# Avoiding Pseudoscience in the Science of Security

## A Case Study of Malware Indicator Analysis

Jonathan Spring
Netsa-contact@cert.org
Symposium on the Science of Security
April 21, 2015

Special thanks: Eric Hatleback, Drew Kompanek

DM-0002025

# Problem

Society can no longer afford to ignore security holes in widely-deployed computer systems

Information security now means understanding this Internet we've created

- "Science of Security" wants to solve this problem
- However it has not
  - Cite roll-call of breaches lately

Root problem: Current methods and frameworks for both investigating security problems and deriving understanding therefrom are not well-defined

# Method & Talk Outline

Define Psuedoscience

Summary of operational practice and challenges

- Malicious software analysis
- Malicious ecosystem analysis

Summary of current published work in security

- We’re doing some pseudoscience
- Examine reasons why or what is missing

Describe effective science tools from other disciplines

- Criteria for good analysis and observation
- Mechanisms and modeling mechanistically

Highlight a way to combine all the good things

# Goal

Fit work into a cohesive world-view of models that appropriately model, explain, and predict ~~a set of phenomena~~ networked-computer security

- Identify pseudoscience as to exclude it
- Pedagogically: method for making reliable contributions

Ultimate goal:

Engineer reliable & safe systems, as we do buildings

- Of course, that has a longer history
- “If a builder build a house for some one, and does not construct it properly, and the house which he built fall in and kill its owner, then that builder shall be put to death.”

Hammurabi's Code of Laws. Translated by L. W. King. http://eawc.evansville.edu/anthology/hammurabi.htm

# What is Pseudoscience?

"spurious or pretended science; study or research that is claimed as scientific but is not generally accepted as such. Chiefly derogatory."

"pseudoscience, n.". OED Online. September 2014. Oxford University Press. http://www.oed.com/view/Entry/153794?redirectedFrom=pseudoscience (accessed November 11, 2014).

But what makes science "spurious" or "pretended" ?

To answer, we need to know what "science" is

- Science is… a checklist? a club? …no

# Turn to Philosophy of Science

Philosophy of Science is not pure philosophy

- Philosophy of Science is about:
    - Foundations (What is science?[1])
    - Methods (How do you design observations to reach proper conclusions? – Important!)
    - Implications (What do these collected observations mean? – Perhaps more important)

[1] See for example: Bunge, Mario. "What is science? Does it matter to distinguish it from pseudoscience? A reply to my commentators." New ideas in psychology 9, no. 2 (1991): 245-283.

# Pitfall – Language is Hard

Wittgenstein's insight

- "Meaning as use"

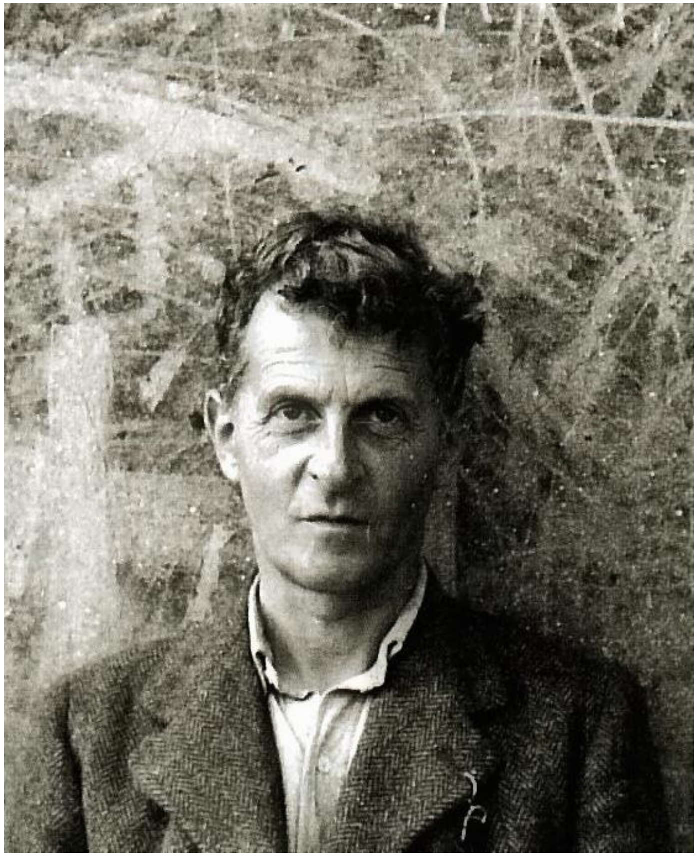
Meaning is not in a dictionary

Context and assumptions (tacit knowledge) are central

Thus, I tend to avoid asking "What is Science?" directly

✓ Methods and Implications

- (doing and learning the right things)

