AGI and the Knight-Darwin Law: why idealized AGI reproduction requires collaboration

Samuel Allen Alexander[1 [0000−0002−7930−110X]]

The U.S. Securities and Exchange Commission `samuelallenalexander@gmail.com`
https://philpeople.org/profiles/samuel-alexander/publications

Abstract. Can an AGI create a more intelligent AGI? Under idealized assumptions, our answer is: "Not without outside help". This is a paper on the mathematical structure of AGI populations when parent AGIs create child AGIs. We argue that such populations satisfy a certain biological law. Motivated by observations of sexual reproduction in seemingly-asexual species, the Knight-Darwin Law states that it is impossible for one organism to asexually produce another, which asexually produces another, and so on forever: that any sequence of organisms (each one a child of the previous) must contain occasional multi-parent organisms, or must terminate. By proving that a certain universal intelligence measure decreases when an idealized parent AGI single-handedly creates a child AGI, we argue that a similar Law holds for AGIs.

Keywords: Intelligence Measurement · Knight-Darwin Law · Ordinal Notations · Intelligence Explosion

1 Introduction

One thing that distinguishes agents with Artificial General Intelligence (AGIs) from weaker AIs is that an AGI ought to be capable of programming AGIs[1]. After all, the first AGIs will, apparently, need to be programmed by humans. If AGIs are at least as smart as humans, and if humans are smart enough to build AGIs, then AGIs should also be smart enough to do so.

It is difficult to reason about AGIs programming AGIs. To get our hands on something solid, we have attempted to find structures that abstractly capture the core essence of AGIs programming AGIs. This led us to discover what we call the *Intuitive Ordinal Notation System* (presented in Section 2), an ordinal notation system that gets directly at the heart of AGIs creating AGIs.

We call an AGI *truthful* if the things it knows (or could eventually deduce[2]) are true, i.e., if it does not know any falsehoods. In [3], we used the Intuitive

[1] Our approach to AGI is what Goertzel [11] describes as the Universalist Approach: we consider "...an idealized case of AGI, similar to assumptions like the frictionless plane in physics", with the hope that by understanding this "simplified special case, we can use the understanding we've gained to address more realistic cases."

[2] Identifying what is known with what could eventually be deduced is of course only valid in the idealized context of this paper. In Section 6 of [3] we suggest a hypothetical way these idealized results could be applied to certain less-idealized AGIs.

Ordinal Notation System to argue that if a truthful AGI X creates (without external help) a truthful AGI Y, in such a way that X knows the truthfulness[3] of Y, then X must be more intelligent than Y in a certain formal sense. The argument is based on the key assumption that if X creates Y, without external help, then X necessarily knows Y's source code.

Iterating the above argument, suppose $X_1, X_2, \ldots$ are truthful AGIs such that each X_i creates, and knows the truthfulness and the code of, X_{i+1}. By the previous paragraph, X_1 would be more intelligent than X_2, which would be more intelligent than X_3, and so on. In Section 3 we will argue that this implies it is impossible for such a list $X_1, X_2, \ldots$ to go on forever: it would have to stop after finitely many elements[4].

At first glance, the above results might seem to suggest skepticism regarding the singularity—or regarding what Hutter [15] calls *intelligence explosion*, the idea of AGIs creating better AGIs, which create even better AGIs, and so on. But there is a loophole. Suppose X and X' are AGIs who collaborate to create Y. Suppose X contributes code for part of Y, but keeps it secret from X', and suppose X' contributes code for another part of Y, but keeps it secret from X. Then neither X nor X' knows Y's full source code, and yet if X and X' trust each other, then both X and X' can trust Y, so the above-mentioned argument breaks down.

Darwin and his contemporaries observed that even seemingly asexual plant species occasionally reproduce sexually. For example, a plant in which pollen is ordinarily isolated, might release pollen into the air if a storm damages the part of the plant that would otherwise shield the pollen[5]. The Knight-Darwin Law [7], named after Charles Darwin and Andrew Knight, is the principle (rephrased in modern language) that there cannot be an infinite sequence $X_1, X_2, \ldots$ of biological organisms such that each X_i asexually parents X_{i+1}. In other words, if $X_1, X_2, \ldots$ is any infinite list of organisms such that each X_i is a biological parent of X_{i+1}, then some of the X_i would need to be multi-parent organisms. The reader will immediately notice a striking parallel between this principle and the discussion in the previous two paragraphs.

