtaining a reasonable non-dictatorial solution in this manner if there were only finitely many environments casting votes, but that is not the case when there are infinitely many environments casting votes. We must instead turn to the infinite-voters version of Arrow’s impossibility theorem (Kirman and Sondermann, 1972) (see also Fishburn, 1970), which *does* admit non-dictatorial solutions in terms of ultrafilters.

The paper is structured as follows:

- 
3. In more practical, less universal contexts, much work has been done on comparing agents, see Balduzzi et al. (2018) (section 1) for a collection of references. Many practical methods for comparing agents turn out to be non-transitive, but the abstract method we introduce is transitive (Lemma 11).
  4. We are not the first to apply election theory to unexpected areas by reinterpreting the voters and candidates. For non-AI uses of this technique, see (Okasha, 2011; Stegenga, 2013).
  5. Technically speaking, the way environments vote in different elections is constrained, with complicated relationships binding how environments vote in different such elections. Thus, it is still possible that other solutions to the voting problem might exist, since Arrow’s impossibility theorem assumes that how a voter votes in one election is independent of how that voter votes in a different election. But because of the complicated nature of the constraints, I conjecture it would be difficult to actually exploit this loophole.

- Section 2 contains preliminaries: a formalization of the agent-environment model, and a definition of ultrafilters.
- In Section 3 we define a family of intelligence comparators—functions that compare intelligence of DLHAs—in terms of ultrafilters, and establish some basic lemmas about these comparators.
- In Section 4, we prove some structural properties about our intelligence comparators.
- In Section 5, we discuss some anticipated objections to our intelligence comparators and to the background framework.
- In Section 6, we state some open questions about intelligence measures defined by other authors.
- In Section 7, we summarize and make concluding remarks.

## 2. Preliminaries

We will begin with a formal definition of deterministic Legg-Hutter agent (DLHA) and of deterministic environment. We envision an agent being placed in an environment, where the agent receives an initial numerical reward, receives an initial observation, takes an initial action, receives a numerical reward for that action, receives an observation of how the environment has changed, takes a second action, and so on *ad infinitum*. This high-level vision basically agrees with Legg and Hutter (2007) and other authors, except that we require both agent and environment to be deterministic, whereas Legg and Hutter allowed both agent and environment to have an element of randomness. At first glance this vision might seem quite removed from real-world agents, but the abstract numbers which we refer to as “rewards” can stand in for any number of realistic quantities such as: amounts of gold extracted from a mine; amounts of dopamine released by a brain; amounts of positive ratings on a website; etc. Similarly, a single numerical “observation” can encode any sort of real-world observation we can think of (just as any computer file is really just a long string of binary, i.e., a single number). This model seems flexible enough to accommodate many, if not all, practical uses to which one would want to put a general-purpose AI. For example, if we want the general-purpose AI to translate documents, we can concoct an environment which incentivizes that; if we want the general-purpose AI to analyze protein-folding, we can concoct an environment which incentivizes that; and so on. To achieve good rewards across the universe of all environments, such an AI would need to have (or appear to have) creativity (for those environments intended to reward creativity), pattern-matching skills (for those environments intended to reward pattern-matching), ability to adapt and learn (for those environments which do not explicitly advertise what things they are intended to reward, or whose goals change over time), etc. See Legg and Hutter (2007) for further discussion of how this model relates to machine intelligence (and for caveats about the agent-environment model).

To simplify the mathematics<sup>6</sup>, we want to avoid the awkward situation where the reward-sequence an agent receives adds up to  $\infty$ , or to  $-\infty$ , or diverges (since we will allow negative rewards). There are different ways to avoid this. For example, one approach would be to require that the  $n$ th reward be bounded between  $\pm 2^{-n}$ . For increased generality<sup>7</sup>, we employ (in its debut appearance) a more abstract approach: we will first define what we call a deterministic pre-environment, with no such constraints, and then we will define a deterministic environment to be a deterministic pre-environment whose reward sequences (for every arbitrary DLHA) converge.

**Definition 1** (*Environment and Agent*)

1. A deterministic pre-environment is a function  $e$  which takes as input a finite (possibly empty) sequence  $a_1, \dots, a_n$  of natural numbers, called actions. It outputs a pair  $e(a_1, \dots, a_n) = (r, o)$ , where  $r \in \mathbb{R}$  is called a reward and  $o \in \mathbb{N}$  is called an observation.
2. A deterministic Legg-Hutter agent (or DLHA) is a function  $A$  which takes as input a finite sequence  $r_1, o_1, \dots, r_n, o_n$  of reward-observation pairs, and outputs an action  $A(r_1, o_1, \dots, r_n, o_n)$ .
3. Suppose  $A$  is a DLHA and  $e$  is a deterministic pre-environment. We define the reward-observation-action sequence determined by letting  $A$  play in  $e$  to be the infinite sequence  $r_1, o_1, a_1, r_2, o_2, a_2, \dots$  defined inductively as follows:
  - The initial reward and initial observation,  $r_1, o_1 = e(\langle \rangle)$ , are obtained by plugging the empty action-sequence into  $e$ .
  - The initial action  $a_1 = A(r_1, o_1)$  is obtained by plugging  $r_1, o_1$  into  $A$ .
  - Assume  $r_1, o_1, a_1, \dots, r_i, o_i$  have been defined. We define  $a_i$ , the action which  $A$  performs in response to reward-observation sequence  $r_1, o_1, \dots, r_i, o_i$ , to be  $a_i = A(r_1, o_1, \dots, r_i, o_i)$ .
  - Assume  $r_1, o_1, a_1, \dots, r_i, o_i, a_i$  have been defined. We define  $r_{i+1}, o_{i+1}$ , the reward and observation produced in response to action-sequence  $a_1, \dots, a_i$ , to be  $r_{i+1}, o_{i+1} = e(a_1, \dots, a_i)$ .
4. A deterministic pre-environment  $e$  is a deterministic environment if, for every DLHA  $A$ , the total reward  $r_1 + r_2 + \dots$  which  $A$  achieves when  $A$  plays in  $e$ , converges.
5. If  $A$  and  $B$  are DLHAs and  $e$  is a deterministic environment, we say  $A$  outperforms  $B$  on  $e$  if the total reward which  $A$  achieves when  $A$  plays in  $e$ , is larger than the total reward which  $B$  achieves when  $B$  plays in  $e$ . We define what it means for  $A$  to underperform  $B$  on  $e$ , or for  $A$  and  $B$  to perform equally well on  $e$ , in similar ways.

---

6. To be clear, the results in this paper could be generalized so as to make this simplification unnecessary. But if we went that route, all the mathematics would be more tedious, and there would be little additional insight gained. See Legg and Hutter (2007) for justification and caveats about the decision to restrict to environments with convergent reward sequences.

7. We will see in Section 4.2 that the more we restrict the universe of environments, the more structural properties will result. We prefer to initially lay the foundations as general as possible, and only specialize as needed.

