sciendo Published by Sciendo under the Creative Commons Attribution 3.0 License.

Journal of Artificial General Intelligence volume(issue) pages, year
DOI: DOI

Submitted submitted
Accepted accepted

Toward structural properties of universal intelligence

Samuel Allen Alexander SAMUELALLENALEXANDER@GMAIL.COM

Editor:

Abstract

Legg and Hutter, and 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 agent, how can we compare the relative intelligence of two agents? We propose an elegant answer based on the following insight: we can view 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

Starting with the landmark paper of Legg and Hutter (2007), various attempts (Hernández-Orallo and Dowe, 2010; Hibbard, 2011) have been made to quantify the intelligence of an idealized intelligent agent, one which interacts with an arbitrary environment based on observations and rewards from that environment[1] (formalized below). These agents' intelligence is measured in various ways using real numbers (for example, by means of infinite series involving Kolmogorov complexities).

Aside from quantifying the intelligence of individual agents, this field of research has the potential to shed light on intelligence itself, abstracted over many agents. To that end, we intend to present a few *structural properties* of intelligence. What do I mean by "structural properties"? Roughly speaking, I mean the following: if I know something about the intelligence of an agent A, a "structural property" is a theorem which allows me 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 teamwork.

1. Essentially, the agents and environments of *general reinforcement learning*, but with a more universal point of view.

Rather than directly compute a real number intelligence measure for any particular agent, we will instead focus on the simpler problem of how to determine whether one agent is more intelligent than another agent[2].

We will introduce an abstract intelligence measure based on the following insight: when considering whether agent A is more intelligent than agent B, we can imagine that different environments are *voters* voting in an election with three candidates. Those three candidates are: "A is more intelligent", "B is more intelligent", and "A and B are equally intelligent". Arrow's impossibility theorem (Arrow, 1963) would crush all hope of easily[3] 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.

2. Preliminaries

We will begin with a definition of agent and 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[4]. 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). So this field really is potentially relevant to real-world agents, and by abstracting technical details away, we can hope to make theoretical progress without being so distracted.

To simplify the mathematics[5], we want to avoid the awkward situation where the reward-sequence an agent receives adds up to ∞, 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

2. 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 10).
3. Technically speaking, the way environments vote in different elections is constrained, with complicated relationships between different agents 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.
4. For simplicity, we require both agent and environment to be deterministic, whereas some authors allow an element of randomness.
5. 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.

require that the nth reward be bounded between $\pm 2^{-n}$. For increased generality[6], we employ (in its debut appearance) a more abstract approach: we will first define what we call a pre-environment, with no such constraints, and then we will define an environment to be a pre-environment whose reward sequences (for every arbitrary agent) converge.

Definition 1 *(Environment and Agent)*

1. *A* pre-environment *is a function e which takes as input a finite (possibly empty) sequence $a_1, \ldots, a_n$ of natural numbers, called* actions. *It outputs a pair $e(a_1, \ldots, a_n) = (r, o)$, where $r \in \mathbb{R}$ is called a* reward *and $o \in \mathbb{N}$ is called an* observation.

2. *An* agent *is a function A which takes as input a finite sequence $r_1, o_1, \ldots, r_n, o_n$ of reward-observation pairs, and outputs an action $A(r_1, o_1, \ldots, r_n, o_n)$.*

3. *If A is an agent and e is a pre-environment, the* reward-observation-action sequence determined by letting A play in e *is the infinite sequence $r_1, o_1, a_1, r_2, o_2, a_2, \ldots$ 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 $r_1, o_1, a_1, \ldots, r_i, o_i$ have been defined. We define a_i, the action which A performs in response to reward-observation sequence $r_1, o_1, \ldots, r_i, o_i$, to be $a_i = A(r_1, o_1, \ldots, r_i, o_i)$.*
   - *Assume $r_1, o_1, a_1, \ldots, r_i, o_i, a_i$ have been defined. We define r_{i+1}, o_{i+1}, the reward and observation produced in response to action-sequence $a_1, \ldots, a_i$, to be $r_{i+1}, o_{i+1} = e(a_1, \ldots, a_i)$.*