In Section 2 we present the Intuitive Ordinal Notation System.

In Section 3 we argue[6] that if truthful AGI X creates truthful AGI Y, such that X knows the code and truthfulness of Y, then Y is less intelligent than X.

In Section 4 we adapt the Knight-Darwin Law from biology to AGI and speculate about what it might mean for AGI.

---

[3] If the created AGI could eventually deduce a contradiction, then, at least in classical mathematics, it could also deduce things like "I should kill my creator".

[4] This may initially seem to contradict some abstract mathematical constructions [18] [22] of infinite descending chains of theories. But those constructions only work for weaker languages, making them inapplicable to AGIs which presumably comprehend linguistically strong second-order predicates.

[5] Even prokaryotes can be considered to occasionally have multiple parents, if lateral gene transfer is taken into account.

[6] This argument appeared in a fully rigorous form in [3], but in this paper we attempt to make it more approachable.

In Section 5 we address some anticipated objections.

2 The Intuitive Ordinal Notation System

Based on our conviction that an AGI should be capable of programming AGIs, we would like to come up with a more concrete structure, easier to reason about, which we can use to better understand AGIs.

What structure would capture the essence of an AGI's AGI-programming capability? Initially, one might try: "computer program that prints computer programs." That seems like the right direction to go, but as written, there is no constraint on the printed computer programs, so this initial attempt seems to capture the essence of an AGI's capability of writing *computer programs*, rather than of writing *AGIs*.

As a second attempt, how about: "computer program that prints computer programs that print computer programs"? This brings us closer, but this second attempt seems to capture the essence of an AGI's capability of writing *program-writing programs*, rather than of writing *AGIs*.

As a third attempt, how about: "computer program that prints computer programs that print computer programs that print computer programs"? This brings us closer still, but this third attempt seems to capture the essence of an AGI's capability of writing *program-writing-program-writing programs*, rather than of writing *AGIs*.

We need to short-circuit the above process. We need to come up with a notion X which is equivalent to "computer program that prints members of X".

Definition 1 *(See the following examples) We define the Intuitive Ordinal Notations to be the smallest set $\mathcal{P}$ of computer programs such that:*

– *Each computer program p is in $\mathcal{P}$ iff all of p's outputs are also in $\mathcal{P}$.*

Example 2 *(Some simple examples)*

1. *Let P_0 be "End", a program which immediately stops without any outputs. Vacuously, all of P_0's outputs are in $\mathcal{P}$ (there are no such outputs). So P_0 is an Intuitive Ordinal Notation.*
2. *Let P_1 be "Print('End')", a program which outputs "End" and then stops. By (1), all of P_1's outputs are Intuitive Ordinal Notations, therefore, so is P_1.*
3. *Let P_2 be "Print('Print('End')')", which outputs "Print('End')" and then stops. By (2), all of P_2's outputs are Intuitive Ordinal Notations, therefore, so is P_2.*

Example 3 *(A more interesting example) Let P_ω be the program:*

Let X = 'End'; While(True) { Print(X); X = "Print('" + X + "')"; }

When executed, P_ω outputs "End", "Print('End')", "Print('Print('End')')", and so on forever. As in Example 2, all of these are Intuitive Ordinal Notations. Therefore, P_ω is an Intuitive Ordinal Notation.

Examples 2 and 3 are reminiscent of Franz's approach of "head[ing] for general algorithms at low complexity levels and fill[ing] the task cup from the bottom up" [9]. For a much larger collection of examples, see [2].

Definition 4 *For any Intuitive Ordinal Notation x, we define an ordinal $|x|$ inductively as follows: $|x|$ is the smallest ordinal α such that $\alpha > |y|$ for every output y of x.*

Example 5
- *Since P_0 (from Example 2) has no outputs, it follows that $|P_0| = 0$, the smallest ordinal.*
- *Likewise, $|P_1| = 1$ and $|P_2| = 2$.*
- *Likewise, P_ω (from Example 3) has outputs notating $0, 1, 2, \ldots$—all the finite natural numbers—and no other outputs. It follows that $|P_\omega| = \omega$, the smallest infinite ordinal.*
- *Let $P_{\omega+1}$ be the program "Print(P_ω)", where P_ω is as in Example 3. It follows that $|P_{\omega+1}| = \omega + 1$, the next ordinal after ω.*

The Intuitive Ordinal Notation System is a more intuitive simplification of an ordinal notation system known as Kleene's $\mathcal{O}$.

