Protocols and Institutions

Michael Zargham and Ilan Ben-Meir

This paper is a pre-print of a book chapter that explores the relationship between protocols and institutions through the lens of control systems engineering. Cross-referencing concepts from Institutional Economics, Cybernetics, Control Theory, and Market Design, we define institutions mathematically through correspondence to stability (in the sense of Lyapunov), demonstrating that behavior can be enduringly and predictably patterned without resorting to coercive control. Grounded in engineering practice, we emphasize field surveys as prerequisites for formal modeling, avoiding purely theoretical constructs. Aligning our terminology with control engineering's canonical forms enables us to direct readers to a broad suite of practical methods, while highlighting emerging work on institutional design through protocolization. The chapter will appear in the second edition of World Scientific's *Web3 Blockchain Economic Theory*, edited by Melanie Swan, Soichiro Takagi, and Frank Witte.

DOI (This Version): 10.5281/zenodo.15116453
DOI (Latest Version): 10.5281/zenodo.15116452

1. Introduction

In recent years, the claim that society's institutions are breaking down has matured from a provocation into a truism. The growing consensus that our institutions are dysfunctional, however, has not yet produced an actionable discourse about *how to fix them* that is commensurate with the complexity of the problem. This deficit is due, at least in part, to the lack of clarity surrounding our understanding of what institutions are, and (therefore) what characterizes their failure. Answering these questions rigorously makes it possible to think about institutions and their failure through the lens of control systems engineering — thus enabling us to apply methods and insights drawn from that discipline to the urgent but achievable work of revitalizing them.

2. Institutions

The natural starting point for our inquiry is the field of "Institutional Economics," where the academic study of institutions *as such* is concentrated. At the foundation of that ongoing conversation, one finds Douglass North's definition: "Institutions are the rules of the game in a society or, more formally, are the humanly devised constraints that shape human interaction. In consequence, they structure incentives in human exchange, whether political, social, or economic."[1] In North's telling, "[t]he major role of institutions in a society is to establish a stable (but not necessarily efficient) structure to human action." Elinor Ostrom sounds a similar note in *Governing the Commons: The Evolution of Institutions for Collective Action*, explaining that "'[i]nstitutions' can be defined as the sets of working rules"[2] that govern interactions within a particular context. "Working rules are those actually used, monitored, and enforced when individuals make choices about the actions they will take," Ostrom elaborates, explaining that when it comes to institutions, "all rules are nested in another set of rules that define how the first set of rules can be changed." This "nesting of rules within rules at several levels is similar to the nesting of computer languages at several levels," she goes on to suggest, and helps institutions to realize their purpose insofar as it reinforces "the stability of mutual expectations among individuals interacting according to a set of rules." Ostrom and North thus appear to agree: Institutions are the rules that human beings develop in order to provide stable structures for our interactions.

Ostrom seems to have revised her understanding of institutions, however, during the five-year interval separating the release of *Governing the Commons* and the publication of "A Grammar of Institutions" (co-authored with Sue E.S. Crawford). The latter paper begins with a declaration that it "is based on a view that institutions are enduring regularities of human action in situations structured by rules, norms and shared strategies, as well as by the physical world."[3] Here, "institutions" are no longer the "rules, norms, and shared strategies" that structure the "situations" in which human beings interact; instead, they have become the "enduring regularities of human action" that are *produced by* or *emerge from* the interactions that take place in such situations. In fact, Crawford and Ostrom's definition of institutions as "enduring regularities of human action in situations structured by rules, norms, and shared strategies, as well as the physical world" involves no fewer than *four* distinct concepts:

(1) The "rules, norms, and shared strategies" that make up the "humanly constituted elements of institutions."
(2) The "physical world" that, along with these "humanly constituted elements," structures the "situations" in which institutions can be found.
(3) The "situations" that are "structured by" both the "components of institutions" and the "physical world."
(4) The "enduring regularities of human action" that are found "in [such] situations."

Insofar as Crawford and Ostrom intend for their proposed "Institutional Grammar" to provide "a theoretical structure for analysis of the humanly constituted elements of institutions (i.e. rules, norms, and shared strategies)," it must also be read as explicitly acknowledging a *difference* between these "humanly constituted elements of institutions" (which Crawford and Ostrom also call "the components of institutions") and the institutions into which they are composed.

Institutional Economics (as a discipline) is still engaged in the work of litigating precisely how these four concepts relate to one another, along with the related task of *assigning each a distinct (and consistent) name.* While North, for example, identifies "[a] crucial distinction [...] between institutions and organizations" corresponding to the separation of "the rules from the players" of a given game, his characterization of "organizations" collapses the conceptual distinction between the loci of activity that Crawford and Ostrom call "situations" and the "physical world" in which various organizational "bodies"[1] concretely interact. Furthermore, Geoffrey M. Hodgson convincingly argues that the imprecision of North's terminology means that "even by North's own definition, organizations must be regarded as a type of *institution.*"[4]

We thus conclude that it is necessary to look beyond its borders to find clarity about institutions, and terminology that is precise enough to facilitate both their description and their design.

3. Protocols & Purposive Systems

Although his language often appears to coincide with that of Ostrom and North, the understanding of institutions put forth by Stafford Beer in *Designing Freedom* offers tools for untangling the conceptual confusions opened up by the Institutional Economics literature.

As Beer explains, "institutions [...] are not just things, entities we recognize and label. They are instead *dynamic and surviving systems* [...] they consist of related parts, and the relations — the connexions — between those parts [... in] a particular organization which produces particular outputs."[5] The "specific modes of organization" in which these parts and their relations are *realized* are concrete systems — not just *specifications*, but particular *implementations.*

Institutions, Beer notes, "are actually *outputs* of this system. They are what they are because the system is organized in the way that it is, and this organization produces an inescapable kind of behavior." Translating Beer's famous acronymic aphorism "POSIWID" (The Purpose of a System is What it Does) to the context at hand, one could say that "the purpose of an organization is the behavior that it *realizes.*"[6]

Beer thus makes clear that the "enduring regularities of human action" that Crawford and Ostrom identify as "institutions" are more precisely understood as *the stable emergent properties* (or the system-level behavior) of the concrete systems that Beer calls *organizations.* These realized systems correspond to both the "physical world" and the *implementations* of any "rules, norms, and shared strategies," thus overlapping both item (1) and item (2) in Crawford and Ostrom's definition of "institutions." The "situations structured by" this physical world (and its interplay with the organization's existing "rules, norms, and shared strategies"), meanwhile, correspond to such systems' *states.* While at first glance it may be confusing that "rules, norms, and shared strategies" are represented in both the *organization* and in the *situation*; this simply means we are dealing with a "second-order system." Second-order dynamical systems have state variables, which represent the way other state variables change.[7,8] This complexity is necessary to accurately represent the ways that an organization's "rules, norms, and shared strategies" both *influence* and *are influenced by* the actions of its actors, albeit along different temporal horizons.

Beer's work helps us to refine our model by mapping three of the four concepts in Crawford and Ostrom's definition to a (second-order) dynamical system. What remains is to find a suitable term for the "rules, norms, and shared strategies" *themselves.*

3.1. *Protocols*

What do "rules, norms, and shared strategies" have in common *other than* their structural relationship to institutions? The answer is that all three are means of patterning behavior which be grouped together under a single category: they are all types of*protocols.*

Derived from the Greek *protokollon* (from *protos* – "first" – and *kolla* – "glue") the first protocols were cover-sheets attesting to the provenance (and thus the authenticity or validity) of the documents to which they were attached, thereby operating as "an early techno-social system of administration that has functional equivalents from Antiquity up until today."[9] Eventually, the term expanded to refer to regularizing practices in areas as diverse as the sciences or international diplomacy — sets of common constraints that function to produce interoperability despite difference, "often independent from time and place." In contemporary usage, "protocol" has also taken on a particular technical meaning (albeit one highly consonant with its broader sense): "a cascade of formalized standards or agreements to be implemented as control regimes for flexible material and/or semiotic organization [... that] predictably structures in an often layered, sometimes hierarchical way the behaviors of data and objects [such that they are able] to participate in infrastructural networks."

The close conceptual association that has formed between the concept of "protocols" and digital technology is not accidental; protocols are, essentially, "distributed algorithms"[10] (the title of Nancy A. Lynch's 1996 textbook describing the protocols underlying the then-emerging internet), or *control mechanisms for the flow of information through decentralized systems.* In *Protocol: How Control Exists After Decentralization*, Alexander Galloway thus identifies protocol as "the principle of organization native to computers in distributed networks."[11] Galloway connects the contemporary technical meaning of "protocol" to its historical antecedents, explaining that inasmuch as both usages refer to "conventional rules that govern the set of possible behavior patterns within a heterogeneous system [...] protocol is a technique for achieving voluntary regulation within a contingent environment [...] a distributed management system that allows control to exist within a heterogeneous material milieu."

