

Surrogation

Table of Contents

- [1.](#) Introduction
- [2.](#) Behavioral vs. introjective surrogation
- [3.](#) Quantitative surrogation: The problem of value capture
 - [3.1.](#) Complexity and compression
 - [3.2.](#) Behavioral surrogation: Motivated interpretation & massaged maps
 - [3.3.](#) Introjective surrogation: Value capture, value clarity
- [4.](#) Surrogation and the crisis in psychology
 - [4.1.](#) Surrogation and the replication crisis
 - [4.2.](#) Surrogation and the generalizability crisis
 - [4.2.1.](#) Appearance-optimization as a cargocult
 - [4.2.2.](#) The sociology of surrogation: Why can't honest actors recover corrupted systems?
- [5.](#) Qualitative surrogation: The tragedy of appearances
 - [5.1.](#) Dylan at Newport, 1965: Artistic authenticity
 - [5.2.](#) Lessons from glamrock
 - [5.3.](#) "Creative" cargocults
- [6.](#) Mitigations
 - [6.1.](#) Supplementation vs. surrogation
 - [6.1.1.](#) Developing an aesthetic
 - [6.1.2.](#) Tracking intuitions
 - [6.2.](#) Minimizing measurement surrogation in Goodhart-Campbell
- [7.](#) Conclusion
- [8.](#) Works Cited
- [9.](#) Footnotes

Spirit takes many forms—rarely can we establish its exact throughlines; we may recognize when we see it, but only in person, in the fullness of its situation.

Letter—the specification of spirit—can attempt to capture some of the shapes and guises in which spirit manifests. Though this translation process cannot succeed in full, spirit on its own cannot be legislated, cannot be uniformly instituted as expectation, cannot tile itself across a superorganism.^[1] Insofar as it can, it is hard, and unaccountable, relies on trust and discretion, defies monitor. So we are stuck with letter.

Insofar as letter invariably fails to capture spirit, each rounding off and uncovered patch of ground, each poor synopsis or compression, shifts the behavior of the organization or individual away from their desired spirit, toward some other, emergent, perverse nonintention. Often in this process, the substitution of letter for spirit is forgotten—there is a slow forgetting that something complex and extraverbal has been surrogated into language, into

And each venture

Is a new beginning, a raid on the inarticulate,

With shabby equipment always deteriorating

In the general mess of imprecision of feeling,

Undisciplined squads of emotion. And what there is to

By strength and submission, has already been discover

Once or twice, or several times, by men whom one can

To emulate—but there is no competition—

There is only the fight to recover what has been lost

And found and lost again and again

"Four Quartets" (Eliot)

measurement, into specification, and the surrogate comes to stand in as *spirit itself*.

In the story of King Midas, Dionysus promises the king any wish he desires, and Midas wishes for all he touches to turn to gold. We know what he “really” was after, but did he even know, precisely, what he wanted when he wished it? Had he specified its parameters and constraints, or assumed a human audience would infer his spirit? Unfortunately, the total space of intelligence is much broader than the space of human minds: Midas met only by a trickster god, a spirit that needs no translating, and his sloppy presentation of spirit in letters became his end.

1. Introduction

Goodhart’s Law, in the (re)phrasing of anthropologist Marilyn Strathern: “When a measure becomes a target, it ceases to be a good measure.” Since we expect humans to behave roughly rationally toward self-interest, and as the incentive structure of an activity *determines* self-interest, the surrogate measure distorts behavior in the direction of optimizing for the *surrogate*, at cost to the surrogated which is proportional to the divergence *between* surrogate and surrogated.^[2]

We will call this general mechanism—where a *representation* of a holistic target generates its own gravitational field, and in some meaningful way *replaces* that original whole—“surrogation.” Choi, Hecht, and Tayler in the early 2010s published papers in management accounting theory (2011, 2012) proposing that managers exhibited a pattern of losing sight of their original strategic target in favor of an instituted proxy measure set up to represent it. The authors are potentially the first to use the term “surrogation” in this context, and they include in their definition the psychological “amnesia” of managers—but it is well-chosen as an umbrella handle for a broader phenomenon comprising multiple theoretic carvings. In management studies, that of Choi, Hecht, and Taylor. In the social sciences, that of Goodhart as well Law and that of Donald Campbell (i.e. “Campbell’s Law”) stating that “[t]he more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor.” (In the context of policing, specifically, Campbell has accused the Nixon administration’s crackdown on crime as having “as its main effect the corruption of crime-rate indicators, achieved through underrecording and downgrading the crimes to less serious offenses.”) In philosophy, we see C. Thi Nguyen’s theory of value capture as outlined in 2020’s *Games and the Art of Agency* (and presented in the larger context of *gamification*, a process similar to surrogation). Nguyen defines *value capture* as the substitution of a simplified metric, or indicator, for a richer holistic value, distinguishing it from Goodhart’s Law in that the substitution is *internalized* by the agents situated within the surrogate incentive structure. There is not just a change in the agents’ behavior but in their actually held values.

The surrogate by definition is chosen because it is somehow easier or more tractable than the surrogated “real” or “original” destination. That previous destination may have been hidden from sight; it may have been too difficult to compare or rank among instances; it may simply have been costlier to track in time or money. A proxy is a surrogate; so is a signal, a metric, a marker, and a representation.

Surrogation is linked in meaningful ways to the concepts of “economic thinking” and “commodification,” where the formalization, compression, or technical specification of a vague or humanistic value is “lossy,” i.e. cannot meaningfully capture the whole. And it is the inverse of *reification*—where features of the map are mistakenly treated as properties of the territory—though sometimes, introjective surrogation ends in what is effectively reification—the two phenomena “meet on the other side.”

But surrogation is not a problem or phenomenon unique to metrics. Setting reality to words is always the first surrogation: we reify our concepts, confuse them with nature. Were we to avoid quantitative analysis, we would not avoid surrogation. Language, like our internal drives and desires, is always vague—our goals are always underspecified, and our words are always unstable and underdefined. One way to put *surrogation* in simple English, losing the unnecessary if common statistical bent, is to say that some simplified marker of *appearance*, eligible for its co-occurrence with an intractable or invisible *reality*, becomes in some way a substitute *for* that reality.^[3] In the language of signaling theory, the sign is reified in place of the hidden, signaled quality. Markers in fashion—an everyday, material instantiation of signaling theory—are famously contextual. The same piece of clothing can signify very different aspects of its wearer depending on the larger inferred complex of intentionality, knowingness, and providence in which its display (is inferred to) originate. A “designer” brand like Lacoste, on its own and out of context, is interpreted as a sign of white wealth only by the excessively naive; like the Silicon Valley hoodie uniform, it is a contextual move which is able to signify only against an understood landscape of signification; it can *only* be considered in context.^[4]

Much related, a [cargocult](#) is the confusion of surface details and instantiation-specific components for substantive or functionally necessary aspects (Feynman 1974, Reason 2016). “The cargoculter builds a motorless airplane from palm fronds, sprinkles it with holy water, and prays to the gods for it to fly” (Reason 2016). Typically the confusion is born of a lack of deep systems understanding of the target domain. As a result, a cargocult imitates superficial and aesthetic elements (markers or ritual indicators) in expectation of their efforts reproducing the operation of the original. These confusions may be attributed in part to confusion over the direction of causality, and the role of the components in the enveloping system. A system which surrogates non-causal attributes, and especially the surface products of deeper causes, can be considered a cargocult.