**Example 1** • *The function  $A(r_1, o_1, \dots, r_i, o_i) = 0$  is a trivial DLHA which totally ignores the environment and blindly performs the same action over and over.*

- *The function  $e(a_1, \dots, a_i) = (1, 0)$  is a trivial deterministic pre-environment which ignores a DLHA’s actions, instead always blindly rewarding the DLHA with a reward of  $r = 1$ , and never allowing any new observations. This deterministic pre-environment might be thought of as a sensory-deprivational paradise where DLHAs receive constant injections of reward from an immutable void. This deterministic pre-environment is not a deterministic environment, because the total reward it grants to any DLHA diverges to  $+\infty$ .*
- *The function  $e(a_1, \dots, a_i) = (0, 0)$  is a deterministic pre-environment similar to the previous one, except that it never grants DLHAs any reward. It is a deterministic environment because the total reward it grants any DLHA converges (to 0).*
- *For a nontrivial example, choose a single-player video-game  $V$  and consider a function  $e(a_1, \dots, a_i) = (r_{i+1}, o_{i+1})$  where  $o_{i+1}$  encodes the image displayed on screen by  $V$  after the player presses buttons encoded by  $a_1, \dots, a_i$ . Assuming  $V$  displays a numerical score on the screen, we can take  $r_{i+1}$  to be the currently displayed score minus the score which was displayed in  $o_i$  (or minus 0 if  $i = 0$ ). This function  $e$  is a deterministic pre-environment. There are various conditions on  $V$  which would suffice to make  $e$  an environment. For example, if  $V$  cannot be played forever (so the score eventually always freezes); or if scores in  $V$  never decrease, and have some limiting max score; etc.*

One might be tempted to wonder whether it is possible to find one single, particularly clever deterministic environment such that each DLHA’s intelligence can be defined simply as the total reward it earns in that environment. The following lemma will show that this is impossible: no single environment  $e$  can, by itself, serve as a good measure of intelligence.

**Lemma 2** *(Inadequacy of any single deterministic environment) Let  $e$  be any deterministic environment. For every DLHA  $A$ , there exists some DLHA  $B$  such that  $A$  and  $B$  achieve the same total reward when they play in  $e$ , and yet  $B$ ’s output never depends on which rewards or observations are input into it—loosely speaking,  $B$  ignores everything.*

**Proof** Let  $r_1^0, o_1^0, a_1^0, \dots, r_i^0, o_i^0, a_i^0, \dots$  be the reward-observation-action sequence determined by letting  $A$  play on  $e$ . Define  $B$  by  $B(r_1, o_1, \dots, r_i, o_i) = a_i^0$ , note that  $B(r_1, o_1, \dots, r_i, o_i)$  only depends on  $i$ . Informally speaking,  $B$  completely ignores everything and instead blindly regurgitates the actions which  $A$  takes on  $e$ . By construction,  $B$  outputs the exact same actions as  $A$  on the particular environment  $e$ , so  $A$  and  $B$  achieve the same total reward when they play in  $e$ . ■

The proof of Lemma 2 recalls the following words from Sanghi and Dowe (2003): “...the reason that I.Q. test success can be automated with comparative ease is that administering an I.Q. test requires little intelligence—it requires comparatively little more than giving a list of questions with known answers.” Lemma 2 shows that if intelligence were measured by

performance on any single deterministic environment, then every DLHA would be exactly as intelligent as some DLHA that ignores the environment—a property which clearly should not hold of a reasonable intelligence measure.

In order to arrive at an abstract method of comparing the intelligence of different DLHAs, we will consider deterministic environments to be voters who vote in an election. The question is how to use these votes to decide the election. Since there are infinitely many deterministic environments, that means there are infinitely many voters, which scenario is thoroughly investigated by Kirman and Sondermann (1972). For the infinite voter case, Kirman and Sondermann showed that reasonable solutions are intimately related to the following mathematical logical device:

**Definition 3** *An ultrafilter on  $\mathbb{N}$  (hereafter simply an ultrafilter) is a collection  $\mathcal{U}$  of subsets of  $\mathbb{N}$  satisfying the following requirements:*

- (Properness)  $\emptyset \notin \mathcal{U}$ .
- (Upward Closure) For every  $X \in \mathcal{U}$  and every  $X' \subseteq \mathbb{N}$ , if  $X' \supseteq X$ , then  $X' \in \mathcal{U}$ .
- ( $\cap$ -closure) For every  $X, Y \in \mathcal{U}$ , the intersection  $X \cap Y \in \mathcal{U}$ .
- (Maximality) For every  $X \subseteq \mathbb{N}$ , either  $X \in \mathcal{U}$  or the complement  $X^c \in \mathcal{U}$ .

An ultrafilter is free if it contains no singleton  $\{n\}$  for any  $n \in \mathbb{N}$ .

For example, for each  $n \in \mathbb{N}$ , the set-of-subsets

$$\{X \subseteq \mathbb{N} : n \in X\}$$

is an ultrafilter, but not a free ultrafilter. It is not difficult to prove that every non-free ultrafilter has the above form.

The following theorem is well-known, and we state it here without proof. All the proofs of this theorem are non-constructive: in a sense which can be made formal, it is impossible to actually exhibit a concrete example of a free ultrafilter.

**Theorem 4** *There exists a free ultrafilter.*

There are two competing intuitions one can use to reason about an ultrafilter, and these intuitions almost seem to contradict each other.

- The first intuition is that “ $X \in \mathcal{U}$ ” can be read as “ $X$  contains almost every natural number”. Through this intuitive lens, the  $\cap$ -closure property seems obvious, whereas the Maximality property seems implausible.
- The second intuition is that “ $X \in \mathcal{U}$ ” can be read as “ $X$  is a winning bloc of natural numbers” (in an electoral sense). Through this lens, the Maximality property seems obvious (when two candidates compete in an election, one of them must win), whereas the  $\cap$ -closure property seems implausible.

Theorem 4 is profound because it says these two intuitions can be reconciled in a non-degenerate way.

### 3. An abstract intelligence comparator

We will arrive at an abstract method of comparing DLHAs' intelligence by means of an imaginary election. The method we arrive at will not itself directly involve elections: elections are merely the heuristic used to obtain our definition, and do not directly feature in our definition. So in motivating the definition, we will intentionally speak about elections in less than full formality. For full formality, see Kirman and Sondermann (1972).

We would like to answer the question, “who is more intelligent,  $A$  or  $B$ ?”, by letting different deterministic environments vote. In the election, there are three candidates: “ $A$  is more intelligent”, “ $B$  is more intelligent”, and “ $A$  and  $B$  are equally intelligent”. A deterministic environment  $e$  is considered to rank these three candidates, from most to least preferable, as follows:

1. If  $A$  earns more reward than  $B$  when run on  $e$ , then  $e$  ranks “ $A$  is more intelligent” most preferable, followed by “ $A$  and  $B$  are equally intelligent”, followed by “ $B$  is more intelligent”.
2. If  $B$  earns more reward than  $A$  when run on  $e$ , then  $e$  ranks “ $B$  is more intelligent” most preferable, followed by “ $A$  and  $B$  are equally intelligent”, followed by “ $A$  is more intelligent”.
3. If  $A$  and  $B$  earn the same reward when run on  $e$ , then  $e$  ranks “ $A$  and  $B$  are equally intelligent” most preferable, followed by “ $A$  is more intelligent”, followed by “ $B$  is more intelligent”.

The order of the last two preferences in the latter case is arbitrary. It would be more natural to simply let  $e$  vote on a single winner, but we make  $e$  rank the three candidates because that is the form of election which Kirman and Sondermann considered. It remains now to specify how to use the votes to determine a winner.

A method of turning voter preferences into a group-preference is called a *social welfare function*. A social welfare function is a *dictatorship* if there is a particular voter (called a *dictator*) whose preference always equals the group-preference given by the social welfare function. Using a dictatorship to settle the election in the above paragraph would amount to defining relative intelligence entirely by performance in one specific fixed deterministic environment, which is undesirable (by Lemma 2).

