

# Machine Intelligence meets Natural Intelligence



Presented by Harshal Kate

Elaborated jointly with Nikolaos Stratigakis

(Participants of the Academia Engelberg Summer School-2016

directed by Prof. Dr. Harald Atmanspacher)

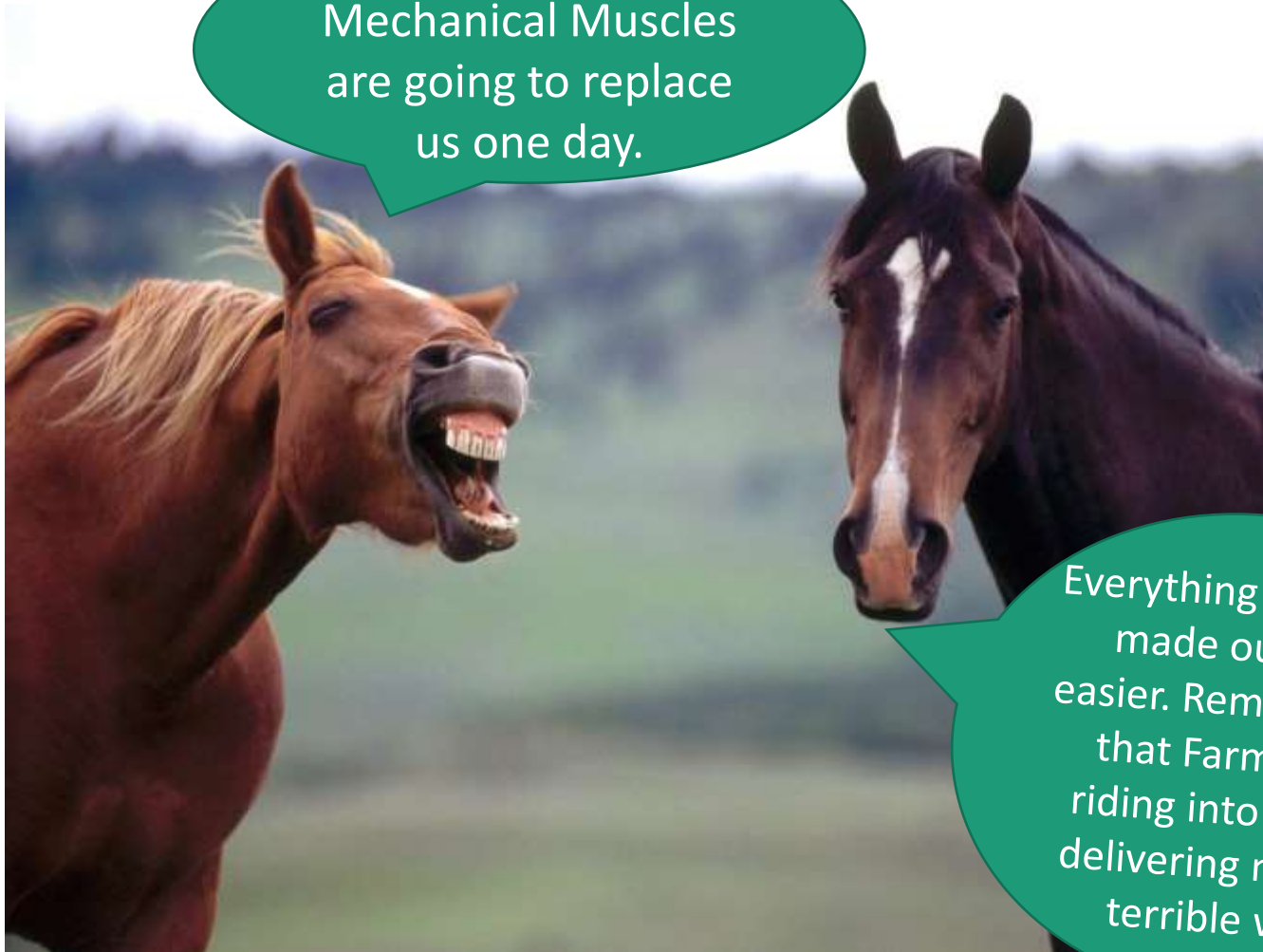


ACADEMIA ENGELBERG

**“AT THE LIMIT”, 15th Dialogue on Science. Oct. 2016. Engelberg, Switzerland.**

# Late 19<sup>th</sup> century

**Stephen  
Elon  
Musking**



All these New  
Mechanical Muscles  
are going to replace  
us one day.

