Measuring intelligence and growth rate: variations on Hibbard's intelligence measure

Samuel Alexander[*†] and Bill Hibbard[‡]

2020

Abstract

In 2011, Hibbard suggested an intelligence measure for agents who compete in an adversarial sequence prediction game. We argue that Hibbard's idea should actually be considered as two separate ideas: first, that the intelligence of such agents can be measured based on the growth rates of the runtimes of the competitors that they defeat; and second, one specific (somewhat arbitrary) method for measuring said growth rates. Whereas Hibbard's intelligence measure is based on the latter growth-rate-measuring method, we survey other methods for measuring function growth rates, and exhibit the resulting Hibbard-like intelligence measures and taxonomies. Of particular interest, we obtain intelligence taxonomies based on Big-O and Big-Theta notation systems, which taxonomies are novel in that they challenge conventional notions of what an intelligence measure should look like. We discuss how intelligence measurement of sequence predictors can indirectly serve as intelligence measurement for agents with Artificial General Intelligence (AGIs).

1 Introduction

In his insightful paper [13], Bill Hibbard introduces a novel intelligence measure (which we will here refer to as the *original Hibbard measure*) for agents who play a game of adversarial sequence prediction [12] "against a hierarchy of increasingly difficult sets of" evaders (environments that attempt to emit 1s and 0s in such a way as to evade prediction). The levels of Hibbard's hierarchy are labelled by natural numbers, and an agent's original Hibbard measure is the maximum $n \in \mathbb{N}$ such that said agent learns to predict all the evaders in the nth level of the hierarchy, or implicitly[1] an agent's original Hibbard measure is ∞ if said agent learns to predict all the evaders in all levels of Hibbard's hierarchy.

[*]The U.S. Securities and Exchange Commission.

[†]Samuel Alexander is the primary author.

[‡]Space Science and Engineering Center, University of Wisconsin, Madison.

[1]Hibbard does not explicitly include the ∞ case in his definition, but in his Proposition 3 he refers to agents having "finite intelligence", and it is clear from context that by this he means agents who fail to predict some evader somewhere in the hierarchy.

The hierarchy which Hibbard uses to measure intelligence is based on the growth rates of the runtimes of evaders. We will argue that Hibbard's idea is really a combination of two orthogonal ideas. First: that in some sense the intelligence of a predicting agent can be measured based on the growth rates of the runtimes of the evaders whom that predictor learns to predict. Second: Hibbard proposed one particular method for measuring said growth rates. The growth rate measurement which Hibbard proposed yields a corresponding intelligence measure for these agents. We will argue that *any* method for measuring growth rates of functions yields a corresponding *adversarial sequence prediction intelligence* measure (or *ASPI* measure for short) provided the underlying number system provides a way of choosing canonical bounds for bounded sets. If the underlying number system does not provide a way of choosing canonical bounds for bounded sets, the growth-rate-measure will yield a corresponding ASPI taxonomy (like the big-O taxonomy of asymptotic complexity).

The particular method which Hibbard used to measure function growth rates is not very standard. We will survey other ways of measuring function growth rates, including some more standard ways, and these will yield corresponding ASPI measures and taxonomies.

The structure of the paper is as follows.

- In Section 2, we review the original Hibbard measure.
- In Section 3, we argue that any method of measuring the growth rate of functions yields an ASPI measure or taxonomy, and that the original Hibbard measure is just a special case resulting from one particular method of measuring function growth rate.
- In Section 4, we consider some of the most standard taxonomies of function growth rate—Big-O notation and Big-Θ notation—and define corresponding ASPI taxonomies.
- In Section 5, we consider several numeric solutions to the problem of measuring the growth rate of functions (using various number systems), and define corresponding ASPI measures and taxonomies.
- In Section 6, we give pros and cons of different ASPI measures and taxonomies.
- In Section 7, we summarize and make concluding remarks.

2 Hibbard's original measure

Hibbard proposed an intelligence measure for measuring the intelligence of agents who compete to predict evaders in a game of adversarial sequence prediction (we define this formally below). A predictor p (whose intelligence we want to measure) competes against evaders e. In each step of the game, both predictor and evader simultaneously choose a binary digit, 1 or 0. Only after both of them have made their choice do they see which choice the other one made, and then the game proceeds to the next step. The predictor's goal in each round is to choose the same digit that the evader will choose; the evader's goal is to choose a different digit than the predictor. The predictor wins the game (and is said to *learn to predict* e, or simply to *learn* e) if, after finitely many initial steps, eventually the predictor always chooses the same digit as the evader.

Definition 1. *By B, we mean the binary alphabet $\{0,1\}$. By B^*, we mean the set of all finite binary sequences. By $\langle\rangle$ we mean the empty binary sequence.*

Definition 2. *(Predictors and evaders)*

1. *By a* predictor*, we mean a Turing machine p which takes as input a finite (possibly empty) binary sequence $(x_1,\ldots,x_n) \in B^*$ (thought of as a sequence of* evasions*) and outputs 0 or 1 (thought of as a* prediction*), which output we write as $p(x_1,\ldots,x_n)$.*

2. *By an* evader*, we mean a Turing machine e which takes as input a finite (possibly empty) binary sequence $(y_1,\ldots,y_n) \in B^*$ (thought of as a sequence of* predictions*) and outputs 0 or 1 (thought of as an* evasion*), which output we write as $e(y_1,\ldots,y_n)$.*