If protocols are mechanisms that control the flow of information (or that structure the substrate by way of which information, matter, energy, and thus also action flow), then the field of **control systems engineering** can be understood as a rigorous and formal study of the interplay between the physical world, specified protocols and realized behavior of the assemblage thereof — thus equipping us with the machinery needed to engineer the stabilization of institutions at all scales, from micro-markets to the most vital functions of a healthy society.

3.2. *Control Systems Engineering*

Design and analysis of control systems is accomplished by representing these assemblages via dynamical systems. In the context of multi-agent systems, control engineering explores how locally-defined and locally-implemented rules interact within dynamic environments to generate emergent behaviors that are both stable and predictable at the system level — in other words, how the protocols that constrain behavior *within* an organization can be composed in order to induce stable properties in the emergent system-level behavior *of* that organization.[12,13]

Cooperative Control of Multi-Agent Systems, edited by Jeff S. Shamma, provides a detailed look at the methods (which is to say, the protocols and collections of interrelated protocols) that contemporary control systems use to compose complex assemblages where individual agents, operating with access to only local information and retaining autonomy over their decisions, collaboratively produce emergent, higher-level patterns of behavior.[14] These stabilized patterns are not static, but instead represent the system's capacity to dynamically and continuously solve complex constrained resource allocation problems in changing environments, all without any agent accessing global information.

3.3. *Cybernetics*

The cybernetics literature extended the (then-contemporary) practices of control systems engineering to assemblages other than human-made systems. As Norbert Wiener explains in the introduction to the volume that christened the field, "[w]e have decided to call the entire field of control and communication theory, whether in the machine or in the animal, by the name Cybernetics, which we form from the Greek $\chi\upsilon\beta\epsilon\rho\nu\eta\tau\eta\varsigma$ or *steersman*."[15] In *Designing Freedom*, Beer revises Wiener's definition, asserting that "[c]ybernetics is the science of effective organization" — or, we might say, of how to organize systems in order to produce specific desired effects. Beer pays particular attention to "the conditions [organizations] must meet in order to remain stable yet adaptive." These latter properties are what is meant by the second part of Beer's definition of institutions as "*dynamic and surviving systems*." In the context of a dynamic system, "survival" is best understood as persistence of function, rather than prevention of demise.

As Beer explains, "[t]he number of possible states of a system is called its *variety*," and "[i]n order to regulate a system, we have to absorb its variety. If we fail in this, the system becomes unstable" (1974). He proceeds to invoke "the Law of Requisite Variety — named Ashby's law, after its discoverer," which asserts that "only variety can absorb variety"; the implication of Ashby's law, Beer explains, is that "[i]f varieties in a regulatory system are disbalanced, the system cannot attain stability. Assuming that the regulator has the smaller variety, there are only two ways of meeting the demand of Ashby's Law. One is to attenuate variety in the system, the other is to amplify variety in the regulator." Beer is careful to note that "these strategies can be mixed" — but in both cases, "[t]he mode of organization adopted for the system is its variety controller." In other words, the *stability* of a dynamical system depends on that system's *adaptive capacity*, which is in turn a function of its "mode of organization," or *the way that its organization is structured.*

Crucially, the characterizations of both Ashby and Beer are implicitly dependent on a "second-order" model, wherein an organization can adapt its own structure.

3.4. *Dynamic Stability*

In control systems engineering, "stability" has a precise technical meaning: it refers to the system's ability to maintain consistent and predictable behavior in the face of disturbances. Importantly, stability does not necessarily imply returning to a fixed, unchanging state; rather, the equilibrium of a system can itself be a dynamic behavior — such as a pattern of oscillation or a stimulus-response pair — as long as the defining properties of that behavior remain stable over time. For a system to be stable, any deviations from this dynamic equilibrium must either decay over time (asymptotic stability) or remain within a bounded region where the system continues to exhibit the same characteristic behavior.[1] These dynamic equilibria manifest as "enduring regularities" in the system-level behaviors of organizations — which is to say, as **institutions**, "patterns of behavior"[5] that persist despite variations in individual behavior or shocks from the environment. The "failure" of an institution occurs when this dynamic stability is lost.[2]

In Beer's time, the ability to apply cybernetics to practical problems was limited — both technically and politically. The methods of implementation required access to expensive computational resources and a mandate from centralized authorities. The discipline's methods reflected these constraints. In recent years, however, significant advances in technology and mathematics have relaxed both the technical and political constraints. Advances in distributed optimization and multi-agent control provide the basis for a political shift towards "ceding control in favor of coordination" (as per Zargham *et al.*, summarizing Swann).[19,20] The rise of open-source software development (particularly peer-to-peer protocols which constitute large public networks) demonstrates the realizability of a shift from top-down cybernetics toward bottom-up (or at least middle-out) alternatives[21] — from the *command and control* cybernetics of the 1980s to a techno-politics whose watchwords are *protocolize and govern.*

3.5. *Engineered Institutions*

The field of Market Design offers a glimpse into the future of emerging institutions constituted by protocols that have been carefully engineered to produce reliable behavior. In his seminal work "The Economist as Engi-

[1]This characterization of stability is also consistent with the work of Robert Axelrod, who studied the emergence and stabilization of equilibria over norms in his seminal work *The Evolution of Cooperation.*[16] Notably, Axelrod's models placed the norms that agents used to make their decisions within the state of his models, allowing them to change over time via turnover in the agent population. It is also related to stability in population games, as developed extensively by Sandholm in *Population Games and Evolutionary Dynamics.*[17] Population games study strategic interactions in large populations where individual agents have negligible influence but collectively shape system-level behavior.

[2]This perspective aligns closely with the thesis of Acemoglu and Robinson's *Why Nations Fail*, in which they argue that the long-term success or failure of nations hinges on the inclusiveness and adaptability of their institutions.[18] In this view, the nation itself is the organization, and its "institutions" constitute the protocols that contour its potential system-level behavior. Nations whose institutions co-regulate each other in a stable and productive way succeed (that is, themselves become institutions), whereas nations whose institutions lack stabilizing feedback loops fail.

neer: Game Theory, Experimentation, and Computation as Tools for Design Economics," Alvin E. Roth discusses how economists can design and implement protocols to create reliable and efficient market behaviors.[22] The canonical example is the development of kidney exchange programs, where incompatible patient-donor pairs are matched through carefully engineered algorithms to facilitate life-saving transplants. These protocols ensure that the exchanges are both efficient and fair, maximizing the number of successful transplants while adhering to ethical standards.[23]

Another example of an engineered economic system is a platform economy. A platform economy provides software and serves as an intermediary to facilitate exchanges between users. Rochet and Tirole's research on platform competition in two-sided markets highlights how these platforms are designed to efficiently match supply and demand, often employing sophisticated algorithms to clear transactions.[24] While academically, market design focuses on efficiency and fairness, it is important to critically examine these platforms through the lens of Beer's POSIWID principle: Despite their stated goals, many platform economies appear empirically to optimize for profit extraction, benefiting shareholders while only questionably serving users, and often inducing other negative externalities.[25] This observation suggests a need for further scrutiny to ensure that these engineered markets serve the broader public interest.[26]

4. An Approach to Institutional Engineering

Supplementing the insights of institutional economics with the rigor and precision of control systems engineering gives us the necessary materials to lay out a general yet *practical* approach for institutional engineering; that is, one that can be *applied* by those creating or governing concrete organizations with specific purposes, insofar as these purposes can be understood as desired "patterns of behavior" which must be rendered "stable" (in the control-theoretic sense) through the design, implementation, and governance of protocols. Despite our use of formal methods, we acknowledge the fundamental subjectivity in the act of representation, as well as the inevitable influence such representations have on the systems they represent. Our confidence that methods for control engineering are appropriate for addressing challenges in institutional engineering emanates from the fact that control systems engineers have successfully integrated economic mechanism design methods into cyber-physical systems such as transportation networks and power grids.[27]

A fundamental challenge in designing systems meant to *pattern* (and thus, to *regulate*) human behavior stems from the need to preserve individual autonomy even as one seeks to stabilize desirable properties in system-level behavior. Inevitably, governance decisions — *i.e.* those which act on an organization's protocols — create winners and losers; "the questions of central relevance to governance of impersonal organizations are those surrounding disagreement or dispute among members," as Eric Alston writes in "Governance As Conflict."[28] In "Aligning Decentralized Autonomous Organizations to Precedents in Cybernetics," Zargham and Nabben adapt the phrase "control surface" to describe the "governance surface" through which decentralized organizations evolve their structures by modifying their protocols — both technically, via smart contract code, and socially, via community covenant documents.[29]

4.1. *Institution as Dynamical System*

In "A State-Space Modeling Framework for Engineering Blockchain-Enabled Economic Systems," Zargham *et al.* introduce two key ideas: (1) blockchain networks can be understood as second-order dynamical systems and (2) that Lyapunov-like Value functions can be used to encode and preserve desirable properties within crypto-economic systems.[30] In "Foundations of Cryptoeconomic Systems," Voshmgir and Zargham argue that cryptoeconomic systems are fundamentally a governance technology, and can be approached like cyber-physical systems.[21] Zargham and Shorish build on this intuition in "Generalized Dynamical Systems: Foundations," generalizing the state-space framework presented in "A State-Space Modeling Framework for Engineering Blockchain-Enabled Economic Systems" to control systems with non-numerical state-spaces, and relaxing the control conditions by restricting admissible actions rather than specifying a unique control action.[31]

This innovation – replacing the control law with a restriction on admissible actions – provides the basis for engineering institutions through the design of protocols. Figure 1 demonstrates the intuition for reasoning about governance in terms of second-order dynamical systems.

