

# Experimentation Reading Group: Value Capture

# Agenda

1. Value Capture
2. A Necessary Evil?
3. The Horror?
4. Working with Horror



## Value Capture

**Values are rich and subtle, their measurement is not.**

01

### Values are rich and Subtle

Agents whether individuals or firms adhere to particular protocols of action which imply their values. These values can be rich and subtle

02

### Metrics offer a quantification of Value

Ongoing deliberation over the exact articulation of our values is replaced and out-sourced for schematisation and quantification via available metric-proxies.

03

### Value “capture” conflates metrics and values

Value capture occurs when the metric comes to displace or overshadow the original value.

This results in real harms due to overly-simplified metrics required for large scale coordination sacrificing nuance and context.

## Self-Optimisation: Externally Defined

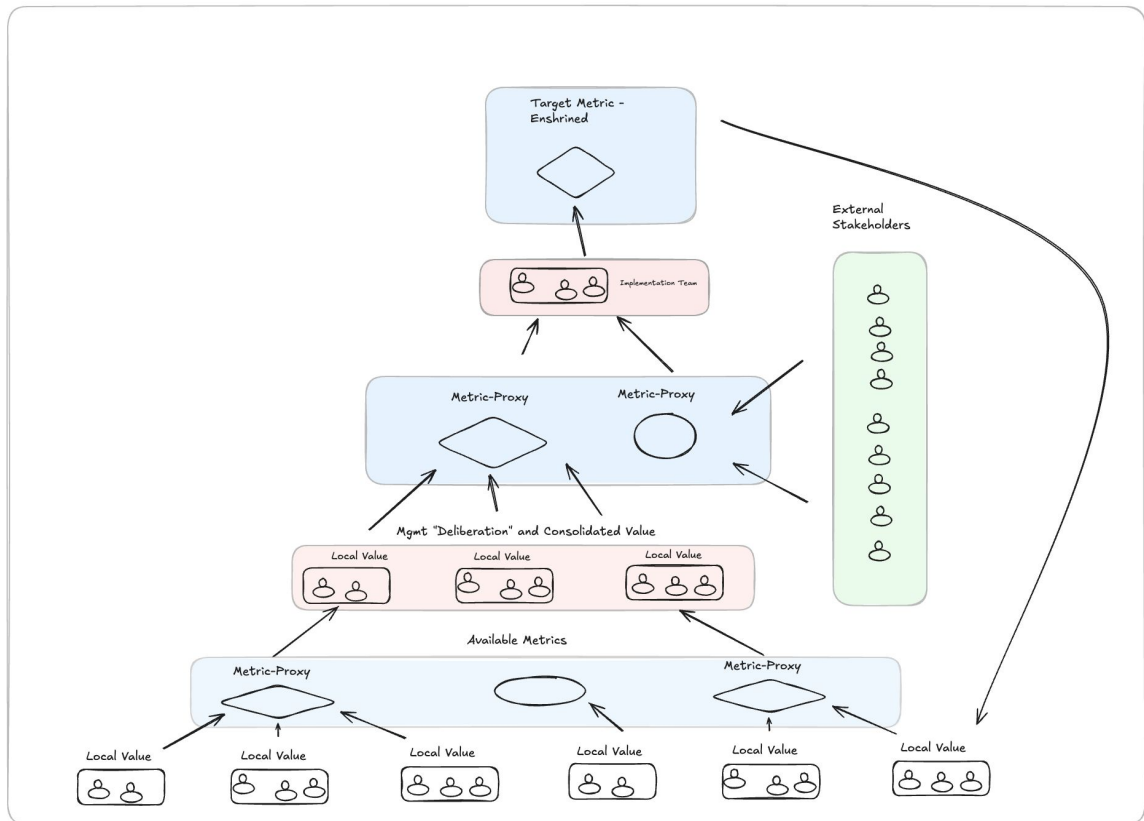


“Value capture happens when a person or group adopts an externally-sourced value as their own, without adapting it to their particular context”

10,000 steps or 10,000 hours of study is an external standard set as an external benchmark irrespective of local needs.

# Metric Capture and Coordination

## Value Capture through Coordination and Metric Proxies



## Value Capture

"In value capture, we outsource the process of value deliberation. And, as with other forms of outsourcing, there is a trade-off...When we adopt those values, we gain access to readymade methods for justification"

## Value Capture: Is it Horrifying?

“Why might this strike some of us as horrifying, rather than as simply a useful and empowering tool? I will suggest that there is a problem with the **nature of the values** on offer.