3. *For any predictor p and evader e, the* result of p playing the game of adversarial sequence prediction against e *(or more simply, the* result of p playing against e*) is the infinite binary sequence $(x_1, y_1, x_2, y_2, \ldots)$ defined as follows:*

   (a) *The first evasion $x_1 = e(\langle\rangle)$ is the output of e when run on the empty prediction-sequence.*

   (b) *The first prediction $y_1 = p(\langle\rangle)$ is the output of p when run on the empty evasion-sequence.*

   (c) *For all $n > 0$, the $(n+1)$th evasion $x_{n+1} = e(y_1,\ldots,y_n)$ is the output of e on the sequence of the first n predictions.*

   (d) *For all $n > 0$, the $(n+1)$th prediction $y_{n+1} = p(x_1,\ldots,x_n)$ is the output of p on the sequence of the first n evasions.*

4. *Suppose $r = (x_1, y_1, x_2, y_2, \ldots)$ is the result of a predictor p playing against an evader e. For every $n \geq 1$, we say* the predictor wins round n in r *if $x_n = y_n$; otherwise,* the evader wins round n in r. *We say that p* learns to predict e *(or simply that p* learns e*) if there is some $N \in \mathbb{N}$ such that for all $n > N$, p is the winner of round n in r.*

Note that if e simply ignores its inputs $(y_1,\ldots,y_n)$ and instead computes $e(y_1,\ldots,y_n)$ based only on n, then e is essentially a sequence. Thus Definition 2 is a generalization of sequence prediction, which many authors have written about (such as Legg [16], who gives many references).

In the following definition, we differ from Hibbard's original paper because of a minor (and fortunately, easy-to-fix) error there.

Definition 3. *Suppose e is an evader. For each $n \in \mathbb{N}$, let $t_e(n)$ be the maximum number of steps that e takes to run on any length-n sequence of binary digits. In other words, $t_e(0)$ is the number of steps e takes to run on $\langle\rangle$, and for all $n > 0$,*

$$t_e(n) = \max_{b_1,\ldots,b_n \in \{0,1\}} (\text{number of steps } e \text{ takes to run on } (b_1, \ldots, b_n)).$$

Example 4. *Let e be an evader. Then $t_e(2)$ is equal to the number of steps e takes to run on input $(0,0)$, or to run on input $(0,1)$, or to run on input $(1,0)$, or to run on input $(1,1)$—whichever of these four possibilities is largest.*

Definition 5. *Suppose $f : \mathbb{N} \to \mathbb{N}$ and $g : \mathbb{N} \to \mathbb{N}$. We say f* majorizes *g, written $f \succ g$, if there is some $n_0 \in \mathbb{N}$ such that for all $n > n_0$, $f(n) > g(n)$.*

Definition 6. *Suppose $f : \mathbb{N} \to \mathbb{N}$. We define E_f to be the set of all evaders e such that $f \succ t_e$.*

Definition 7. *(The original Hibbard measure) Let $g_1, g_2, \ldots$ be the enumeration of the primitive recursive functions given by Liu [17]. For each $m > 0$, define $f_m : \mathbb{N} \to \mathbb{N}$ by*

$$f_m(k) = \max_{0<i\leq m} \max_{j\leq k} g_i(j).$$

For any predictor p, we define the original Hibbard intelligence *of p to be the maximum $m > 0$ such that p learns to predict e for every $e \in E_{f_m}$ (or 0 if there is no such m, or ∞ if p learns to predict e for every $e \in E_{f_m}$ for every $m > 0$).*

2.1 Predictor intelligence and AGI intelligence

Definition 7, and similar measures and taxonomies which we will define later, as written, only quantify the intelligence of one very specific type of agent, namely, predictors in the game of adversarial sequence prediction. But any method for quantifying the intelligence of such predictors can also approximately quantify the intelligence of (suitably idealized) agents with Artificial General Intellience (that is, the intelligence of AGIs).

Presumably, a suitably idealized AGI should be capable of understanding, and obedient in following or trying to follow, commands issued in everyday human language[2]. For example, if an AGI were commanded, "until further notice, compute and list the digits of pi," the AGI should be capable of understanding that command, and should obediently compute said digits until commanded otherwise[3].

[2]It is somewhat unclear how explicitly an AGI would obey certain commands. To use an example of Yampolskiy [24], if we asked a car-driving AGI to stop the car, would the AGI stop the car in the middle of traffic, or would it pull over to the side first? We assume this ambiguity does not apply when we ask the AGI to perform tasks of a sufficiently abstract and mathematical nature.

[3]Our thinking here is reminiscent of some remarks of Yampolskiy [26].

It is unclear how an AGI ought to respond if given an impossible command, such as "write a computer program that solves the halting problem", or Yampolskiy's [24] "Disobey!" But an AGI should be capable of understanding and attempting to obey an open-ended command, provided it is not impossible. For example, we could command an AGI to "until further notice, write an endless poem about trees," and the AGI should be able to do so, writing said poem line-by-line until we tell it to stop. This is despite the fact that the command is open-ended and under-determined (there are many decisions involved in writing a poem about trees, and we have left all these decisions to the AGI's discretion). The AGI's ability to obey such open-ended and under-determined commands exemplifies its ability to "adapt with insufficient knowledge and resources" [21]. One well-known example of an open-ended command which an AGI should be perfectly capable of attempting to obey (perhaps at great peril to us all) is Bostrom's "manufacture as many paperclips as possible" [4].