3 Intuitive Ordinal Intelligence

Whatever an AGI is, presumably an AGI knows (or could eventually deduce) certain mathematical facts. The following is a universal notion of an AGI's intelligence based solely on said facts. In [3] we argue that this notion captures key components of intelligence such as pattern recognition, creativity, and the ability to abstract or generalize. We will offer further justification in Section 5.

Definition 6 *The* Intuitive Ordinal Intelligence *of a truthful AGI X is the smallest ordinal $|X|$ such that $|X| > |p|$ for every Intuitive Ordinal Notation p such that X knows (or could eventually deduce) that p is an Intuitive Ordinal Notation.*

The following theorem provides a relationship[7] between Intuitive Ordinal Intelligence and AGI creation of AGI.

Theorem 7 *Suppose X is a truthful AGI, and X creates a truthful AGI Y, in such a way that X knows Y's code and truthfulness. Then $|X| > |Y|$.*

Proof. Suppose Y were commanded to spend eternity enumerating the biggest Intuitive Ordinal Notations Y could think of. This would result in some list L of Intuitive Ordinal Notations enumerated by Y. Since Y is an AGI, L must be computable, because L can be enumerated by a robot (namely by Y). Thus,

[7] Formalizing a relationship implied offhandedly by Chaitin, who suggests ordinal computation as a mathematical challenge intended to encourage evolution, "and the larger the ordinal, the fitter the organism" [6].

there is some computer program P whose outputs are exactly L. Since X knows Y's code, and presumably X is capable of reasoning about code, it follows that X can infer a program P that[8] lists L. Having constructed P this way, X would certainly know: "P outputs L, the list of things Y would output if Y were commanded to spend eternity trying to enumerate large Intuitive Ordinal Notations". Since X knows Y is truthful, X can deduce that L contains nothing except Intuitive Ordinal Notations, thus X can deduce that P's outputs are Intuitive Ordinal Notations, and so X knows (or could deduce) that P is an Intuitive Ordinal Notation. So $|X| > |P|$. But $|P|$ is the least ordinal $> |Q|$ for all Q output by L, in other words, $|P| = |Y|$. $\square$

Theorem 7 is mainly intended for the situation where parent X creates independent child Y, but can also be applied in case X self-modifies, viewing the original X as being replaced by the new self-modified Y (assuming X has prior knowledge of the code and truthfulness of the modified result).

4 The Knight-Darwin Law

> "...it is a general law of nature that no organic being self-fertilises itself for a perpetuity of generations; but that a cross with another individual is occasionally—perhaps at long intervals of time—indispensable." (Charles Darwin)

In his Origin of Species, Darwin devotes many pages to the above-quoted principle, later called the Knight-Darwin Law [7]. In [1] we translate the Knight-Darwin Law into mathematical language.

Principle 8 *(The Knight-Darwin Law) There cannot be an infinite sequence $x_1, x_2, \ldots$ of organisms such that each x_i is the lone biological parent of x_{i+1}. If each x_i is a parent of x_{i+1}, then some x_{i+1} must have multiple parents.*

A key fact about the ordinals is they are *well-founded*: there is no infinite sequence $o_1, o_2, \ldots$ of ordinals such that[9] each $o_i > o_{i+1}$. In Theorem 7 we showed that if truthful AGI X creates truthful AGI Y, in such a way as to know the truthfulness and code of Y, then X has a higher Intuitive Ordinal Intelligence than Y. Combining this with the well-foundedness of the ordinals yields a theorem extremely similar to the Knight-Darwin Law.

[8] For example, X could write a general program $Sim(c)$ that simulates an input AGI c waking up in an empty room and being commanded to spend eternity enumerating Intuitive Ordinal Notations. This program $Sim(c)$ would then output whatever outputs AGI c outputs under those circumstances. Having written $Sim(c)$, X could then obtain P by pasting Y's code into Sim (a string operation—not actually running Sim on Y's code). Nowhere in this process do we require X to actually execute Sim (which might be computationally infeasible).

[9] This is essentially true by definition, unfortunately the formal definition of ordinals is beyond the scope of this paper.