Kahneman and Tversky have also theorized *attribution substitution*, in which an agent tasked with a difficult question may resort—unwittingly—to answering a related, proxying but distinct question that is easier to answer. Perhaps most famous as an example is the bat and ball cognitive reflection test, where subjects are asked to calculate the cost of a ball, given that the bat and ball together cost \$1.10, and the bat costs \$1 more than the ball: respondents appear to most immediately answer that the ball costs ten cents, which the behavioral economists speculate is the result of subjects substituting the real task for the task of merely parsing large and small quantities (e.g. as a fast-and-frugal heuristic for making financial decisions). Whether subjects come to the correct response is largely reliant on whether they use System 2 thinking to monitor and correct their System 1 intuition. While analogous in underlying mechanism to *surrogation*, attribution substitution in Kahneman’s factoring is performed quickly and unconsciously, rather than at the level of conscious institutional or individual structuring. Still, we can think of them as similar in kind.

Finally, in artificial intelligence research (which frequently mobilizes Goodhart's Law), there are the concepts of *underspecification* and *nearest unblocked strategy* (Manheim 2019, Arbital 2020). Specifying a telos in code proves a hard task: any behaviors not explicitly prohibited may be exploited; incentives turn perverse; roadblocks prove insufficient and—like Midas—the goal literal turns out not to be the goal actual.

In his December 2018 article on the origins of Goodhart's and Campbell's laws, Jeff Rodamar makes the case that differences in use of the different terms, field to field, "harms communication, creates barriers to science, and hinders improvements in practice." This is to say nothing of the many other similar concepts, enumerated above, in management accounting, behavioral economics, artificial intelligence, and the philosophy of games.

2. Behavioral vs. introjective surrogation

Roughly, there are two kinds, or possible stages, of surrogation: the alteration of behavior towards a surrogate, and/or the psychological internalization of the surrogate—the reification of the surrogate *as if* it were the thing itself; an amnesia surrounding the switch.

Perhaps the major problem of surrogation is that it alters and corrupts human behaviors, moving their telos away from the originally desired behavior and into those which, while rational at an individual level (as exploitations of the incentive structure), are inefficient at the level of institution and society. Moreover, in cases in which individuals are aware that their incentivized behavior diverges from pro-social goals, the activity loses meaning and the individuals become "cynical"—they are aware of the performative aspects of their activity. (See, by way of example, Michael Inzlicht's reflections on his disillusionment with social psychology [2016].) In selectively rewarding individuals or institutions who optimize *away* from the real target and toward the instituted surrogate, it (1) discourages play interested in the real target, (2) discourages players interested in the real target, who engage in exit from the game, (3) increasingly promotes and advances individuals or institutions who optimize (undesirably) to the surrogate at the cost of the real target.

To understand just how *perverse* (cf. "perverse incentives") surrogate incentives can be, we can look to Robert Jackall's sociological study of institutional ethics, *Moral Mazes*:

[A]t Covenant Corporation the story is told about a plant that produced a useful by-product at no extra cost. One simply had to store it until it was needed for other internal operations. Covenant, however, works with an accounting system that considers by-products as inventory; moreover, inventory counts against one at the end of a fiscal year. In order to cut costs, managers decided to throw out the by-product at the end of a financial cycle. But a sudden shortage of the material trebled its cost two months later. To service their own operations, managers had to go hat in hand to their competitors to buy the material at the premium prices.

The introjective kind of surrogation is performed (or "happens to") not just those are expected themselves to optimize toward the surrogate—as in the

behavioral kind—but also those who have constructed the incentive structure, who pass down the surrogate and, in institutional contexts, may even have designed it. In more decentralized social settings, subject to cultural inheritance and the ongoing, distributed negotiation of norms—realms where we “are organized beings, but are not the authors of our organization” (paraphrasing Noë 2015)—values, preferences, and norms are more amorphous, enacted but often only half-consciously known; their slow replacement by a surrogate can occur without intentionality or conscious recognition.

Goodhart and Campbell’s Law approach surrogation *from the perspective of the system*. Nguyen’s intervention, with his concept of “value capture,” is to shift the perspective away from the system and toward the individual inhabiting it.

3. Quantitative surrogation: The problem of value capture

By setting a metric as a target, and by linking that target to a reward structure, we create an incentive for that metric to be gamed in some way. (Rodamar 2018)

The central problem in any superorganism or institution is that of aligning values between members so as to coordinate action toward a shared purpose; this dilemma is known as the principal-agent problem. Solving principal-agent problems requires preferential treatment—rewards or punishments doled out on the basis of performance (fixed-rate salary with bonuses is a classic example of financial incentive—though prestige and reputational incentives have proved efficacious on their own).

Broadly speaking there are three main advantages to instituting measurement across these systems, David Manheim writes in his essay series on measurement for *Ribbonfarm*: “[It] replaces intuition, which is often fallible. It replaces trust, which is often misplaced. [And it] finesses complexity, which is frequently irreducible”—where *irreducible* entails *intractable* ^[5] (2016a, 2016b). In other words, the organization wishes to *supervise* and *monitor* the behavior of its subagents, to ensure honest and high quality performance. Complexity must somehow be reduced to a synopsis, or indicator, in order to effectively evaluate performance. Additionally, a desire to make the basis of this preference consistent across the organization, and transparent and legible for involved parties—as is frequently expected in a society that values equal opportunity—involves instituting public bases for advancement. In some arenas, performance plays out in a way that is easily quantitatively tractable, such as sales figures, but where there is divergence—where a number or statistic is preferred over “the real deal”—the publicity makes the surrogate gameable, and rational self-interest adjusts its targets accordingly. Our accounting of the motivations for quantitative surrogation thus includes not only the replacement of intuition and the reduction of complexity, but the production of legible, transparent, consistent, fair, and objective-seeming bases to ensure better management of the organization, and (contiguously) the principal-agent solving preferential treatment of organization members. Nguyen, in *Games and the Art of Agency*, illustrates how the reduction of qualitative values to quantitative metrics ensures the

“units” or demonitator of evaluation are consistent across both time and space:

large-scale institutions often need quantified measures of their various functionings for management purposes. High-level administrators in large institutions need to be able to compare, say, productivity, customer satisfaction, and worker satisfaction across various departments. This requires quantified representations of values. An administrator might first need to aggregate productivity numbers across different departments in, say, their Tokyo and Los Angeles locations, or aggregate productivity numbers from all locations to compare institutional productivity over years.