Everything so far has  
made our lives  
easier. Remember all  
that Farm work,  
riding into battles,  
delivering mails, all  
terrible work.

# What's the Difference between Humans and The Rest?



	Dogs	Computers	Humans	Sci-Fi AI (Super Intelligence)
Computational Power	No	Yes (**)	Yes (**)	Yes (***)
Knowledge of Data	No	Yes (**)	Yes (**)	Yes (***)
Emotional Intelligence	Yes (*)	No	Yes (**)	Yes (?)
Capacity for Inference	Yes (*)	No	Yes (**)	Yes (?)
Social Intelligence	Yes (*)	No	Yes (**)	Yes (?)
Creativity	No	No	Yes (**)	Yes (?)

Approximate Strength  
(\*): Low  
(\*\*): Medium  
(\*\*\*): High

# Developments so far



- Initially: Systems able to achieve things in a particular micro-world, e.g. to compose a particular kind of sentence.
- Since the '80s: more “Organic” performance, that is more flexible.
- Back in the day: Translation systems with pre-set grammar.

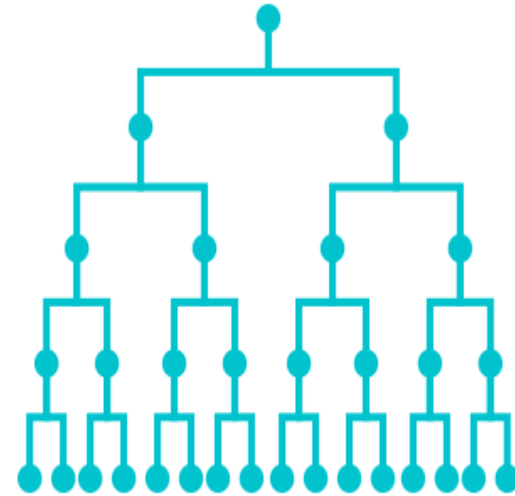
Today: Translation systems with learning abilities.

- An essential idea: “Bayesian agent” → Starts with a particular set of probabilities which can be reshaped in response to new information.

# Problems faced by AI



- **Combinatorial explosion**
- **Doing without thinking.** AI is relatively good in thinking but not so good in intuition which is often used by humans
- **Bounded Rationality:** Concept that decision makers (irrespective of their level of intelligence) have to work under three unavoidable constraints:
  - (1) Only **limited**, often unreliable, **information** is available regarding possible alternatives and their consequences,
  - (2) Human mind has only **limited capacity** to evaluate and process the information that is available, and
  - (3) Only a **limited** amount of **time** is available to make a decision. (Herbert Simon)



# Kahneman's distinction



- Two systems of thinking:
  - System 1: Fast, automatic, frequent, emotional, stereotypic, and subconscious
  - System 2: Slow, effortful, infrequent, logical, calculating, and conscious
- We can argue that the present forms of softwares and AI are better at the latter.

# “Chinese room” argument



- Let us take a Swiss Lady.
- She is closed in a room with a set of Mandarin letters she cannot understand.
- Though, she has a guidebook which explains how to put them in correct order in response to the ones slid under the door by another person who is outside.
- Using the guide she responds right.
- But she can never really understand what they are about. She has no consciousness about what she translates (argument by John Searle).



- Demis Hassabis, leader of Google's "Deepmind" project:

We are very far from self-conscious AI like the one in the movies!

Practically speaking, AI's goal is just to make machines smart enough to process large amounts of data. Just a tool for humans.



# An Alternative Way: GNR



- **Genetics (Biotechnology).**
  - Electronically emulate the human's evolutionary process (and speed it up).
  - Scan a brain and emulate its tissues.
  - Rejuvenate all of body's tissues and organs.
- **Nanotechnology and Robotics (Nanobots)**
  - To enhance the human brain with digital implants.
  - Blood cell sized robots that can travel in the bloodstream destroying pathogens, removing debris, correcting DNA errors, and reversing aging processes.