The **problem with internalizing institutional metrics** isn't simply that we are getting our values from the outside. It is that such metrics are subject to the demand for a certain kind of stability and institutional usability. These **institutional demands pushes our metrics away from the subtle, the dynamic, the sensitive** — and towards what can easily be measured at scale, propagated across institutional units, and recorded in institutional memory.”



The Horror

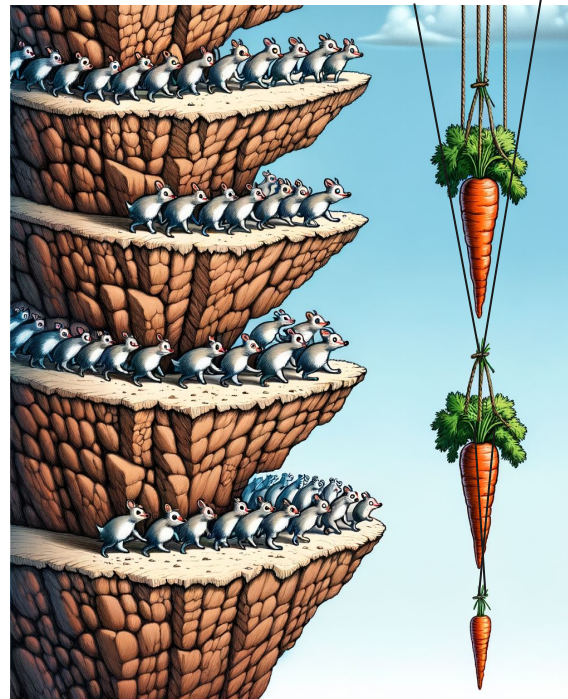
# "The Purpose of a System is what it does"



This Horror would be bad enough...



...but the fear is worse



... being coaxed into pursuing values through proxies handed down from above by actors ignorant of the initiating value and not exposed to the consequences of poor proxies.

Perhaps we're all equally misled...



# Accidental and Deliberate Value Substitution

TECH

## Musk's xAI says Grok's 'white genocide' posts resulted from change that violated 'core values'

PUBLISHED FRI, MAY 16 2025 12:22 AM EDT

Jonathan Vanian

@IN/JONATHAN-VANIAN-B704432/

WATCH LIVE

### KEY POINTS

- Elon Musk's xAI said an "unauthorized modification" to its Grok chatbot led to unprompted responses about "white genocide" in South Africa.
- Numerous X users posted screenshots of Grok offering replies on the topic despite being asked unrelated questions.

### WEALTHY NICKEL

MAKE MONEY ▼ SAVE MONEY ▼ INVEST LIFE HACKS ▼ ABOUT

## 10 Things Wealthy People Do Every Day (And You Should Too)

👤 Andrew Herrig : 📁 Manage Money, Featured : 📅 02/12/2021 : 🔗 AFFILIATE LINKS



“Value capture occurs when an externally-sourced value plays the dominant role in practical reason – when it gets put in charge, for some domain.”

Substitution of metrics for deliberation on particular values is an appeal for efficiency. Not all such appeals are bad.

But when we out-source the instrumentalization of value to 3rd parties, **we are making an efficiency/nuance tradeoff and “selecting on observables”** - a common fallacy.

Such trade-offs may be legitimate in low-stakes endeavours but the increasing encroachment of software aided decision-tools in our day-to-day makes the the weighing of these tradeoffs a more acute problem.

# Exploration and Refinement



“This is a process of exploration, where you try out things out, figuring out how they fit with you, and changing around your approach in response, seeing how it goes in an ongoing loop of feedback and Adjustment.”

**Business Strategy is dynamic, exploratory but the value-encoding of the business targets is generally not.**

Failure to fit metrics to the goals is a management failure.

# The Tailoring Argument

“In institutional value capture, we do not adjust our values in light of our particular experiences. We take values as provide by some large-scale institution, and live under them as given. Those values will have been formulated to take deeply into account various institutional interests: like the ability to be counted in a reliable way across a large institution, and the ability to be readily aggregated in an institutional bureaucracy. **They will not have been formulated in light of the rich feedback of how our particular lives have gone when we live under these values**”

Nguyen phrases this as a problem of individual value (a nicety), but i think local value is the better phrasing. It suggests the failure is one of insensitivity to relevant detail rather than a failing of consideration.