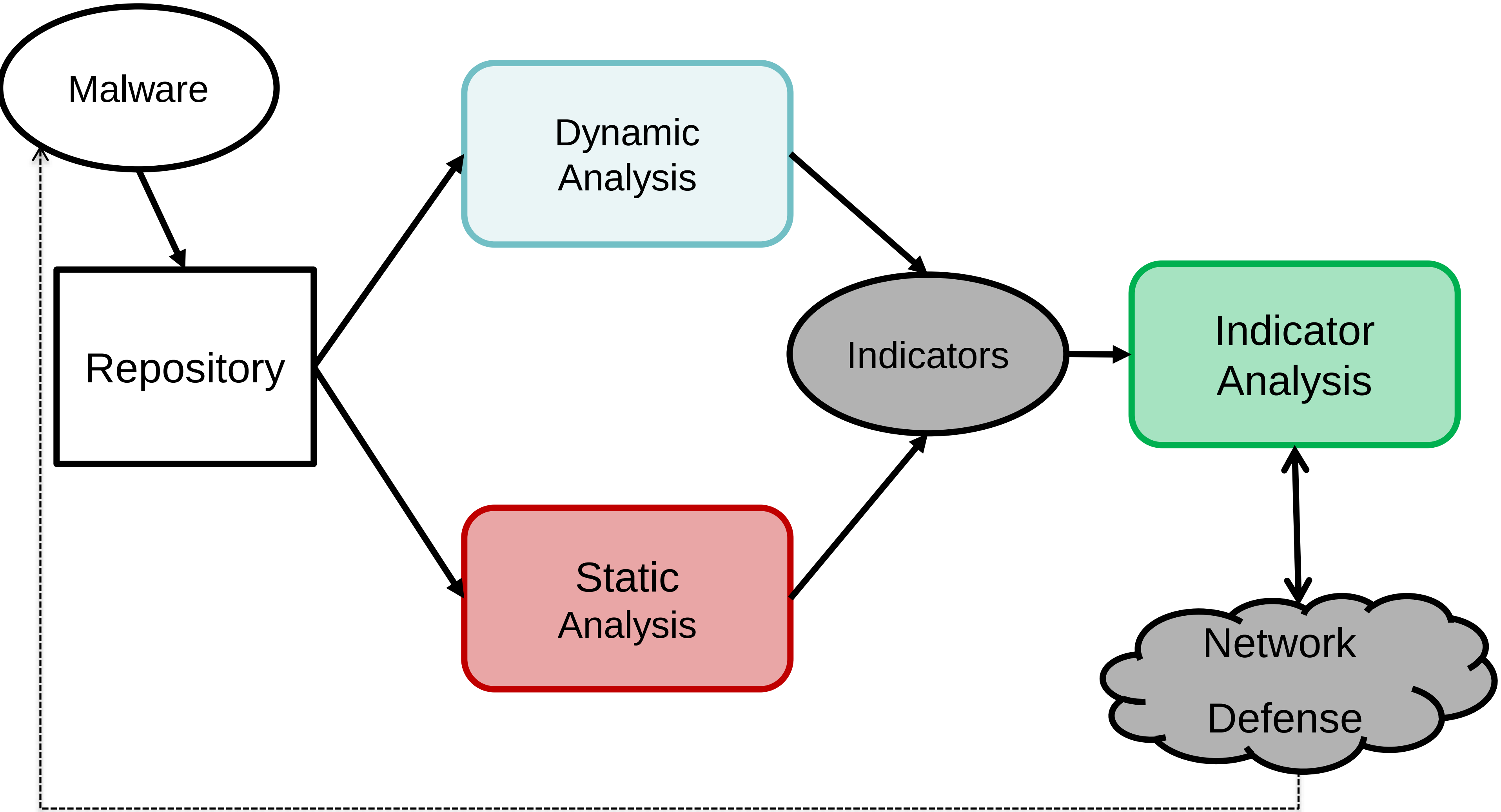


Ludwig Wittgenstein, Wales, 1947
photo by Ben Richards

# SUMMARY OF OPERATIONAL PRACTICE AND CHALLENGES

At least, as we at CERT see it

# Case Study: Malicious Software Analysis and Network Defense



# Context

1. **Static Analysis**
   - Inspecting binary code without execution
     - Hash, fuzzy hash, string extraction, object extraction, source code recovery
2. **Dynamic Analysis**
   - Running code to see what it does
     - Log registry changes, file changes, network access, code execution coverage, inspect memory for assembled code
3. **Indicator Analysis**
   - Analysis of derived indicators to improve defense
     - Indicators extracted from above, such as domain names, IP address, MD5 hash, file names, etc.

# Features We Want in Well-structured Observation

1. **Internal Validity**: the mechanism under observation is of suitable scope to achieve the reported results.