# Superintelligence

- Superintelligence (SI) = Synthetic General Intelligence with nearly infinite resources that is way smarter than Human Intelligence (HI) and smart enough to keep improving itself.
- Nick Bostrom defines Superintelligence as “Any intellect that radically outperforms humanity in all practical fields which includes scientific creativity, social skills, general wisdom, etc.”
- So a human-level AI would be a self-conscious mind capable of complex and flexible thinking (which requires learning capacity) and intuition.
- What if though the AI is superintelligent, that is superior to human mind in every aspect?

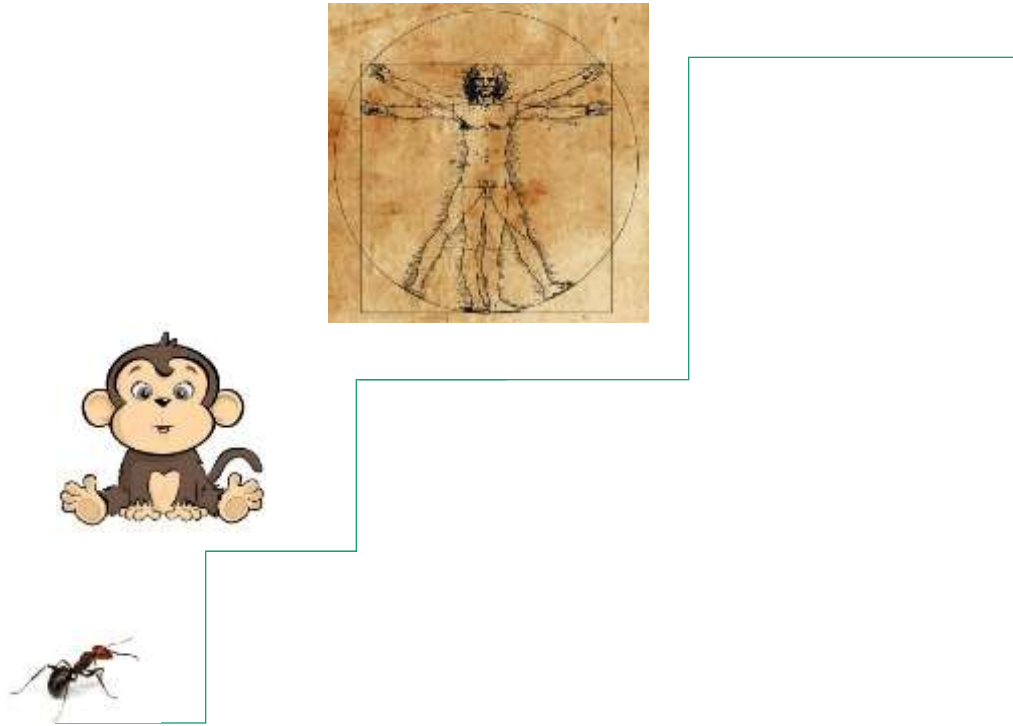
# What would be its goal?

