When Do Measures Fail? Understanding the Pitfalls of Technical Measures in Engineering Design Through Case Studies Illustrating Goodhart's Law

Casey Eaton and Bryan Mesmer, PhD

COLLEGE OF ENGINEERING
THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

ISEEM

What is Goodhart's Law?

Goodhart's Law focuses on how measures can lose their meaningfulness:

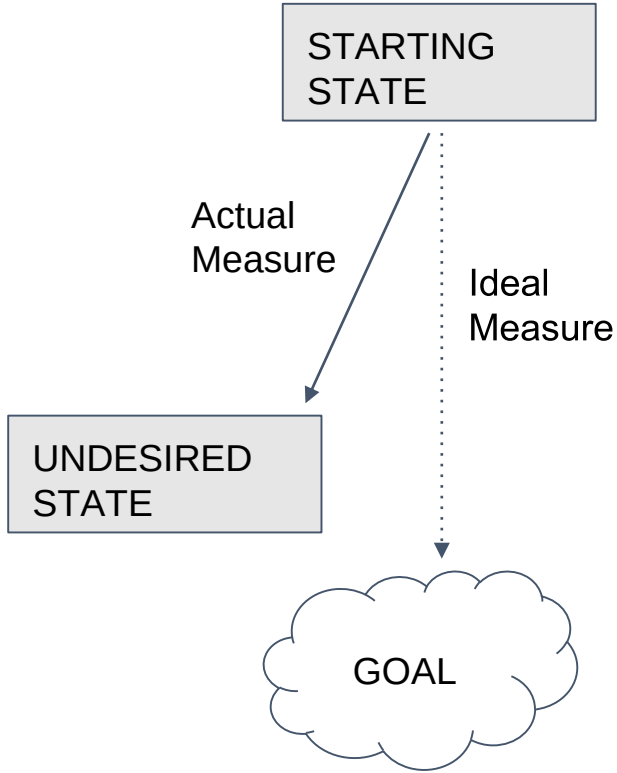
"Any observed statistical regularity will tend to collapse once pressure is placed upon it for control purposes"
(Goodhart, 1984)

Marilyn Strathern's 1997 summary is often quoted:

"When a measure becomes a target, it ceases to be a good measure"
(Rodamar, 2018)

What is Goodhart's Law?

Many researchers have contributed to this idea and variants appear including Goodhart's Law, Campbell's Law, Lucas Critique, the Cobra Effect, Perverse Incentives, Gaming the System, and McNamara Fallacy.



A Common Example of Goodhart's Law

Goal: reduce the population of venomous snakes

Measure: bounty for dead snakes

Unintended Result: breeding venomous snakes for profit

Where is Goodhart's Law Studied?

Goodhart's Law (and its variants) has predominantly been explored in economics and social science.

Academic Testing (Berliner, 2011)

Healthcare (Poku, 2005)