2. **External Validity**: the mechanism under observation is not solely an artifact of the laboratory; the observed mechanism is faithful to mechanisms “in the wild.”
3. **Containment**: no pre-mechanism causes threaten to confound the results, and no post-mechanism effects are a threat to safety.
4. **Transparency**: there are no explanatory gaps in the observed mechanism or observation design; the diagram for the experimental mechanism is complete.

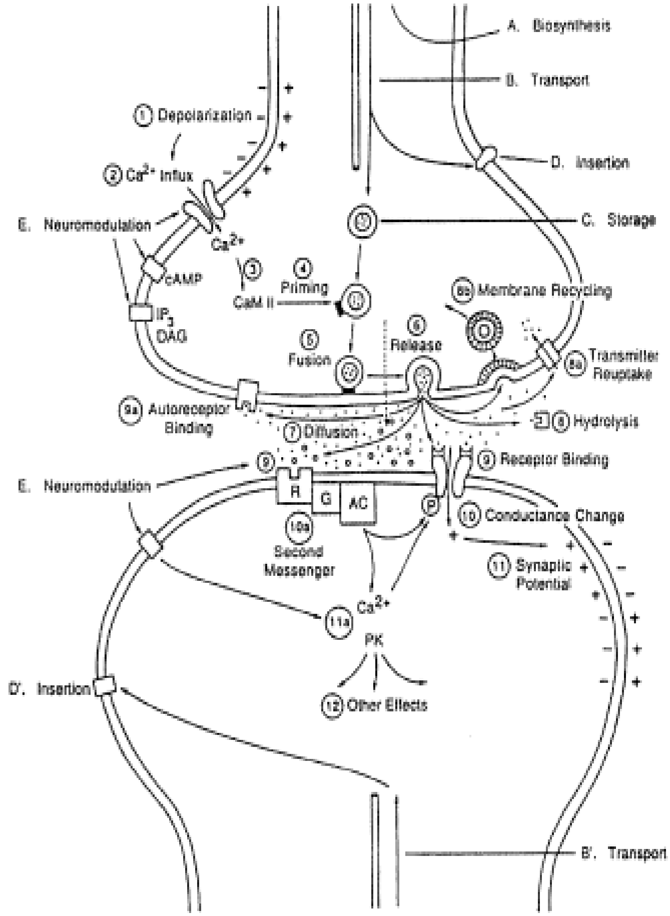
Hatleback, E., and J. M. Spring, “Exploring a mechanistic approach to experimentation in computing,” Philosophy & Technology, vol. 27, no. 3, pp. 441–459, 2014.

# State of Science: Mechanistic Approach

Models are useful representations

Scientists model mechanisms

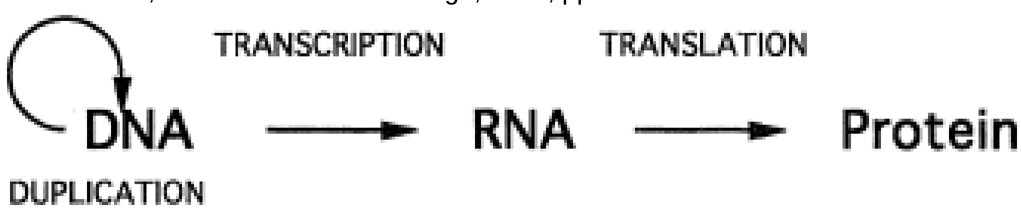
- "A mechanism for a phenomenon consists of entities and activities organized in such a way that they are responsible for the phenomenon."