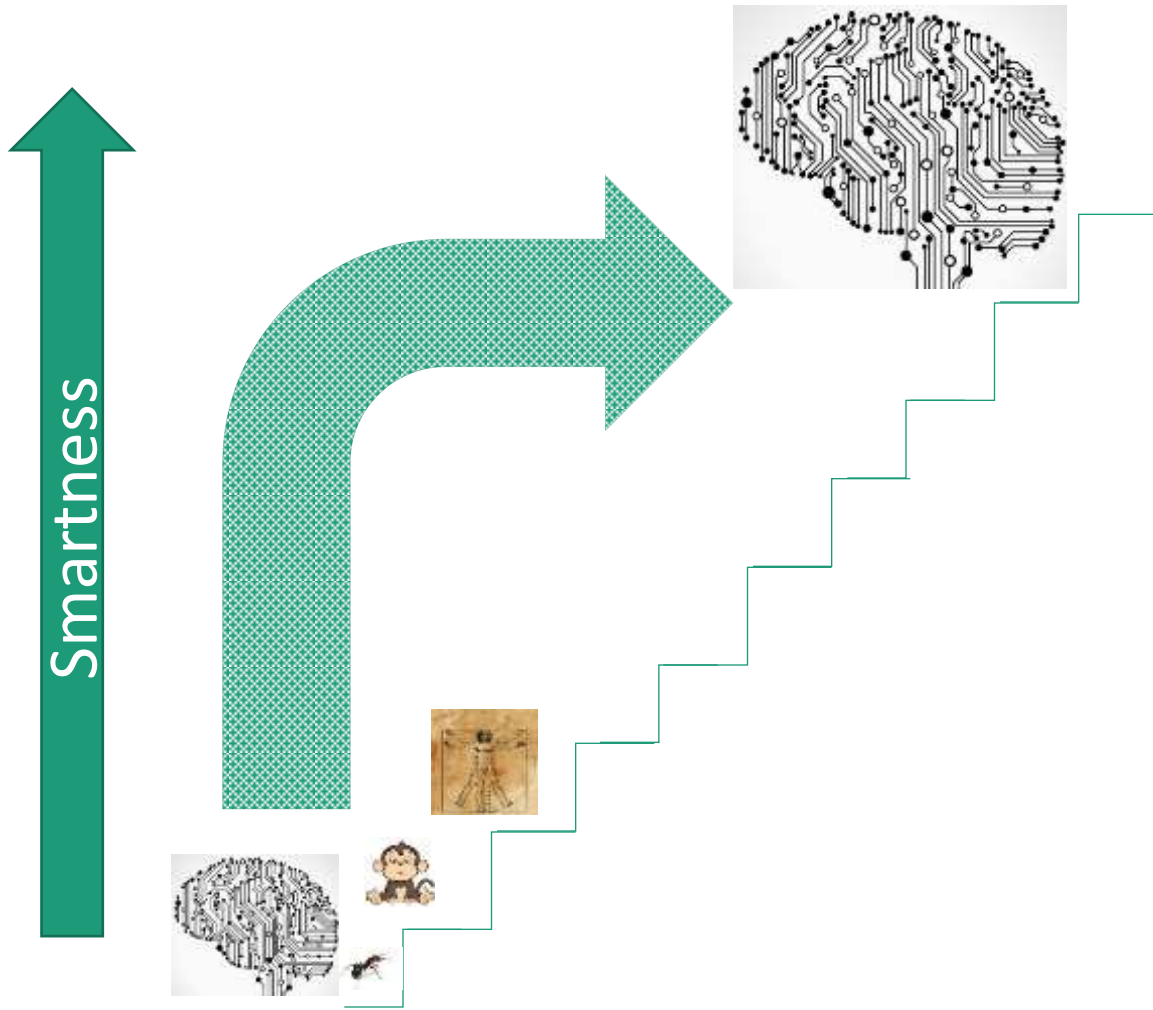


- It is not necessary that they (Super Intelligent Systems) have human goals like the AI in the films.
- They may have *instrumental goals* and *final goals*.