Two other desirable properties of a social welfare function are:

- (Unanimity) If every voter agrees to prefer candidate  $A$  over candidate  $B$ , then the group-preference prefers candidate  $A$  over candidate  $B$ .
- (Independence) Given any two voter-preference-sets and any two candidates, if the relative voter-preferences between the two candidates are the same in the two voter-preference-sets, then the corresponding group-preferences between the two candidates are the same.

Arrow's impossibility theorem states that if there are finitely many voters, then no social welfare function can satisfy (Unanimity) and (Independence) without being a dictatorship. See Morreau (2016) to get a better feel for what is at stake here. Fortunately, non-dictatorial

solutions are possible when there are infinitely many voters. The following theorem is an immediate corollary of Kirman and Sondermann (1972):

**Theorem 5** *For any countably infinite sequence  $\vec{e} = (e_0, e_1, \dots)$  of deterministic environments, suppose the members of  $\vec{e}$  shall rank the following three candidates: “A is more intelligent”, “B is more intelligent”, and “A and B are equally intelligent”. Let  $\Sigma$  be the corresponding set of social welfare functions satisfying (Unanimity) and (Independence).*

1. *For every social welfare function  $\sigma \in \Sigma$ , there is exactly one ultrafilter  $\mathcal{U}_\sigma$  with the following property: Any way  $e_0, e_1, \dots$  rank the three candidates, for every two candidates  $x$  and  $y$ , if*

$$\{n : e_n \text{ prefers } x \text{ over } y\} \in \mathcal{U}_\sigma,$$

*then  $\sigma$  says that  $x$  is preferred over  $y$  in the corresponding group-preference.*

2. *The map  $\sigma \mapsto \mathcal{U}_\sigma$  is a surjection onto the set of ultrafilters (in other words, every ultrafilter  $\mathcal{U}$  is equal to  $\mathcal{U}_\sigma$  for some social welfare function  $\sigma \in \Sigma$ ).*
3. *A social welfare function  $\sigma \in \Sigma$  is a dictatorship if and only if  $\mathcal{U}_\sigma$  is not free.*

**Proof** By Theorem 1 and Proposition 2 of Kirman and Sondermann (1972). ■

In particular, since Theorem 4 says there exists a free ultrafilter, Theorem 5 implies there is a non-dictatorship social welfare function satisfying (Unanimity) and (Independence).

Motivated by Theorem 5, we will define a family of intelligence comparators.

**Definition 6** *By an electorate, we mean a pair  $E = (\vec{e}, \mathcal{U})$  where  $\vec{e} = (e_0, e_1, \dots)$  is any countably infinite sequence of deterministic environments and  $\mathcal{U}$  is an ultrafilter.*

**Definition 7** *For every electorate  $E = (\vec{e}, \mathcal{U})$ , we define relations  $>_E$ ,  $<_E$  and  $=_E$  on DLHAs as follows. Let  $A$  and  $B$  be any two DLHAs.*

1. *If*

$$\{n : A \text{ outperforms } B \text{ on environment } e_n\} \in \mathcal{U},$$

*we declare  $A >_E B$  (and say  $A$  is more intelligent than  $B$  according to  $E$ ).*

2. *If*

$$\{n : B \text{ outperforms } A \text{ on environment } e_n\} \in \mathcal{U},$$

*we declare  $A <_E B$  (and say  $A$  is less intelligent than  $B$  according to  $E$ ).*

3. *If*

$$\{n : A \text{ earns the same reward as } B \text{ on } e_n\} \in \mathcal{U},$$

*we declare  $A =_E B$  (and say  $A$  and  $B$  are equally intelligent according to  $E$ ).*



If  $\mathcal{U}$  is non-free, then Definition 7 is trivial: it prescribes that we compare intelligence of DLHAs by comparing their performance in one fixed deterministic environment (because the corresponding social welfare function is a dictatorship). We are mainly interested in the case where  $\mathcal{U}$  is free<sup>8</sup>.

Note that Definition 7 already compares all possible DLHAs. Thus we avoid a common intelligence comparison problem, wherein one compares some limited set of agents, but then the addition of another agent changes the intelligence order of the original agents.

An important difference between our approach and that of Legg and Hutter (2007) is that when comparing the relative intelligence of two agents, we only concern ourselves with the sets of environments where each agent outperforms the other, and we do not concern ourselves with the numerical difference of the two agents' performance in any environment. Thus, to us, if DLHA  $A$  gains 10000 more reward than DLHA  $B$  on some deterministic environment, that only helps  $A$  the same amount as if  $A$  gained 0.00001 more reward than  $B$  on that environment. This is similar to how Elo ratings of chess-players are based only on final outcomes (win, loss, draw), ignoring any notion of the magnitude of a win or loss. We would defend this choice by pointing out that for any particular deterministic environment  $e$ , there is an equivalent deterministic environment  $e'$  which is identical to  $e$  in every way except that its rewards are all multiplied by 10000, and there is also an equivalent deterministic environment  $e''$  which is identical to  $e$  in every way except that its rewards are all multiplied by 0.00001.

Note that we could have stated Definition 7 without the precise structure of Definition 1. All that is needed to state Definition 7 is that agents outperform each other on environments, regardless of what exactly that actually means. However, the precise structure of DLHAs and deterministic environments becomes important when one wants to actually obtain nontrivial *structural properties* about the resulting comparators (Section 4).

**Lemma 8** *For every electorate  $E = (\vec{e}, \mathcal{U})$  and DLHAs  $A$  and  $B$ , exactly one of the following is true:*

1.  $A >_E B$ .
2.  $A <_E B$ .
3.  $A =_E B$ .

**Proof** Let

$$\begin{aligned} X_1 &= \{n : A \text{ earns more reward than } B \text{ on } e_n\}, \\ X_2 &= \{n : A \text{ earns less reward than } B \text{ on } e_n\}, \\ X_3 &= \{n : A \text{ earns the same reward as } B \text{ on } e_n\}. \end{aligned}$$

By Maximality (from Definition 3), either  $X_1 \in \mathcal{U}$  or  $X_1^c \in \mathcal{U}$ .

Case 1:  $X_1 \in \mathcal{U}$ . Then  $X_2 \notin \mathcal{U}$  lest  $X_1 \cap X_2 = \emptyset$  be in  $\mathcal{U}$  (by  $\cap$ -closure), which would violate Properness of  $\mathcal{U}$ . Similarly,  $X_3 \notin \mathcal{U}$ . Altogether,  $A >_E B$ ,  $A \not<_E B$ , and  $A \neq_E B$ .

---

8. In particular, the strategy of this paper would break down if the universe of deterministic environments were restricted to a finite set.

Case 2:  $X_1 \notin \mathcal{U}$ . By Maximality, either  $X_2 \in \mathcal{U}$  or  $X_2^c \in \mathcal{U}$ .

Subcase 1:  $X_2 \in \mathcal{U}$ . Then  $X_3 \notin \mathcal{U}$ , lest  $X_2 \cap X_3 = \emptyset$  be in  $\mathcal{U}$ , violating Properness. Altogether,  $A <_E B$ ,  $A \not>_E B$ , and  $A \neq_E B$ .