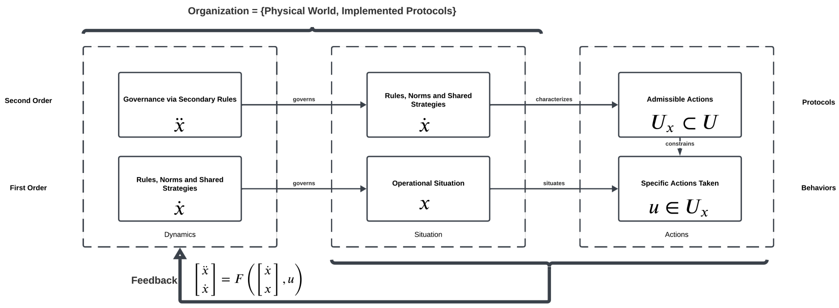

Fig. 1. Control framework for governance, rules, norms, and actions. The system evolves dynamically through feedback and constraints. For simplicity, *laws of motion* representing the physical world are compressed into the differential topology characterized by $F : Z \times U \to \partial Z$.

The framing allows us to reason about the systems as a differential inclusion. Let $z = \begin{bmatrix}\dot{x}\\ x\end{bmatrix}$, then the dynamics $\dot{z} = \begin{bmatrix}\ddot{x}\\ \dot{x}\end{bmatrix}$ are given by

$$\dot{z} \in \{F(z, u) | u \in U_{x(z)}\}$$

where the set of admissible actions $U_x(z)$ are still state-dependent in the full state z. The resulting system is non-deterministic, but is nonetheless *structured by* the "organization," and it realizes behavior in the form of a sequence of values u and z. Such a system potentially has many stable properties, despite the variation in time and non-determinism; these stable properties can be described via Lyapunov-like value functions V, as long as those functions can be thought of as loss functions (to borrow a concept from Machine Learning), and the properties they encode are oriented at their zero value. We are then interested in how the protocol regulates the differential of the loss function:

$$\dot{V} \in \nabla V(z)^\top F(z, u), \quad \forall u \in U_{x(z)}$$

and ultimately, we can claim the property represented by the value function or loss function V encodes a stable emergent property only if

$$\sup_{u \in U_{x(z)}} \nabla V(z)^\top F(z, u) \leq 0.$$

That is to say, if even without knowing the individual choices u, the introduction of the protocol into the organization stabilizes the property encoded by V, in the sense of Lyapunov.

4.2. *Institutional Dynamics*

The intuition is presented in continuous time because the dynamical systems canon is developed for continuous systems. Our work on protocols and institutions, however, is more naturally situated in discrete-event systems for two reasons: (1) the protocols we have studied are implemented as computer software, which is natively discrete and (2) organizations adapt themselves through discrete changes to their protocols, *i.e.* passing governance proposals and/or merging updated code into production environments. Therefore, for the remainder of this chapter, we will use the discrete-time characterization from "Generalized Dynamical Systems: Foundations."[31] Furthermore, while we retain the assumption that our model is second-order — that is, that a state contains within itself the rules, norms and shared strategies that determine actors' admissible action sets — we will collapse our notation to simply denote that state $x \in X$, with the discrete-event dynamics represented as $f : X \times U \to X$.

Let an **Organization** be the concrete action-situation, structured by *implemented* protocols and the physical world, in which actors take actions, and the consequences of those actions are realized as changes to that organization's state. We *represent* the organization as a generalized dynamical system (GDS):

$$x^+ = f(x, u)$$

where $x \in X$ represents the state of the organization $u \in U$ represents the actions taken by actors, and $x^+ \in X$ represents the state of the organization resulting from the actions u. The unconstrained U represents the set of all the *physically possible* actions u, which must be distinguished from the set U_x representing the *admissible* actions in any given state — those that satisfy the organization's already-implemented norms, rules and shared strategies.[3] For such a dynamical system there is realization

$$(x_0, u_0), (x_1, u_1), (x_2, u_2), \ldots (x_t, u_t), \ldots$$

which may exhibit desirable (or undesirable) patterns of behavior. Some of the patterning is attributable to the structure of the organization (which we have represented as $f(x, u)$, but some of the patterning is attributable to regularities in the choices of u given x, and knowledge or beliefs *about* $f(x, u)$. In control engineering, one asserts a control law $u = g(x)$ with

[3]While some systems, such as cryptographically enforced peer-to-peer networks, can strictly enforce their protocols, we do not assume perfect compliance in practice. Rather, we rely on control-theoretic techniques for *bounding the consequences of non-compliance.*

the purpose of achieving a property represented as $V(x_t) \leq \epsilon$. This is accomplished by learning or deriving a model of "the plant" — in our case the "organization" — and then ensuring the control rule satisfies the condition that

$$V(x^+) = V(f(x, g(x))) \leq V(x).$$

Although our model $f(x, u)$ is surely an imperfect representation, and our control action is imperfectly translated into the world, these representations help us to create the circumstances under which the organization exhibits the property $V(x) \leq \epsilon$. That is to say that each of

$$V(x_0), V(x_1), V(x_2), \ldots, V(x_t), \ldots$$

satisfy the property $V(x) \leq \epsilon$, and that we expect the property to be satisfied in the foreseeable future.

Let us define an **Institution** to be an enduring regularity in the behavior of an organization. In our framework, the institution is a property $V(x) \leq \epsilon$ for an organization represented by $x^+ = f(x, u)$. We do not wish to "control" the behavior of the actors within the organization, however, so we will not assert a control law $g(x)$ that the actors within that organization are commanded to abide.

But how, then, can we ensure the stability of the property $V(x) \leq \epsilon$. Are we simply at the whim of the actors choices? We are — *unless* we institute a **Protocol**. A protocol is a set of constraints introduced into an organization to induce a desirable property; in other words, it is a state-dependent subset $U_x \subset U$ that offers our property-preserving guarantee

$$\Delta V(x) = V(f(x, u)) - V(x) \leq 0 \quad \forall u \in U_x, \forall x \in X.$$

This expression does not guarantee a reduction in $V(x)$ if we cannot guarantee $u \in U_x$, but it sets us up to write supermartingale and Lyapunov drift-based stability proofs if we can bound the nature, frequency, and/or magnitude of violations of the protocols within a particular organization being modeled. Control theory gives us the necessary equipment to achieve stability of the desired properties in the concrete system *because* it recognizes that there is never a perfect correspondence between a model's abstractions and the realized behavior of the system being modeled, and thus offers machinery that enables us to take this asymmetry into account.

4.3. *Ontology*

- **Individual Behavior** ($u \in U$): Individual actors make their own choices based on factors that are not known or knowable to system modelers or designers. These choices influence the evolution of the organization but are not strictly determined by any external control.

- **Pattern of Behavior** ($g : X \to U$): A representation of *observed patterns* in individual actions as state-dependent policies. This captures how individual behavior tends to correlate with specific states of the organization.
- **Situation** ($x \in X$): The *state of the organization* at a given time. Actors take their concrete choices with **partial awareness** of the system's state and dynamics. They have beliefs but not full knowledge of the consequences of their actions.
- **Dynamics** ($f : X \times U \to X$): A representation of how the organization's *state evolves* as a result of the actions taken by individuals. It models the transition process from one situation to another.
- **Protocol** ($U_x \subset U$): A set of *humanly constructed rules, norms, or shared strategies* that specify a subset of actions available to individuals in a given situation, thereby coordinating individual behavior.
- **Property** ($V : X \to \mathbb{R}_+$): A function that can be understood as a *loss function encoding a pattern* in the system's state. It provides a measure of how well the system aligns with some desired or undesired behavior.
- **Institution** ($\Delta V(x) \leq 0$): An *enduring pattern of behavior* where $V(x)$ remains stable (in the sense of Lyapunov) over time, indicating that the system maintains certain properties despite variations in specific situations and individual actions.