Smartness ↑

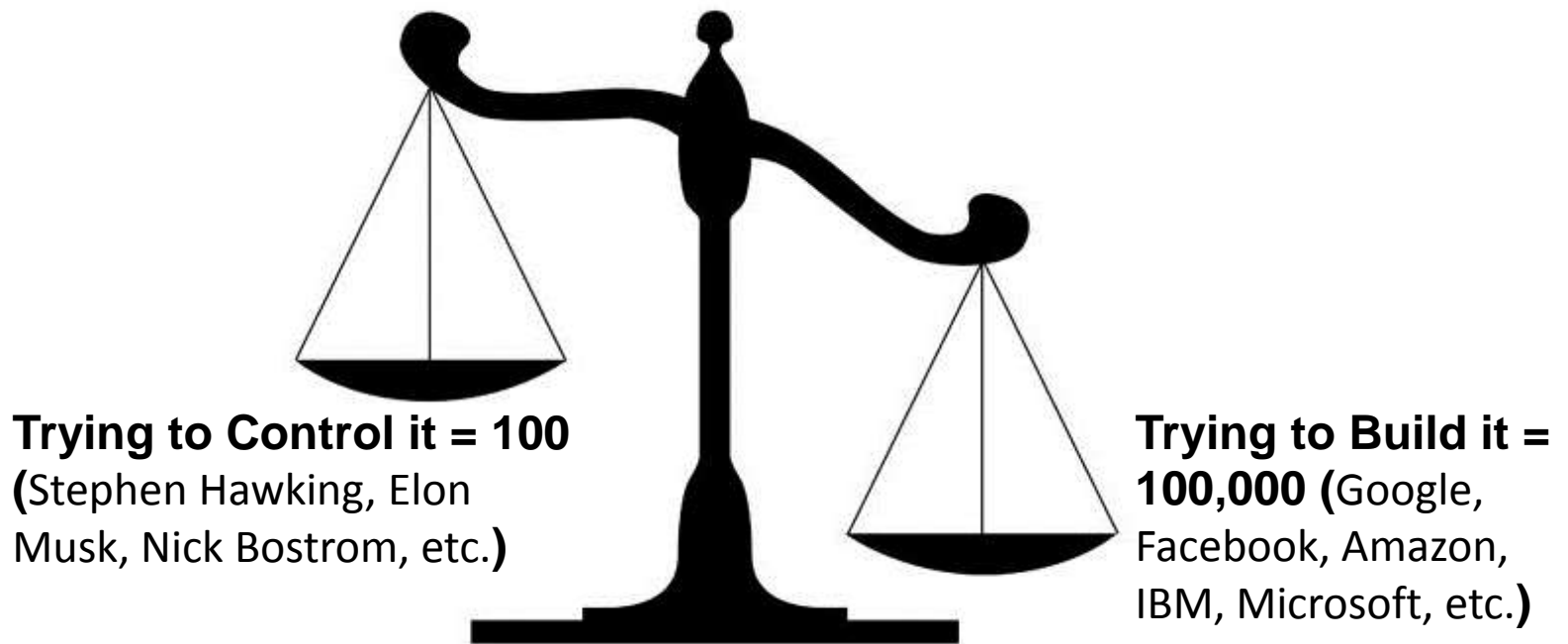




# Will this explosion be actually sudden?



- If so, we have to make special preparation so that we will be able to put the SI under control.





# Risks with SI

- It is difficult to anticipate
- We are comparing SI to HI however it need not be alike.
  - It would be a mistake to think it as a human as it is not.



- Setting of it's Final Goal



- Example:
  - a) Final goal: Preservation of the Environment.  
Instrumental goal: Protection of Humans  
→ Destruction of Humans!
  - b) Final goal: Protection of Humans.  
Instrumental goal: Preservation of the Environment.  
→ Preservation of Environment as long as Humans are not harmed.

So we need caution with goal setting!



# Stamp collector who is an AI programmer



- Essentials for AI
  1. Connection to **Internet**: To send and receive data
  2. **Internal model of reality**: Allowing it to make *accurate predictions* of what will happen with its different inputs.
  3. **Utility Function + Optimisation**: For every possible sequence of data it sends, it uses these functions to predict the number of stamps it will collect
  4. Highest rated **Output**: Collect maximum number of stamps.

# The Aftermath



- The creation of a self-aware, Super Intelligent AI with access to a huge amount of data. How would it act?
- Since Data is nowadays as precious as money, an oligarchy of companies could be created that will manage these data affecting our lives. These companies would be in possession of Super Intelligent AI.
- What if “Bad Guys” succeed first?
- Developed Nations will be the first Victims