Subcase 2:  $X_2^c \in \mathcal{U}$ . By Maximality, either  $X_3 \in \mathcal{U}$  or  $X_3^c \in \mathcal{U}$ . We cannot have  $X_3^c \in \mathcal{U}$ , or else  $X_1^c \cap X_2^c \cap X_3^c = \emptyset$  would be in  $\mathcal{U}$ , violating Properness. So  $X_3 \in \mathcal{U}$ . Altogether,  $A =_E B$ ,  $A \not>_E B$ , and  $A \not<_E B$ . ■

**Lemma 9** *Let  $E = (\vec{e}, \mathcal{U})$  be an electorate and let  $A$  and  $B$  be DLHAs.*

1. *The following are equivalent:  $A =_E B$ ,  $B =_E A$ .*
2. *The following are equivalent:  $A >_E B$ ,  $B <_E A$ .*

**Proof** Straightforward. ■

**Lemma 10** *For every electorate  $E = (\vec{e}, \mathcal{U})$  and DLHA  $A$ ,  $A =_E A$ ,  $A \not>_E A$ , and  $A \not<_E A$ .*

**Proof** Straightforward (use the facts that  $\emptyset \notin \mathcal{U}$  by Properness, and  $\mathbb{N} \in \mathcal{U}$  by Properness plus Maximality). ■

When one defines a numerical intelligence measure for a single agent (as Legg and Hutter), the corresponding comparison of two agents is automatically transitive. If one instead chooses to directly compare two agents rather than measure a single agent, one runs the risk of losing transitivity. The following lemma shows that we have avoided that danger.

**Lemma 11** (*Transitivity*) *Let  $E = (\vec{e}, \mathcal{U})$  be an electorate and let  $A, B, C$  be DLHAs.*

1. *If  $A >_E B$  and  $B >_E C$ , then  $A >_E C$ .*
2. *If  $A <_E B$  and  $B <_E C$ , then  $A <_E C$ .*
3. *If  $A =_E B$  and  $B =_E C$ , then  $A =_E C$ .*

**Proof** We prove the  $>_E$  claim, the others are similar. Let

$$\begin{aligned} X_{AB} &= \{n : A \text{ earns more reward than } B \text{ on } e_n\}, \\ X_{BC} &= \{n : B \text{ earns more reward than } C \text{ on } e_n\}, \\ X_{AC} &= \{n : A \text{ earns more reward than } C \text{ on } e_n\}. \end{aligned}$$

Since  $A >_E B$ ,  $X_{AB} \in \mathcal{U}$ . Since  $B >_E C$ ,  $X_{BC} \in \mathcal{U}$ . Now, for any  $n$ , if  $A$  earns more reward than  $B$  on  $e_n$ , and  $B$  earns more reward than  $C$  on  $e_n$ , then  $A$  earns more reward than  $C$  on  $e_n$ . This shows  $X_{AC} \supseteq X_{AB} \cap X_{BC}$ . By  $\cap$ -closure,  $X_{AB} \cap X_{BC} \in \mathcal{U}$ , so by Upward Closure,  $X_{AC} \in \mathcal{U}$ , that is,  $A >_E C$ . ■

### 3.1 Comparing specialized intelligence

Definition 7 is flexible in that it allows one to choose which deterministic environments to consider. By considering specialized environments, one could obtain comparisons of specialized intelligence. For example, suppose  $P_0, P_1, \dots$  is a list of chess-playing computer programs. For each  $i$ , let  $e_{2i}$  (resp.  $e_{2i+1}$ ) be the environment: “play the white (resp. black) pieces in a game of chess against  $P_i$ , receiving +1 reward on the move (if any) in which you win,  $-1$  reward on the move (if any) in which you lose, and 0 reward on all other moves” (we can extend chess into an infinite game by obliging the players to play dummy moves forever after usual gameplay ends). For each ultrafilter, Definition 7 yields an abstract comparison of performance at the game of chess.

A couple nuances about this paper will be illuminated by contrasting the above-mentioned comparison with practical chess-rating systems such as Elo ratings. Rating systems such as Elo cast chess-players as agents who can improve (or get worse) from game to game. By contrast, the DLHAs of Definition 1 are fixed, and do not change from one environment to the next<sup>9</sup>. This illustrates that Definition 7 is not really appropriate for the intelligence of human beings; it is more appropriate for the intelligence of a general-purpose utility-maximizing AI intended to be used in many different well-defined, rational, emotionless tasks.

A primary reason we care about Definition 7 is because it is simple enough to allow the discovery of nontrivial, non-contrived structural properties. Something like Elo rating is infinitely more practical, but we doubt that it is possible to prove any nontrivial, non-contrived structural properties about performance as measured by Elo. Indeed, it is partly because our abstract definition treats agents as fixed and unchanging and deterministic that structural properties become obtainable. A chess-player’s Elo rating at any given time depends not only on the outcomes that player has achieved, but also on the order of those outcomes, and on the orders of the outcomes achieved by the opponents at each play.

## 4. Structural properties

In this section, we will exhibit some nontrivial structural properties of our intelligence comparators. The properties we have been able to come up with are humble and few, but we believe that in the context of Legg-Hutter-style intelligence measurement, they are first of their kind.

### 4.1 Properties of Teams

The following definition is a special case of a more general definition which will follow shortly.

---

9. This is reminiscent of an observation of Good (1955): in an ideal chess-rating system, assuming players were accurately rated, the expected change to the ratings after any particular chess-gameplay ought to be 0.

**Definition 12** For any DLHAs  $A$  and  $B$ , we define a new DLHA  $A \oplus B$  such that for every observation-reward sequence  $(r_1, o_1, \dots, r_n, o_n)$ ,

$$(A \oplus B)(r_1, o_1, \dots, r_n, o_n) = \begin{cases} A(r_1, o_1, \dots, r_n, o_n) & \text{if } o_1 \text{ is even,} \\ B(r_1, o_1, \dots, r_n, o_n) & \text{if } o_1 \text{ is odd.} \end{cases}$$

One can think of  $A \oplus B$  as a DLHA who plans to act as  $A$  or  $B$ , but has not yet decided which one. This DLHA will wait until seeing the first observation in the environment before committing to act as  $A$  or committing to act as  $B$ . In an intuitive sense,  $A \oplus B$  is a type of “team” formed by  $A$  and  $B$ .

**Proposition 13** Let  $E = (\vec{e}, \mathcal{U})$  be an electorate and let  $A, A', B, B'$  be DLHAs. If  $A >_E A'$  and  $B >_E B'$ , then  $A \oplus B >_E A' \oplus B'$ .

**Proof** Let  $V$  be the set of  $n$  such that  $e_n$  has first observation even, and let  $D$  be the set of  $n$  such that  $e_n$  has first observation odd. By Maximality, either  $V \in \mathcal{U}$ , or  $V^c = D \in \mathcal{U}$ . We will assume  $V \in \mathcal{U}$ , the other case is similar.

Let

$$\begin{aligned} X_1 &= \{n : A \text{ outperforms } A' \text{ on } e_n\}, \\ X_2 &= \{n : A \oplus B \text{ outperforms } A' \oplus B' \text{ on } e_n\}. \end{aligned}$$

Since  $A >_E A'$ ,  $X_1 \in \mathcal{U}$ . By  $\cap$ -closure,  $X_1 \cap V \in \mathcal{U}$ . Now, for every  $n \in X_1 \cap V$ , we have the following facts:

1.  $A$  outperforms  $A'$  on  $e_n$  (since  $n \in X_1$ ).
2.  $e_n$ 's first observation is even (since  $n \in V$ ).
3.  $A \oplus B$  acts exactly like  $A$  on  $e_n$  (since  $e_n$ 's first observation is even).
4.  $A' \oplus B'$  acts exactly like  $A'$  on  $e_n$  (since  $e_n$ 's first observation is even).