Theorem 9 *(The Knight-Darwin Law for AGIs) There cannot be an infinite sequence $X_1, X_2, \ldots$ of truthful AGIs such that each X_i creates X_{i+1} in such a way as to know X_{i+1}'s truthfulness and code. If each X_i creates X_{i+1} (intending X_{i+1} to be truthful), then occasionally certain X_{i+1}'s must be co-created by multiple creators (assuming that creation by a lone creator implies the lone creator would know X_{i+1}'s code).*

Proof. By Theorem 7, the Intuitive Ordinal Intelligence of $X_1, X_2, \ldots$ would be an infinite strictly-descending sequence of ordinals, violating the well-foundedness of the ordinals. □

Note that it is perfectly consistent with Theorem 7 that Y might operate faster than X, thereby performing better in realtime environments (as in [10]). Theorems 7 and 9 are profound because they suggest that descendants might initially appear more practical (faster, stronger, etc.), yet, without outside help, their knowledge must degenerate. This parallels the *hydra game* of Kirby and Paris [16], where a hydra seems to become stronger as the player hacks off its heads, yet inevitably dies if the player keeps hacking.

If AGI Y has distinct parents X and X', neither of which fully knows Y's code, then Theorem 7 does not apply to X, Y or X', Y and does not force[10] $|Y| < |X|$ or $|Y| < |X'|$. This does not necessarily mean that $|Y|$ can be arbitrarily large, though. If X and X' were themselves created single-handedly by a lone parent X_0, then similar reasoning to Theorem 7 would force $|Y| < |X_0|$ (assuming X_0 could infer the code and truthfulness of Y from the codes of X and X')[11].

In the remainder of this section, we will non-rigorously speculate about three implications Theorem 9 might have for AGIs and for AGI research.

4.1 Motivation for Multi-agent Approaches to AGI

In the Introduction, we argued that an AGI ought to be capable of programming AGIs. If so, Theorem 9 suggests that a fundamental aspect of AGI should be the ability to collaborate with other AGIs in the creation of new AGIs[12].

One interesting idea along these lines is that there should be no such thing as a *solipsistic* AGI[13], or at least, solipsistic AGIs would be limited in their

[10] By this loophole, our conviction "AGIs must be capable of creating AGIs" does not technically rule out an AGI having Intuitive Ordinal Intelligence 0: such an AGI could not single-handedly create a child AGI (in such a way as to know the child's truth and code), but could still co-create children with other AGIs.

[11] This suggests possible generalizations of the Knight-Darwin Law such as "There cannot be an infinite sequence $x_1, x_2, \ldots$ of biological organisms such that each x_i is the lone grandparent of x_{i+1}," and AGI versions of same.

[12] It is particularly interesting that our intelligence measure naturally motivates multi-agent considerations in spite of the fact that, a priori, Definition 6 measures intelligence completely independently of the surrounding collective of other agents (contrast [12]) or even the outside world.

[13] That is, an AGI which believes itself to be the only entity in the universe.

ability to reproduce. For, if an AGI were solipsistic, it seems like it would be difficult for this AGI to collaborate with other AGIs to create child AGIs. To quote Hernández-Orallo et al: "The appearance of multi-agent systems is a sign that the future of machine intelligence will not be found in monolithic systems solving tasks without other agents to compete or collaborate with" [12].

More practically, Theorem 9 might suggest prioritizing research on multi-agent approaches to AGI, such as [5], [12], [14], [21], [19], [17], and similar work.

4.2 Motivation for AGI Variety

Darwin used the Knight-Darwin Law as a foundation for a broader thesis that the survival of a species depends on the inter-breeding of many members. By analogy, if our goal is to create robust AGIs, perhaps we should focus on creating a wide variety of AGIs, so that those AGIs can co-create more AGIs.

On the other hand, if we want to reduce the danger of AGI getting out of control, perhaps we should *limit* AGI variety. At the extreme end of the spectrum, if humankind were to limit itself to only creating one single AGI[14], then Theorem 9 would constrain the extent to which that AGI could reproduce.

4.3 AGI Genetics

If, as Theorem 9 seems to suggest, AGI collaboration is a fundamental requirement for an AGI "population" to propagate, then it might someday be possible to view AGI through a genetic lens. If AGIs X and X' co-create child AGI Y by means of X contributing proprietary code for one submodule, concealed from X', while X' contributes proprietary code for another submodule, concealed from X, perhaps there might be some notion of AGI "genes". For example, if X runs operating system O, and X' runs operating system O', perhaps their joint offspring will somehow exhibit traces of both O and O'.

5 Discussion

In this section, we discuss some anticipated objections.