Illari, P. M. and J. Williamson (2012). What is a mechanism? thinking about mechanisms across the sciences. European Journal for Philosophy of Science 2(1), 119–135.

See also: S. Glennan, "Mechanisms," in The Routledge Companion to Philosophy of Science, M. Curd and S. Psillos, Eds. New York: Routledge, 2013, pp. 420–428.

# Additional Wants Across Observations

Reproducibility

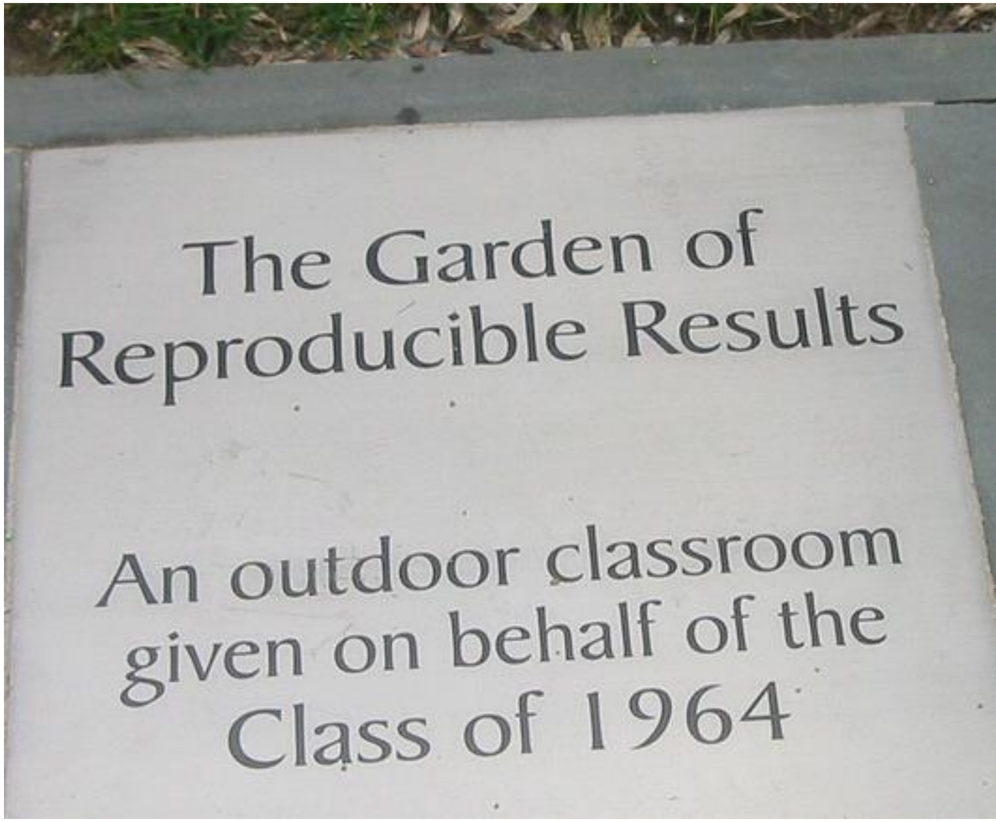
- Experimental
- Computational
- Statistical

Stodden, V. (2013). Resolving Irreproducibility in Empirical and Computational Research. IMS Bulletin Online. http://bulletin.imstat.org/2013/11/resolving-irreproducibility-in-empirical-and-computational-research/

Corroboration

Feitelson, D. G. (2015). From repeatability to reproducibility and corroboration. ACM SIGOPS Operating Systems Review, 49(1), 3-11.

&c.



# SUMMARY OF STATE OF PUBLISHED WORK

Spoiler: There’s some pseudoscience now

# State of Malicious Software Analysis

Survey of 36 malware analysis publications (2012)

Proposed criteria (next slide)

Evaluated publications against criteria with detailed rubric

“Identified shortcomings in most papers”

- I.e. only 1 or 2 passed inspection
- Both top-tier and less prominent venues failed equally

Rossow, Christian, et al. "Prudent practices for designing malware experiments: Status quo and outlook." Security and Privacy (SP), 2012 IEEE Symposium on. IEEE, 2012.