3.1. Complexity and compression

What this involves, necessarily, is the reduction or summarization of complex and fuzzy realities.^[6] Manheim:

Complex systems have complex problems that need to be solved. Measures can summarize, but they don't reduce the complexity. This means that measures hide problems, or create them, instead of solving them. This concept is related to imposed legibility, but we need to clarify how in a bit more detail than the 'recipe for failure' discussed in the linked piece. In place of that recipe, I suggest another triad to explain how complexity is hidden and legibility is imposed by metrics, leading to Goodhart's law failures. These failures are especially probable when dimensionality is reduced, causation is not clarified, and the reification of metrics into goals promotes misunderstanding. (2016a)

In the language of computation, the surrogate *lossily compresses* some complex whole, and bears inverse fidelity to its divergence from said whole, and conflates many possible worlds into a single measure. For instance, one's appearance as a student with a GPA of 3.15 may mask two very different realities—on the one hand, a straight-B student; on the other hand, an enormously talented physicist who flunked his compulsory literature course. The Australian counterinsurgency expert David Kilcullen writes in “Measuring Progress in Afghanistan”:

Violence tends to be high in contested areas and low in government-controlled areas. But it is also low in enemy-controlled areas, so that a low level of violence indicates that someone is fully in control of a district but does not tell us who.

When quantitative metrics obscure meaningful valence differences in the compressed whole, selection based on these metrics produces disastrous, counterproductive results which are difficult to monitor precisely *because* the disaster is invisible to the system of monitoring. Historian Muller, in *The Tyranny of Metrics*, describes the downfall of simple counting:

From [commanding officers'] point of view—and from the point of view of the politicians to whom they reported—every arrest was of the same value. The course of action that produced the

best performance indicators did little to diminish the sale of narcotics. (2018)

We can adapt the excerpt for psychology—whose use of surrogation, as we will soon see, has led to its generalizability and replication crises:

From the point of view of researchers' hiring committees and grant foundations, every publication was of the same value after controlling for the prestige of the journal. The course of action that produced the best performance indicators did little to advance the discipline's larger project of quality research.

3.2. Behavioral surrogation: Motivated interpretation & massaged maps

While quantitative surrogates are often instituted to provide an objective oversight (to prevent being fooled by an employee's "spun" self-representation, for instance), in practice, when the interpretation of reality as statistics—the choice of how to compress that reality—is left to the monitored agents, statistics are easily "massaged." Muller reports:

In 2014, a whistle-blower from the London police force told a parliamentary committee that massaging statistics had become "an ingrained part of policing culture": serious crimes such as robbery were downgraded to "theft snatch," and rapes were often underreported so as to hit performance targets. As a retired detective chief superintendent put it, "When targets are set by offices such as the Mayor's Office for Policing and Crime, what they think they are asking for are 20% fewer victims. That translates into 'record 20% fewer crimes' as far as... senior officers are concerned." Such underreporting and downgrading of crimes "are common knowledge at every level in every police force within England and Wales," he added. (2018)

Muller quotes a Chicago detective on the ease of "juiking the stats" (i.e., orienting "the activity of the department toward seemingly impressive outcomes"):

"It's so easy [to massage figures]." First, the responding officer can intentionally misclassify a case or alter the narrative to record a lesser charge. A house break-in becomes "trespassing"; a garage break-in becomes "criminal damage to property"; a theft becomes "lost property."

This massaging occurs at all levels, so long as the information is being passed upward in command, i.e. to the agents responsible for doling out preferential treatment. Since in many cases even the highest-ranking members of an organization are responsible to shareholders or a public, massaging occurs at all levels. And while superiors often prefer accurate over massaged information in order to make better strategic decisions and fill out the organization with competent workers, in some situations, statistics massaged at lower levels may be preferable or even knowingly demanded, since "keeping their hands clean" in this way allows higher-ups plausible deniability in passing their own claims forward.

3.3. Introjective surrogation: Value capture, value clarity

Games extract pleasure from what C. Thi Nguyen, in 2020's *Games and the Art of Agency*, calls “value clarity”:

Life is a confusing welter of subtle values, in a vast and confusing plurality. Living our lives, as fully sensitive valuing agents, involves making painful judgments, tough decision calls, and agonizing comparisons.

In game life, our temporary agency's values are usually extremely clear. That clarity is encoded into a game's specification of its goals. The values we take on in games are clearer, easier to apply, and easier to evaluate than our enduring values.

Game play, in other words, involves an “all-consumingly instrumental mode of practical reasoning.” The legibility, meanwhile, allows public ranking, encourages improvements in productivity and performance by establishing common knowledge of relative performance, fostering competition among members.

The appeal of value clarity can lead human superorganisms into what Nguyen calls *accidental gamification*, where game-like features—such as clear metrics, often introduced top-down with the explicit aim of motivating employees through public competition:

[A]cademic life has recently come to be ruled by quantified metrics for research quality—like citation rates and impact factors. These metrics may not have explicitly been designed to produce gamification among researchers. Conceivably, they arose from the bureaucratic need to collate information, or in university administrators' quest to make more object-sounding decisions about faculty hiring and promotion. But the clear, simple, and quantified nature of such metrics can foster game-like motivation... We could be drawn to redefine our notion of success in the newly clear terms specified by those metrics. (2020)

The gamification of academia, science, and the “global knowledge game” is discussed in the following section, “Surrogation and the crisis in psychology.”

Value capture occurs when:

1. Our values are, at first, rich and subtle.
2. We encounter simplified (often quantified) versions of those values.
3. Those simplified versions take the place of our richer values in our reasoning and motivation.
4. Our lives get worse.

In “simplifying the specification of the target” we end up pursuing, “with ever more fervor and ferocity, the wrong target.” Often, by the laws of complexity—that is, the inevitability of perverse incentives—surrogated efforts even make the situation worse than passivity—this being part of the case made by Michael Huemer in his defenses of policy passivity (2012).

by simplifying the specification of the target, we may bring ourselves to pursue, with ever more fervor and ferocity, the wrong target.

Such measures are useful, but we must always recall that they are merely abbreviations—usefully portable simplifications of something larger and subtler. But when our values are captured, we are motivationally caught by a simplified measure.

4. Surrogation and the crisis in psychology

See also entry [Against Psychology](#).

As we have seen, surrogation permeates distributed human projects, or “superorganisms”—institutions like the military, police department; the medical or justice system; diplomacy and trade; but also what Sarah Perry (2020) dubs the “global knowledge game”—the ongoing process of attempting discovery of global truths spanning scientific and, to a lesser extent, humanities work. In other words, a lifting of knowledge *out of context* and into some human (aspirational-)universal or generalization.

There are numerous surrogation-caused problems in the global knowledge game (GKG). Because the GKG has become a vast enterprise characterized by information overload—by the simultaneous production of millions of members—and because there is a vast, distributed incentive structure designed to reward certain behaviors ostensibly in the service of knowledge production, we should expect it to have the same institutional issues of stats-gaming (e.g. p-hacking) already discussed with respect to the police and military. (Moreover, surrogation is common across knowledge-oriented fields, such as education, where in the United States we’ve seen controversies over “teaching to the test” as well as more blatantly corrupt Goodhartian actions such as teachers manually altering students’ Scantron forms.)