5.1 What does Definition 6 have to do with intelligence? Shouldn't intelligence depend on environmental interaction?

We do not claim that Definition 6 is the "one true measure" of intelligence. Maybe there is no such thing: maybe intelligence is inherently multi-dimensional. Definition 6 measures a type of intelligence based on mathematical knowledge[15].

[14] Or to perfectly isolate different AGIs away from one another—see [25].

[15] Wang has correctly pointed out [23] that an AGI consists of much more than merely a knowledge-set of mathematical facts. Still, we feel mathematical knowledge is at least one important aspect of an AGI's intelligence.

An AGI could be good at video games but poor at ordinals. But the broad AGIs we are talking about in this paper should be capable (if properly instructed) of attempting any reasonable well-defined task, including that of notating ordinals. So Definition 6 does measure one aspect of an AGI's abilities. Perhaps a different word, such as "knowledge-level", would fit better: but that would not qualitatively change the Knight-Darwin Law implications.

Whatever intelligence is, it has core components like pattern-matching, creativity, and the ability to abstract and to generalize. We claim that these core parts of intelligence are indispensable if one wants to competitively name large ordinals. If p is an Intuitive Ordinal Notation obtained using certain techniques, then *any* AGI who used those techniques to construct p should also be able to iterate those same techniques. Thus, to advance from p to a larger ordinal which not just *any* p-knowing AGI could obtain, must require the creative invention of some new technique, and this invention requires some amount of creativity, pattern-matching, etc. This becomes clear if the reader tries to notate ordinals qualitatively larger than Example 3; see the more extensive examples in [2].

For analogy's sake, imagine a ladder which different AGIs can climb, and suppose advancing up the ladder requires exercising intelligence. Having so little other idea how to measure intelligence, one way to measure (or at least estimate) it would be to measure how high an AGI can climb said ladder.

Not all ladders are equally good. A ladder would be particularly poor if it had a top rung which many AGIs could reach: for then it would fail to distinguish between AGIs who could reach that top rung, even if one AGI reaches it with ease and another with difficulty. Even if the ladder was infinite and had no top rung, it would still be suboptimal if there were AGIs capable of scaling the whole ladder (i.e., of ascending however high they like, on demand)[16]. A good ladder should have, for each particular AGI, a rung which that AGI cannot reach.

Definition 6 offers a good ladder. The rungs which an AGI manages to reach, we have argued, require core components of intelligence to reach. And no particular AGI can scale the whole ladder, because no AGI can enumerate all the Intuitive Ordinal Notations: it can be shown that they are not computably enumerable[17].

[16] Hibbard's intelligence measure [13] is an infinite ladder which is nevertheless short enough that many AGIs can scale the whole ladder. Those whole-ladder-scalers are the AGIs which do not "have finite intelligence" in Hibbard's words (see Hibbard's Proposition 3). It should be possible to use a *fast-growing hierarchy* [8] [24] to extend Hibbard's ladder into the transfinite and reduce (or eliminate?) the set of whole-ladder-scalers. This would make Hibbard's measurement ordinal-valued (perhaps Hibbard intuited this, since his abstract uses the word "ordinal" in its everyday sense as a synonym for "natural number").

[17] Thus, this ladder avoids a common problem that arises when trying to measure machine intelligence using IQ tests, namely, that for any IQ test, an algorithm can be designed to dominate that test, despite being otherwise unintelligent [4].

5.2 Can't an AGI just print a copy of itself?

If an AGI knows its own code, then it can certainly print a copy of itself. But if so, then it necessarily cannot know the truthfulness of that copy, lest it would know the truthfulness of itself. Versions of Gödel's incompleteness theorems adapted [20] to mechanical knowing agents imply that a suitably idealized AGI cannot know its own code and its own truthfulness.

5.3 Prohibitively expensive simulation

The reader might object that Theorem 7 breaks down if Y is prohibitively expensive for X to simulate. But Theorem 7 and its proof have nothing to do with simulation. In functional languages like Haskell, functions can be manipulated, filtered, formally composed with other functions, and so on, without needing to be executed. Likewise, if X knows the code of Y, then X can manipulate and reason about that code without executing a single line of it.

5.4 Computable enumerability of mathematical theorems

"Coundn't an AGI just enumerate all mathematical theorems by brute force, and wouldn't that AGI have unbeatable intelligence according to Definition 6?" The set of mathematical theorems is only computably enumerable when a computably enumerable base axiom-set is fixed in the background.