So for every  $n \in X_1 \cap V$ ,  $A \oplus B$  outperforms  $A' \oplus B'$  on  $e_n$ . This shows  $X_2 \supseteq X_1 \cap V$ . By Upward Closure,  $X_2 \in \mathcal{U}$ , so  $A \oplus B >_E A' \oplus B'$ .  $\blacksquare$

Proposition 13 depends on the fact that we compare intelligence only by where an agent outperforms, without regard for the magnitude difference in rewards. Otherwise, we could imagine a DLHA  $A$  who performs slightly worse than  $A'$  on even-numbered environments but makes up for it by clobbering  $A'$  on odd-numbered environments, making  $A$  more intelligent than  $A'$ . And we could imagine a DLHA  $B$  who performs slightly worse than  $B'$  on odd-numbered environments but clobbers  $B'$  on even-numbered environments. Then  $A \oplus B$  would perform worse than  $A' \oplus B'$  everywhere, despite each team-member being more intelligent than its counterpart.

The following definition generalizes Definition 12.

**Definition 14** Let  $X$  be any set of reward-observation sequences and let  $A$  and  $B$  be DLHAs. The team combination of  $A$  and  $B$  given by  $X$  is the DLHA  $A \oplus_X B$  defined as follows. Suppose  $r_1, o_1, \dots, r_n, o_n$  is any observation-reward sequence. If for all  $m \leq n$ ,  $A(r_1, o_1, \dots, r_m, o_m) = B(r_1, o_1, \dots, r_m, o_m)$ , then we declare  $(A \oplus_X B)(r_1, o_1, \dots, r_n, o_n) = A(r_1, o_1, \dots, r_n, o_n)$ . Otherwise, let  $m \leq n$  be minimal such that  $A(r_1, o_1, \dots, r_m, o_m) \neq B(r_1, o_1, \dots, r_m, o_m)$ . We declare

$$(A \oplus_X B)(r_1, o_1, \dots, r_n, o_n) = \begin{cases} A(r_1, o_1, \dots, r_n, o_n) & \text{if } (r_1, o_1, \dots, r_m, o_m) \in X, \\ B(r_1, o_1, \dots, r_n, o_n) & \text{otherwise.} \end{cases}$$

In other words,  $A \oplus_X B$  is the DLHA which has decided either to act as  $A$ , or to act as  $B$ , but refuses to commit to one or the other until it is forced to. In any particular deterministic environment, as long as the observations and rewards are such that  $A$  and  $B$  would act identically, then  $A \oplus_X B$  acts in that way, without committing to either one. Only when the observations and rewards are such that  $A$  and  $B$  would choose different actions, does  $A \oplus_X B$  finally decide which DLHA to follow in that environment, and it makes that decision based on whether or not  $X$  contains the observation-reward sequence which caused  $A$  and  $B$  to disagree. Again,  $A \oplus_X B$  is intuitively a type of “team” formed by  $A$  and  $B$ .

The reader can easily check that if  $X$  is the set of reward-observation sequences with first observation even, then  $\oplus_X$ , as given by Definition 14, is the same as  $\oplus$  from Definition 12.

**Proposition 15** For every electorate  $E = (\vec{e}, \mathcal{U})$ , DLHAs  $A$  and  $B$ , and set  $X$  of observation-reward sequences,  $A \oplus_X B =_E A$  or  $A \oplus_X B =_E B$ .

**Proof** Let

$$S = \{n : A \oplus_X B \text{ acts identically to } A \text{ on } e_n\}.$$

By Maximality, either  $S \in \mathcal{U}$  or  $S^c \in \mathcal{U}$ .

Case 1:  $S \in \mathcal{U}$ . By Upward Closure,

$$\{n : A \oplus_X B \text{ gets the same reward as } A \text{ on } e_n\} \supseteq S$$

is also in  $\mathcal{U}$ , so  $A \oplus_X B =_E A$ .

Case 2:  $S^c \in \mathcal{U}$ . By construction, for every  $n \in S^c$ ,  $A \oplus_X B$  acts identically to  $B$  on  $e_n$ . So by similar reasoning as in Case 1,  $A \oplus_X B =_E B$ . ■

Recalling that our intelligence comparators were ultimately motivated by social welfare functions, Proposition 15 might best be understood through the aphorism: “Looking like Lincoln to Lincoln-voters is enough to get you elected, regardless how you look to Douglas-voters.”  $A \oplus_X B$  looks exactly like  $A$  to one set of voters, and exactly like  $B$  to the opposite set of voters.

At first glance, Proposition 15 seems incompatible with proposals based on weighing environments by Kolmogorov complexity. Surprisingly, the incompatibility is not as big as initially appears. Kolmogorov complexity depends on a reference universal Turing machine, and one could contrive a reference universal Turing machine that gives unfair Kolmogorov

complexity to (say) environments with odd-numbered initial observations. This would cause such environments to be under-represented in Kolmogorov-complexity-based intelligence, so that the intelligence of  $A \oplus B$  would be approximately the same as that of  $A$ .

**Proposition 16** *(Compare Proposition 13) Let  $E = (\vec{e}, \mathcal{U})$  be an electorate, let  $X$  be an observation-reward sequence set, and let  $A, A', B, B'$  be DLHAs. Assume the following:*

$$\begin{aligned} A &>_E A', \\ A &>_E B', \\ B &>_E A', \\ B &>_E B'. \end{aligned}$$

*Then  $A \oplus_X B >_E A' \oplus_X B'$ .*

**Proof** By Proposition 15,  $A \oplus_X B =_E A$  or  $A \oplus_X B =_E B$ , and  $A' \oplus_X B' =_E A'$  or  $A' \oplus_X B' =_E B'$ . In any of the four cases,  $A \oplus_X B >_E A' \oplus_X B'$  by one of the four corresponding hypotheses. ■

Comparing Propositions 13 and 16, we see that as the team notion becomes more general, the necessary hypotheses seem to become more demanding.

One can imagine many other ways of forming teams; the teamwork notions we have considered here are narrow. We have not managed to obtain structural properties for other notions of teamwork. This might reflect the non-monolithic nature of real-world intelligence, of which our intelligence comparators are a rather monolithic approximation.

## 4.2 Properties of Quitters

In this section, we will prove a couple of structural properties about DLHAs who quit playing once they achieve a certain total reward. In order to arrive at these results, some preliminary definitions are needed. First, we will formalize a notion of a DLHA skipping a turn and what it means for a deterministic environment to respect that.

**Definition 17** *(Action-Skipping)*

1. By “skip”, we mean the natural number 0 (considered as an action).
2. A deterministic environment  $e$  is said to respect skipping if for every action-sequence  $a_1, \dots, a_n$ , if  $a_n = \text{“skip”}$  and  $e(a_1, \dots, a_n) = (r, o)$ , then  $r = 0$  (informally:  $e$  always gives 0 reward in response to a “skip” action).
3. An electorate  $E = (\vec{e}, \mathcal{U})$  is said to respect skipping if each  $e_i$  respects skipping.

The structural properties we are aiming for in this section also require a notion of deterministic environments having a limit on how big of a reward they can give at any one time (for simplicity we will make that limit be 1, although this is not particularly important).