4.4. *Implications of this Approach*

The precise innovation of GDS is the observation that, mathematically speaking, it is not necessary to assert a control law $g(x)$ to produce a desirable property—it suffices to introduce a protocol that limits actions *just enough* to ensure the property of interest. It is worth noting that although institutional economists often identify patterns in $g(x)$ as institutions, we can can use $V(x)$ without loss of generality simply by defining $V(x)$ as a loss function expressing the deviation of the $V(x_t) = \|u_t - u_t^*\|^2$ where u_t is the observed behavior and $u_t^* = g(x_t)$, the expected behavior.

The benefit of building on the control engineering literature to develop this approach is that doing so gives us the composing and nesting properties of control systems practically for free. For general results, see Jan C. Willems's "The Behavioral Approach to Open and Interconnected Systems."[32] For a GDS-specific discussion of composability, "Block Diagrams for Categorical Cybernetics."[33]

While this method provides a clear path for the analysis of institutions as persistent patterns in the behavior exhibited by organizations, and for the design of protocols via control-theoretic methods in order to modify or stabilize those institutions, it also invites important questions regarding the degrees of restriction to be tolerated, relative to the benefits of the desired stable properties.

As shown in "A State-Space Modeling Framework for Engineering Blockchain-Enabled Economic Systems,"[30] users of the Bitcoin Network tolerate the restriction "no double-spends" to induce the global property of a stable total supply of 21 million Bitcoin. Cryptographically-secured software systems, however, are significantly more susceptible to this type of formal analysis than organizations where the *rules in use* may deviate significantly from the *rules in form*.[34]

Furthermore, some of our most important institutions emerge from organizations in which the choices actors make with their degrees of freedom contribute meaningfully to the institution. Markets, for example, do not exist to *set* prices; they clear transactions and *discover* prices. In "Economic Games as Estimators," Zargham *et al.* show that a particular set of protocols implemented in automated market makers have the property of being "price estimators"; the price estimate is realized when individual actors exercise their freedoms within the protocol.[35] In the kidney exchange example from Roth, the key properties are fairness and efficiency, but there is no notion of price.[23]

In practice, organizations are heterogeneous and distributed — entangled with each other such that we, as modelers, must make choices about what is in and out of scope. Modeling involves subjective decisions, such as how to divide stakeholders into groups with similar action-situations as well as similar incentives and preferences.[4] Continuing with our control engineering approach, suppose the actors are denoted $i \in S$ and that each actor has a set of observables $y_i \in Y_i$ and controllables $u_i \in U_i$, representing what they *can see* and what they *can do* respectively. The realized behav-

[4]While the use of advanced modeling techniques can sometimes be viewed as problematic for questions in political science, we appeal to Primo and Clark's "A Model Discipline" for the argument that models are an integral part of political science.[36] Keeping with an engineering ethos, we strive to use models as tools in service to the public, rather than as a means of manipulation. Our engineering-informed approach to modeling is documented in "Method for Functional Decomposition of Organizations and Their Environments."[37]

ior at any time is the Cartesian product of all the individual actor behavior $U = \bigotimes_{i \in S} U_i$, as informed by their observables and constrained by their respective controllables. This consideration attends to North's insistence on the separation of the actors from the rules of the game; it also provides the basis for the dynamics to be distributed across local action situations k, such that $f(x, u) = [\cdots, f_k(x, u), \cdots]$ to be set up in a manner that represents complex and varied local actuation-situations, including topologies with non-trivial power disparities between actor groups. The topology of the network can take on arbitrary causal patterns such as those described in Multi-Agent Influence Diagrams (MAIDs).[38] Furthermore, modeling overlapping action-situations within such a representation $f(x, u)$ provides a more formal perspective on what Vincent and Elinor Ostrom's work refers to as *polycentricity*[39,40] — if we also introduce a plethora of distinct $V_j(x)$ representing a range of dynamically entangled properties, consistent with North's "institutional matrix" and with Abigail Devereaux's work in "Multiple Games Analysis: A Petri Dish for Growing Polycentric Orders."[41]

Equipped with these additional concepts, we can turn our attention to developing models of organizations that enable us to explore the impact of potential protocols on those organizations' system-level behaviors. Leaping directly to modeling, however, would be akin to building a house without blueprints or a stable foundation; instead, a rigorous and responsible engineering process must begin by surveying the field and mapping the critical loci at which meaningful interventions might be designed. To demonstrate this approach, we shall therefore generate preliminary mappings of the distributed organizations of *science* and *engineering* — first enumerating a set of actor groups, then characterizing their local action-situations, and finally describing how those situations are coupled through dynamics. Doing so will enable us to identify some candidate solutions that later stages of the design process could explore through formal modeling.

5. Preliminary Mappings

In accordance with our ontology, science and engineering are actually large-scale distributed *organizations*; those organizations exhibit long-term stable patterns of behavior, which are their associated institutions. Science studies phenomena in the world and produces, validates, and distributes reliable shared knowledge about the world. Engineering applies the reliable shared knowledge to create or otherwise modify the world. In this section, we will factor the organizations apart to examine how their protocols contribute to the (in)stability of their defining properties.

5.1. *Science*

Our first example, **science**, is inextricably linked with a defining protocol: the "Scientific Method." We shall thus begin by examining this protocol, and work backwards to characterize first the system-level property that it is meant to induce or stabilize, and next the concrete organization that we hope will exhibit this property in its realized behavior. Then, we shall consider what forces are currently causing the relevant property to be destabilized, and examine how new protocols might be designed specifically to counteract these forces.

According to the Oxford English Dictionary, the scientific method is generally understood to consist of "(a) systematic observation, measurement, and experimentation, (b) induction and the formulation of hypotheses, (c) the making of deductions from the hypotheses, (d) the experimental testing of the deductions, and (if necessary) (e) the modification of the hypotheses."[42] In *Scientific Method in Practice*, however, Hugh G. Gauch Jr. quotes the American Association for the Advancement of Science in order to highlight the ways that "[t]he scientific method 'is often misrepresented as a fixed sequence of steps,' rather than being seen for what it truly is, 'a highly variable and creative process.'"[43] Despite the adherence to the scientific method across "the various scientific disciplines," the AAAS carefully explains elsewhere, "scientists differ greatly from one another in what phenomena they investigate and in how they go about their work; in the reliance they place on historical data or on experimental findings and on qualitative or quantitative methods; in their recourse to fundamental principles; and in how much they draw on the findings of other sciences."[44]

The degrees of freedom that the scientific method affords its practitioners in terms of *process* exist in order to induce a stable *property* in the concrete behaviors through which that process is realized — specifically, the property of *producing reliable shared knowledge.* The scientific method thus functions as a minimal set of common constraints adopted by actors within what we shall call "the organization of science" *for the purpose* of stably inducing this property in the organization's realized system-level behavior; as Gauch reminds us, it is "[p]recisely because general principles are general [... that] the entire scientific community has a single, shared set of principles" related to a fundamental unifying aspiration. For our purposes, then, we shall characterize the "organization of science" — that is, the concrete system whose realized behavior we are concerned with — very broadly.

5.1.1. *Actors and Local Action-Situations*

Table A.1 in **Appendix A** identifies the key **actors** within the scientific ecosystem, and describes their **local action-situations**.Each actor operates based on a set of *observables*—information they can access about the system—and *controllables*—actions they can take to influence the system. Understanding these relationships helps clarify the decision-making processes of different stakeholders in scientific research, from researchers and peer reviewers to funding agencies and government regulators.

Researchers, for example, can *observe* **published literature, raw data, experimental results, peer reviews, and funding opportunities**, all of which influence the set of decisions under their *control*: **which funding opportunities they pursue, how they design their experiments, what papers they seek to publish (and in which publications), how they review the work of peers, and which peers they collaborate with**. Meanwhile, **journal editors** can *observe* **submitted papers, reviewer feedback, journal impact factors, and trending topics in their fields**, and can *control* **what sorts of submissions to solicit, whether they accept or reject a given paper, which revisions they suggest, and how they enforce formatting and other standards**.

Modeling **actors** and their **local action-situations** in this way gives us the materials that we need in order to begin describing the system's **dynamics** — the "laws of motion" that determine how situation-level actions interact to produce system-level results. By identifying these connections, we gain insight into how institutional behavior emerges from the interplay of individual actions and external constraints.

5.1.2. *Dynamics Coupling Action-Situations*

Scientific dynamics do not function in isolation; rather, they exhibit interconnected decision-making processes where actions taken by one group influence others. **Table A.2** in **Appendix A** explores how **action-situations are dynamically coupled**, detailing how various roles interact through **systemic feedback loops**. **The process of peer review**, for example, links **researchers** (who conduct experiments and submit papers for publication, **peer reviewers** (who evaluate research quality and validity, influencing what gets published), and **journal editors** (who assign peer reviewers to submissions, as part of the process of deciding which submissions meet their journals' standards, and should therefore be published). Similarly, the decisions made by **funding agencies** (about what research to fund) has an impact on how **university administrators** allocate budgets — influencing universities' hiring decisions, departments' research priorities, and even the resources that are available to specific labs.