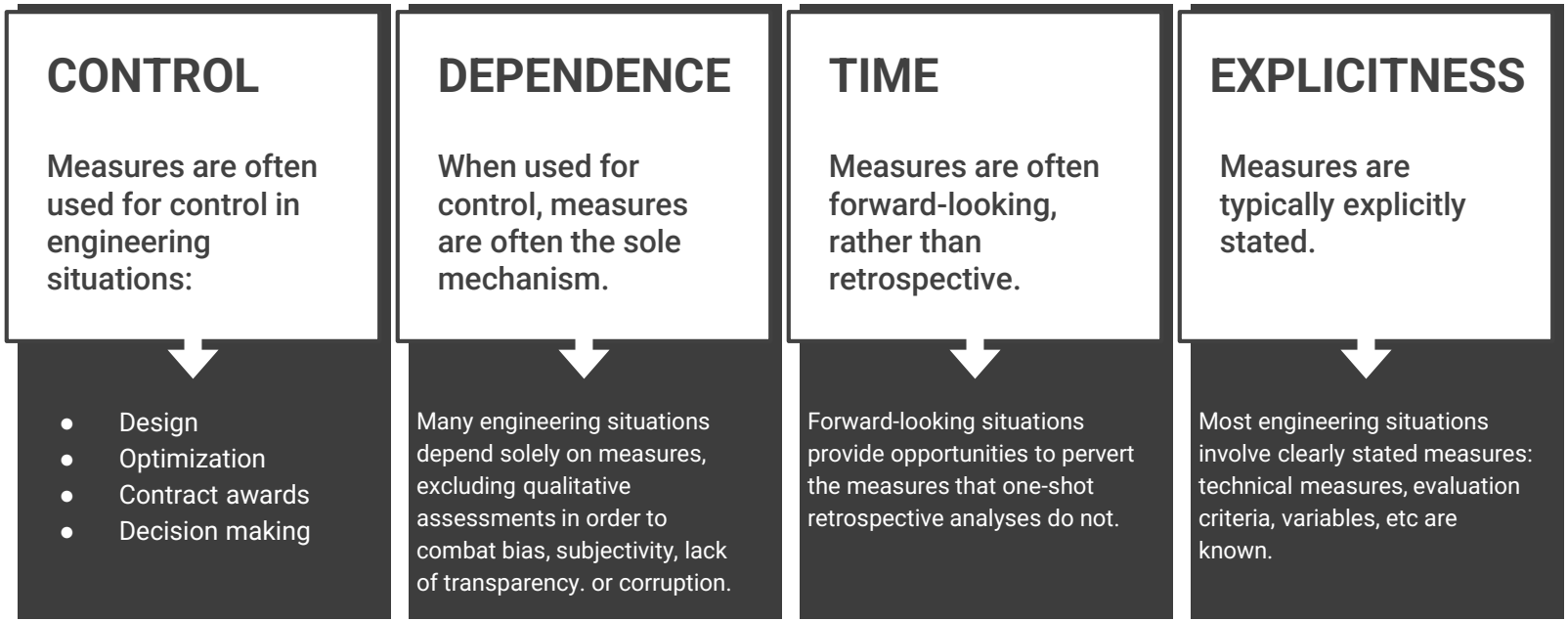
Criminal Justice (Brennan and Surprenant, 2020)

Engineering situations frequently invite the pitfalls of Goodhart's Law.

"The problem appears when we try to use the direct, unmediated data in decision-making"

(Sidorkin, 2016)

Is Goodhart's Law Applicable to Engineering Situations?



Case Studies

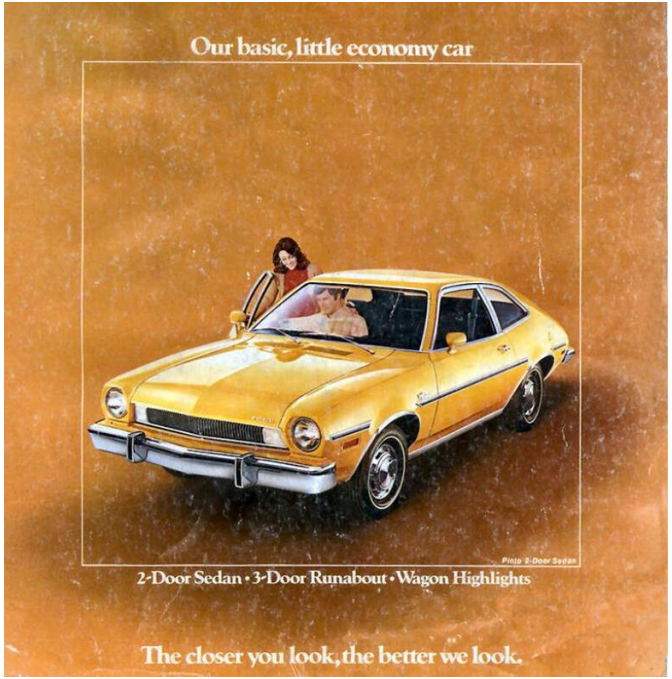
Ford Pinto: Design Decision

Background:

- A design flaw was discovered within Ford before production: low speed rear collisions were prone to fuel tank leaks, and therefore fires (Wojdyla, 2011).