**Definition 18** (*Bounded rewards*)

1. A deterministic environment  $e$  is said to have bounded rewards if for every action-sequence  $a_1, \dots, a_n$ , if  $e(a_1, \dots, a_n) = (r, o)$ , then  $r \leq 1$ .
2. An electorate  $E = (\vec{e}, \mathcal{U})$  is said to have bounded rewards if each  $e_i$  has bounded rewards.

**Definition 19** Suppose  $A$  is a DLHA and  $r$  is a real number. We define a new DLHA  $A \upharpoonright_r$  as follows. For every reward-observation sequence  $r_1, o_1, \dots, r_n, o_n$ ,

$$A \upharpoonright_r(r_1, o_1, \dots, r_n, o_n) = \begin{cases} A(r_1, o_1, \dots, r_n, o_n) & \text{if } r_1 + \dots + r_n < r, \\ \text{“skip”} & \text{if } r_1 + \dots + r_n \geq r. \end{cases}$$

We can think of  $A \upharpoonright_r$  as a version of  $A$  that becomes satisfied as soon as it has achieved a total reward of at least  $r$ , after which point  $A \upharpoonright_r$  takes it easy, ignores the environment, and performs nothing but “skip” forever after.

**Proposition 20** Let  $E = (\vec{e}, \mathcal{U})$  be an electorate with bounded rewards, and assume  $E$  respects skipping. Suppose  $A$  and  $B$  are DLHAs with  $A >_E B$ . Let  $r \in \mathbb{R}$ . If  $B \upharpoonright_r =_E B$ , then  $A \upharpoonright_{r+1} >_E B$ .

**Proof** Let

$$X_1 = \{n : B \upharpoonright_r \text{ and } B \text{ get the same reward on } e_n\}.$$

Since  $B \upharpoonright_r =_E B$ ,  $X_1 \in \mathcal{U}$ . Let

$$X_2 = \{n : A \text{ outperforms } B \text{ on } e_n\}.$$

Since  $A >_E B$ ,  $X_2 \in \mathcal{U}$ . By  $\cap$ -Closure,  $X_1 \cap X_2 \in \mathcal{U}$ . If I can show that  $A \upharpoonright_{r+1}$  outperforms  $B$  on  $e_n$  whenever  $n \in X_1 \cap X_2$ , the proposition will be proved.

Let  $n \in X_1 \cap X_2$ . Since  $n \in X_1$ ,  $B$  and  $B \upharpoonright_r$  get the same reward on  $e_n$ . By construction,  $B \upharpoonright_r$  must get less than  $r + 1$  total reward on  $e_n$  (because  $E$  has bounded rewards, so  $e_n$  never gives a reward larger than 1, and as soon as  $B \upharpoonright_r$  gets  $r$  or more total reward, it begins playing “skip” forever after, which causes  $e_n$  to give reward 0 forever after, since  $E$  respects skipping). Therefore,  $B$  must get less than  $r + 1$  total reward on  $e_n$ . Since  $n \in X_2$ , we know that  $A$  outperforms  $B$  on  $e_n$ . There are two cases.

Case 1:  $A$  does not get  $r + 1$  or more reward on  $e_n$ . Then by construction  $A \upharpoonright_{r+1}$  acts exactly like  $A$  on  $e_n$ , so outperforms  $B$  on  $e_n$ , as desired.

Case 2:  $A$  does get  $r + 1$  or more reward on  $e_n$ . It follows that  $A \upharpoonright_{r+1}$  gets  $r + 1$  or more reward on  $e_n$ . Since we already established that  $B \upharpoonright_r$  must get less than  $r + 1$  reward on  $e_n$ , this shows that  $A \upharpoonright_{r+1}$  outperforms  $B \upharpoonright_r$  on  $e_n$ , and hence outperforms  $B$  on  $e_n$  since  $B$  and  $B \upharpoonright_r$  get the same total reward on  $e_n$ .  $\blacksquare$

Intuitively, the way to think of Proposition 20 is as follows. The fact that  $B =_E B \upharpoonright_r$  implies that in “almost every environment” (or “in an election-winning bloc of environments”),  $B$ ’s total reward is less than  $r + 1$ . So if  $A$ ’s only objective is to beat

$B$ ,  $A$  might as well relax and take it easy any time  $A$  has already achieved at least  $r + 1$  total reward: by that point,  $A$  has already beaten  $B$  on the environment in question (with “almost no environments” being exceptions).

**Definition 21** *A deterministic environment  $e$  is merciful if the rewards it outputs are never negative. An electorate  $E = (\vec{e}, \mathcal{U})$  is merciful if each  $e_i$  is merciful.*

**Proposition 22** *Let  $E = (\vec{e}, \mathcal{U})$  be a merciful electorate that respects skipping. For every DLHA  $A$  and real number  $r$ ,  $A \not\prec_E A|_r$ .*

**Proof** Since  $E$  is merciful, it follows that  $A$  cannot possibly get less total reward than  $A|_r$  on any  $e_n$ , and the proposition trivially follows. ■

The proof of Proposition 22 would not go through if the deterministic environments were not merciful. This brings the following interesting fact to our attention: if deterministic environments are allowed to give out punishments, then the relative intelligence of two DLHAs really ought to depend on context, namely, on how risk-averse we are. After all, if  $A$  is a DLHA, then who is more intelligent:  $A$  or  $A|_1$ ? If it is better to safely achieve a total reward of at least 1, at the price of forgoing higher rewards with a higher risk, then  $A|_1$  should be more intelligent, because in some deterministic environments,  $A$  could temporarily achieve rewards adding to  $\geq 1$ , but then later receive large punishments. On the other hand, if it is better to achieve highest-possible rewards, even at risk of being punished in some environments, we might prefer  $A$ , because there may be many deterministic environments where  $A|_1$  will fall well short of  $A$ . Risk-aversion is, itself, already a potentially multi-dimensional parameter, and we expect it is probably only one of many hidden parameters underlying intelligence. The choice of an ultrafilter in Definition 7 serves as a convenient catch-all parameter which, we think, might serve to wrap up all these different complicated parameters into one.

## 5. Discussion

In this section, we will discuss some objections to our comparators and to Legg and Hutter’s agent-environment model. We thank an anonymous reviewer for these objections.

### 5.1 What does this definition really have to do with intelligence?

Out of context, Definition 7 might not appear to have anything to do with intelligence. There are two steps to see its relationship to intelligence. First, that the agent-environment model (Definition 1) is closely related to a type of intelligence; second, that Definition 7 is closely related to the problem of using Definition 1 to compare intelligence.

Legg and Hutter (2007) have done a fine job arguing for the close relationship between the agent-environment model and a type of intelligence. We will add the following. Whatever intelligence is, it surely involves components such as creativity, pattern recognition, the ability to abstract and to extrapolate, etc. Whether or not a DLHA is capable of creativity and so on is beyond the scope of this paper, but it seems clear that



there are deterministic environments *intended*<sup>10</sup> to reward creativity, or to reward pattern recognition, or to reward other specific components of intelligence. To consistently extract good rewards from many such environments, a DLHA would either need to possess those components of intelligence, or at least appear to possess them for practical intents and purposes. To quote Socrates: “But, not possessing right judgment, you would not realize that you are enjoying yourself even while you do, and, being unable to calculate, you could not figure out any future pleasures for yourself” (Plato, 1997).

It remains to show how Definition 7 is closely related to the problem of using Definition 1 to compare intelligence. This is a matter of summarizing the argument from Section 3.