The Scientific Method serves as a protocol within this system of dynamically-coupled action-situations, constraining the set of admissible actions in local action-situations in ways that stabilize specific properties in the system's global behavior.

5.1.3. *Actors, Incentives, and Regulation by Scientific Method*

The stable system-level behavior of knowledge production does not emanate from any one action-situation or any one actor group's pattern of behavior. Instead, the Scientific Method constrains decisions within individual action-situations, incentivizing actors to behave in ways that stabilize specific system-level properties. **Table A.3** in **Appendix A** details the incentives of various actor-groups within the distributed system of science, and shows how the scientific method structures the ways that they pursue these incentives. **Researchers**, for example, are incentivized to pursue **career advancement, recognition, funding, and meaningful impact**; the scientific method, however, requires research to be **reproducible, validated by peers, and evidence-based in its conclusions**. These requirements **discourage researchers from faking their findings, as they pursue prestige and/or funding**.

5.1.4. *Failures and Interventions*

To claim that the institution of science is breaking down is to suggest that the distributed organization described above is struggling to stabilize the production of reliable knowledge as a property of its realized behavior — and one need look no further than the "reproduction crisis" in the sciences to see the truth of this description.[45–47] While the scientific method itself continues to function as an effective protocol, other protocols within the organization of science — for example, those related to how prestige is factored, and what kinds of work is therefore incentivized — are falling out of alignment with *the property that we are attempting to stabilize.* For example, "publish or perish" dynamics enforce a paradigm in which work is valued in terms of its *quantity* rather than its *quality*.[48,49] Meanwhile, vital but unexciting reproducibility studies are neglected in favor of a steady stream of incompletely-verified "breakthroughs" or densely cross-cited debates over terminology.

Stabilizing "the institution of science" (i.e. the property of the organization's system-level behavior that we are seeking to reinforce) can thus be undertaken through the implementation of new protocols designed to bring such incentives back into alignment with the purpose they are meant to serve. Funding and evaluation protocols can be restructured to prioritize reproducibility, collaboration, and long-term contributions over short-term metrics; researchers could agree, for example, to undertake one reproducibility study for every three novel investigations in which they participate.[50] Similarly, journals and funding agencies could implement new peer-review standards focused on societal impact and methodological soundness, and open science initiatives, shared data repositories, and multi-institutional collaborations could be scaled to foster rigor and transparency.[51] Protocols such as these would constrain the choices of local actors in ways that stabilize the institution of science — and, crucially, would do so while *preserving* the degrees of freedom through which creativity, innovation, and individuality are able to enter the broader system. Such interventions are already being undertaken from the bottom up, as citizen science and decentralized science movements are gaining cultural momentum (and in some cases access to capital); others, meanwhile, are trying to restructure the organization of science from the top down, as the government of the United States attempts to freeze all federal research funding and massively refactor the relevant administrations.

5.2. *Engineering*

Let us now turn our attention to **engineering**. As was the case with science, we shall first suppose that the property that characterizes the institution is the same property that its defining protocol — in this case, the *Engineering Design Process* (EDP) — is intended to stabilize. Let us, therefore, examine the protocol, and reason our way to a characterization of its function from an analysis of the constraints that it imposes.

As Morgan *et al.* note, "[t]here are many variations [of the EDP] in practice today, but most of them include the same basic steps": "Identify Problem and Constraints," "Research," "Ideate," "Analyze Ideas," "Build," "Test and Refine," and "Communicate and Reflect."[52] In their account, "the design process is a systematic approach followed when developing a solution for a problem with a well-defined outcome." Furthermore, they explain, "the process is, by nature, iterative [...] until the final design solution is identified." Fidai *et al.* add that "[t]he overarching goal of engaging in the engineering design process is to create solutions to any given problem while staying within the boundaries of provided constraints, resources, time, and budget."[53] Meanwhile, Nancy Leveson's work offers an urgent reminder that *safety* and *reliability* should always be considered critical properties that implemented systems must seek to preserve.[54]

At its core, engineering is not only about the creation of physical structures or devices but also about the careful alignment of multiple actors, interests, and constraints. As Lucy Suchman demonstrates in her analysis of bridge building, engineering work involves continuous negotiation between material affordances, human intentions, and institutional frameworks in ways that make engineering as much a process of *organizing alignment* as it is one of designing objects.[55] Engineering systems, then, are not merely artifacts; they are social and technical assemblages whose stability emerges from the coordination of physical materials with shared strategies, norms, and formalized rules.

This complexity is formally acknowledged in **systems engineering**, which the International Council on Systems Engineering (INCOSE) defines as "an interdisciplinary approach and means to enable the realization of successful systems." Systems engineering concerns itself with the integration of various subsystems—each with its own constraints and dynamics—into a coherent whole that functions reliably in its intended operational environment.[56] Relatedly, INCOSE defines **cyber-physical systems** as "engineered systems that are built from, and depend upon, the seamless integration of computational algorithms and physical components."[57] These definitions underscore how modern engineering does not merely apply scientific knowledge; it actively structures the interaction between humans, human-designed artifacts, automated systems, and the broader physical world.

The iterative nature of the engineering design process is reflected in the rigorous practices of **Verification and Validation (V&V)**, which ensure that engineered systems perform their intended functions both correctly (*verification*) and appropriately within their intended context (*validation*).[58] As Dieter and Schmidt explain, "[t]he design method is very similar to the scientific method if we allow for differences in viewpoint and philosophy," foremost among which is the fact that engineering is motivated primarily by "the needs of society," as opposed to "scientific curiosity."[59]

Insofar as we have characterized the primary function of the institution of science as *the production of reliable shared knowledge*, let us therefore characterize engineering's societal function as *the transformation of knowledge into practical, safe, and reliable systems that meet human needs*. Like the scientific method, the engineering design process does not prescribe the concrete actions of individual actors, but instead operates as a set of shared constraints adopted by actors within the organization of engineering for the purpose of stabilizing its ability to realize this function.

5.2.1. *Actors and Local Action-Situations*

Table B.1 in **Appendix B** defines the key **actors** within the public infrastructure engineering ecosystem, and describes their **local action-situations**. Since public infrastructure serves the general public, key roles include **public involvement activities**, **policy-setting functions**, and **community engagement**. **Design Engineers**, for example, can *observe* **infrastructure requirements, environmental constraints, safety standards, and material properties**, which influence the decisions under their control, such as **how to develop specifications, which materials to select, and whether to iterate on a particular design**. Similarly, **project managers** can *observe* their projects' **budget, timeline, stakeholder requirements, and political constraints**, in order to decide **how to allocate funding, manage risks, and adjust the project's scope**.

5.2.2. *Dynamics Coupling Action-Situations*

Public infrastructure projects involve multiple interacting decision-making processes. **Table B.2** in **Appendix B** explores how such **action-situations are dynamically coupled** — particularly, how engineers, regulators, and the public interact. The architectures defined by **systems engineers**, for example, are coupled with the detailed specifications developed by **design engineers** insofar as **a system architecture constrains subsystem design**; design engineers must work within the framework it provides. Similarly, when infrastructure enters operational use, **the public provides feedback on its usability and performance**, which feeds back to **engineers** in the form of future design improvements.

5.2.3. *Actors, Incentives, and Regulation by the Engineering Design Process*

Table B.3 in **Appendix B** shows how distributed organization of engineering's stable system-level behavior of *creating and maintaining safe, reliable, and publicly accountable infrastructure* emerges from the interactions between engineers, policymakers, regulators, and the general public. These different actors are motivated by distinct incentives, but their behaviors are regulated by shared constraints within the Engineering Design Process. **Quality Assurance (QA) Engineers**, for example, are incentivized to **reduce failures** and **ensure regulatory compliance**; in the process, however, they are constrained both by **legal safety requirements** and by **the quality control procedures and standard testing protocols required by the EDP**. Similarly, **Regulatory Agencies** are incentivized to **ensure public safety and enforce legal compliance** as they go about the processes of **defining infrastructure codes, conducting inspections, and approving projects**. **The public**, meanwhile, is incentivized to **help engineers and regulators develop safe, reliable, and cost-effective infrastructure** — which they can do by participating in public hearings, or seeking to influence policy through democratic processes.

5.2.4. *Failures and Interventions*

To claim that the institution of engineering is breaking down is to suggest that it is failing to stabilize the transformation of knowledge into practical, safe, and reliable systems that meet human needs. Unlike science, where failure manifests as a reproducibility crisis, engineering failures emerge through misaligned incentives, accountability gaps, and prestige distortions. One critical issue is the entrenchment of *CYA ("cover your ass") norms*, where avoiding liability outweighs ensuring safety.[54,60] Defensive documentation and bureaucratic risk-shifting often take precedence over robust, proactive design, leading to under-maintained but legally insulated infrastructure. Meanwhile, *political incentives favor vanity projects over essential maintenance*, leaving roads, bridges, and public utilities to deteriorate while funds are diverted to high-visibility initiatives.[61] The *cultural prestige of engineering* has also shifted. Traditional infrastructure and civil service roles have been devalued, while *software development* — aggressively distancing itself from engineering licensure and regulation — has redefined "engineering" in ways that undermine professional accountability.[62,63] As prestige migrates toward startup founders building consumer apps, public-focused engineering loses talent and influence, exacerbating the neglect of vital systems.