# Critera of Rossow, *et. al.*

Correctness: the use of good datasets to ensure that the experiment tests what it is intended to test.

Internal Validity — The mechanism under experimentation is of suitable scope to achieve the reported results.

Realism: the maintenance and use of widely-varied, currently-relevant malware families and operating systems in the experiment.

External Validity — The experimental mechanism is faithful to the mechanisms "in the wild."

Safety: the use of proper containment policies to prevent the experiment from causing harm to others.

Containment — No pre-mechanism causes threaten to confound the results, and no post-mechanism effects are a threat to safety.

Transparency: the clear, unambiguous description of the various components of the experimental setup.

Transparency — There are no explanatory gaps in the experimental mechanism; the diagram for the experimental mechanism is complete.

# State of Internet Measurement

Largely the same as malware community

“Although the Internet has been studied for decades with increasing diversity in the set of measurements collected and entities studied, there has been a **notable lack of precisely articulated standards** for such measurement-driven studies.” [emph added]

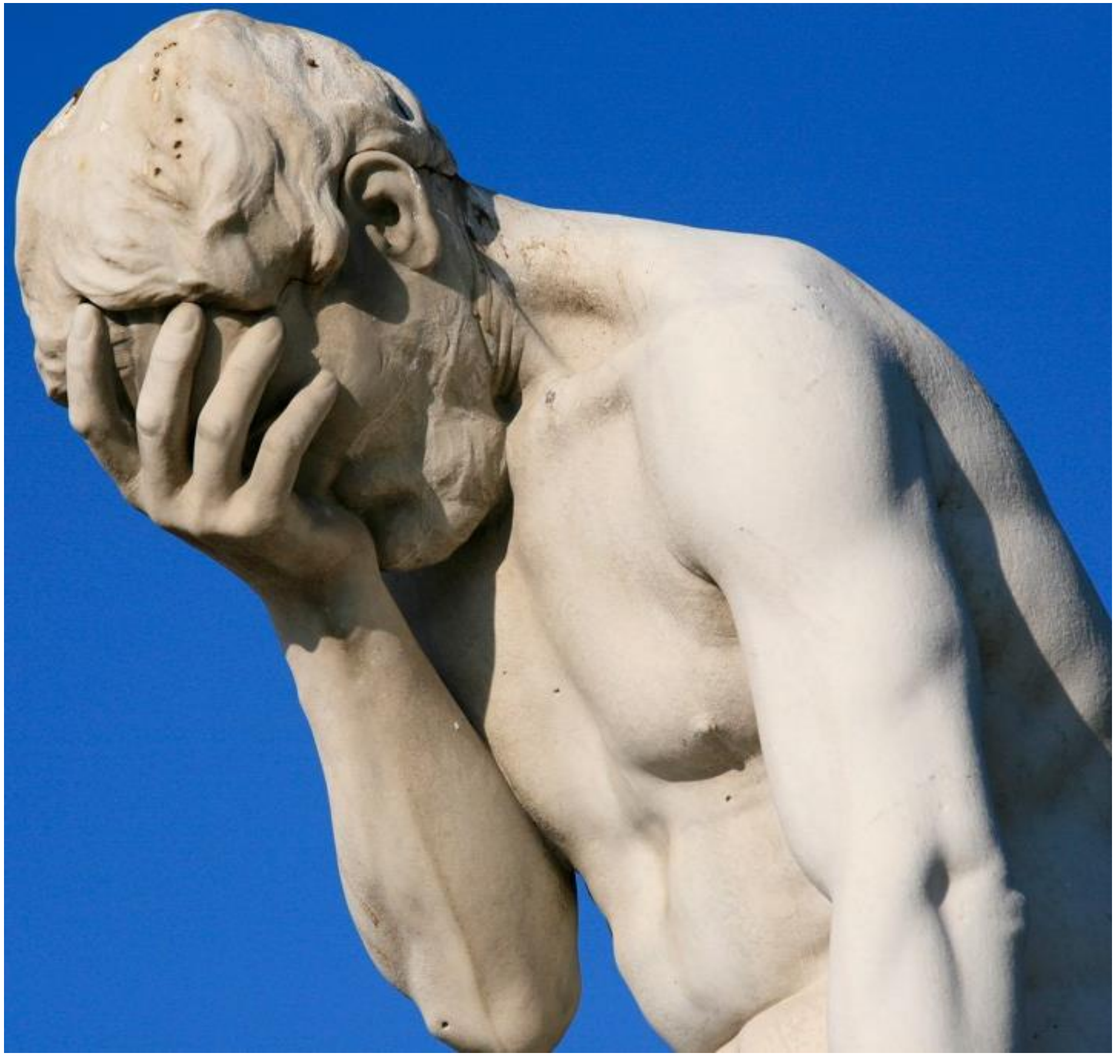
Krishnamurthy, Balachander, et al. "A Socratic method for validation of measurement-based networking research." Computer Communications 34.1 (2011): 43-53.