Social psychologist Pamela Smith in an interview on the academic psychology podcast *Two Psychologists Four Beers*, discusses the compressive, context-stripping, quantitative surrogation of quality work into publication statistics:^[7]

We are still rewarding people based on publications. It is true now more than ever, if you want a publication, maybe people are paying more attention to sample, but they’re very happy to let people do online studies that don’t necessarily map well onto behavior, and just run a bunch of them. You get penalized if you want to do careful work. You get penalized if you want to do work on people other than college undergraduates or people who are willing to do online surveys for fifty cents a shot.

Additionally, in the “inexact sciences”—that is, those which are attempting to mature past their qualitative roots and into a more quantitative or empirical science, for instance psychology’s abandonment of phenomenology and psychoanalysis in favor of statistical lab studies—there is a problem of wanting to grow up too fast. In their rush to “objectify” and rigorize themselves, many of the social sciences have hastily abandoned old methods, replacing them entirely with a more performatively “scientific” surrogate.

Here, I'll use Tal Yarkoni's recent assault on social psychology, "The generalizability crisis" (2019), as a launching pad to discuss the phenomenology or psychology *of* surrogation, as well as some of the sociological reasons that institutions deep in surrogated divergence (i.e., away from the "real" target) are so difficult to correct.

4.1. Surrogation and the replication crisis

[TK: Gigerenzer on the "surrogate idol" of a universal method; p-hacking and gamification of paper submission; the incentive structure that discouraged replication in the first place. Certain markers, e.g. p-values, surrogate critical thinking about generalizability and significance, and instead, statistical analysis is performed automatically in a ritual against the uncertainties of complexity.]

4.2. Surrogation and the generalizability crisis

The broad argument Yarkoni advances is that psychology studies' ability to generalize—for the narrow bounds of a lab study done with "just one video, one target face, and one set of foils" to provide evidence for the existence of some broad psychological construct like ego depletion—is orders of magnitudes lower than traditionally assumed in the field. Yarkoni's critiques are not new—as he himself notes, many thinkers in the inexact sciences have been raising the alarm on similar issues, including Gerd Gigerenzer and Paul Meehry, in some cases for upwards of half a century—but they compile and make sense of the scope of the problem social psychology faces.

First, a psychological construct, in order to gather evidence as to its "existence" or "nonexistence"—and even here there is a whiff of conceptual confusion—must be operationalized:

things like cognitive dissonance, language acquisition, and working memory capacity—cannot be directly measured with an acceptable level of objectivity and precision. What *can* be measured objectively and precisely are operationalizations of those constructs—for example, a performance score on a particular digit span task, or the number of English words an infant has learned by age 3. Trading vague verbal assertions for concrete measures and manipulations is what enables researchers to draw precise, objective quantitative inferences; however, the same move also introduces new points of potential failure, because the validity of the original verbal assertion now depends not only on what happens to be true about the world itself, but also on the degree to which the chosen proxy measures successfully capture the constructs of interest—what psychometricians term construct validity.

Yarkoni himself has characterized the surrogative aspects of operationalization: the validity of any findings depend, post-operationalization, on "the degree to which the chosen proxy measures successfully capture the constructs of interest."

Once the study is completed, a second stage follows: the discovered quantitative or operationalized reality is re-translated back into language via generalization or loose induction. The coarse metrics to some extent "disappear," as we re-enter the realm of language where knowledge is hosted

and decisions made. The context is further stripped as the narrow lab finding is “generalized” into a larger claim about human behavior: “Papers should be given titles like ‘Transient manipulation of self-reported anger influences small hypothetical charitable donations,’ and not ones like ‘Hot head, warm heart: Anger increases economic charity.’,” Yarkoni writes. (This behavior is an example of the [Extend and Retreat](#) strategy, or “style of being,” in public life, as seen elsewhere in *motte and bailey* arguments and Daniel Dennet’s concept of “deepities.”)

4.2.1. Appearance-optimization as a cargocult



Pictured: A cargocult with a wooden, full-scale model airplane.

Recall that to *cargocult* is to imitate a work’s surface structures while lacking a proper understanding of the actual mechanisms behind its power. This kind of behavior can be either opportunistic and knowing, putting on a show of appearances for others—as in the cult leader, cynic, or grifter—or else merely a kind of magical thinking and wish fulfillment: “The cargocult builds a motorless airplane from palm fronds, sprinkles it with holy water, and prays to the gods for it to fly.” The psychologist builds up all the meticulous appearances of real science, and prays that his findings contribute to human knowledge. What’s more, since we live in a society that unwittingly or uncaringly surrogates appearance for reality in decision-making and evaluation—in other words, an *optikratic* society^[8] that lives and dies by appearances—these performances frequently *do* succeed in “flying,” perpetuating the optikratic incentive structure.

Yarkoni himself uses the phrase “cargocult science” to refer to the performative aspects of empiricism in psychology, and its concurrent optimization of metrics à la p-hacking:

It’s hard to think of a better name for this kind of behavior than what Feynman famously dubbed *cargo cult science* (Feynman, 1974)—an obsessive concern with the superficial form of a scientific activity rather than its substantive empirical and logical content.

Here, the “superficial” stands as the actually-incentivized surrogate, and the “substantive” the surrogated destination which organizations and players in the global knowledge game self-purport to navigate toward.

Ironically, it may be the case that the inexact sciences, rather than abandoning qualitative research, have merely cloaked it in the grand rhetoric of empiricism; Yarkoni himself contends that, since generalization and inference

cannot be statistically defended, researchers' de facto work involves "a good deal of what currently passes for empirical psychology is already best understood as insightful qualitative analysis dressed up as shoddy quantitative science." In the process, "enormous resources" are expended on this "window dressing" or rhetoric, but what it has purchased for psychology is an enormous amount of public influence and disciplinary prestige.

4.2.2. The sociology of surrogation: Why can't honest actors recover corrupted systems?

Researchers, Yarkoni writes, are driven in psychology and related fields "to expend enormous resources on studies that are likely to have very little informational value even in cases where results can be consistently replicated." Statistically and inferentially unfounded claims are passed up, from psychology, to the highest levels of public and private decision-making, altering the behavior of governments, corporations, and public institutions alike (see the recent *amicus curae* brief on stereotype threat), in large part because this performance of empiricism is highly effective in lending legitimacy to psychological hypotheses. Books are published, and become bestsellers, or talks given that go viral, by psychologists who lead the public to claims and generalities that their studies do not support. There is widespread abuse and gamification of statistics of legitimization, the most well-known being p-hacking. Yarkoni delivers the takeaway of his devastating piece with a quotation from Paul Meehl:

I think that for most faculty in soft psychology the full acceptance of my line of thought would involve a painful realization that one has achieved some notoriety, tenure, economic security and the like by engaging, to speak bluntly, in a bunch of nothing.

