

# Re112: *The Attention Hazard and The Attention (Distraction) Economy*

retraice.com

*Drawing attention to dangerous information can increase risk, but the attention economy tends to draw attention toward amusement.*

Information hazards; formats include data, idea, attention, template, ‘signaling’ and ‘evocation’; increasing the number of information locations; adversaries, agents, search, heuristics; the dilemma of attention; suppressing secrets; the Streisand effect; the attention economy as elite ‘solution’; Liu’s ‘wall facers’.

Air date: Tuesday, 10th Jan. 2023, 10:00 PM Eastern/US.

## Attention hazard of information

Bostrom (2011): “**Information hazard:** A risk that arises from the dissemination or the potential dissemination of (true) information that may cause harm or enable some agent to cause harm.”<sup>1</sup>

Attention is one format (or ‘mode’) of information transfer:<sup>2</sup>

“**Attention hazard:** The mere drawing of attention to some particularly potent or relevant ideas or data increases risk, even when these ideas or data are already ‘known’.”<sup>3</sup>

This increase is because ‘attention’ is physically increasing the number of locations where the hazard data or idea are instantiated.

## Adversaries and agents

“Because there are countless avenues for doing harm, an **adversary** faces a **vast search task** in finding out which avenue is most likely to achieve his goals. Drawing the adversary’s **attention to a subset of especially potent avenues can greatly facilitate the search**. For example, if we focus our concern and our discourse on the challenge of defending against viral attacks, this may signal to an adversary that viral weapons—as distinct from, say, conventional explosives or chemical weapons—constitute an especially promising domain in which to search for destructive applications. **The better we manage to focus** our defensive deliberations on our greatest vulnerabilities, **the more useful our conclusions may be to a potential adversary.**”<sup>4</sup>

Consider the parallels in Russell & Norvig (2020):

- ‘adversarial search and games’ (chpt. 5);
- ‘intelligent agents’ (chpt 2);
- ‘solving problems by searching’ (chpt. 3);
- drawing attention can facilitate search: heuristics (sections 3.5, 3.6);

**The dilemma:** We focus on risk, and also lead adversary-agents to our vulnerabilities.

Cf. the ‘vulnerable world hypothesis’<sup>5</sup> on the policy implications of unrestrained technological innovation given the unknown risk of self-destructing innovators.

“Still, one likes to believe that, on balance, investigations into existential risks and most other risk areas will tend to reduce rather than increase the risks of their subject matter.”<sup>6</sup>

## Secrets and suppression

“Clumsy attempts to suppress discussion often backfire. An adversary who discovers an attempt to conceal an idea may infer that the idea could be of great value. Secrets have a special allure.”<sup>7</sup>

[https://en.wikipedia.org/wiki/Streisand\\_effect](https://en.wikipedia.org/wiki/Streisand_effect): “[T]he way attempts to hide, remove, or censor information can lead to the unintended consequence of increasing awareness of that information.”

## The attention (distraction) economy

Might the attention economy, one day or even already, be a ‘solution’ (an elite solution) to the attention hazard? Would it work against AI? Or buy us time? What about ‘wall facers’?<sup>8</sup>

Cf. Re30, Retraice (2022/10/26), on things being done. □

<sup>1</sup>Bostrom (2011) p. 2.

<sup>2</sup>The others he distinguishes are: data, idea, template, ‘signaling’ and ‘evocation’.

<sup>3</sup>Bostrom (2011) p. 3.

<sup>4</sup>Bostrom (2011) p. 3.

<sup>5</sup>Bostrom (2019).

<sup>6</sup>Bostrom (2011) p. 4.

<sup>7</sup>Bostrom (2011) p. 3.

<sup>8</sup>Liu (2016).

## References

- Bostrom, N. (2011). Information Hazards: A Typology of Potential Harms from Knowledge. *Review of Contemporary Philosophy*, 10, 44–79. Citations are from Bostrom’s website copy:  
<https://www.nickbostrom.com/information-hazards.pdf> Retrieved 9th Sep. 2020.
- Bostrom, N. (2019). The vulnerable world hypothesis. *Global Policy*, 10(4), 455–476. Nov. 2019. Citations are from Bostrom’s website copy:  
<https://nickbostrom.com/papers/vulnerable.pdf> Retrieved 24th Mar. 2020.
- Liu, C. (2016). *The Dark Forest*. Tor Books. ISBN: 978-0765386694. Searches:  
<https://www.amazon.com/s?k=9780765386694>  
<https://www.google.com/search?q=isbn+9780765386694>  
<https://lccn.loc.gov/2015016174>
- Retraice (2022/10/26). Re30: AI Progress and Surrender. *retraice.com*.  
<https://www.retraice.com/segments/re30> Retrieved 27th Oct. 2022.
- Russell, S., & Norvig, P. (2020). *Artificial Intelligence: A Modern Approach*. Pearson, 4th ed. ISBN: 978-0134610993. Searches:  
<https://www.amazon.com/s?k=978-0134610993>  
<https://www.google.com/search?q=isbn+978-0134610993>  
<https://lccn.loc.gov/2019047498>