# This is a Poor State of Affairs

Smells of Pseudoscience

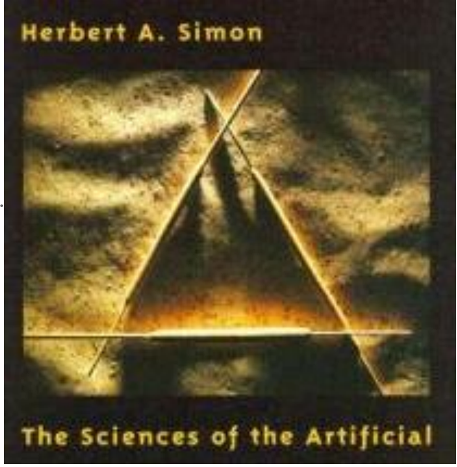
- Lacks validity
- Lacks containment
- Lacks transparency
- Appeals to authority

Why?



Flickr/Alex Proimos, cropped, CC BY-NC

# Philosophy of Science in CS?




Current state is essentially Herb Simon

- Consistent with famous philosophy of science like Kuhn and Popper
- i.e., it hasn't been updated since the 60s.

Computer Science and Philosophy of Science appear to have parted ways in the 60s

- There has been substantial progress in both
- It's time for them to update each other

# EFFECTIVE SCIENCE TOOLS FROM OTHER DISCIPLINES

Other people have had this problem before

The History of Science is helpful

# What Can We Do?

1. Find what can be borrowed
   - Implement into cybersecurity
     - Expectations for experimental design
     - Study of physical mechanisms
     - Many specifics from various fields
2. Describe where additional tools are needed
   - Build those philosophical and methodological tools
     - Study of engineered mechanisms
     - Specific novel aspects in, e.g., game theory, economics, risk analysis, etc.

# Standard Roles for Experimentation from Physics

Allan Franklin's classification:

1. Theory choice
2. Theory articulation
3. Demonstration that entities involved in our accepted theories exist
4. Measurement of physical quantity
5. Life of its own

We think philosophers need to add one:

6. Demonstration that activities involved in our accepted theories occur

# Disputes with Learning from Other Fields

Dispute what computing, and cybersecurity, can learn

- Computing has a “intrinsically different disciplinary nature, scientific and, at the same time, engineering.”
- “Experiments about artifacts in computing tell us more about the people that have done the job, than the way the world is.”

Schiaffonati, Viola, and Mario Verdicchio (2013). "Computing and Experiments: A Methodological View on the Debate on the Scientific Nature of Computing." *Philosophy & Technology*, 1-18. (p. 14)

There is an important kernel of truth here:

- Computing is different, it’s a separate field for a reason

But it is *not* hopeless or disjoint, as the above seems

# Science vs. Engineering

Building bridges uses scientific results and method

Science and engineering are not disjoint

- Rather, they are  related on a spectrum

Flickr/Josepha, cropped, CC BY-NC-SA 2.0

For some examples of this spectrum, see: Vincenti, Walter G. What Engineers Know and How They Know It : Analytical Studies from Aeronautical History, Johns Hopkins Studies in the History of Technology [New. Ser., No. 11]. Baltimore: Johns Hopkins University Press, 1990

# Example Challenge: Static Malicious Software Analysis

*Given an unknown file, tell me what it does* (activities)

Millions of samples – alarm accuracy is key

- False alarm rate! Not false positive rate
- Avoid base rate fallacy (see S. Axelsson 1999)

And malware authors try to lower your accuracy

- Object-oriented code alone is hard to recover
- Analysis must cover whole control flow graph

Jin, W., Chaki, S., Cohen, C., Gennari, J., Gurfinkel, A., Havrilla, J., Hines, C., Narasimhan, P.: Recovering C++ Objects From Binaries Using Inter-Procedural Data-Flow Analysis. 3rd ACM SIGPLAN Program Protection and Reverse Engineering Workshop (PPREW 2014). 2014.