The paper presents a number of "next steps," given the outlined state of affairs, but they are addressed to individuals: leave the field, practice slower science, present one's findings more modestly. As a result, they miss the crucial sociological situation which enables and perpetuates these problems. There are game-theoretic forces at play here, and the structure of incentives (i.e. "reward function") in which the problematic behavior originates is not much altered by individual decision-making.^[9]

First, it must be established, however obvious, that making more modest epistemological claims would come at the loss of power, prestige, and reputation. Not only would the field as a whole cede much of its previously claimed credibility, but that credibility would ostensibly drop even further on the basis of the prior deception. It would arguably take quite some time for the field's place in public discourse to recover—this may be to the benefit of society, but would come at an immense cost for stakeholders in the field, thus no field-wide admission of impotence is unlikely.

Since no such coordinated admission or exit will happen, a second problem emerges which is already occurring in the discipline, at least insofar as graduate students high in integrity are turned off from psychology's performative and surrogative aspects.^[10] An individual psychologist who leaves the field, or ceases to advise public policy, or ceases to make grand claims on-stage, will inevitably be replaced by those willing to. Those who replace will on average have less integrity, less interest in rigorous skepticism, and less knowledge as to the limitations of their practice than

those whom they replace (since the basis of self-selection *out* of the field is an understanding of its problems). They will then train PhD students in their techniques. In other words, as knowledgeable insiders slowly leave the field (or choose never to join it in the first place), psychology will become increasingly dangerous and destructive until its public credibility collapses. This mode of vicious cycle (“brain drain”) is endemic to disciplines that over-surrogate, creating (1) elaborate credential requirements that distract from real work—“hoops” to jump through, in other words, and (2) specializing in the production of models of models in the Baudrillardian simulacra sense, such that work increasingly diverges from contact with ground—what is known informally as “bullshit” work.

Those psychologists who choose to stay will be out-competed, out-hired, and out-tenured compared to those who are willing to play ball with p-hacking regimes, with performative pseudoempiricism, with statistically baseless inference, and with the publish-or-perish emphasis on quantity over quality. Other social scientists’ abuse of metrics, and misrepresentation of findings, ends up raising the bar of expectation; those who optimize toward “real” science—in other words, the surrogated target—are penalized in their competition with those who more efficiently and directly optimize toward the actual basis of promotion, advancement, and recognition—the surrogate that is “optics.” This affects not just the career prospect of individuals but the larger efficacy and service of the distributed activity.

Finally, psychology—insofar as it can be meaningfully said to “freeride” the reputation of legitimate science, by enjoying the benefits of its perceived reputation while showing little obligation to the same standard of rigor, will increasingly harm the overall perception of legitimacy of the sciences. We can see some taste of this in the so-called Science Wars of the late 20th C, where the failings and hubris of social sciences contributed to a loss of faith in the naturalized science (in part because the problem of psychology is the problem of inference, a problem which, while of a very different *scale* in physics, is of the same general kind—the age-old and possibly intractable problem of inference).

5. Qualitative surrogation: The tragedy of appearances

Though surrogation is typically described in terms of the lossy compression of qualitative, intuitive, and holistic judgment into quantitative metrics, it may arguably occur any time a *marker* is substituted for what it de-marcates. In [signaling theory](#), classically, signals are external, public-facing attributes that indicate, to other organisms, a probabilistic presence of some hidden, private trait. As in language, with the connection between the *signified* and the *signifier* (the reference and the referent), this ability to “stand proxy for,” and represent publically, some private, incommunicable, hard-to-verify truth is built up through associative learning. An individual experiences the coincidence of some prominent physical marker and some attribute, which instill a relationship that can be meaningfully used as the basis for future inference. This associative landscape, which I’ve previously called the “fashion” landscape (Reason 2018), is heavily mediated and deeply social: it is passed along that e.g. a certain item of clothing, or type of behavior, “sends a certain message,” or “gives a particular impression.” To illustrate just how common this kind of behavior is in our social lives, consider how we size up

an strangers's socioeconomic class, extrapolate a candidate's future performance in a hiring interview, or choose an interlocutor at a social event. [\[11\]](#)

5.1. Dylan at Newport, 1965: Artistic authenticity

In summary, the metonymic surrogation and reification we see play out in the sphere of quantitative metrics plays out in the qualitative signaling sphere as well. To take an example from the history of pop music, the concept of authenticity—a vague, hard-to-measure, and complex trait—has been historically instantiated, which is to say recognized, in different forms; for instance, the folk scene in Greenwich Village in the 1950s held this reputation, as did “lo-fi aesthetics” in the late 20th and early 21st century—that is, music recorded on relatively inexpensive amateur equipment. The logic for this association between quality and form was relatively straightforward: it is premised on the lack of incentives present in these domains. Folk singers typically lone, itinerant performers making very little money; their performances required only an acoustic guitar and a small performance venue (e.g. a bar or comedy club). Indie musicians in the late 20th C, home-recording from a Tascam 4-track tape recorder, likewise, were liberated by not needing to pay a studio or producer's fee (entailing freedom from label support). In either case, there is a lack of financial pressure, with the recognition that such pressure tends result in a Faustian compromise of the artist's “aesthetic integrity”—his “vision” or “soul”—for work that caters toward an audience (and hence drives sales, delivering a higher return on investment for labels).

When these fields of aesthetic production were ignored, in their early years, with little money involved, there was a meaningful sense in which these associations-as-signals *were* costly: artists who cared more about decisional autonomy than commercial success would forego the income and reputation that label support might afford them. Thus audiences could, at least to some extent, be assured they were getting “the real deal.” When the scenes began to attract attention, a quick freerider economy emerged around *acting as if*: since there was nothing intrinsically honest about performing on an acoustic guitar, or having distorted audio from poor quality recording hardware, and by imitating all the aesthetic residue and markers—the associated surface signals—of authenticity, acts could see the associated authenticity conferred on them in turn. (At least until the established association “collapsed” in the 1960s, *cf.* “frequency-dependent selection” in signaling theory.)

This burgeoning fetishization of surface aesthetics still permeates the independent music scene, where “slick” overproduction is *déclassé*, and tape warble and white noise, vocal clipping and compression, are tactically deployed to give a certain affective impression. And since the affect is so fleeting, who could make an accusation of falsehood “stick”? This is a prototypal case of surrogation: the public, by incentivizing compliance to a set of surface qualities in a purported bid to monitor and secure authenticity, have in reality incentivized musicians to falsify their own material origins, visions, and capacities.