4. *A pre-environment e is an* environment *if, for every agent A, the* total reward $r_1 + r_2 + \cdots$ *which A* achieves when A plays in e, *converges.*

5. *If A and B are agents and e is an 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.*

Example 1

- *The function $A(r_1, o_1, \ldots, r_i, o_i) = 0$ is a trivial agent which totally ignores the environment and blindly performs the same action over and over.*

- *The function $e(a_1, \ldots, a_i) = (1, 0)$ is a trivial pre-environment which ignores an agent's actions, instead always blindly rewarding the agent with a reward of $r = 1$, and never allowing any new observations. This pre-environment might be thought of as a sensory-deprivational paradise where agents receive constant injections of reward from an immutable void. This pre-environment is* not *an environment, because the total reward it grants to any agent diverges to $+\infty$.*

6. 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.

- *The function* $e(a_1,\ldots,a_i)=(0,0)$ *is a pre-environment similar to the previous one, except that it never grants agents any reward. It is an environment because the total reward it grants any agent converges (to* 0*).*
- *For a nontrivial example, choose a single-player video-game* V *and consider an environment* $e(a_1,\ldots,a_i)=(r_{i+1},o_{i+1})$ *where* o_{i+1} *encodes the image displayed on screen by* V *after the player presses buttons encoded by* $a_1,\ldots,a_i$*. 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 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.*

In order to arrive at an abstract method of comparing the intelligence of different agents, we will consider 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 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 2 *An* ultrafilter on $\mathbb{N}$ *(hereafter simply an* ultrafilter*) is a collection* $\mathscr{U}$ *of subsets of* $\mathbb{N}$ *satisfying the following requirements:*

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

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 3 *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 "$X \in \mathscr{U}$" 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 "$X \in \mathscr{U}$" 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 3 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 agents' intelligence by means of an imaginary election. The method we arrive at will not itself directly involve elections, so in motivating it, 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 environments vote. In the election, there are three candidates: "A is more intelligent", "B is more intelligent", and "A and B are equally intelligent". An 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 an environment vote on a single winner, but we make the environments rank the three candidates because that is the form of election which Kirman and Sondermann considered.) It remains 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 environment, which would clearly be undesirable[7].

Two other desirable properties of a social welfare function are:

7. For example, suppose relative intelligence was completely determined by performance in one environment e. Let A be a very intelligent agent. Suppose the reward-observation-action sequence determined

- (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.

The following theorem is an immediate corollary of Kirman and Sondermann (1972):

Theorem 4 *For any countably infinite sequence $\vec{e} = (e_0, e_1, \ldots)$ of environments, suppose the environments 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 Σ 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 $\mathscr{U}_\sigma$ with the following property: Any way the environments rank the three candidates, for every two candidates x and y, if*
$$\{n : e_n \text{ prefers } x \text{ over } y\} \in \mathscr{U}_\sigma,$$
then σ says that x is preferred over y in the corresponding group-preference.
2. *The map $\sigma \mapsto \mathscr{U}_\sigma$ is a surjection onto the set of ultrafilters (in other words, every ultrafilter $\mathscr{U}$ is equal to $\mathscr{U}_\sigma$ for some social welfare function $\sigma \in \Sigma$).*
3. *A social welfare function $\sigma \in \Sigma$ is a dictatorship if and only if $\mathscr{U}_\sigma$ is not free.*

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

In particular, since Theorem 3 says there exists a free ultrafilter, Theorem 4 implies there is a non-dictatorship social welfare function satisfying (Unanimity) and (Independence).

Motivated by Theorem 4, we will define a family of intelligence comparators.