In particular, an AGI X should be perfectly capable of obeying the following command: "Act as a predictor in the game of adversarial sequence prediction". By giving X this command, and then immediately filtering out all X's sensory input except only for input about the digits chosen by an evader, we would obtain a formal predictor in the sense of Definition 2. This predictor might be called "the predictor generated by X". Strictly speaking, if the command is given to X at time t, then it would be more proper to call the resulting predictor "the predictor generated by X at time t": up until time t, the observations X makes about the universe might have an effect on the strategy X chooses to take once commanded to act as a predictor; but as long as we filter X's sensory input immediately after giving X the command, no further such observations can so alter X's strategy. In short, to use Yampolskiy's terminology [25], the act of trying to predict adversarial sequence evaders is *AI-easy*.

Thus, any intelligence measure (or taxonomy) for predictors also serves as an intelligence measure (or taxonomy) for AGIs. Namely: the intelligence level of an AGI X is equal to the intelligence level of X's predictor. Of course, a priori, X might be very intelligent at various other things while being poor at sequence prediction, or vice versa, so this only approximately captures X's true intelligence.

3 Quantifying growth rates of functions

The following is a general and open-ended problem.

Problem 8. *Quantify the growth-rate of functions from $\mathbb{N}$ to $\mathbb{N}$.*

The definition of the original Hibbard measure (Definition 7) can be thought of as implicitly depending on a specific solution to Problem 8, which we make explicit in the following definition.

Definition 9. *For each $m > 0$, let f_m be as in Definition 7. For each $f : \mathbb{N} \to \mathbb{N}$, we define the* original Hibbard growth rate *$H(f)$ to be $\min\{m > 0 : f_m \succ f\}$ if there is any such $m > 0$, and otherwise $H(f) = \infty$.*

Lemma 10. *For every natural $m > 0$ and every $f : \mathbb{N} \to \mathbb{N}$, $H(f) \leq m$ if and only if $f_m \succ f$.*

Proof. Straightforward. $\square$

Definition 11. *For every $m \in \mathbb{N}$, let E^H_m be the set of all evaders e such that $H(t_e) \leq m$.*

Lemma 12. *For every natural $m > 0$, $E^H_m = E_{f_m}$.*

Proof. Let e be an evader. By Definition 11, $e \in E^H_m$ if and only if $H(t_e) \leq m$. By Lemma 10, $H(t_e) \leq m$ if and only if $f_m \succ t_e$. But by Definition 6, this is the case if and only if $e \in E_{f_m}$. $\square$

Corollary 13. *For every predictor p, the original Hibbard measure of p is equal to the maximum natural $m > 0$ such that p learns e whenever $e \in E^H_m$, or is equal to ∞ if p learns e whenever $e \in E^H_m$ for all $m > 0$.*

Proof. Immediate by Lemma 12 and Definition 7. $\square$

In other words, if S is the set of all the m as in Corollary 13, then the original Hibbard measure of p is the "canonical upper bound" of S, where by the "canonical upper bound" of a set of natural numbers we mean the maximum element of that set (or ∞ if that set is unbounded).

Remark 14. *Corollary 13 shows that the definition of the original Hibbard measure can be rephrased in such a way as to show that it depends in a uniform way on a particular solution to Problem 8, namely on the solution proposed by Definition 9. For* any *solution H' to Problem 8, we could define evader-sets $E^{H'}_m$ in a similar way to Definition 11, and, by copying Corollary 13, we could obtain a corresponding intelligence measure given by H' (provided there be some way of choosing canonical bounds of bounded sets in the underlying number system—if not, we would have to be content with a taxonomy rather than a measure, a predictor's intelligence falling into many nested taxa corresponding to many different upper bounds, just as in Big-O notation a function can simultaneously be $O(n^2)$ and $O(n^3)$). This formalizes what we claimed in the Introduction, that Hibbard's idea can be decomposed into two sub-ideas, firstly, that a predictor's intelligence can be classified in terms of the growth rates of the runtimes of the evaders it learns, and secondly, a particular method (Definition 9) of measuring those growth rates (i.e., a particular solution to Problem 8).*

4 Big-O and Big-Θ intelligence

One of the most standard solutions to Problem 8 in computer science is to categorize growth rates of arbitrary functions by comparing them to more familiar functions using Big-O notation or Big-Θ notation. Knuth defines [15] these as follows (we modify the definition slightly because we are only concerned here with functions from $\mathbb{N}$ to $\mathbb{N}$).

Definition 15. *Suppose $f : \mathbb{N} \to \mathbb{N}$. We define the following function-sets.*

- *$O(f(n))$ is the set of all $g : \mathbb{N} \to \mathbb{N}$ such that there is some real $C > 0$ and some $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, $g(n) \leq Cf(n)$.*
- *$\Theta(f(n))$ is the set of all $g : \mathbb{N} \to \mathbb{N}$ such that there are some real $C > 0$ and $C' > 0$ and some $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, $Cf(n) \leq g(n) \leq C'f(n)$.*

Note that Definition 15 does not measure growth rates, but rather categorizes growth rates into Big-O and Big-Θ taxonomies. For example, the same function can be both $O(n^2)$ and $O(n^3)$, the former taxon being nested within the latter.

By Remark 14, Definition 15 yields the following elegant taxonomy of predictor intelligence.