It is against this backdrop we can understand Dylan's 1965 performance at the Newport Folk Festival—an incident with its own encyclopedia page, the “Electric Dylan controversy,” and which stands as the flipside to this surrogation. We can see footage today of the set: of Dylan performing the

same songs that had been heralded—indeed, borderline sanctified by the public and press—for their honesty and activism—but now they were being performing with an electric, rather than acoustic guitar, with a full-band setup. Dylan had “plugged in,” abandoning the old associative, surrogated markers of the authentic. The widespread sentiment was that in doing so, Dylan had “sold out,” was no longer a performer of integrity, *on the basis of a new guitar sound*. Substance confused with symbol. Without playing down the complexities of the historical situation—without denying that there is something legitimate about anger over abandoned symbols, or that the mythologization of this event undoubtedly led to its exaggeration—how else can we make sense of the outrage that followed, than as the reification of an associated but causally distinct measure, as the surrogation of a complex trait like “authenticity” for a much simpler one, the way one speaks or the instrument one plays? The reception lasted for years, jeers of “Judas” from the crowd on Dylan’s international tours.

5.2. Lessons from glamrock

This is the context—the nefarious surrogation of real efforts into cardboard cutouts, surface signaling replacing genuine embodiment—in which we can properly understand the emergence of showy, fantasy-ridden, egoic and “artificial” glam rock in the early 1970s—just a few years after Newport—as well as the disdain that it raised.^[12] The pop studies scholar Simon Reynolds, in his book-length study of glam, *Shock & Awe*, sets the scene for us with an illustration of surrogation in 60s theater, and the prevailing regime of “naturalness” against which glam reacted (see the entry on [Signals and Correctives](#)):

a post-Method school of actors and directors aspired to a de-theatricalised form of naturalistic acting, all mumbling and tics, that inevitably spawned a new set of mannerisms that today look as stagey and trapped in time as the Hollywood golden age of poise and elocution. In all the arts, in fact, every attempt of realism, no matter how stringently stripped down or crude, seems to birth a new repertoire of stylised conventions and stock gestures. Bowie, for one, was acutely aware of this in relation to rock, which he precociously grasped was a *performance* of realness rather than a straightforward presentation of reality onstage. (Reynolds 2016)

This is both in the sense that all naturalness is “technically” a performance, and also that the performance had become increasingly and meaningfully more conscious, strategic, and commercial. Glam, as Reynolds shows, took the strongest symbols of Sixties “natural honesty”—hair and nudity—and mocked them with makeup, costume, and dye. What it really mocked was surrogation—the dangerously cheerful illusion that we can fetishize a measure and it will continue to tell us the truth about its subject. This is the brilliance of glam. How else can we understand these archetypal progressions in the history of pop other than as products of freeriding and surrogation, as symbols reified as the things themselves?

5.3. “Creative” cargocults

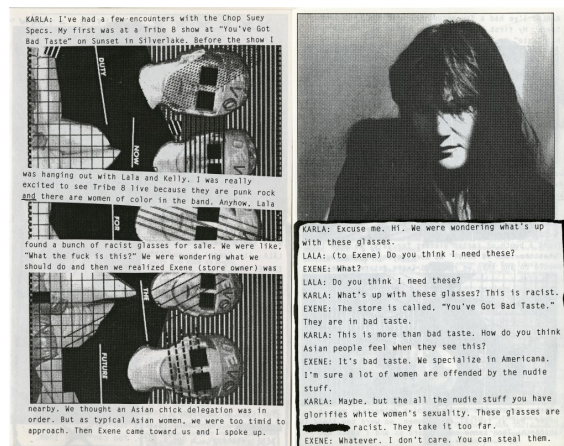
The imitation of surface attributes, rather than causal mechanisms, is a common one in beginning artists. In Arthur Danto’s book-length profile of

pop artist Andy Warhol, we encounter Warhol's early imitation of abstract expressionism's "paint drips," and his belief that it was somehow critical to the painting project:

[Warhol] applies paint the way an Abstract Expressionist artist would, allowing it to drip. "You can't do a painting without a drip," he told Ivan Karp, who was director of the Castelli Gallery. This is what I meant by saying that he used Abstract Expressionist gestural painting as *protective coloration*. The drips did not come from some *inner conviction*... (or, we might interpret, an *internal logic*) ...they did not refer to that moment of trance when the A. E. painter moved the paint around without tidying up. "The drip"... for Warhol... [was] an *affectation*... [Danto (2009), *emphasis added*]

Whereas, for the original Abstract Expressionists, paint drips were a byproduct of a *technique* that embodied an *ideology* of art (an ideology much in line with the emphasis on spontaneity and honesty found also in folk music). Here that very byproduct is lifted out of its context and treated as a goal in its own right. Amidst these performances, which are often enough to fool average critics, genuine embodiments of qualities like innovation or integrity go unrecognized, while regurgitation disguised by savvy signaling is showered by praise. ("The payoffs of today determine the behavior of tomorrow.")

Today in many visual art cultures, the aesthetics of a "zine"—as pioneered by groups like Riot grrrl, and whose appearance is partially an artifact of copymachine and wordprocessing technologies from the 1990s—surrogates the proxied-for qualities and is perceived as somehow "more DIY" than those projects made with contemporary technologies.



A spread from a Riot grrrl zine.

This historical residue is all around us—it is the lingering ooze of prestige past, available for any who care more about said prestige than the status of the distributed project undertaken by the superorganism. In this way, the presence of this residue in works ought to be treated as a *reverse* indicator, of zombie art animated by the hungover associations of eras past. As A.D. Jameson describes, training filmmakers who wish to be perceived as experimental will engage in the now-antiquated techniques of avant-gardes past, in order to seem "of a kind" with their hallowed paters:

The canonical works define the style and range of [what is considered “proper” U.S. experimental] cinema: It is non-narrative (favoring surreal logic or structural organizing principles), abstract, often incorporates found footage, and also frequently involves directly treating the film itself (scratching it, painting it, growing mold on it, and so on). It often demonstrates some aspect of the film apparatus or filmmaking process, sometimes by taking a self-reflexive approach (foregrounding the use of the camera) or a conceptual approach (projecting through alternate substances, or projecting plain black leader, or projecting nothing but the projector light itself). (Jameson 2010)

Imitation of a canon is obviously antithetical to the spirit of experimentalism. And yet “the film students of today frequently make work that employs those techniques [associated with historical experimentalism]. The question then becomes: Are they making experimental films?” We can leave quibbling over handles to art historians and analytic philosophers, while confidently assessing that the original target of experimental practice has been lost, surrogated for those techniques which are known, in the critical and public sphere, to have accompanied it—and which are still met, by critics and elite audiences, with the prestige accorded to the originals.

6. Mitigations

On the one hand, Kahneman found that decisions are subject to cognitive biases and can be systematically improved once we move past our intuition. On the other, despite our systematic biases, as Gary Klein originally noted when studying firefighters, many decisions don’t use metrics, and are incredibly effective despite that. In fact, this success isn’t despite the lack of cognition, but because of it. Klein’s “recognition-primed decision making” works exactly where our intuition beats measurement. As Klein and Kahneman now agree, there are domains in which “raw intuition” beats reflection. (Manheim 2016a)

6.1. Supplementation vs. surrogation

Crucially, measurement alone cannot devastate; it merely constitutes a second source of information. Rather, it is when other forms of evaluation are surrogated (“substituted for”) partial measurements, in a way that affects incentive structures (“reward functions”) downstream that causes trouble. Post-surrogation, decision-makers have *less* information rather than more.