The Design Decision:

- The design flaw was not corrected and the Ford Pinto went to production in 1970.



1973 Ford Pinto Manual

Ford Pinto: The Measures

What Measures Were Used To Make the Design Decision?

REPAIR COST PER CAR - estimated at $11 (Irwin, 2020).

TIME - Delays would be significant if the design flaw was corrected (Irwin, 2020).

1976 Ford Pinto Manual

Ford Pinto: Unintended Results

What Was the Unintended Result?

- A 1977 expose brought attention to the flaw being known by Ford before production. Negative public perception and lawsuits followed.
- In 1978 all 1971-1976 Ford Pintos were recalled.
- Despite claims that the situation was blown out of proportion, Ford spent upwards of $30 million. Ford also would be plagued with a damaged reputation even when cars safety ratings were on par with competitors.: “Fix Or Repair Daily” and “Found On the Road Dead”.

1972 Ford Pinto Manual

Ford Pinto: Just Amoral Decision Makers?

Did Measures Impact the Unintended Results or Did Amoral Decision Makers?

The outcome of the Ford Pinto has been described as:

“a normal outcome of organizational and institutional processes”

(Lee and Ermann, 1999)

Decision makers were using common measures in expected ways. While the decision is now typically seen as amoral, the measures created a situation where an amoral outcome was indicated as better than a moral one.

Ford Pinto: Potential Solutions?

1. *Using Additional Measures in the Design Decision*

   RECALL COSTS - How much would fixing the issue later cost?

   LAWSUIT COSTS - What if Ford was liable?

   COST OF LOSING CUSTOMERS - What if the issue caused bad press?

   INJURIES - safety was not a common consideration (Lee and Ermann, 1999)

1. *Considering Other Mechanisms for the Design Decision*

   ETHICAL CONSIDERATIONS - National U.S. automobile safety standards had just been introduced in 1966. The Ford Pinto case brought attention to the standards, asking if it is unethical to leave a known design flaw if the system still meets the safety standard.

F-111: Design Concept

Background:

- In the early 1960s, the Air Force and Navy each needed to develop new combat aircraft (Bernier, 2018).
- Despite the Air Force and the Navy having very different goals, designers were directed to pursue a single aircraft concept in 1961 (Richey, 2005).

https://www.google.com/url?sa=i&url=https%3A%2F%2Fen.wikipedia.org%2Fwiki%2FGeneral_Dynamics_F-111_Aardvark&psig=AOvVaw3fKv-B1mSaoXWtci5ntzbn&ust=1636693514186000&source=images&cd=vfe&ved=0CAsQjRxqFwoTCKi2jafGj_QCFQAAAAAdAAAAABAD

F-111: The Measures

What Measure was Used To Inform the Design?

MAXIMUM COMMONALITY between Air Force and Navy versions was a primary measure. Other measures were combined, altered, and sacrificed by both parties in order to preserve maximum commonality of over 80% in the airframe, engines, subsystems and avionics in terms of structural weight and parts count (Richey, 2005).

"The "forced commonality" of the Air Force and Navy versions… drove the design" (Richey, 2005).

"McNamara selected the General Dynamics entry, despite strenuous objections from a military selection board that favored a Boeing proposal, mainly because the General Dynamics idea promised that the commonality would provide greater savings." (Bernier, 2018)

F-111: Unintended Results

What was the Unintended Result?

- Cost per system raised from an estimate $3 million to $8 million (Bernier, 2018).
- Despite plans for 1,726 total aircraft only 562 were produced (Richey, 2005).
- Additional systems had to be developed since the F-111 could not achieve all it was planned to (Axe, 2021).

https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.lockheedmartin.com%2Fen-us%2Fnews%2Ffeatures%2Fhistory%2Ff-111.html&psig=AOvVaw0OppYhUUWG1Dci7dH8sqwu&ust=1636681737983000&source=images&cd=vfe&ved=0CAsQjRxqFwoTCOCMi6TFj_QCFQAAAAAdAAAAABAe

F-111: Potential Solutions?