Quinlan, D. “ROSE: A Preprocessor Generation Tool for Leveraging the Semantics of Parallel Object-Oriented Frameworks to Drive Optimizations via Source Code Transformations,” 383-397. *Proc. Eighth Int’l Workshop on Compilers for Parallel Computers (CPC ‘00).* Aussois, France, Jan. 2000.

# The Problem: Engineered Mechanisms

"Engineered mechanisms are susceptible to having their entities or their activities changed during the course of the investigation at the will of a rational decision maker."

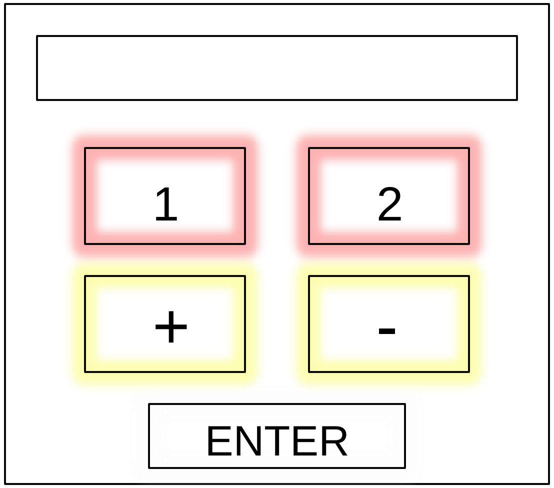
Warhol's calculator:

1 + 1 = 2

1 + 1 = 4

1 + 1 = 0

Imagine entities and activities swapped during operation




Engineered mechanisms exhibit a lack of generality.

- No "Laws of Nature" (Logical Empiricism does not help)
- Discernible entities and activities (Mechanisms do help)

Hatleback, E., and J. M. Spring, "Exploring a mechanistic approach to experimentation in computing," Philosophy & Technology, vol. 27, no. 3, pp. 441–459, 2014.

# ∴ Mechanism Type => Research Strategy

Physical mechanisms – borrow investigative method

Lots of good advice in existing literature

- Experiment design and set up
- Statistical analysis, induction, and results analsyis
- Case study design and analysis

# ∴ Mechanism Type => Research Strategy (II)

Engineered mechanisms require new methods

- Use existing examples of good work
- This is a question of proper design
  - Look to other sciences for requirements on what counts as satisfactory elements

Malicious software analysis is discovery of engineered mechanisms in a sample

- In this context, scientific analysis should be fruitful

# Other Disputes

Claim:

- Adversaries make modeling impossible

That's just not true.
Other fields do it.

- Economics
- Game Theory

# Another Challenge: Assessing Validity

Cannot assess external validity without knowing the external environment: the ecosystem

Malicious ecosystem analysis investigated blacklists

- Tracked indicators known publicly over 30 months
- Over 150 million unique indicators tracked in 2014

The results tell us:

- Most lists are mostly unique indicators
- There is no convergence
- There is therefore no ground truth about what is bad
- External validity cannot be justified by matching a list


Metcalf, L. & Spring, J. Everything You Wanted to Know About Blacklists but Were Afraid to Ask (CERT/CC 2013-39). Software Engineering Institute, Carnegie Mellon University, 2013. http://resources.sei.cmu.edu/library/asset-view.cfm?assetID=83438

Metcalf, L. & Spring, J. Blacklist Ecosystem Analysis Update: 2014. Software Engineering Institute, Carnegie Mellon University, 2014. http://resources.sei.cmu.edu/library/asset-view.cfm?assetID=428609

# Will This Work? Why the Need?

Science has been marvelously successful by implementing these principles

Adversaries are not going away

Experiment design and model building are skills

- People can learn skills when trained properly

Merely practice does *not* make perfect

- Perfect practice makes perfect

# To Avoid Pseudoscience

Apply & expect principles from philosophy of science

- Internal Validity
- External Validity
- Containment
- Transparency
- Corroboration

Manage and assess uncertainty due to engineered mechanisms

Manage and assess uncertainty due to adversaries

# Questions/comments?

# References I


Axelsson, S. (1999, November). The base-rate fallacy and its implications for the difficulty of intrusion detection. In Proceedings of the 6th ACM Conference on Computer and Communications Security (pp. 1-7). ACM.