Employees self-narrativizing may be high in qualitative richness, but fail in providing accountable, grounded, no-bullshit understandings of performance. Metrics ground them in reality, and shrink the space of fabricated unrealities.

Further, expertise is still needed to situate and contextualize metrics. There is good reason that Yarkoni and Gigerenzer (2014) both emphasize that much of the current crisis in psychology comes from the conventional, automatic, and uncritical surrogation of statistical measures (i.e. the *context-independent* application of statistics). Muller (2018):

Interpretation of indicators is critically important, and requires informed expert judgment. It is not enough merely to count

incidents or conduct quantitative or statistical analysis—interpretation is a qualitative activity based on familiarity with the environment, and it needs to be conducted by experienced personnel who have worked in that environment for long enough to detect trends by comparison with previous conditions.

In other words, the alternative to surrogating the qualitative-holistic is a combinatorial approach that uses the metric to monitor the qualitative-holistic, while at the same time using the qualitative-holistic to contextualize the metric within. The two operate as mutual agitation, as dialectic.

This is more or less the approach which Kahneman—whose behavioral economics work is founded in “debugging” human biases, and emphasizing System 2 thought—as well as Gerd Gigerenzer and Gary Klein—whose work on “ecological rationality” pushes back on behavioral economics, emphasizing the value of intuition—have come to agree on. System 2, which is explicit, legible, and top-down “rational” in the sense typically purported of quantitative surrogates, is used for *monitoring* System 1; it is a check against error and bias, but cannot fully replace it.

6.1.1. Developing an aesthetic

To some extent this approach can be summarized as “rigorous taste”—an intuitive or qualitative evaluative core that is structured by systematicity. Ben Connable’s Rand study *Embracing the Fog of War: Assessment and Metrics in Counterinsurgency*, which examines the problems of surrogation in combat situations, writes that counterinsurgency is “both art and science, but mostly art,” requiring the honed, intuitive judgment—the *guiding aesthetic*, insofar as its pattern-matching is unconscious—of an artform. In other words, the “inexact sciences” tried to grow up too fast.

6.1.2. Tracking intuitions

Another potential synthesis of qualitative and quantitative, System 1 and System 1 thought—and an established effective approach to measurement—is to track locals’ subjective, acted-upon, skin-in-the-game assessments of complex situations, e.g. tracking the stability of Afghanistan through the price of market goods. Kilcullen (2010):

Afghanistan is an agricultural economy, and crop diversity varies markedly across the country. Given the free-market economics of agricultural production in Afghanistan, risk and cost factors—the opportunity cost of growing a crop, the risk of transporting it across insecure roads, the risk of selling it at market and of transporting money home again—tend to be automatically priced in to the cost of fruits and vegetables. Thus, fluctuations in overall market prices may be a surrogate metric for general popular confidence and perceived security. In particular, exotic vegetables—those grown outside a particular district that have to be transported further at greater risk in order to be sold in that district—can be a useful telltale marker.

(This is similar to the approach in [ethnomethodology](#), where social organization is understood primarily through *demonstrated behavioral adaptation* and spontaneous self-organization, response to stimuli rather than surveyed self-interpretation or external metric.)

6.2. Minimizing measurement surrogation in Goodhart-Campbell

Here we can understand G as the surrogated original and G^* as the surrogate.

From *LessWrong* (2010):

Hansonian Cynicism^[13]

Pointing out what most people would have in mind as G and showing that institutions all around are not following G , but their own convoluted G^* s. Hansonian cynicism is definitely the second step to mitigation in many many cases (Knowing about Goodhart's law is the first). Most people expect universities to be about education and hospitals to be about health. Pointing out that they aren't doing what they are supposed to be doing creates a huge cognitive dissonance in the thinking person.

Better measures

Balanced scorecards

Taking multiple factors into consideration, trying to make G^* as strong and spoof-proof as possible. The Scorecard approach is mathematically, the simplest solution that strikes a mind when confronted with Goodhart's law.

Optimization around the constraint

There are no generic solutions to bridging the gap between G and G^* , but the body of knowledge of theory of constraints is a very good starting point for formulating better measures for corporates.

Extrapolated Volition

CEV tries to mitigate Goodhart's law in a better way than mechanical measures by trying to create a complete map of human morality. If G is defined fully, there is no need for a G^* . CEV tries to do it for all humanity, but as an example, individual extrapolated volition should be enough. The attempt is incomplete as of now, but it is promising.

Solutions centred around Human discretion

Human discretion is the one thing that can presently beat Goodhart's law because the constant checking and rechecking that G and G^* match. [...] However, this is not scalable in a strict sense because of the added testing and quality control requirements.

Left Anarchist ideas

Left anarchist ideas about small firms and workgroups are based on the fact that hierarchy will inevitably introduce goodhart's law related problems and thus the best groups are small ones doing simple things.

Hierarchical rule

On the other end of the political spectrum, Molbuggian hierarchical rule completely eliminates the mechanical aspects of the law. There is no letter of the law, its all spirit. I am supposed to take total care of my slaves and have total obedience to my master. The scalability is ensured through hierarchy.

7. Conclusion

Surrogation is a problem arguably endemic to daily life and the functioning of human superorganisms. It has been noticed by various fields including philosophy, management theory, social psychology, and artificial intelligence research. Developing strategies for minimizing the harms of surrogation is aided by a cross-disciplinary understanding of a general problem as it ripples across human social organization.

8. Works Cited

Amodei and Clark (2016). “Faully Reward Functions in the Wild.” *Open AI Blog*. <https://openai.com/blog/faulty-reward-functions/>.

Arbital (2020). “Nearest unblocked strategy.” https://arbital.com/p/nearest_unblocked/.

blogospheroid (2010). “The Importance of Goodhart’s Law.” *AI Alignment Forum*. <https://www.alignmentforum.org/posts/YtvZxRpZjcFNwJecS/the-importance-of-goodhart-s-law>.

Choi, Hecht, and Tayler (2012). “Strategy Selection, Surrogation, and Strategic Performance Measurement Systems”. *Journal of Accounting Research*. 51 (1): 105–133. <https://doi:10.1111/j.1475-679X.2012.00465.x>.

Choi, Hecht, and Tayler (2011). “Lost in Translation: The Effects of Incentive Compensation on Strategy Surrogation.” *The Accounting Review*. 87 (4): 1135–1163. <https://doi:10.2308/accr-10273>.

Connable, Ben (2012). *Embracing the Fog of War: Assessment and Metrics in Counterinsurgency*. RAND Corporation. <https://www.rand.org/pubs/monographs/MG1086.html>.

Danto, Arthur (2009). *Andy Warhol*. Yale University Press.

Feynman, Richard (1974). Caltech commencement address.