Definition 16. *Suppose p is a predictor, and suppose $f : \mathbb{N} \to \mathbb{N}$.*

- *We say p* has Big-O ASPI measure *$O(f(n))$ if p learns every evader e such that t_e is $O(f(n))$.*
- *We say p* has Big-Θ ASPI measure *$\Theta(f(n))$ if p learns every evader e such that t_e is $\Theta(f(n))$.*

5 ASPI measures and taxonomies using various number systems

In this section we will consider various solutions to Problem 8 in various different number systems, and the ASPI measures and taxonomies they produce.

5.1 The hyperreal ASPI taxonomy

Hyperreal numbers, studied in the field of non-standard analysis [18] [9], are equivalence classes of infinite sequences of reals, so for every sequence $r = (r_0, r_1, r_2, \ldots)$ of reals, there is a corresponding hyperreal $\hat{r}$ represented by that sequence. Arithmetic is performed pointwise, so if $r = (r_0, r_1, \ldots)$ and $s = (s_0, s_1, \ldots)$, then $\hat{r} + \hat{s}$ is the hyperreal represented by $(r_0 + s_0, r_1 + s_1, \ldots)$, and $\hat{r} \cdot \hat{s}$ is the hyperreal represented by $(r_0s_0, r_1s_1, \ldots)$.

The difficulty in constructing the hyperreals is how to compare two hyperreals. We would like to compare them pointwise, so for example with r, s as above, we would want to have $\hat{r} < \hat{s}$ if every $r_i < s_i$, or $\hat{s} < \hat{r}$ if every $s_i < r_i$. Having done this, we would immediately obtain an elegant abstract solution to Problem 8 using hyperreal numbers, namely: the growth rate of $f(n)$ could be defined to be the hyperreal number represented by $(f(0), f(1), f(2), \ldots)$. But[4] how should $\hat{r}$ and $\hat{s}$ compare if there are infinitely many i such that $r_i < s_i$ and

[4] This is similar to the problem of deciding which of two reinforcement learning agents is more intelligent, if one agent performs better in infinitely many environments, but the other agent performs better in infinitely many other environments. For a treatment of that problem using free ultrafilters, see [1].

infinitely many i such that $s_i < r_i$? The answer is to declare $\hat{r} < \hat{s}$ if the set $\{i \in \mathbb{N} : r_i < s_i\}$ is "large", but it is not clear what it should mean for a set of natural numbers to be "large".

It turns out that the best answer—the only answer which leads to hyperreals with the properties we would want an extension of the reals to have—is to let the question of largeness of natural numbers be decided by an object called a *free ultrafilter*. Loosely speaking, a free ultrafilter can be thought of as a black box which makes judgments about which subsets of $\mathbb{N}$ are "large" (and does so in a particularly consistent way). Unfortunately, free ultrafilters are non-constructive: mathematicians have shown that free ultrafilters exist, but that it is impossible to concretely exhibit one. For the remainder of the section, we assume a free ultrafilter has been chosen, and we work in the resulting construction of the hyperreal numbers.

Definition 17. *(The hyperreal solution to Problem 8) For any function $f : \mathbb{N} \to \mathbb{N}$, the* hyperreal growth rate of f *is the hyperreal number $\hat{r}$ where $r = (f(0), f(1), f(2), \ldots)$.*

Because of the non-constructive nature of free ultrafilters, the following notion is computationally impractical. However, it could potentially be useful for proving theoretical properties about the intelligence of predictors. In the following definition, rather than assigning a particular hyperreal number intelligence to every predictor, rather, we categorize predictors into a taxonomy. This is necessary because there is no way of choosing canonical bounds of bounded sets of hyperreal numbers in general. For lack of a way of choosing a particular bound, we are forced to consider many taxa corresponding to many bounds.

Definition 18. *(The hyperreal ASPI taxonomy) Let p be a predictor and let $\hat{r}$ be a hyperreal number. We say that p* has hyperreal ASPI intelligence at least $\hat{r}$ *if and only if the following condition holds: for every evader e, if the hyperreal growth rate of t_e is $< \hat{r}$, then p learns e.*

5.2 The surreal ASPI measure

The surreal numbers [6] [14] [7] are an even larger extension of the real numbers into which the hyperreals can be embedded.

Definition 19. *(The surreal solution to Problem 8) Let ι be the embedding of the hyperreal numbers into the surreal numbers. For any function $f : \mathbb{N} \to \mathbb{N}$, the* surreal growth rate of f *is $\iota(\hat{r})$ where $\hat{r}$ is the hyperreal growth rate of f (Definition 17).*

The advantage of the surreal numbers is that for any set L of surreal numbers, there is a canonical way to choose a surreal upper bound on L, which upper bound is written $\{L \mid \}$. This upper bound is similar to a supremum of L in a certain sense: the surreal numbers possess a simplicity-hierarchical structure, and $\{L \mid \}$ is the *simplest* strict upper bound of L. This allows us to exactly measure intelligence of predictors, rather than merely classify predictors in a taxonomy as in the case of the hyperreal ASPI measures (Definition 18).

Definition 20. *(The surreal ASPI measure) For every predictor p, the* surreal ASPI measure *of p is defined to be $\{L\,|\,\}$, the simplest surreal strict upper bound of L, where L is the set of all surreal numbers ℓ such that the following condition holds: for every evader e, if the surreal growth rate of t_e is $< \ell$, then p learns e.*

5.3 ASPI measures based on majorization hierarchies

Majorization hierarchies [23] provide ordinal-number-valued measures for the growth rates of certain functions. A majorization hierarchy depends on many infinite-dimensional parameters. We will describe two majorization hierarchies up to the ordinal ϵ_0, using standard choices for the parameters, and the ASPI measures which they produce.