1. *Using More Direct Measures in the Design Decision*

   DESIGN COSTS - while cost reduction was the ultimate *goal* in measuring maximum commonality, was it actually a measure? Could measuring estimated cost directly have resulted in a different design than using maximum commonality as a surrogate measure for cost reduction?

1. *Considering Other Mechanisms for the Design Decision*

   USEFULNESS CONSIDERATIONS - maximum commonality was emphasized, without considering if, beyond that measure, the system produced would be valuable to each party.

   LISTENING TO STAKEHOLDERS - many stakeholders voiced protests to the maximum commonality metric (Richey, 2005).

Takeaways

Capturing unintended consequences in measures is difficult.

Some red flags include:

Unbalanced measures
Single or few measures
Sole use of measures

Acknowledgements and References

Special thanks to Christopher White and David Perner for their direction in identifying case studies.

Axe, D. (n.d.). *The F-111 Was The 'F-35 Of Its Day,' But Its Failure Was A Boon To U.S. Air Power*. Forbes. Retrieved November 10, 2021, from https://www.forbes.com/sites/davidaxe/2021/02/26/the-f-111-was-the-f-35-of-its-day-but-its-failure-was-a-boon-to-us-air-power/

Berliner, D. (2011). Rational responses to high stakes testing: The case of curriculum narrowing and the harm that follows. *Cambridge Journal of Education*, *41*(3), 287–302. https://doi.org/10.1080/0305764X.2011.607151

Bernier, R. (n.d.). *Was the Navy's F-111 Really That Bad?* Air & Space Magazine. Retrieved November 15, 2021, from https://www.airspacemag.com/military-aviation/13_sep2018-cancelled-f111b-1-180969916/

Brennan, J., & Surprenant, C. (2020, July 14). *Perverse Incentives Created Our Terrible Criminal Justice System*. Cato Unbound. https://www.cato-unbound.org/2020/07/14/jason-brennan-chris-w-surprenant/perverse-incentives-created-our-terrible-criminal

Goodhart, C. A. E. (1984). Problems of Monetary Management: The UK Experience. In C. A. E. Goodhart (Ed.), *Monetary Theory and Practice: The UK Experience* (pp. 91–121). Macmillan Education UK. https://doi.org/10.1007/978-1-349-17295-5_4

Lee, M., & Ermann, D. (1999). *Pinto "Madness" as a Flawed Landmark Narrative: An Organizational and Network Analysis*. https://academic.oup.com/socpro/article-abstract/46/1/30/2925335

Poku, M. (2016). Campbell's Law: Implications for health care. *Journal of Health Services Research & Policy*, *21*(2), 137–139. https://doi.org/10.1177/1355819615593772

Richey, G. K. (2005). *F-111 systems engineering case study. Air Force Center for Systems Engineering, Dayton*.

Rodamar, J. (2018). There ought to be a law! Campbell versus Goodhart. *Significance (Oxford, England)*, *15*(6), 9–9. https://doi.org/10.1111/j.1740-9713.2018.01205.x

Sidorkin, A. M. (2016). Campbell's Law and the Ethics of Immensurability. *Studies in Philosophy and Education*, *35*(4), 321–332. https://doi.org/10.1007/s11217-015-9482-3

Irwin, C. (2020). 50 Years Ago: The Pinto Becomes Ford's "Embarrassment." Ultimate Classic Rock. Retrieved November 15, 2021, from https://ultimateclassicrock.com/ford-pinto-history/

Wojdyla, B. (2011, May 20). *Ford Pinto Fuel Tanks: Epic Auto Failures*. Popular Mechanics. https://www.popularmechanics.com/cars/news/industry/top-automotive-engineering-failures-ford-pinto-fuel-tanks

Thank you! Questions or Examples?