- We would like to use the relative performance of DLHAs in deterministic environments to determine which DLHAs are more intelligent. More intelligent DLHAs should win higher rewards, not in every deterministic environment, but in some aggregate sense across a whole universe of deterministic environments. This idea sounds great, but it is unclear how to realize it.
- We view the problem from a different angle: rather than saying a DLHA with higher intelligence achieves higher rewards on average from deterministic environments, instead say that deterministic environments *elect* more intelligent machines. For, what else can it possibly be when a winner is chosen by the aggregated judgments of many judges? In this way, we obtain a bridge linking AGI to election theory, a literature from which we can import centuries of research. The question “Who is more intelligent?” is transformed into the question “Who wins the election?”

Kirman and Sondermann exhaustively characterized ways of determining election winners, under additional constraints. It is debateable whether those additional constraints are appropriate, but if we assume they are, then Definition 7 is immediate. Whether or not the constraints are appropriate, Kirman and Sondermann can be used as a heuristic in order to obtain a definition. If the constraints are appropriate, our definition might be the only possible definition, subject to those constraints! If the constraints are not appropriate, then our definition is still better than nothing, even if someday a better definition is found (one which would necessarily have to violate those constraints).

## 5.2 Why give a new definition if Legg and Hutter already gave one?

The intelligence measure defined by Legg and Hutter is important because it showed that it is in principle possible to formalize an abstract measurement of intelligence. And it is more computationally tractable than our comparators. There are at least two problems with the specific measure which Legg and Hutter proposed.

1. Legg and Hutter’s measure depends on a reference universal Turing machine (Leike and Hutter, 2015).
2. Legg and Hutter’s measure is biased toward agents who perform well in simple environments. For example, almost all environments have Kolmogorov complexity

---

10. We emphasize the intention part of this argument because to assume that an environment can *succeed* in rewarding creativity would be begging the question.

$> 100$ , but the Legg-Hutter intelligence of an agent is almost completely determined by its performance on environments with Kolmogorov complexity  $\leq 100$ .

We are skeptical that any nontrivial, non-contrived *structural properties* will be proven about Legg and Hutter’s intelligence measure. Any such property would either have to fundamentally hinge on the specific choices that went into their definition (which we think unlikely), or else be so general as to work in spite of those specific choices (any such property would surely be weak).

### 5.3 What about emotions, emotional intelligence, irrational intelligence, etc?

This paper is concerned with pragmatic intelligence, appropriate for the situation where one is evaluating a general-purpose utility-maximizing AI intended to be used in many different well-defined, rational, emotionless tasks. Our definition does not account for emotions (if any) in the machine, nor for what the machine knows or believes or understands (if anything), except indirectly to the extent that such emotions, knowledge, beliefs, or understanding help the machine to better extract rewards from reward-giving environments. It could potentially be dangerous to apply these definitions to human beings.

### 5.4 Not all instances of intelligence are ordinarily comparable

For any particular electorate, Definition 7 makes all DLHAs comparable, but we should stress that if  $E$  and  $E'$  are distinct electorates, then  $A >_E B$  does not generally have any bearing on whether or not  $A >_{E'} B$ . Thus, the fact that not all instances of intelligence are ordinarily comparable is actually evidence in *favor* of the parametrized-family approach of this paper. Different electorates might measure different types of intelligence, and a particular DLHA might be more intelligent according to one electorate and less intelligent according to another.

### 5.5 What about scholars who do not consider intelligence amenable to definition?

We understand that some scholars consider intelligence immune to definition, end-of-story. At first glance this paper might seem opposed against those scholars. But in actuality, this paper *reconciles* those scholars with scholars who think intelligence can be defined. Except for degenerate edge-cases, the intelligence comparators of Definition 7 depend on a *free ultrafilter*. Mathematical logicians have formally proven that, although free ultrafilters exist, it is impossible to exhibit one. Thus in a real and formal sense we vindicate the non-definition scholars, while at the same time showing (by proving some structural properties) that all is not lost.

### 5.6 Concerns about infinity

In the Legg-Hutter model, an agent interacts with an environment for an eternity. Readers might be concerned that this is inaccurate, considering that our universe will eventually end, etc. Likewise, we work over the whole infinite space of all hypothetical environments, even though only finitely many environments can exist in reality (if our universe has only finitely many atoms and those atoms can only be arranged in finitely many ways). We

acknowledge these shortcomings, but we defend them by comparing the situation to that of Turing machines. Turing machines, almost universally accepted as a model of computation, are equipped with infinite memory-tapes and allowed to run for eternity. And computability theorists routinely consider the entire infinite space of hypothetical Turing machines, even though a finite, discrete universe can only instantiate finitely many of them. Infinity serves as an important simplifying assumption. Without such simplifying assumptions, computer science would be filled with arbitrary decisions about how many bits of memory the idealized computer actually has, how long algorithms can run before the heat-death of the universe, etc. Considering how little we really know about these questions, and how rapidly our opinions about them change, we opine that the resulting science would be weakened.

For additional remarks about the justification and usefulness of infinity-based simplifying assumptions in (biological) science, see section 2 of Alexander (2013).

### 5.7 Drawing conclusions from environments that don't actually exist

As a followup to the previous subsection about infinity, the reader might question how relevant it is to consider environments that cannot possibly occur (due to, e.g., the finiteness of our universe) when comparing intelligence. We would like to propose a thought experiment. Suppose a passenger trapped in the sinking Titanic is a world-champion chess-player. Arguably, in the moments before the Titanic sinks, it is virtually impossible that this passenger will ever play another game of chess. Should that mean that their chess abilities are irrelevant to their intelligence? The answer could go either way, and it might indeed be interesting to study notions of intelligence in which the model takes into account what sort of scenarios the agent can possibly find itself in. In this paper, we take the tack that being a world-champion chess-player *is* relevant to one's intelligence, even when one is in a predicament that makes any future chess games impossible. We should stress that this thought experiment only serves as an analogy, and this paper is not really appropriate for comparing the intelligence of human beings, but rather the intelligence of DLHAs.

### 5.8 Shouldn't the model consider future discounting?

Readers might object that rewards early-on in a DLHA's encounter with a deterministic environment should carry more weight than rewards later-on.

In our opinion, time discounting is a complicated proposition, packed with arbitrary decisions of an infinite-dimensional nature, none of which (according to what we know so far) have any intrinsic relationship to intelligence. Just as, in physics, an equation can be simplified by replacing a product of many physical constants by one single new constant, in the same way, the ultrafilter in Definition 7 serves as an elegant master-parameter which hides within it many complicated choices and parameters, including those related to future discounting. For example, an ultrafilter might be chosen which gives preference to environments whose rewards taper off exponentially. As a general principle, any time a thing must unavoidably depend on arbitrary choices, the philosopher should define that thing with those choices parametrized away, rather than embedding them within the definition itself.

### 5.9 Could this definition be used to compare intelligence of weather systems, stars, etc.?

Definition 7 is a formal definition which applies to DLHAs (i.e., to functions with specific domain and range), not to things like weather systems or stars. Possibly, a DLHA could be contrived which takes actions somehow derived from the behavior of a formalized weather system or star. We conjecture that according to most non-contrived electorates, such a DLHA would compare unfavorably against any decent AGI.

## 6. Open Questions about other intelligence proposals

Even if the reader completely disagrees that our proposal has anything to do with intelligence, nevertheless, the results we have obtained still serve as a useful source of open questions which can be asked about other proposals for measuring the intelligence of Legg-Hutter-style agents. If nothing else, answers to these questions could ground a rich comparative vocabulary for the discussion of intelligence measurement systems.