Definition 21. *(Classification of ordinal numbers) Ordinal numbers are divided into three types:*

1. *Zero: The ordinal* 0.
2. *Successor ordinals: Ordinals of the form $\alpha + 1$ for some ordinal α.*
3. *Limit ordinals: Ordinals which are not successor ordinals nor* 0.

For example, the smallest infinite ordinal, ω, is a limit ordinal. It is not zero (because zero is finite), nor can it be a successor ordinal, because if it were a successor ordinal, say, $\alpha + 1$, then α would be finite (since ω is the *smallest* infinite ordinal), but then $\alpha + 1$ would be finite as well.

Ordinal numbers have an arithmetical structure: two ordinals α and β have a sum $\alpha + \beta$, a product $\alpha \cdot \beta$, and a power α^β. It would be beyond the scope of this paper to give the full definition of these operations. We will only remark that some care is needed because although ordinal arithmetic is associative—e.g., $(\alpha+\beta)+\gamma = \alpha+(\beta+\gamma)$, and similarly for multiplication—it is not generally commutative: $\alpha+\beta$ is not always equal to $\beta+\alpha$, and $\alpha\cdot\beta$ is not always equal to $\beta\cdot\alpha$. For this reason, one often sees products like $\alpha\cdot 2$, which are not necessarily equivalent to the more familiar $2 \cdot \alpha$.

The ordinal ϵ_0 is the smallest ordinal bigger than the ordinals $\omega, \omega^\omega, \omega^{\omega^\omega}, \ldots$. It satisfies the equation $\epsilon_0 = \omega^{\epsilon_0}$ and can be intuitively thought of as

$$\epsilon_0 = \omega^{\omega^{\omega^{\cdot^{\cdot^{\cdot}}}}}.$$

Ordinals below ϵ_0 include such ordinals as ω, $\omega^{\omega+1} + \omega^\omega + \omega^5 + 3$,

$$\omega^{\omega^{\omega^{\omega^\omega}}} + \omega^{\omega^{\omega^\omega}+\omega^{\omega\cdot 2+1}+\omega^4+3} + \omega^{\omega^5+\omega^3} + \omega^8 + 1,$$

and so on. Any ordinal below ϵ_0 can be uniquely written in the form

$$\omega^{\alpha_1} + \omega^{\alpha_2} + \cdots + \omega^{\alpha_k}$$

where $\alpha_1 \geq \cdots \geq \alpha_k$ are smaller ordinals below ϵ_0—this form for an ordinal below ϵ_0 is called its *Cantor normal form.* For example, the Cantor normal form for $\omega^{\omega\cdot 2} \cdot 2 + \omega \cdot 3 + 2$ is

$$\omega^{\omega\cdot 2} \cdot 2 + \omega \cdot 3 + 2 = \omega^{\omega\cdot 2} + \omega^{\omega\cdot 2} + \omega^1 + \omega^1 + \omega^1 + \omega^0 + \omega^0.$$

Definition 22. *(Standard fundamental sequences for limit ordinals $\leq \epsilon_0$) Suppose λ is a limit ordinal $\leq \epsilon_0$. We define a* fundamental sequence for λ, *written* $(\lambda[0], \lambda[1], \lambda[2], \ldots)$*, inductively as follows.*

- *If $\lambda = \epsilon_0$, then $\lambda[0] = 0$, $\lambda[1] = \omega^0$, $\lambda[2] = \omega^{\omega^0}$, and so on.*
- *If λ has Cantor normal form $\omega^{\alpha_1} + \cdots + \omega^{\alpha_k}$ where $k > 1$, then each*

$$\lambda[i] = \omega^{\alpha_1} + \cdots + \omega^{\alpha_{k-1}} + (\omega^{\alpha_k}[i]).$$

- *If λ has Cantor normal form $\omega^{\alpha+1}$, then each $\lambda[i] = \omega^\alpha \cdot i$.*
- *If λ has Cantor normal form ω^{λ_0} where λ_0 is a limit ordinal, then each $\lambda[i] = \omega^{\lambda_0[i]}$.*

Example 23. *(Fundamental sequence examples)*

- *The fundamental sequence for $\lambda = \omega = \omega^1 = \omega^{0+1}$ is $\omega^0 \cdot 0, \omega^0 \cdot 1, \omega^0 \cdot 2, \ldots$, i.e., $0, 1, 2, \ldots$.*
- *The fundamental sequence for $\lambda = \omega^5$ is $0, \omega^4, \omega^4 \cdot 2, \omega^4 \cdot 3, \ldots$.*
- *The fundamental sequence for $\lambda = \omega^\omega$ is $\omega^0, \omega^1, \omega^2, \ldots$.*
- *The fundamental sequence for $\lambda = \omega^\omega + \omega$ is $\omega^\omega + 0, \omega^\omega + 1, \omega^\omega + 2, \ldots$.*

Definition 24. *(The standard slow-growing hierarchy up to ϵ_0) We define functions $g_\beta : \mathbb{N} \to \mathbb{N}$ (for all ordinals $\leq \epsilon_0$) by transfinite induction as follows.*

- *$g_0(n) = 0$.*
- *$g_{\alpha+1}(n) = g_\alpha(n) + 1$ if $\alpha + 1 \leq \epsilon_0$.*
- *$g_\lambda(n) = g_{\lambda[n]}(n)$ if $\lambda \leq \epsilon_0$ is a limit ordinal.*

Here are some early levels in the slow-growing hierarchy, spelled out in detail.