Definition 5 *By an* electorate, *we mean a pair $E = (\vec{e}, \mathscr{U})$ where $\vec{e} = (e_0, e_1, \ldots)$ is any countably infinite sequence of environments and $\mathscr{U}$ is an ultrafilter.*

Definition 6 *For every electorate $E = (\vec{e}, \mathscr{U})$, we define relations $>_E$, $<_E$ and $=_E$ on agents as follows. Let A and B be any two agents.*

1. *If*
$$\{n : A \text{ outperforms } B \text{ on environment } e_n\} \in \mathscr{U},$$
we declare $A >_E B$ (and say A is more intelligent than *B* according to *E).*

by letting A play on e is $r_1^0, o_1^0, a_1^0, \ldots, r_i^0, o_i^0, a_i^0, \ldots$. We can consider a new agent A' defined by $A'(r_1, o_1, \ldots, r_i, o_i) = a_i^0$. In other words, A' is the agent which completely ignores its environment and instead blindly regurgitates the actions A takes on e. Then A' should presumably not be very intelligent, since A' never even pays attention to its environment, and yet, A' performs exactly as well as A on the particular environment e (by construction), and so, by choice of e, A' would be just as intelligent as A. This would clearly be absurd.

2. *If*

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

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

3. *If*

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

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

If $\mathscr{U}$ is non-free, then Definition 6 is trivial: it prescribes that we compare intelligence of agents by comparing their performance in one fixed environment (because the corresponding social welfare function is a dictatorship). We are mainly interested in the case where $\mathscr{U}$ is free[8].

Note that Definition 6 already compares all possible agents. 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 agent A gains 10000 more reward than agent B on some environment, that only helps A the same amount as if A gained 0.00001 more reward than B on that environment. We would defend this choice by pointing out that for any particular environment e, there is an equivalent 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 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 6 without the precise structure of agents and environments from Definition 1. All that is needed to state Definition 6 is that agents outperform each other on environments, regardless of what exactly that actually means. However, the precise structure of agents and environments becomes important when one wants to actually obtain nontrivial *structural properties* about the resulting comparators.

Lemma 7 *For every electorate $E = (\vec{e}, \mathscr{U})$ and agents A and B, exactly one of the following is true:*

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

Proof Let

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

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

By Maximality (from Definition 2), either $X_1 \in \mathscr{U}$ or $X_1^c \in \mathscr{U}$.

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

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

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

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

Lemma 8 *Let $E = (\vec{e}, \mathscr{U})$ be an electorate and let A and B be agents.*

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 9 *For every electorate $E = (\vec{e}, \mathscr{U})$ and agent A, $A =_E A$, $A \not>_E A$, and $A \not<_E A$.*

Proof Straightforward (use the facts that $\emptyset \notin \mathscr{U}$ by Properness, and $\mathbb{N} \in \mathscr{U}$ 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 10 *(Transitivity) Let $E = (\vec{e}, \mathscr{U})$ be an electorate and let A, B, C be agents.*

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 $A >_E B$, $X_{AB} \in \mathscr{U}$. Since $B >_E C$, $X_{BC} \in \mathscr{U}$. 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, $X_{AB} \cap X_{BC} \in \mathscr{U}$, so by Upward Closure, $X_{AC} \in \mathscr{U}$, that is, $A >_E C$. ■

4. Structural properties of intelligence

In this section, we will exhibit some nontrivial structural properties of intelligence. The properties we have been able to come up with are humble and few, but we believe that in the context of universal intelligence, 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.

Definition 11 *For any agents A and B, we define a new agent $A \oplus B$ such that for every observation-reward sequence $(r_1, o_1, \ldots, r_n, o_n)$,*

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

One can think of $A \oplus B$ as an agent who plans to act as A or B, but has not yet decided which one. This agent 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 12 *Let $E = (\vec{e}, \mathscr{U})$ be an electorate and let A, A', B, B' be agents. 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 $V \in \mathscr{U}$, or $V^c = D \in \mathscr{U}$. We will assume $V \in \mathscr{U}$, 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 $A >_E A'$, $X_1 \in \mathscr{U}$. By $\cap$-closure, $X_1 \cap V \in \mathscr{U}$. 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, $X_2 \in \mathscr{U}$, so $A \oplus B >_E A' \oplus B'$. ■

Proposition 12 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 an agent 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 an agent 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 11.