Gigerenzer and Marewski (2014). “Surrogate Science: The Idol of a Universal Method for Scientific Inference.” *Journal of Management*. <https://doi.org/10.1177/0149206314547522>.

Huemer, Michael (2012). “In Praise of Passivity.” *Studia Humana* 1,2 (2012): 12-28.

Inzlicht, Michael (2016). “Reckoning with the past.” *Getting Better*. <http://michaelinzlicht.com/getting-better/2016/2/29/reckoning-with-the-past>.

Jackall, Robert (1988). *Moral Mazes: The World of Corporate Managers*. Oxford University Press.

Jameson, A.D. (2010). "Experimental Fiction as Genre and as Principle." *Big Other*. <https://bigother.com/2010/02/03/experimental-fiction-as-genre-and-as-principle/>.

Kahneman and Tversky (1973). "On the Psychology of Prediction". *Psychological Review*. 80 (4): 237–51. <https://doi.org/10.1037/h0034747>.

Manheim, David (2016a). "Goodhart's Law and Why Measurement is Hard." *Ribbonfarm*. <https://www.ribbonfarm.com/2016/06/09/goodharts-law-and-why-measurement-is-hard/>.

Manheim, David (2016b). "Overpowered Metrics Eat Underspecified Goals." *Ribbonfarm*. <https://www.ribbonfarm.com/2016/09/29/soft-bias-of-underspecified-goals/>.

Manheim, David (2019). "Multipart Dynamics and Failure Modes for Machine Learning and Artificial Intelligence." *Artificial Superintelligence: Coordination & Strategy*.

Hanson and Simler (2017). *The Elephant in the Brain: Hidden Motives in Everyday Life*. Oxford University Press.

Perry, Sarah (2020). "Ignorance: a skilled practice." *Carcinisation*. <https://carcinisation.com/2020/01/27/ignorance-a-skilled-practice/>.

Nguyen, C. Thi (2020). *Games And The Art Of Agency*. Oxford University Press.

Noë, Alva (2015). *Strange Tools: Art and Human Nature*. Hill and Wang.

Ortega and Maini et al (2018). "Building safe artificial intelligence: specification, robustness, and assurance." *DeepMind Safety Research*.

Reason, Suspended (2016). "Intro to Cargocult." *Suspended Reason*. <https://suspendedreason.com/2016/11/16/intro-to-cargocult/>.

Reason, Suspended (2018). "Oscillation / Fashion." *Suspended Reason*. <https://suspendedreason.com/2018/07/19/oscillation-fashion/>.

Reason, Suspended (2020). "The Dark Miracle of Optics." *Less Wrong*. <https://www.lesswrong.com/posts/zzt448rSfwdydinbZ/the-dark-miracle-of-optics>.

Reynolds, Simon (2016). *Shock and Awe: Glam Rock and Its Legacy, from the Seventies to the Twenty-first Century*. Dey Street Books.

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

Simler, Kevin (2016). "Minimum Viable Superorganism." *Ribbonfarm*. <https://www.ribbonfarm.com/2016/02/11/minimum-viable-superorganism/>.

Yarkoni, Tal (2019). "The Generalizability Crisis." *PsyArXiv*. November 22. <https://doi.org/10.31234/osf.io/jqw35>.

9. Footnotes

1. Cooperative enterprises organized by prestige economies to a common purpose (Simler 2016). ↩
2. It “is the mismatch between our generalization intention and the model specification”—in other words, between holism and quantitative representative—“that introduces an inflated risk of inferential error, and not the model specification alone” (Yarkoni 2019). ↩
3. In this way, the realm of social reputation is prone to a range of related dynamics in which the gap between appearance—assumed signals or markers of some hidden reality—and that reality itself causes problems. ↩
4. Partially this is described by the “barberpole” metaphor of fashion, “where lower classes continually imitate higher classes, who are themselves engaged in a continual quest for “distinction” from the chasing masses... Its cyclical nature is the result of limited options and a continual evasion of freeriders who exploit an associative proxy: clothing for caste” (Reason 2020). ↩
5. One reason intuition and subjective judgment have been given over, so easily, to quantitative measures is the case made by behavioral econ against such judgments. (In other words, the defense of pseudoempiricism comes from pseudoempirical work.) And yet our ability to parse highly indexical social signals, or linguistic expressions, within an *ecological context*, and to an extent that consistently defies our best computational models, testifies to our ignored “indexical genius.” As Perry puts it in “Ignorance, a skilled practice”: “Overconfidence in the global knowledge game, especially [among the] social sciences, threatens the production and appreciation of the genuine kind of indexical knowledge that humans are geniuses at producing and using” (2020). ↩
6. The connection with legibility: a lack of modesty toward complexity, and a lack of faith in the hard-to-scrutinize. In surrogation, the lack of faith lies in judgment; in legibility, in the forces of cultural evolution and the local emergence of sense-making and self-order. ↩
7. Recall Bourdieu’s idea that the gossip of a field makes up some of its essential wisdom ↩
8. Looking to the most sclerotic and dysfunctional arenas of our society—politics, policing, institutionalized art, among others—suggests that one of the main problems of the modern world is a much softer type than the traditional corruption. Our society is not so much meritocratic as it is *optikratic*: to be seen *is to have* power, just as being seen to have power is also to have it, and power is awarded not on the basis of “actual” (in the ideal sense) merit or value, but on the basis of sporting their appearance. This is at once utterly obvious—one’s impression of an object is all one can, finally, operate off, and thus there “is no other way”—and at the same time seriously non-trivial: the translation of holistic quality to public appearance is lossier than usually assumed or acted-upon. ↩

9. Arguably actions like Yarkoni's, which alter the common knowledge of the field and thus potentially alter its internal prestige economy, may improve the situation. [↵](#)
10. Psychologists Yoel Inbar and Michael Inzlicht report multiple occasions of "bright undergraduates" voicing complaints similar to Yarkoni's, and we can only imagine that psychology's inability to convincingly answer such concerns discourages participation those with the foresight to see it from entering. [↵](#)
11. Recall, from [signaling theory](#), that those associations which endure—which are robust to destruction by free-riding—must be built on a logic of *cost disparity*: the public marker must be disproportionately cheap to exhibit for those who really possess the private quality, and accordingly, be disproportionately expensive to those who do not. A middle-class individual might be able to *afford* a Rolex, or a very expensive car, but it so impacts his finances and freedoms, requires such significant sacrifices that it is only rarely and in specific circumstances worth it, in the final cost-benefit analysis, to "purchase" a signal. For the upper-class individual, such expenditures involve very little sacrifice at all, and the small gain in status from parading such luxury goods will outweigh their cost. There are many associative signals that are *not* robust, however; though they may be ridden into oblivion over months or years, they still exert a presence in daily life. [↵](#)
12. The rockist anti-disco movement shows similar tendencies. [↵](#)
13. In Hanson & Simler's view, surrogation, signaling, and prestige economies so disrupt major human enterprises that many of them—especially the sufficiently institutionalized—can reasonably be described as serving a very different telos than self-purported (2017). [↵](#)