Example 25. *(Early examples of functions in the slow-growing hierarchy)*

1. *$g_1(n) = g_{0+1}(n) = g_0(n) + 1 = 0 + 1 = 1$.*
2. *$g_2(n) = g_{1+1}(n) = g_1(n) + 1 = 1 + 1 = 2$.*
3. *More generally, for all $m \in \mathbb{N}$, $g_m(n) = m$.*
4. *$g_\omega(n) = g_{\omega[n]}(n) = g_n(n) = n$.*

5. $g_{\omega+1}(n) = g_\omega(n) + 1 = n + 1$.

6. $g_{\omega+2}(n) = g_{\omega+1}(n) + 1 = (n+1) + 1 = n + 2$.

7. *More generally, for all* $m \in \mathbb{N}$, $g_{\omega+m}(n) = n + m$.

8. $g_{\omega\cdot 2}(n) = g_{(\omega\cdot 2)[n]}(n) = g_{\omega+n}(n) = n + n = n \cdot 2$.

Following Example 25, the reader should be able to fill in the details in the following example.

Example 26. *(More examples from the slow-growing hierarchy)*

1. $g_{\omega^2}(n) = n^2$.

2. $g_{\omega^3}(n) = n^3$.

3. $g_{\omega^\omega}(n) = n^n$.

4. $g_{\omega^{\omega\cdot 3+1}+\omega+5}(n) = n^{3n+1} + n + 5$.

5. $g_{\omega^{\omega^\omega}}(n) = n^{n^n}$.

What about g_{ϵ_0}? Thinking of ϵ_0 as

$$\omega^{\omega^{\omega^{\cdot^{\cdot^{\cdot}}}}},$$

one might expect $g_{\epsilon_0}(n)$ to be

$$n^{n^{n^{\cdot^{\cdot^{\cdot}}}}},$$

but such an infinite tower of natural number exponents makes no sense if $n > 1$. Instead, the answer defies familiar mathematical notation.

Example 27. *(Level ϵ_0 in the slow-growing hierarchy) The values of g_{ϵ_0} are as follows:*

- $g_{\epsilon_0}(0) = 0$.
- $g_{\epsilon_0}(1) = 1^1$.
- $g_{\epsilon_0}(2) = 2^{2^2}$.
- $g_{\epsilon_0}(3) = 3^{3^{3^3}}$.
- *And so on.*

Examples 25–27 illustrate how the slow-growing hierarchy systematically provides a family of reference functions against which any particular function can be compared. This yields a solution to Problem 8: we can declare the growth rate of an arbitrary function $f : \mathbb{N} \to \mathbb{N}$ to be the smallest ordinal $\beta < \epsilon_0$ such that $g_\beta \succ f$ (or ∞ if there is no such β). For any bounded set S of ordinals, there is a canonical upper bound for S, namely, the supremum of S. Thus we obtain an ASPI measure (not just a taxonomy).

Definition 28. *If p is a predictor, the* ASPI measure of p given by the standard slow-growing hierarchy up to ϵ_0 *is defined to be the supremum of S (or ∞ if $\epsilon_0 \in S$), where S is the set of all ordinals $\alpha \leq \epsilon_0$ such that the following condition holds: for every predictor e, if $g_\alpha \succ t_e$, then p learns e.*

In Definition 24, in the successor ordinal case, we chose to define $g_{\alpha+1}(n) = g_\alpha(n) + 1$. The resulting majorization hierarchy is referred to as *slow-growing* because in some sense this makes $g_{\alpha+1}$ just barely faster-growing than g_α. Different definitions of $g_{\alpha+1}$ would yield different majorization hierarchies, such as the following.

Definition 29. *(The standard fast-growing hierarchy up to ϵ_0, also known as the Wainer hierarchy) We define functions $h_\beta : \mathbb{N} \to \mathbb{N}$ (for all ordinals $\beta \leq \epsilon_0$) by transfinite induction as follows.*

- $h_0(n) = 0$.
- $h_{\alpha+1}(n) = h^n_\alpha(n)$, *where h^n_α is the nth iterate of h_α (so $h^1_\alpha(x) = h_\alpha(x)$, $h^2_\alpha(x) = h_\alpha(h_\alpha(x))$, $h^3_\alpha(x) = h_\alpha(h_\alpha(h_\alpha(x)))$, and so on).*
- $h_\lambda(n) = h_{\lambda[n]}(n)$ *if λ is a limit ordinal $\leq \epsilon_0$.*

The functions in the fast-growing hierarchy grow quickly as α grows. It can be shown [20] that for every computable function f whose totality can be proven from the axioms of Peano arithmetic, there is some $\alpha < \epsilon_0$ such that $h_\alpha \succ f$.