Several interventions targeted at these failure points are already gaining traction. *Public frustration over infrastructure neglect* has driven reinvestment in maintenance and resilience projects,[61] and *liability reforms are increasingly debated* as regulators grapple with failures in both physical and digital systems.[54] Efforts to *reintroduce professional accountability into software engineering* — including discussions around licensing requirements for work related to AI safety and cybersecurity — also reflect broader *concerns over deregulation in critical systems engineering*.[62] Whether these changes will succeed in realigning the profession with its fundamental mission, however, remains an open question, with its answer contingent on the implementation of *fit for purpose protocols* that constrain misaligned incentives while preserving the variety necessary to serve diverse stakeholders.

6. Conclusion

Rehabilitating society's institutions is not just about fixing what is broken — it is about designing for a future where stability, variety, and adaptability co-exist. Historically, organizations have been controlled through document-based policy paradigms, where governance relied on formalized rules, laws, and regulations.[64] Today, we are witnessing a shift toward a data-driven, software-mediated protocol paradigm, where intra- and inter-organizational processes are increasingly embedded in digital systems, algorithms, and automated governance mechanisms.[65] This transformation offers both opportunities and risks, as it reconfigures how institutions stabilize behavior while allowing for adaptive evolution.

Science has entered the digital era in an *ad hoc* manner, leaning heavily on digital publishers, citation metrics, and algorithmic evaluation measures that have arguably destabilized, rather than reinforced, its core epistemic properties. Promising endeavors such as NumFOCUS's "OpenSource Science Initiative" and decentralized scientific infrastructures like the DeSci Network, however, suggest that a more deliberate and constructive transition is possible.[66,67] Similarly, engineering has confronted the challenges of digital transformation with model-based systems engineering (MBSE), successfully integrating system design into digital workflows, executable models, and dynamic document generation.[68,69] Meanwhile, open-source software communities are pioneering governance mechanisms using git protocol-based coordination tools. These approaches offer compelling templates for protocol-driven institutional design — where modularity, version control, and transparency can enhance institutional resilience.

At the frontiers of institutional design, we find work that challenges us to think beyond the constraints of legacy governance models. Kei Kreutler's exploration of designing for unprecedented situations urges us to engineer institutions that can function in radically novel environments.[70] Ellie Rennie's ethnographic study of SourceCred introduces the concept of "terraforming" organizations, and shows how new protocols can reconstitute economic landscapes and redefine value coordination.[71] Potts *et al.* extend this line of inquiry with their theory of contribution goods, where value is dictated by protocols of production and exchange, rather than traditional market pricing.[72] In a similar vein, Austin Wade Smith conceptualizes ecological institutions, whose explicit function is to stabilize the fragile and visibly unhealthy (and unstable) relationship between human society and ecological systems.[73]

These emerging frameworks suggest that protocolization provides a unifying lens for institutional design. By decomposing institutions into their constituent protocols, we can pinpoint failure points, reinforce stability, and enhance adaptability in a modular, recursive fashion. This shift from monolithic structures to dynamic, reconfigurable systems enables organizations to respond to complex challenges like climate change, public health crises, and the digitization of civic engagement — without becoming brittle or obsolete.

We began this chapter by reviewing literature from institutional economics, cybernetics, and other related fields. We then presented a control-theoretic approach to institutional design that focuses on stabilizing desirable properties via the introduction of protocols. We explored this conception of institutions through preliminary mappings of the distributed systems of science and engineering, but acknowledge these are still document-based rather than model-based organizations. Finally, we extrapolated the digitalization trend to hypothesize that a control-theoretic approach to designing and governing institutions via protocols is not only possible but necessary, given the complexity of our society and the severity of its issues.

In the end, we believe our conception of institutions enables a more rigorous approach to their design, and opens up new possibilities for their renovation. We have found that top-down control and rigid rules are not required in order to implement organizations that can withstand shocks while supporting diverse forms of local and regional action; the urgent (but achievable) work of re-stabilizing our institutions can be accomplished with *protocols*. By aligning organizations with their intended purposes while embedding adaptability at their core, we can lay the groundwork for a society that is resilient, cooperative, and capable of thriving amid uncertainty.

References

1. D. C. North, *Institutions, Institutional Change and Economic Performance*. Cambridge University Press, Cambridge, UK (1990).
2. E. Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press, Cambridge, UK (1990).
3. S. E. S. Crawford and E. Ostrom, A grammar of institutions, *American Political Science Review*. **89**(3), 582–600 (1995).
4. G. M. Hodgson, What are institutions?, *Journal of Economic Issues*. **40**(1), 1–25 (2006). doi: 10.1080/00213624.2006.11506879.
5. S. Beer, *Designing Freedom*. John Wiley & Sons, London (1974).
6. S. Beer, What is cybernetics?, *Kybernetes*. **31**(2), 209–219 (2002). doi: 10.1108/03684920210417283.
7. H. K. Khalil, *Nonlinear Systems*, 3rd edn. Prentice Hall (2001). ISBN 0-13-067389-7.
8. E. D. Sontag, *Mathematical Control Theory: Deterministic Finite Dimensional Systems*, 2nd edn. Springer (1999). ISBN 0-387-98489-5. URL `https://link.springer.com/book/10.1007/978-1-4757-3048-7`.
9. G. Beuster, O. Leistert, and T. Röhle, Protocol, *Internet Policy Review*. **11**(1) (2022). doi: 10.14763/2022.1.1651. URL `https://doi.org/10.14763/2022.1.1651`.
10. N. A. Lynch, *Distributed Algorithms*. Morgan Kaufmann, San Francisco (1996).
11. A. R. Galloway, *Protocol: How Control Exists After Decentralization*. MIT Press, Cambridge, MA (2004).
12. S. S. Kia, B. Van Scoy, J. Cortes, R. A. Freeman, K. M. Lynch, and S. Martinez, Tutorial on dynamic average consensus: The problem, its applications, and the algorithms, *IEEE Control Systems Magazine*. **39**(3), 40–72 (2019). doi: 10.1109/MCS.2019.2900783.
13. H. Uchiyama, H. Ishii, and K. Sakurama, *Emergent Behavior Detection and Task Coordination for Multiagent Systems*. Springer (2022). ISBN 978-3-030-86895-6. doi: 10.1007/978-3-030-86896-3. URL `https://link.springer.com/book/10.1007/978-3-030-86896-3`.
14. J. S. Shamma, ed., *Cooperative Control of Multi-Agent Systems: A Consensus Region Approach*. Automation and Control Engineering, Springer, New York (2007). ISBN 978-0-387-69171-0. doi: 10.1007/978-0-387-69171-0.
15. N. Wiener, *Cybernetics: Or Control and Communication in the Animal and the Machine*. MIT Press, Cambridge, MA (1949).
16. R. Axelrod, *The Evolution of Cooperation*. Basic Books, New York, NY (1984). ISBN 9780465021215. URL `https://www.basicbooks.com/titles/robert-axelrod/the-evolution-of-cooperation/9780465005642/`.