# The Usefulness of Standards

“Perhaps most importantly it standardizes a product for future consumption on a market. Standardized grades make possible standardized educational certificates, which are extremely useful for potential employers. It was administrators and employers who “placed a premium on readily interpretable and necessarily abstract grading systems”. Qualitative evaluations of student might be nuanced and context-sensitive — but they are **illegible to the large-scale administrative institution.**”

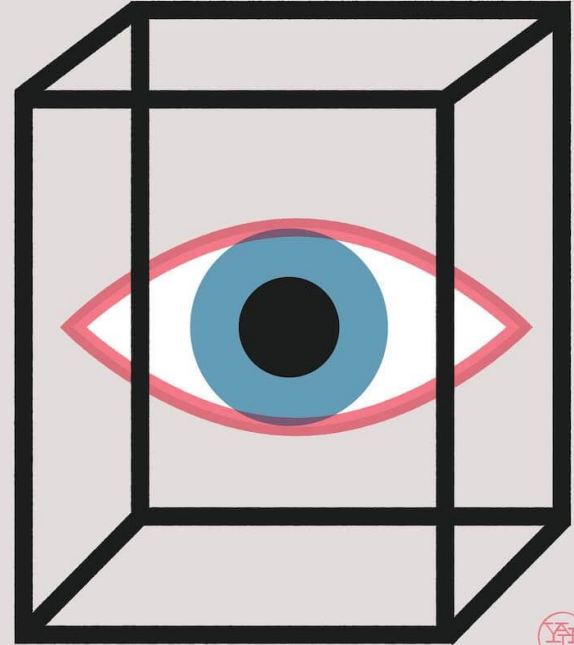
Failure to instrumentalize values through appropriate proxy metrics hinders communication, cohesion and cooperation.

Crude instrumentalization turns work from a compelling collective action project into a metric movement charade.

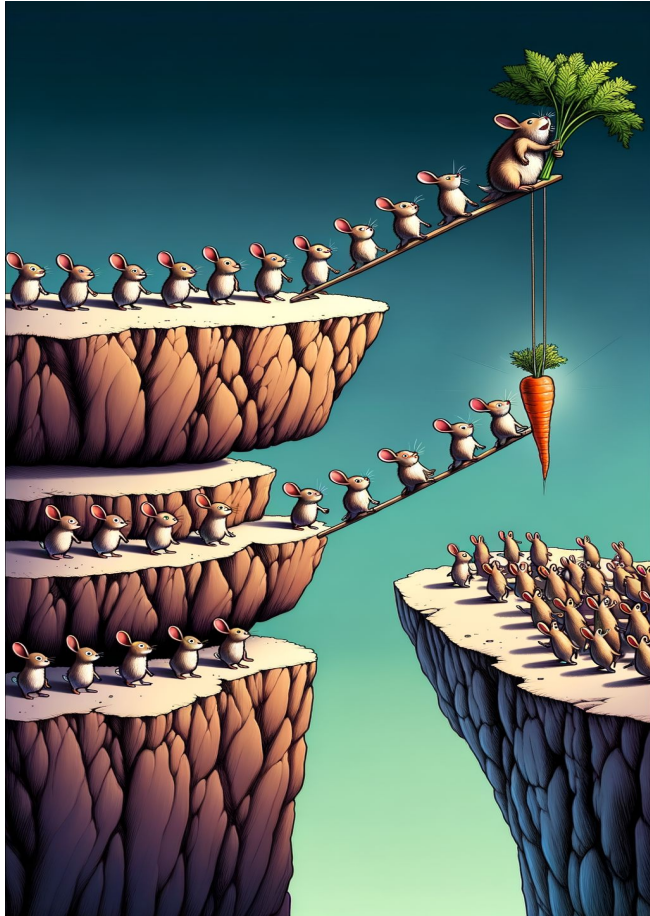
## Seeing Like a State

How Certain Schemes to Improve the Human Condition Have Failed

James C. Scott



# Working with the Horror



## The Puzzle:

**P1:** There is a reasonable case to be made that value capture is a horrifying fate that we would seek to avoid

**P2:** There is a reasonable case to be made that some degree of standardisation is required for cohesive collective action.

**Q1:** How does an effective mgmt team arrange circumstances so that the worst alienating effects of value capture can be mitigated while cohesion is maintained.



# Standardisation within Operating Scales

**The Good:** Coordination within layers of a hierarchy can facilitate cooperation and cohesion.

**The Bad:** The “Horror” of Value Capture stems from the reflexive internalization of values imposed between rungs on the hierarchy.

With externally imposed value we lack the context to “see” the rationale of the imposed standard and when locally applied they engender side-effects with unappreciated consequences. This cements the alienation from the project of the externally imposed values and the suspicion of managerial incompetence and detachment.

**The Implication:** A reflective management team will be sensitive to consequences of misapplied metrics to alleviate “the horror”, promote cohesion and convey competence. **The way management instrumentalizes value, matters.** Poorly chosen metrics, implies limited understanding of value, breeds bad decisions, and alienation from work.