Definition 30. *If p is a predictor, the* ASPI measure of X given by the standard fast-growing hierarchy up to ϵ_0 *is defined to be the supremum of S (or ∞ if $\epsilon_0 \in S$), where S is the set of all ordinals $\alpha \leq \epsilon_0$ such that the following condition holds: for every predictor e, if $h_\alpha \succ t_e$, then p learns e.*

Between Definitions 28 and 30, the former offers a higher granularity intelligence measure for the predictors which it assigns non-∞ intelligence to, but the latter assigns non-∞ intelligence to a much larger set of predictors.

Definitions 24 and 29 are only two examples of majorization hierarchies. Both the slow- and fast-growing hierarchies can be extended by extending the fundamental sequences of Definition 22 to larger ordinals[5], however, the larger the ordinals become, the more difficult it is to do this, and especially the less clear it is how to do it in any sort of canonical way. There are also other choices for how to proceed at successor ordinal stages besides $g_{\alpha+1}(n) = g_\alpha(n) + 1$ or $h_{\alpha+1}(n) = h^n_\alpha(n)$—for example, one of the oldest majorization hierarchies is the Hardy hierarchy [11], where $H_{\alpha+1}(n) = H_\alpha(n + 1)$. And even for ordinals up to ϵ_0, there are other ways to choose fundamental sequences besides how we defined them in Definition 22—choosing non-canonical fundamental sequences can drastically alter the resulting majorization hierarchy [22]. All these different majorization hierarchies yield different ASPI measures.

[5] Remarkably, the slow-growing hierarchy eventually catches up with the fast-growing hierarchy if both hierarchies are extended to sufficiently large ordinals [19] [8], a beautiful illustration of how counter-intuitive large ordinal numbers can be.

5.4 A remark about ASPI measures and AGI intelligence

All the ASPI measures and taxonomies we have defined so far double as indirect intelligence measures and taxonomies for an AGI, by the argument we made in Subsection 2.1.

For a given AGI X, a priori, we cannot say much for certain about the predictor X would act as if X were commanded to act as a predictor. But there is one particularly elegant and parsimonious strategy which X might use, a *brute force strategy*, namely:

- Enumerate all the computable functions f which X knows to be total, and for each one, attempt to predict the evader e by assuming that the evader's runtime t_e satisfies $f \succ t_e$. If the evader proves not to be so majorized (by differing from every computable function whose runtime is so majorized), then move on to the next known total function f, and continue the process.

We do not know for certain which predictor X would imitate when commanded to act as a predictor, but it seems plausible that X would use this brute force strategy or something equivalent.

For an AGI X who uses the above brute force strategy, ASPI measures of X's intelligence would be determined by X's knowledge, namely, by the runtime complexity of the computable functions X knows to be total. Furthermore, the most natural way for X to know totality of functions with large runtime complexity, is for X to know fundamental sequences for large ordinal numbers, and produce said functions by means of majorization hierarchies[6]. This suggests a connection between

1. ASPI measures like that of Definition 30, and
2. intelligence measures based on which ordinals the AGI knows [2].

Indeed, Alexander has argued [3] that the task of notating large ordinals is one which spans the entire range of intelligence. This is reminiscent of Chaitin's proposal to use ordinal notation as a goal intended to facilitate evolution—"and the larger the ordinal, the fitter the organism" [5]—and Good's observation [10] that iterated Lucas-Penrose contests boil down to contests to name the larger ordinal.

6 Pros and cons of different ASPI measures and taxonomies

Here are pros and cons of the ASPI measures and taxonomies which arise from different solutions to the problem (Problem 8) of measuring the growth rate of functions.

[6] It may be possible for an AGI to be contrived to know totality of functions that are larger than the functions produced by majorization hierarchies up to ordinals the AGI knows about, but we conjecture that that is not the case for AGIs not so deliberately contrived.

- The original Hibbard measure (Definition 7), which arises by measuring growth rate by comparing a function with Liu's enumeration [17] of the primitive recursive functions:
  - Pro: Relatively concrete.
  - Pro: Measures intelligence using a familiar number system (the natural numbers).
  - Con: The numbers which the measure outputs are not very meaningful, in that predictor p having a measure of $+1$ higher than predictor q tells us little about how *much* more computationally complex the evaders which p learns are, versus the evaders which q learns.
  - Con: Only distinguishes sufficiently non-intelligent predictors; all predictors sufficiently intelligent receive measure ∞.
- Big-O/Big-Θ (Definition 16), in which, rather than directly measuring the intelligence of a predictor, instead, we would talk of a predictor's intelligence being $O(f(n))$ or $\Theta(f(n))$ for various functions $f : \mathbb{N} \to \mathbb{N}$:
  - Pro: Nearly perfect granularity (slightly coarser than perfect granularity because of the constants C, C' in Definition 15).
  - Pro: Computer scientists already use Big-O/Big-Θ routinely and are comfortable with them.
  - Con: A non-numerical taxonomy.
- Hyperreal intelligence (Definition 18):
  - Pro: A taxonomy like Big-O/Big-Θ, but with the added benefit that the taxons are numerical.
  - Perfect granularity.
  - Con: Depends on a free ultrafilter (rendering it computationally impractical).
- Surreal intelligence (Definition 20):
  - Pro: An actual numerical measure (not just a taxonomy), with perfect granularity.
  - Con: The numbers which the measure outputs are surreal numbers, which are relatively new and thus unfamiliar, and are difficult to work with in practice.
- Intelligence based on a majorization hierarchy such as the standard slow- or fast-growing hierarchy up to ϵ_0 (Definitions 28 and 28):
  - Pro: A numerical measure, albeit less granular than the Big-O/Big-Θ taxonomies.
  - Pro: Relatively concrete.