17. W. H. Sandholm, *Population Games and Evolutionary Dynamics*. MIT Press, Cambridge, MA (2010). ISBN 9780262195874. URL https://mitpress.mit.edu/9780262195874/population-games-and-evolutionary-dynamics/.
18. D. Acemoglu and J. A. Robinson, *Why Nations Fail: The Origins of Power, Prosperity, and Poverty*. Crown Business, New York (2012).
19. M. Zargham, J. Zartler, K. Nabben, R. Goldberg, and J. Emmett, Disambiguating autonomy, *Zenodo* (2023). doi: 10.5281/zenodo.8239311. URL https://doi.org/10.5281/zenodo.8239311.
20. T. Swann, *Anarchist Cybernetics: Control and Communication in Radical Politics*. Bristol University Press (2020). ISBN 978-1-5292-0879-5. URL https://bristoluniversitypress.co.uk/anarchist-cybernetics.
21. S. Voshmgir and M. Zargham, Foundations of cryptoeconomic systems, *Research Institute for Cryptoeconomics Working Paper Series* (2020).
22. A. E. Roth, The economist as engineer: Game theory, experimentation, and computation as tools for design economics, *Econometrica*. **70**(4), 1341–1378 (2002). doi: 10.1111/1468-0262.00335. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-0262.00335.
23. A. E. Roth, T. Sönmez, and M. U. Ünver, Kidney exchange, *The Quarterly Journal of Economics*. **119**(2), 457–488 (2004). doi: 10.1162/0033553041382157.
24. J.-C. Rochet and J. Tirole, Platform competition in two-sided markets, *Journal of the European Economic Association*. **1**(4), 990–1029 (2003). doi: 10.1162/154247603322493212. URL https://academic.oup.com/jeea/article/1/4/990/2280902.
25. N. Srnicek, *Platform Capitalism*. Polity Press (2017). ISBN 978-1-5095-0486-2. URL https://www.politybooks.com/bookdetail/?isbn=9781509504862.
26. J. Shorish. Market design lessons for 'safety first' institutional engineering. https://blog.block.science/market-design-lessons-for-safety-first-institutional-engineering/ (2023). Accessed: 2025-01-29.
27. I. V. Chremos and A. A. Malikopoulos, Mechanism design theory in control engineering: A tutorial and overview of applications in communication, power grid, transportation, and security systems, *IEEE Control Systems Magazine*. **44**(1), 57–78 (2024). doi: 10.1109/MCS.2023.3329919. URL https://ieeexplore.ieee.org/document/10384599/.
28. E. Alston, Governance as conflict: Constitution of shared values defining future margins of disagreement, *MIT Computational Law Report* (dec 5, 2022). https://law.mit.edu/pub/governanceasconflict.
29. M. Zargham and K. Nabben, Aligning 'decentralized autonomous organization' to precedents in cybernetics, *MIT Computational Law Report* (nov 20, 2023). https://law.mit.edu/pub/dao-precedents-cybernetics.
30. M. Zargham, Z. Zhang, and V. Preciado, A state-space modeling framework for engineering blockchain-enabled economic systems, *arXiv preprint* (2018). URL https://arxiv.org/abs/1807.00955.
31. M. Zargham and J. Shorish, Generalized dynamical systems: Foundations,

WU Vienna University of Economics and Business. Working Paper Series (2022). Pre-publication working paper.

32. J. C. Willems, The behavioral approach to open and interconnected systems, *IEEE Control Systems Magazine*. **27**(6), 46–99 (2007). doi: 10.1109/MCS.2007.906923.
33. M. Zargham and J. Shorish, Block diagrams for categorical cybernetics, *Available at SSRN 4569037* (2023).
34. E. Ostrom, R. Gardner, and J. Walker, *Rules, Games, and Common-Pool Resources*. University of Michigan Press, Ann Arbor (1994). ISBN 978-0472065462. doi: 10.3998/mpub.9739.
35. M. "Zargham, K. Paruch, and J. Shorish. Economic games as estimators. In eds. P. Pardalos, I. Kotsireas, Y. Guo, and W. Knottenbelt, *Mathematical Research for Blockchain Economy*, pp. 125–142, Springer International Publishing, Cham (2020). ISBN 978-3-030-53356-4.
36. D. M. Primo and K. M. Clarke, *A Model Discipline: Political Science and the Logic of Representations*. Oxford University Press, New York, NY (2012). ISBN 9780195382204. URL https://global.oup.com/academic/product/a-model-discipline-9780195382204.
37. M. Zargham and I. Ben-Meir, Method for functional decomposition of organizations and their environments, *SSRN Electronic Journal* (2023). doi: 10.2139/ssrn.4606672. URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4606672.
38. D. Koller and B. Milch, Multi-agent influence diagrams for representing and solving games, *Games and economic behavior*. **45**(1), 181–221 (2003).
39. V. Ostrom, C. M. Tiebout, and R. Warren, The organization of government in metropolitan areas: A theoretical inquiry, *American Political Science Review*. **55**(4), 831–842 (1961). doi: 10.2307/1952530. URL https://www.jstor.org/stable/1952530.
40. E. Ostrom, Beyond markets and states: Polycentric governance of complex economic systems, *American Economic Review*. **100**(3), 641–672 (2010). doi: 10.1257/aer.100.3.641. URL https://www.aeaweb.org/articles?id=10.1257/aer.100.3.641.
41. A. Devereaux, Multiple games analysis: A petri dish for growing polycentric orders, *SSRN Electronic Journal* (2019). doi: 10.2139/ssrn.3477461. URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3477461.
42. Oxford English Dictionary. scientific method, n. URL http://www.oed.com (2024). Oxford University Press. Accessed January 2024.
43. H. G. Gauch Jr, *Scientific Method in Practice*. Cambridge University Press (2003).
44. American Association for the Advancement of Science. The nature of science. Online. URL https://www.project2061.org/publications/sfaa/online/chap1.htm (1990). Chapter 1 in Science for All Americans. Project 2061.
45. O. S. Collaboration, Estimating the reproducibility of psychological science, *Science*. **349**(6251), aac4716 (2015). doi: 10.1126/science.aac4716.
46. J. P. A. Ioannidis, Why most published research findings are false, *PLoS Medicine*. **2**(8), e124 (2005). doi: 10.1371/journal.pmed.0020124.

47. L. P. Freedman, I. M. Cockburn, and T. S. Simcoe, The economics of reproducibility in preclinical research, *PLOS Biology*. **13**(6), e1002165 (2015). doi: 10.1371/journal.pbio.1002165.
48. D. Fanelli, Do pressures to publish increase scientists' bias? an empirical support from us states data, *PLoS ONE*. **5**(4), e10271 (2010). doi: 10.1371/journal.pone.0010271.
49. M. Fire and C. Guestrin, Over-optimization of academic publishing metrics: Observing goodhart's law in action, *arXiv preprint* (2018). URL `https://arxiv.org/abs/1809.07841`.
50. B. Nosek and Y. Bar-Anan, Scientific utopia ii: Restructuring incentives and practices to promote truth over publishability, *Perspectives on Psychological Science*. **7**(6), 615–631 (2012). doi: 10.1177/1745691612459058.
51. F. Fang, P. Stone, and M. Tambe. When security games go green: designing defender strategies to prevent poaching and illegal fishing. In *Proceedings of the 24th International Conference on Artificial Intelligence*, IJCAI'15, p. 2589–2595, AAAI Press (2015). ISBN 9781577357384.
52. J. R. Morgan, A. M. Moon, and L. R. Barroso, *Engineering Better Projects*, In *STEM Project-Based Learning*, pp. 29 – 39. Brill, Leiden, The Netherlands (2013). ISBN 9789462091436. URL `https://brill.com/view/book/edcoll/9789462091436/BP000005.xml`.
53. A. Fidai, L. R. Barroso, M. M. Capraro, and R. M. Capraro. Effects of engineering design process on science and mathematics. In *2020 IEEE Frontiers in Education Conference (FIE)*, pp. 1–4 (2020). doi: 10.1109/FIE44824.2020.9274167.
54. N. G. Leveson, *Engineering a Safer World: Systems Thinking Applied to Safety*. Engineering Systems, MIT Press, Cambridge, MA (2012). ISBN 978-0-262-01662-9. doi: 10.7551/mitpress/8179.001.0001.
55. L. Suchman, Organizing alignment: A case of bridge-building, *Organization*. **7**(2), 311–327 (2000). doi: 10.1177/135050840072007.
56. International Council on Systems Engineering (INCOSE), *Systems Engineering Handbook: A Guide for System Life Cycle Processes and Activities*, 5th edn. Wiley (2021).
57. International Council on Systems Engineering (INCOSE), *Cyber-Physical Systems: Engineering a Converged World of Humans, Software, and Hardware*. INCOSE Publications (2020).
58. K. Y. Rozier, Verification and validation in engineering systems: An overview of formal methods, *IEEE Transactions on Systems, Man, and Cybernetics*. **52**(3), 256–270 (2022).
59. G. E. Dieter and L. C. Schmidt, *Engineering Design*, 5 edn. McGraw-Hill, New York (2012). ISBN 978-0073398143.
60. D. Vaughan, *The Challenger Launch Decision: Risky Technology, Culture, and Deviance at NASA*. University of Chicago Press (1996).
61. B. Frischmann, *Infrastructure: The Social Value of Shared Resources*. Oxford University Press (2012).
62. D. Mindell, *Our Robots, Ourselves: Robotics and the Myths of Autonomy*. Viking (2015).