6 Conclusion

The Intuitive Ordinal Intelligence of a truthful AGI is defined to be the supremum of the ordinals which have Intuitive Ordinal Notations the AGI knows to be Intuitive Ordinal Notations. We discussed why we think this notion measures (a type of) intelligence. We proved that if a truthful AGI single-handedly creates a child truthful AGI, in such a way as to know the child's truthfulness and code, then the parent must have greater Intuitive Ordinal Intelligent than the child. This allowed us to establish a structural property for AGI populations, resembling a property of biological populations, the Knight-Darwin Law. We speculated about implications of this biology-AGI parallel. We hope that by better understanding how AGIs create new AGIs, we can make progress toward understanding methods of AGI-creation by humans.

Acknowledgments

We gratefully acknowledge Jordi Bieger, Thomas Forster, José Hernández-Orallo, Bill Hibbard, Mike Steel, and Albert Visser for comments and feedback.

References

1. Alexander, S.A.: Infinite graphs in systematic biology, with an application to the species problem. Acta Biotheoretica **61**(2), 181–201 (2013)
2. Alexander, S.A.: Intuitive ordinal notations (ions). GitHub repository, https://github.com/semitrivial/ions (2019)
3. Alexander, S.A.: Measuring the intelligence of an idealized mechanical knowing agent. In: CIFMA (2019)
4. Besold, T., Hernández-Orallo, J., Schmid, U.: Can machine intelligence be measured in the same way as human intelligence? KI-Künstliche Intelligenz **29**(3), 291–297 (2015)
5. Castelfranchi, C.: Modelling social action for ai agents. AI **103**(1-2), 157–182 (1998)
6. Chaitin, G.: Metaphysics, metamathematics and metabiology. In: Hector, Z. (ed.) Randomness through computation. World Scientific (2011)
7. Darwin, F.: The knight-darwin law. Nature **58**, 630–632 (1898)
8. Fairtlough, M., Wainer, S.S.: Hierarchies of provably recursive functions. In: Handbook of proof theory, pp. 149–207. Elsevier Amsterdam (1998)
9. Franz, A.: Toward tractable universal induction through recursive program learning. In: ICAGI. pp. 251–260 (2015)
10. Gavane, V.: A measure of real-time intelligence. JAGI **4**(1), 31–48 (2013)
11. Goertzel, B.: Artificial general intelligence: concept, state of the art, and future prospects. JAGI **5**(1), 1–48 (2014)
12. Hernández-Orallo, J., Dowe, D.L., España-Cubillo, S., Hernández-Lloreda, M.V., Insa-Cabrera, J.: On more realistic environment distributions for defining, evaluating and developing intelligence. In: ICAGI. pp. 82–91 (2011)
13. Hibbard, B.: Measuring agent intelligence via hierarchies of environments. In: ICAGI. pp. 303–308 (2011)
14. Hibbard, B.: Societies of intelligent agents. In: ICAGI. pp. 286–290 (2011)
15. Hutter, M.: Can intelligence explode? JCS **19**(1-2), 143–166 (2012)
16. Kirby, L., Paris, J.: Accessible independence results for peano arithmetic. Bulletin of the London Mathematical Society **14**(4), 285–293 (1982)
17. Kolonin, A., Goertzel, B., Duong, D., Ikle, M.: A reputation system for artificial societies. arXiv preprint arXiv:1806.07342 (2018)
18. Kripke, S.A.: Ungroundedness in tarskian languages. JPL **48**(3), 603–609 (2019)
19. Potyka, N., Acar, E., Thimm, M., Stuckenschmidt, H.: Group decision making via probabilistic belief merging. In: 25th IJCAI. AAAI Press (2016)
20. Reinhardt, W.N.: Absolute versions of incompleteness theorems. Nous **19**, 317–346 (1985)
21. Thórisson, K.R., Benko, H., Abramov, D., Arnold, A., Maskey, S., Vaseekaran, A.: Constructionist design methodology for interactive intelligences. AI Magazine **25**(4), 77–90 (2004)
22. Visser, A.: Semantics and the liar paradox. In: Handbook of philosophical logic, pp. 149–240. Springer (2002)
23. Wang, P.: Three fundamental misconceptions of artificial intelligence. Journal of Experimental & Theoretical Artificial Intelligence **19**(3), 249–268 (2007)
24. Weiermann, A.: Slow versus fast growing. Synthese **133**, 13–29 (2002)
25. Yampolskiy, R.V.: Leakproofing singularity-artificial intelligence confinement problem. JCS **19** (2012)