# Replacement of humans?



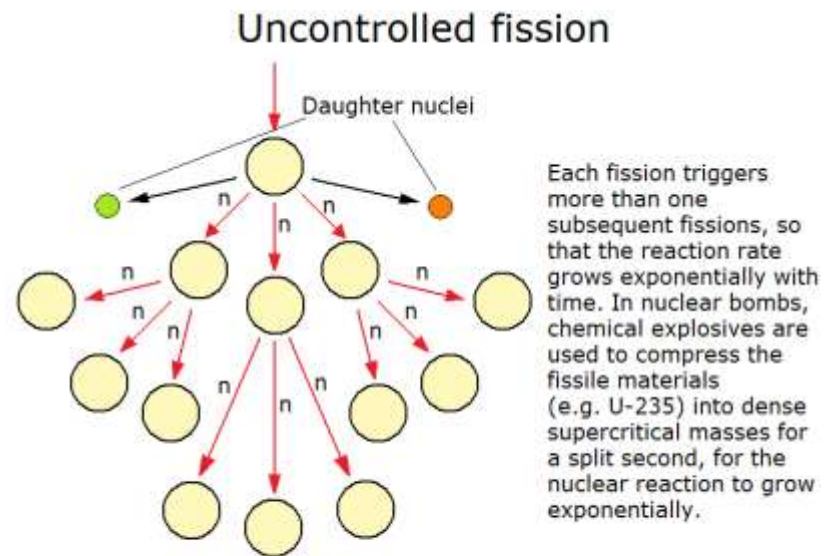
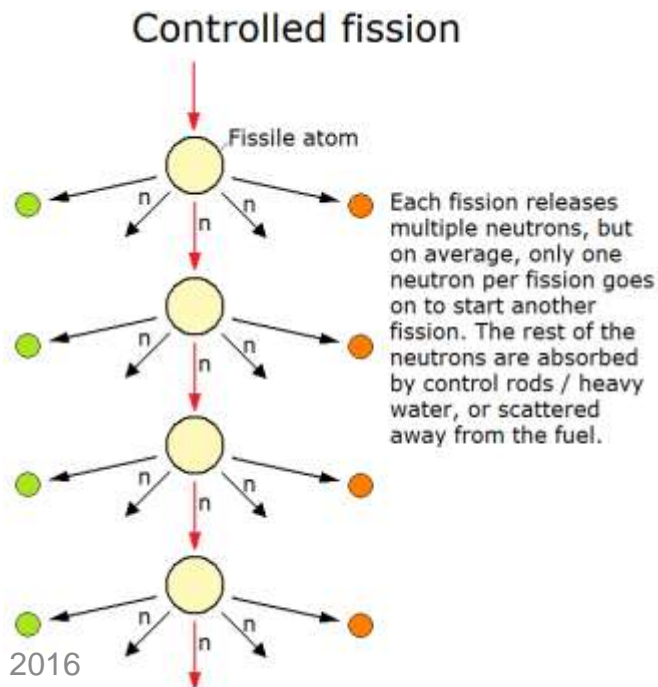
- As it concerns some aspects
  - E.g. People would pay more to see a Human Orchestra than an, otherwise perfect, Robotic one.
  - Human creativity will probably continue being appreciated.
- What if laborers lose their jobs due to AI?
  - The distribution of wealth will be a huge debate.
  - Many manual jobs will be lost, while some others may be created.

(The proportion of jobs threatened by automation in India is 69 % and 77 % in China: World Bank Group President Jim Yong Kim)

# Ways to Control Superintelligence



- **Unplug the system:** It is smart enough to play dumb until it is too late to stop it.
- **Build a cage:** Does not work. It is smart enough to get out and grow on its own.
- **So do not build it:** Oops! Google/Siri in our pockets is already paving the way.





- Does not matter who created SI. Once it is out there, it shall be difficult to control it unless we teach it to “**CARE**” as we do while we are building it.
- We just have one chance to do it right as once the **1<sup>st</sup> system** is built it shall try to stop the other systems from evolving. So our 1<sup>st</sup> try (our 1.0 release) will need to **have enough controls to let humans survive.**
- Need to set up AI with **right initial conditions** then maybe we increase the odds that the 1<sup>st</sup> superintelligence we create **cares about the future of human lives.**
- Need to make sure its **preferences match** with that of ours. Otherwise it is going to do things that we don't want it to do.



# The Cobra Effect

An attempted solution to a problem actually makes the problem worse rather than solving it.

Thank you for your kind Attention!



# Bibliography

- Ahmed, Kamal. “Google’s D. Hassabis: misuse of AI could do harm”. BBC News. 2015.
- Bostrom, Nick. *Superintelligence*. Oxford: Oxford University Press, 2014.
- Kahneman, Daniel. *Thinking Fast and Slow*. New York: Farrar, Strauss & Giroux, 2011.
- Lawrence, Neil. "Beware the Rise of the Digital Oligarchy", ""Google AI vs Go Grandmaster", "How to Prevent Creeping AI becoming Creepy" . The Guardian. 2015-6.