63. E. Mollick, The evolution of engineering as a profession in the age of software, *Management Science* (2018).
64. R. W. Stone, *Controlling Institutions: International Organizations and the Global Economy*. Cambridge University Press, New York, NY (2011). ISBN 978-0521183062. URL https://www.cambridge.org/core/books/controlling-institutions/AB18F080C5212E9D145223144E286EA5.
65. M. Zargham and K. Nabben. Algorithms as policy. URL https://zenodo.org/record/8283145 (2023). Preprint.
66. NumFOCUS. Open source science initiative (ossci). URL https://numfocus.org/open-source-science-initiative-ossci Accessed: 2025-01-29.
67. M. Zargham, D. Sisson, C. Vandevoorde, and I. Ben-Meir. Digital public infrastructure for decentralized science. Technical report, DeSci Labs (September, 2024). URL https://nodes.desci.com/dpid/275/v4/root/DeSci_Labs_Report___9_27_24.pdf.
68. R. Karban, F. G. Dekens, S. Herzig, M. Elaasar, and N. Jankevičius. Creating system engineering products with executable models in a model-based engineering environment. In eds. G. Z. Angeli and P. Dierickx, *Modeling, Systems Engineering, and Project Management for Astronomy VII*, vol. 9911, p. 99110B, SPIE (2016). doi: 10.1117/12.2232785. URL https://doi.org/10.1117/12.2232785.
69. C. Delp, D. Lam, E. Fosse, and C.-Y. Lee. Model based document and report generation for systems engineering. In *2013 IEEE Aerospace Conference*, pp. 1–11 (2013). doi: 10.1109/AERO.2013.6496926.
70. K. Kreutler, How to design a space station, *Journal of Institutional Design* (2024).
71. E. Rennie, The credsperiment: An ethnography of institutional design, *Digital Culture & Society* (2024).
72. J. Potts, F. de la Rosa, and P. Dixon, Contribution goods: Beyond public and private goods, *Journal of Institutional Economics*. **20**(1), 1–15 (2024). Pre-publication working paper.
73. A. W. Smith, Ecological institutions: Designing for environmental stability, *Environmental Governance* (2024).

Appendix A. Preliminary Mappings: Science

Table A.1. Key Stakeholders in the Scientific Research Ecosystem

Stakeholder	Observables	Controllables
Researchers	Published literature, raw data, experimental results, peer reviews, funding opportunities	Design experiments, publish papers, review others' work, apply for funding, collaborate with peers
Peer Reviewers	Manuscripts, supporting data, author reputation (if not blinded)	Accept/reject papers, suggest revisions, highlight flaws or strengths
Journal Editors	Submitted papers, reviewer feedback, journal impact factors, trending topics in field	Accept/reject papers, desk-reject manuscripts, suggest changes, enforce formatting and standards
University Administrators	Faculty performance metrics, research grants, student feedback, university rankings	Hire/fire faculty, allocate funding, set policies on tenure and promotion
Funding Agencies	Research proposals, applicant CVs, past publications, societal needs, government policies	Approve/reject grants, allocate budgets, set funding priorities
Industry (Private Sector R&D)	Scientific advancements, market trends, IP opportunities, government regulations, profit and loss statements	Fund internal/external research, apply for patents, commercialize discoveries
Government Regulators	Research findings, public concerns, policy proposals, priorities of the current administration	Regulate ethical compliance, allocate research funding, enforce safety guidelines
General Public	News, media reports on science, educational materials	Support scientific initiatives, vote on policies, engage in citizen science

Table A.2. Key Interactions Between Stakeholders in Scientific Research

Primary Action	**Coupled With**	**Dynamics**
Researchers conduct experiments & publish	Peer Reviewers evaluate submissions	Peer reviewers assess research quality and validity, influencing what gets published
Researchers apply for funding	Funding Agencies allocate grants	Funding agencies select proposals based on merit, societal impact, and alignment with priorities
Researchers submit papers	Journal Editors decide acceptance	Editors select reviewers and determine whether a paper meets journal standards
Researchers reference past literature	Published literature evolves	New research is built on previous studies, shaping future investigations
Peer Reviewers evaluate papers	Editors decide publication	Editors rely on peer reviews to make acceptance decisions
Journal Editors select papers	University Administrators evaluate faculty	Faculty career advancement depends on publication records, reinforcing academic pressures
Funding Agencies distribute resources	University Administrators allocate budgets	Research funding influences university hiring, department priorities, and lab resources
University Administrators evaluate faculty	Researchers adapt research agendas	Faculty optimize research topics and collaborations based on tenure and promotion criteria
Industry funds academic research	Researchers conduct industry-driven projects	Industry-backed projects steer research towards commercial applications
Government Regulators enforce policies	Researchers comply with ethical and safety standards	Ethical guidelines shape research methods, limiting certain experimental approaches
Public reacts to science in media	Government Regulators & Funding Agencies	Public opinion influences policy and funding priorities for scientific fields

Table A.3. Incentives and Regulatory Functions of Key Stakeholders

Stakeholder	**Incentives**	**Regulation**
Researchers	Career advancement, recognition, funding, discovery, impact	Requires reproducibility, peer validation, and evidence-based conclusions
Peer Reviewers	Academic reputation, community respect, maintaining field standards	Enforces critical evaluation and accountability through anonymous or open reviews
Journal Editors	Increasing journal impact factor, prestige, field influence	Requires adherence to rigorous standards for publication, ensuring credibility
University Administrators	Institutional reputation, rankings, funding	Encourages adherence to research ethics, integrity, and tenure policies
Funding Agencies	Maximizing impact of research funding, political alignment	Promotes transparent evaluation of proposals, aligning with scientific priorities

Appendix B. Preliminary Mappings: Engineering

Table B.1. Key Stakeholders in Infrastructure Development

Stakeholder	**Observables**	**Controllables**
Design Engineers	Infrastructure requirements, environmental constraints, safety standards, material properties	Develop specifications, select materials, iterate on designs
Systems Engineers	System architecture, network integration, interdependencies with existing infrastructure	Define system-wide requirements, integrate multiple subsystems, ensure scalability
Project Managers	Budget, timeline, stakeholder requirements, political constraints	Allocate funding, manage risks, adjust project scope
Quality Assurance (QA) Engineers	Performance simulations, safety compliance reports, material test results	Conduct system validation, enforce safety standards, ensure public safety compliance
Manufacturing & Supply Chain Managers	Availability of materials, supply chain disruptions, cost of production	Source materials, optimize supply chain logistics, maintain production quality
Operations & Maintenance (O&M) Engineers	Infrastructure wear rates, maintenance logs, repair costs, failure reports	Schedule repairs, implement upgrades, recommend decommissioning timelines
Regulatory Agencies	Public health & safety reports, environmental impact studies, engineering assessments	Set legal standards, approve permits, enforce compliance
Requirements Engineers	Public input from Requests for Comment (RFCs), stakeholder needs, feasibility constraints	Translate public needs into engineering requirements, refine project scope
Community Engagement Coordinators	Public sentiment, community concerns, legislative input	Organize public hearings, solicit feedback, shape policy decisions
The Public (End-Users)	Infrastructure reliability, accessibility, safety, environmental impact	Participate in public forums, vote on infrastructure funding, provide feedback

Table B.2. Key Dependencies in Infrastructure Development Process

Primary Action	**Coupled With**	**Dynamics**
Systems Engineers define architecture	Design Engineers develop detailed specifications	The system architecture constrains subsystem design; design engineers must work within the defined framework
Design Engineers develop detailed designs	QA Engineers assess compliance with requirements	Designs must be finalized before they can be tested and validated
Infrastructure is constructed and deployed	Operations & Maintenance Engineers begin system upkeep	No maintenance can occur until infrastructure exists in a deployed state
Design Engineers incorporate public feedback	Requirements Engineers translate feedback into specifications	Public consultations influence engineering trade-offs and project priorities
Regulatory Agencies enforce safety and policy requirements	QA Engineers ensure compliance	Regulatory standards dictate quality control benchmarks
QA / V&V Engineers conduct testing and certification	Manufacturing teams begin production	Manufacturing can only proceed once designs pass validation checks
Infrastructure enters operational use	Public (End-Users) provide feedback on usability and performance	Real-world usage data feeds back into future design improvements
Operations & Maintenance Engineers track failures and maintenance needs	Future design cycles adapt based on operational performance	Maintenance logs inform upgrades and future engineering decisions

Table B.3. Incentives and Regulatory Functions in Infrastructure Development

Stakeholder	**Incentives**	**Regulation**
Systems Engineers	Ensuring functional integration, optimizing performance across subsystems	Constrained by architectural frameworks, safety requirements, and maintainability standards
Design Engineers	Innovation, technical efficiency, ensuring design feasibility	Must adhere to system requirements, material limitations, and operational constraints
Quality Assurance (QA) Engineers	Reducing failures, ensuring regulatory compliance	Bound by legal safety requirements, quality control procedures, and standard testing protocols
Manufacturing Engineers & Supply Chain Managers	Reducing costs, ensuring timely production, maintaining product quality	Must comply with material sourcing policies, production quality standards, and environmental laws
Operations & Maintenance (O&M) Engineers	Extending system lifespan, reducing downtime, optimizing performance	Must follow preventive maintenance schedules, safety protocols, and performance monitoring guidelines
Regulatory Agencies	Ensuring public safety, enforcing legal compliance	Define infrastructure codes, conduct inspections, approve projects
Requirements Engineers	Translating stakeholder needs into actionable specifications	Must balance feasibility constraints with user needs and technical standards
Community Engagement Coordinators	Ensuring equitable development, addressing public concerns	Must align with public consultation mandates, accessibility laws, and transparency policies
The Public (End-Users)	Safe, reliable, and cost-effective infrastructure	Participate in public hearings, influence policies through democratic processes