### 6.1 Legg and Hutter universal intelligence

For every agent  $A$ , let  $\Gamma(A)$  denote the so-called universal intelligence of  $A$  as defined by Legg and Hutter (2007).

**Question 23** (Compare Proposition 13) *Let  $A, A', B, B'$  be agents with  $\Gamma(A) > \Gamma(A')$ ,  $\Gamma(B) > \Gamma(B')$ . Is it necessarily true that  $\Gamma(A \oplus B) > \Gamma(A' \oplus B')$ ?*

**Question 24** (Compare Proposition 13) *Let  $A, A', B, B'$  be agents with  $\min(\Gamma(A), \Gamma(B)) > \max(\Gamma(A'), \Gamma(B'))$ . Is it necessarily true that  $\Gamma(A \oplus B) > \Gamma(A' \oplus B')$ ?*

**Question 25** (Compare Proposition 16) *Let  $A, A', B, B'$  be agents with  $\min(\Gamma(A), \Gamma(B)) > \max(\Gamma(A'), \Gamma(B'))$ , and let  $X$  be a set of reward-observation sequences. Is it necessarily true that  $\Gamma(A \oplus_X B) > \Gamma(A' \oplus_X B')$ ?*

**Question 26** (Compare Proposition 20) *Let  $\Gamma'$  be the same as  $\Gamma$  except that  $\Gamma'$  only considers environments which respect skipping and have bounded rewards. Suppose  $A$  and  $B$  are agents with  $\Gamma'(A) > \Gamma'(B)$ . Let  $r \in \mathbb{R}$  be such that  $\Gamma'(B \downarrow_r) = \Gamma'(B)$ . Is it necessarily true that  $\Gamma'(A \downarrow_{r+1}) > \Gamma'(B)$ ?*

### 6.2 Hernández-Orallo and Dowe universal intelligence

Various proposals for so-called universal intelligence are given by Hernández-Orallo and Dowe (2010).

**Question 27** *What are the answers to Questions 23–26 if  $\Gamma$  denotes the various intelligence definitions given by Hernández-Orallo and Dowe (2010)?*

### 6.3 Hibbard universal intelligence

Hibbard has proposed (Hibbard, 2011) a definition of so-called universal intelligence of quite a different flavor.

**Question 28** *What are the answers to Questions 23–26 if  $\Gamma$  denotes the intelligence measure given by Hibbard (2011)?*

## 7. Conclusion

Following Legg and Hutter (2007), we considered an abstraction in which agents take actions within environments, based on the observations and rewards they receive from those environments. Unlike Legg and Hutter, we constrained both agent and environment to be deterministic. Various authors (Legg and Hutter, 2007; Hernández-Orallo and Dowe, 2010; Hibbard, 2011) have proposed ingenious means of quantifying the intelligence of such agents. Informally, the goal of such a measurement is that an agent with higher universal intelligence should realize larger rewards over the (infinite) space of environments, in some aggregate sense.

We proposed a new approach to comparing the intelligence of deterministic Legg-Hutter agents (DLHAs), a more abstract approach which lends itself to proving certain structural properties about intelligence. Rather than measure the universal intelligence of a DLHA, we focused on a closely related problem: how to compare the relative intelligence of two DLHAs. Our approach is based on the following insight: deterministic environments can be treated as voters in an election to determine which DLHA (if either) is more intelligent. Each environment votes for that DLHA which scores the highest reward in that environment (or votes that the two DLHAs are equally intelligent, if both DLHAs score equally in it). This realization provides an exciting bridge connecting the budding field of Legg-Hutter-style intelligence to the mature field of election theory. In particular, Kirman and Sondermann (1972) have completely characterized reasonable solutions to the infinite-voter election problem in terms of *ultrafilters*, a device from mathematical logic. This led us to an elegant means of using ultrafilters to compare intelligence of DLHAs.

The intelligence comparators we arrived at are elegant enough that we were able to prove some structural properties about how DLHAs’ intelligences are related *in general*. For example, if  $A$  and  $B$  are DLHAs, more intelligent than DLHAs  $A'$  and  $B'$  respectively, we proved (Proposition 13) that  $A \oplus B$  is more intelligent than  $A' \oplus B'$ , where  $\oplus$  is an operator which takes two DLHAs and outputs a new DLHA which can roughly be thought of as a “team” made up of  $A$  and  $B$  (Definition 12). In short (although this is an oversimplification): “If a team’s members are more intelligent, then that team is more intelligent.”

Our comparators are purely theoretical and not useful for doing concrete computations. Nevertheless, the structural properties we are able to prove about these theoretical comparators are a useful source of open questions about other Legg-Hutter-style intelligence approaches (Questions 23–28). Our hope is that this will inspire more research on Legg-Hutter-style intelligence even among readers who disagree about our particular proposal.

## Acknowledgements

We acknowledge Adam Bloomfield, Jordan Fisher, José Hernández-Orallo, Bill Hibbard, Marcus Hutter, Peter Sunehag, and the reviewers for feedback and discussion.

## References

- Alexander, S. A. 2013. Infinite graphs in systematic biology, with an application to the species problem. *Acta Biotheoretica* 61(2):181–201.
- Balduzzi, D.; Tuyls, K.; Perolat, J.; and Graepel, T. 2018. Re-evaluating evaluation. In *Advances in Neural Information Processing Systems*, 3268–3279.
- Fishburn, P. C. 1970. Arrow’s impossibility theorem: Concise proof and infinite voters. *Journal of Economic Theory* 2(1):103–106.
- Gavane, V. 2013. A measure of real-time intelligence. *Journal of Artificial General Intelligence* 4(1):31–48.
- Good, I. J. 1955. On the marking of chess-players. *The Mathematical Gazette* 39(330):292–296.
- Hernández-Orallo, J., and Dowe, D. L. 2010. Measuring universal intelligence: Towards an anytime intelligence test. *Artificial Intelligence* 174(18):1508–1539.
- Hibbard, B. 2011. Measuring agent intelligence via hierarchies of environments. In *International Conference on Artificial General Intelligence*, 303–308. Springer.
- Kirman, A. P., and Sondermann, D. 1972. Arrow’s theorem, many agents, and invisible dictators. *Journal of Economic Theory* 5(2):267–277.
- Legg, S., and Hutter, M. 2007. Universal intelligence: A definition of machine intelligence. *Minds and machines* 17(4):391–444.
- Legg, S., and Veness, J. 2013. An approximation of the universal intelligence measure. In *Algorithmic Probability and Friends. Bayesian Prediction and Artificial Intelligence*. Springer. 236–249.
- Leike, J., and Hutter, M. 2015. Bad universal priors and notions of optimality. *Journal of Machine Learning Research* 40:1244–1259.
- Morreau, M. 2016. Arrow’s Theorem. In Zalta, E. N., ed., *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2016 edition.
- Okasha, S. 2011. Theory choice and social choice: Kuhn versus Arrow. *Mind* 120(477):83–115.
- Plato. 1997. Philebus. In Cooper, J. M.; Hutchinson, D. S.; et al., eds., *Plato: complete works*. Hackett Publishing.
- Sanghi, P., and Dowe, D. L. 2003. A computer program capable of passing IQ tests. In *4th Intl. Conf. on Cognitive Science (ICCS03), Sydney*, 570–575.
- Stegenga, J. 2013. An impossibility theorem for amalgamating evidence. *Synthese* 190(12):2391–2411.