– Pro: The numbers which the measure outputs are meaningful, in the sense that the degree to which a predictor p is more intelligent than a predictor q is reflected in the degree to which p's intelligence-measure is larger than q's.
– Con: The numbers which the measure outputs are ordinal numbers, which may be unfamiliar to some users.
– Con: Only distinguishes sufficiently non-intelligent predictors; for any particular majorization hierarchy, all predictors sufficiently intelligent receive measure ∞.

7 Conclusion

To summarize:

- Hibbard proposed [13] an intelligence measure for predictors in games of adversarial sequence prediction.

- We argued that Hibbard's idea actually splits into two orthogonal subideas. First: that intelligence can be measured via the growth-rates of the run-times of evaders that a predictor can learn to predict. Second: that such growth-rates can be measured in one specific way (involving an enumeration of the primitive recursive functions). We argued that there many other ways to measure growth-rates, and that each method of measuring growth-rates yields a corresponding adversarial sequence prediction intelligence (ASPI) measure or taxonomy.

- We considered several specific ways of measuring growth-rates of functions, and exhibited corresponding ASPI measures and taxonomies. The growth-rate-measuring methods which we considered were: Big-O/Big-Θ notation; hyperreal numbers; surreal numbers; and majorization hierarchies.

- We also discussed how the intelligence of adversarial sequence predictors can be considered as an approximation of the intelligence of idealized AGIs.

Acknowledgments

We acknowledge Philip Ehrlich for correcting a mistake, and we acknowledge Roman Yampolskiy for providing literature references.

References

[1] Samuel Allen Alexander. Intelligence via ultrafilters: structural properties of some intelligence comparators of deterministic Legg-Hutter agents. *Journal of Artificial General Intelligence*, 10(1):24–45, 2019.

[2] Samuel Allen Alexander. Measuring the intelligence of an idealized mechanical knowing agent. In *CIFMA*, 2019.

[3] Samuel Allen Alexander. AGI and the Knight-Darwin law: why idealized AGI reproduction requires collaboration. In *ICAGI*, 2020.

[4] Nick Bostrom. Ethical issues in advanced artificial intelligence. *Science fiction and philosophy: from time travel to superintelligence*, pages 277–284, 2003.

[5] Gregory Chaitin. Metaphysics, metamathematics and metabiology. In Zenil Hector, editor, *Randomness through computation*. World Scientific, 2011.

[6] John H Conway. *On Numbers and Games*. CRC Press, 2nd edition, 2000.

[7] Philip Ehrlich. The absolute arithmetic continuum and the unification of all numbers great and small. *Bulletin of Symbolic Logic*, 18:1–45, 2012.

[8] Jean-Yves Girard. Π^1_2-logic, part 1: Dilators. *Annals of Mathematical Logic*, 21(2-3):75–219, 1981.

[9] Robert Goldblatt. *Lectures on the hyperreals: an introduction to nonstandard analysis*. Springer, 2012.

[10] Irving John Good. Gödel's theorem is a red herring. *The British Journal for the Philosophy of Science*, 19(4):357–358, 1969.

[11] Godfrey H Hardy. A theorem concerning the infinite cardinal numbers. *Quarterly journal of Mathematics*, 35:87–94, 1904.

[12] Bill Hibbard. Adversarial sequence prediction. In *ICAGI*, pages 399–403, 2008.

[13] Bill Hibbard. Measuring agent intelligence via hierarchies of environments. In *ICAGI*, pages 303–308, 2011.

[14] Donald Ervin Knuth. *Surreal numbers: a mathematical novelette*. Addison-Wesley, 1974.

[15] Donald Ervin Knuth. Big Omicron and big Omega and big Theta. *ACM Sigact News*, 8(2):18–24, 1976.

[16] Shane Legg. Is there an elegant universal theory of prediction? In *International Conference on Algorithmic Learning Theory*, pages 274–287. Springer, 2006.

[17] Shih-Chao Liu. An enumeration of the primitive recursive functions without repetition. *Tohoku Mathematical Journal*, 12(3):400–402, 1960.

[18] Abraham Robinson. *Non-standard analysis*. Princeton University Press, 1974.

[19] Stanley Wainer. Slow growing versus fast growing. *The Journal of symbolic logic*, 54(2):608–614, 1989.

[20] Stanley Wainer and Wilfried Buchholz. Provably computable functions and the fast growing hierarchy. In Stephen G Simpson, editor, *Logic and Combinatorics*. AMS, 1987.

[21] Pei Wang. On defining artificial intelligence. *Journal of Artificial General Intelligence*, 10(2):1–37, 2019.

[22] Andreas Weiermann. Sometimes slow growing is fast growing. *Annals of Pure and Applied Logic*, 90(1-3):91–99, 1997.

[23] Andreas Weiermann. Slow versus fast growing. *Synthese*, 133:13–29, 2002.

[24] Roman Yampolskiy. On controllability of artificial intelligence. *Technical report*, 2020.

[25] Roman V Yampolskiy. AI-complete, AI-hard, or AI-easy–classification of problems in AI. In *The 23rd Midwest Artificial Intelligence and Cognitive Science Conference*, 2012.

[26] Roman V Yampolskiy. Turing test as a defining feature of AI-completeness. In *Artificial intelligence, evolutionary computing and metaheuristics*, pages 3–17. Springer, 2013.