

University of Louisville

ThinkIR: The University of Louisville's Institutional Repository

Faculty Scholarship

5-21-2015

Corrigendum: Responses to catastrophic AGI risk: A survey (2015 Phys. Scr. 90 018001)

Kaj Sotala

Machine Intelligence Research Institute

Roman V. Yampolskiy

University of Louisville, roman.yampolskiy@louisville.edu

Follow this and additional works at: <https://ir.library.louisville.edu/faculty>



Part of the [Computer Engineering Commons](#)

Original Publication Information

Kaj Sotala and Roman V Yampolskiy 2015 *Phys. Scr.* 90 069501

[Download Article PDF](#)

ThinkIR Citation

Sotala, Kaj and Yampolskiy, Roman V., "Corrigendum: Responses to catastrophic AGI risk: A survey (2015 Phys. Scr. 90 018001)" (2015). *Faculty Scholarship*. 589.

<https://ir.library.louisville.edu/faculty/589>

This Article is brought to you for free and open access by ThinkIR: The University of Louisville's Institutional Repository. It has been accepted for inclusion in Faculty Scholarship by an authorized administrator of ThinkIR: The University of Louisville's Institutional Repository. For more information, please contact thinkir@louisville.edu.

CORRIGENDUM

**Corrigendum: Responses to catastrophic AGI risk:
a survey (2015 *Phys. Scr.* [90 018001](#))**

To cite this article: Kaj Sotala and Roman V Yampolskiy 2015 *Phys. Scr.* **90** 069501

Recent citations

- [João Guerreiro](#)
- [Global Solutions vs. Local Solutions for the AI Safety Problem](#)
Alexey Turchin *et al*

View the [article online](#) for updates and enhancements.



Corrigendum: Responses to catastrophic AGI risk: a survey (2015 Phys. Scr. 90 018001)

Kaj Sotala¹ and Roman V Yampolskiy²

¹Machine Intelligence Research Institute, Berkeley, CA, USA

²University of Louisville, KY, USA

Parts of the reference list are corrected to the following:

- [105] Goertzel B 2002 Thoughts on AI morality *Dynamical Psychology* (www.goertzel.org/dynapsyc/2002/AIMorality.htm)
- [106] Goertzel B 2004 Encouraging a positive transcension *Dynamical Psychology* (www.goertzel.org/dynapsyc/2004/PositiveTranscension.htm)
- [107] Goertzel B 2004 Growth, choice and joy *Dynamical Psychology* (www.goertzel.org/dynapsyc/2004/GrowthChoiceJoy.htm)
- [108] Goertzel B 2006 Apparent limitations on the ‘AI friendliness’ and related concepts imposed by the complexity of the world (www.goertzel.org/papers/LimitationsOnFriendliness.pdf)
- [109] Goertzel B 2010 Coherent aggregated volition *The Multiverse According to Ben* (<http://multiverseaccordingtoben.blogspot.ca/2010/03/coherent-aggregated-volition-toward.html>)
- [110] Goertzel B 2010 GOLEM (<http://goertzel.org/GOLEM.pdf>)
- [111] Goertzel B 2012 Should humanity build a global AI nanny to delay the singularity until it’s better understood? *J. Consciousness Stud.* **19** 96–111
- [112] Goertzel B 2012 When should two minds be considered versions of one another? *Int. J. Mach. Consciousness* **4** 177–85
- [113] Goertzel B 2012 CogPrime (http://wiki.opencog.org/w/CogPrime_Overview)
- [114] Goertzel B and Bugaj S V 2008 Stages of ethical development in artificial general intelligence systems *Artificial General Intelligence (Frontiers in Artificial Intelligence and Applications no. 171)* (Amsterdam: IOS) pp 448–59
- [115] Goertzel B and Pitt J 2012 Nine ways to bias open-source AGI toward friendliness *J. Evol. Technol.* **22** 116–31
- [133] Hanson R 1994 If uploads come first *Extropy* **6** 10–15
- [134] Hanson R 1998 Economic growth given machine intelligence (<http://hanson.gmu.edu/aigrow.pdf>)
- [135] Hanson R 2007 Shall we vote on values, but bet on beliefs? (<http://hanson.gmu.edu/futarchy.pdf>)
- [136] Hanson R 2008 Economics of the singularity *IEEE Spectr.* **45** 45–50
- [137] Hanson R 2009 Prefer law to values *Overcoming Bias* (www.overcomingbias.com/2009/10/prefer-law-to-values.html)
- [138] Hanson R 2012 Meet the new conflict, same as the old conflict *J. Consciousness Stud.* **19** 119–25
- [146] Hibbard B 2001 Super-intelligent machines ACM SIGGRAPH *Comput. Graph.* **35** 13–5
- [147] Hibbard B 2005 The ethics and politics of super-intelligent machines (https://sites.google.com/site/whibbard/g/SI_ethics_politics.doc)
- [148] Hibbard B 2005 Critique of the SIAI collective volition theory (www.ssec.wisc.edu/~billh/g/SIAI_CV_critique.html)
- [149] Hibbard B 2008 Open source AI *Artificial General Intelligence Frontiers (Artificial Intelligence and Applications no. 171)* ed P Wang, B Goertzel and S Franklin (Amsterdam: IOS) pp 473–7
- [150] Hibbard B 2012 Model-based utility functions *J. Artificial Gen. Intell.* **3** 1–24
- [151] Hibbard B 2012 Decision support for safe AI design *Artificial General Intelligence (Lecture Notes in Artificial Intelligence no. 7716)* ed J Bach, B Goertzel and M Ikl (New York: Springer) pp 117–25
- [152] Hibbard B 2012 The error in my 2001 VisFiles column (www.ssec.wisc.edu/~billh/g/visfiles_error.html)
- [153] Hibbard B 2012 Avoiding unintended AI behaviors *Artificial General Intelligence (Lecture Notes in Artificial Intelligence no. 7716)* ed J Bach, B Goertzel and M Ikl (New York: Springer) pp 107–16
- [306] Yudkowsky E 1996 Staring into the singularity (<http://yudkowsky.net/obsolete/singularity.html>)
- [307] Yudkowsky E 2001 Creating friendly AI 1.0 (<http://intelligence.org/files/CFAI.pdf>)
- [308] Yudkowsky E 2004 Coherent extrapolated volition (<http://intelligence.org/files/CEV.pdf>)
- [309] Yudkowsky E 2011 Artificial intelligence as a positive and negative factor in global risk *Global Catastrophic Risks* ed N Bostrom and M M Cirkovic (Oxford: Oxford University Press) pp 308–45
- [310] Yudkowsky E 2008 Hard takeoff *Less Wrong* (http://lesswrong.com/lw/wf/hard_takeoff/)
- [311] Yudkowsky E 2009 Value is fragile *Less Wrong* (http://lesswrong.com/lw/y3/value_is_fragile/)
- [312] Yudkowsky E 2011 Complex value systems are required to realize valuable futures (<http://intelligence.org/files/ComplexValues.pdf>)
- [313] Yudkowsky E 2012 Reply to Holden on tool AI *Less Wrong* (http://lesswrong.com/lw/cze/reply_to_holden_on_tool_ai/)