Definition 13 *Let X be any set of reward-observation sequences and let A and B be agents. The* team combination of A and B given by X *is the agent $A \oplus_X B$ defined as follows. Suppose $r_1, o_1, \ldots, r_n, o_n$ is any observation-reward sequence. If for all $m \leq n$, $A(r_1, o_1, \ldots, r_m, o_m) = B(r_1, o_1, \ldots, r_m, o_m)$, then we declare $(A \oplus_X B)(r_1, o_1, \ldots, r_n, o_n) = A(r_1, o_1, \ldots, r_n, o_n)$. Otherwise, let $m \leq n$ be minimal such that $A(r_1, o_1, \ldots, r_m, o_m) \neq B(r_1, o_1, \ldots, r_m, o_m)$. We declare*

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

In other words, $A \oplus_X B$ is the agent 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 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 agent 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 whose first observation is even, then $\oplus_X$, as given by Definition 13, is the same as $\oplus$ from Definition 11.

Proposition 14 *For every electorate $E = (\vec{e}, \mathscr{U})$, agents 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 $S \in \mathscr{U}$ or $S^c \in \mathscr{U}$.

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

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

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

Case 2: $S^c \in \mathscr{U}$. 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 14 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 14 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 15 *(Compare Proposition 12) Let $E = (\vec{e}, \mathscr{U})$ be an electorate, let X be an observation-reward sequence set, and let A, A', B, B' be agents. 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 14, $A \oplus_X B =_E A$ 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 12 and 15, 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 agents 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 an agent skipping a turn and what it means for an environment to respect that.

Definition 16 *(Action-Skipping)*

1. *By "skip", we mean the the natural number* 0 *(considered as an action).*
2. *An environment e is said to* respect skipping *if for every action-sequence $a_1, \ldots, a_n$, if $a_n =$ "skip" and $e(a_1, \ldots, a_n) = (r, o)$, then $r = 0$. (Informally: e always gives* 0 *reward in response to a "skip" action.)*

3. *An electorate* $E = (\vec{e}, \mathscr{U})$ *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 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 17 *(Bounded rewards)*

1. *An environment* e *is said to have* bounded rewards *if for every action-sequence* $a_1, \ldots, a_n$, *if* $e(a_1, \ldots, a_n) = (r, o)$, *then* $r \leq 1$.

2. *An electorate* $E = (\vec{e}, \mathscr{U})$ *is said to have* bounded rewards *if each* e_i *has bounded rewards.*

Definition 18 *Suppose* A *is an agent and* r *is a real number. We define a new agent* $A\restriction_r$ *as follows. For every reward-observation sequence* $r_1, o_1, \ldots, r_n, o_n$,

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

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

Proposition 19 *Let* $E = (\vec{e}, \mathscr{U})$ *be an electorate with bounded rewards, and assume* E *respects skipping. Suppose* A *and* B *are agents with* $A >_E B$. *Let* $r \in \mathbb{R}$. *If* $B\restriction_r =_E B$, *then* $A\restriction_{r+1} >_E B$.