Franklin, A., 1981. What Makes a 'Good' Experiment?. *The British Journal for Philosophy of Science,* 32(4), pp. 367-374.

Franklin, A., 1990. *Experiment, Right or Wrong.* Cambridge: Cambridge University Press.

Franklin, A., 2012. *Experiment in Physics.* [Online] Available at: http://plato.stanford.edu/archives/win2012/entries/physics-experiment/

Gagliardo, A., *et. al.* (2013). "Oceanic Navigation in Cory's Shearwaters: Evidence for a Crucial Role of Olfactory Cues for Homing After Displacement." *The Journal of Experimental Biology* 216, 2798-2805.

Glennan, S. (2014). "Mechanisms." In M. Curd & S. Psillos, eds. *The Routledge Companion to Philosophy of Science,* Second edition. New York: Routledge, pp. 420-428.

Hatleback, E., and J. M. Spring, "Exploring a mechanistic approach to experimentation in computing," Philosophy & Technology, vol. 27, no. 3, pp. 441–459, 2014.

Jin, W., Chaki, S., Cohen, C., Gennari, J., Gurfinkel, A., Havrilla, J., Hines, C., Narasimhan, P.: Recovering C++ Objects From Binaries Using Inter-Procedural Data-Flow Analysis. 3rd ACM SIGPLAN Program Protection and Reverse Engineering Workshop (PPREW 2014). 2014.

# References II

Killourhy, K.S., and R.A. Maxion (2009). "Comparing Anomaly-Detection Algorithms for Keystroke Dynamics." *IEEE/IFIP International Conference on Dependable Systems & Networks,*125-134.

Killhoury, K.S., and R.A. Maxion (2011). “Should Security Researchers Experiment More and Draw More Inferences?” *CSET 2011*, 1-8.

Machamer, P., L. Darden, and C.F. Craver (2000). “Thinking About Mechanisms.” *Philosophy of Science* 67, 1-25.

Maxion, R.A., T.A. Longstaff, and J. McHugh (2010). “Why Is There No Science in Cyber Science?” *NPSW 2010*, 1-5.

Metcalf, L., J.M. Spring. Everything You Wanted to Know About Blacklists but Were Afraid to Ask (CERT/CC Whitepaper 2013-39). Software Engineering Institute, Carnegie Mellon University, 2013. http://www.cert.org/netsa/publications/blacklists_CERTCC-2013-39.pdf

Metcalf, L., J.M. Spring. Passive Detection of Misbehaving Name Servers (CMU/SEI-2013-TR-010). Software Engineering Institute, Carnegie Mellon University, 2013. http://resources.sei.cmu.edu/library/asset-view.cfm?AssetID=65269

National Security Agency (2013). “NSA Announces Best 2013 Cybersecurity Paper Competition.” [Press Release]. Retrieved from http://www.nsa.gov/public_info/press_room/2013/2013_best_cybersecurity_paper_competition.shtml.

# References III


Rossow, C., et. al. (2012). "Prudent Practices for Designing Malware Experiments: Status Quo and Outlook." 2012 IEEE Symposium on Security and Privacy, 65-79.

Schiaffonati, V., M. Verdicchio (2013). "Computing and Experiments: A Methodological View on the Debate on the Scientific Nature of Computing." Philosophy & Technology, 1-18.

Spring, J. "A Notation for Describing the Steps in Indicator Expansion", IEEE eCrime Researchers Summit, 2013. URL http://resources.sei.cmu.edu/library/asset-view.cfm?assetID=73560.

Spring, J. "Modeling Malicious Domain Name Take-down Dynamics: Why eCrime Pays", IEEE eCrime Researchers Summit, 2013. URL http://resources.sei.cmu.edu/library/asset-view.cfm?assetID=88265.

Spring, J. "Toward Realistic Modeling Criteria of Games in Internet Security." Journal of Cyber Security & Information Systems. Vol 2, num 2. CSIAC. 2014.

Schwartz, E.J., J. Lee, M. Woo, and D. Brumley, "Native x86 decompilation using semantics-preserving structural analysis and iterative control-flow structuring," USENIX Security