Proof Let

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

Since $B\restriction_r =_E B$, $X_1 \in \mathscr{U}$. Let

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

Since $A >_E B$, $X_2 \in \mathscr{U}$. By $\cap$-Closure, $X_1 \cap X_2 \in \mathscr{U}$. If I can show that $A\restriction_{r+1}$ 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 $B\restriction_r$ get the same reward on e_n. By construction, $B\restriction_r$ 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 $B\restriction_r$ 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 $A\restriction_{r+1}$ 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 $A\restriction_{r+1}$ gets $r+1$ or more reward on e_n. Since we already established that $B\restriction_r$ must get less than $r+1$ reward on e_n, this shows that $A\restriction_{r+1}$ outperforms $B\restriction_r$ on e_n, and hence outperforms B on e_n since B and

$B\restriction_r$ get the same total reward on e_n. ■

Intuitively, the way to think of Proposition 19 is as follows. The fact that $B =_E B\restriction_r$ 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 20 *An environment e is* merciful *if the rewards it outputs are never negative. An electorate $E = (\vec{e}, \mathscr{U})$ is* merciful *if each e_i is merciful.*

Proposition 21 *Let $E = (\vec{e}, \mathscr{U})$ be a merciful electorate that respects skipping. For every agent A and real number r, $A \not<_E A\restriction_r$.*

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

The proof of Proposition 21 would not go through if environments were not merciful. This brings the following interesting fact to our attention: if environments are allowed to give out punishments, then the relative intelligence of two agents really ought to depend on context, namely, on how risk-averse we are. After all, if A is an agent, then who is more intelligent: A or $A\restriction_1$? 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 $A\restriction_1$ should be more intelligent, because in some environments, A could temporarily achieve rewards adding to ≥ 1, but then later receive large punishments from the environment. 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 environments where $A\restriction_1$ will fall well short of A.

5. Open Questions about other intelligence proposals

Even if the reader completely disagrees that our proposal has anything to do with universal intelligence, nevertheless, the results we have obtained still serve as a useful source of open questions which can be asked about any other proposal for universal intelligence.

5.1 Legg and Hutter universal intelligence

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

Question 22 *(Compare Proposition 12) 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 23 *(Compare Proposition 12) 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 24 *(Compare Proposition 15) 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 25 *(Compare Proposition 19) Let* Γ' *be the same as* Γ *except that* Γ' *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* $\Gamma'(B\restriction_r) = \Gamma'(B)$*. Is it necessarily true that* $\Gamma'(A\restriction_{r+1}) > \Gamma'(B)$*?*

5.2 Hernández-Orallo and Dowe universal intelligence

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

Question 26 *What are the answers to Questions 22–25 if* Γ *denotes the various universal intelligence definitions given by Hernández-Orallo and Dowe (2010)?*

5.3 Hibbard universal intelligence

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

Question 27 *What are the answers to Questions 22–25 if* Γ *denotes the universal intelligence measure given by Hibbard (2011)?*

6. 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. Various authors (Legg and Hutter, 2007; Hernández-Orallo and Dowe, 2010; Hibbard, 2011) have proposed ingenious means of quantifying the so-called *universal intelligence* of these 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 quantifying universal intelligence, a more abstract approach which lends itself to proving certain structural properties about intelligence. Rather than measure the universal intelligence of an agent, we focused on a closely related problem: how to compare the relative universal intelligence of two agents. Our approach is based on the following insight: environments can be treated as voters in an election to determine which agent (if either) is more intelligent. Each environment votes for that agent which scores the highest reward in that environment (or votes that the two agents are equally intelligent, if both agents score equally in it). This realization provides an exciting bridge connecting the budding field of universal 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 relative universal intelligence of agents.

The intelligence comparators we arrived at are elegant enough that we were able to prove some structural properties about how certain agents' intelligence is related *in general*. For example, if A and B are agents, more intelligent than agents A' and B' respectively, we proved (Proposition 12) that $A \oplus B$ is more intelligent than $A' \oplus B'$, where $\oplus$ is an operator which takes two agents and outputs a new agent which can roughly be thought of as a "team" made up of A and B (Definition 11). 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 universal intelligence approaches (Questions 22–27). Our hope is that this will inspire more research on universal intelligence even among readers who disagree about our particular proposal for universal intelligence comparison.

Acknowledgements

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

References

Arrow, K. J. 1963. *Social choice and individual values.* Wiley, 2nd edition.

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.

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.