Architecture, constraints, and behavior

John Doyle
Control and Dynamical Systems
California Institute of Technology
Pasadena, CA 91125
doyle@caltech.edu

Marie Csete
Anesthesiology
University of California, San Diego
San Diego CA 92103
mcsete@ucsd.edu

Classification:

Dual: PHYSICAL SCIENCES: Engineering and BIOLOGICAL SCIENCES: Neuroscience, Systems Biology

Abstract

This paper aims to bridge progress in neuroscience involving sophisticated quantitative analysis of behavior, including the new use of robust control, with other relevant conceptual and theoretical frameworks from systems engineering, systems biology, and mathematics. Familiar and accessible case studies are used to illustrate concepts of robustness, organization, and architecture (modularity and protocols) that are central to understanding complex networks. These essential organizational features are hidden during normal function of a system, but fundamental for understanding the nature, design, and function of complex biologic and technologic systems.

Introduction

Systems approaches to biology, medicine, engineering, and neuroscience face converging challenges, as modern science, technology, and culture create dauntingly complex but similar and overlapping problems in these domains. Our goal is to develop more integrated theory and methods applicable to all systems including neuroscience, by concentrating on organizational principles of complex systems. Beyond scientific understanding of systems, practitioners want to avoid and fix network errors, failures, and fragilities. This practical necessity requires mechanistic and often domain-specific explanations, not vague generalities. So "universal" theories must facilitate the inclusion of domain mechanisms and details, and manage rather than trivialize their complexity.

Here we aim to put recent progress in both experimental and theoretical neuroscience (1-15) in the context of a shared conceptual and mathematical framework (7,16-32), in which a main theme is that complexity is driven by robustness, not by minimal functionality. We will emphasize robustness and efficiency tradeoffs and constraints and the control systems that balance them, their highly organized architecture (16-18), and its resulting side effects and fragilities. A confounding commonality we must both overcome and exploit is that the most robust and powerful mechanisms are also the most cryptic, hidden from introspection or simple investigation. These mechanisms can give rise to a host of illusions, errors, and confusion, but are also the essential keys to reverse engineering hidden network complexity.

This paper is inspired by several complementary research themes in behavioral neurosciences. Marder (1) systematically perturbs both experimental and math models of small circuits to explore robustness and fragility properties of neural hardware in mechanistic detail. In humans, cleverly constructed experiments to unmask the workings of the brain can elicit visual (2) and other (3) illusions, suggesting hidden, automatic subconscious functions (4-6). A theoretical framework consistent with other empirical observations treats the brain as an integrated, robust control system (7) in which components for sensing, communication, computation, simulation, and decision are useful primarily to the extent they effect action (8-12). Each theme (1-12) provides a separate constraint on the system as a whole, so seemingly dissimilar viewpoints can prove complementary and synergistic.

Our initial focus is how circuit (1) and system (2-3) fragilities are necessarily the consequence of (not merely consistent with) implementing robust controllers (7) in such circuits. If brains evolved for sensorimotor control and retain much of that evolved architecture, then the apparent distinctions between perceptual, cognitive, and motor processes may be another form of illusion (9), reinforcing the claim that robust control and adaptive feedback (7,11) rather than more conventional "serial signal processing" might be more useful in interpreting neurophysiology data (9). This view also appears broadly consistent with the arguments from grounded cognition that modal simulations, bodily states, and situated action underlie not only motor control, but cognition in general (12), including language (13). Furthermore, the myriad constraints involved in the evolution of circuit and network mechanisms efficiently implementing robust control are essential to explaining the resulting fragilities, which vary from largely benign illusions (2) to dangerous dysfunction (3,4,16,33-37) to potential catastrophes (16, 34-40).

In parallel to its broadening application in neuroscience, control theory and technology have expanded widely into networked, distributed, nonlinear, and hybrid systems in engineering and systems biology (e.g. see (16) and references therein, and (19,20)). All these systems are of potentially great but as yet unrealized relevance to neuroscience, as a source of both metaphors and new mathematics. Unfortunately, there is little shared language and few popular expositions (41). Thus our next focus is to more broadly relate (1-12) with (16-20) while minimizing math and technical details. Using familiar case studies, we aim for accessible and concrete treatment of concepts such as constraints, tradeoffs, and layered architectures. Here layering is functional, not necessarily mapping directly on to brain anatomy or physical architecture. An important example of layering is between computer hardware and software, but additional layering is a ubiquitous and essential architectural feature in complex networks of all types.

Neuroscience and robust control

A recent claim (7) is that human motor control is better explained as a robust rather than optimal controller, an explanation with a long tradition in neuroscience. Controllers optimal on average to only additive noise can be arbitrarily fragile to other uncertainties (21), motivating the development of robust control theory (22- 24). Robust control is risk-sensitive, optimizing worst-case (rather than average or risk-neutral) performance to a variety of disturbances and perturbations. Robust control theory formalized and extended best practice in

control engineering, and coincided with a massive expansion into active control in robots, airplanes, automobiles, smart weapons, communication networks, etc. So robust control is now ubiquitous but hidden, evidenced largely by what does not happen, such as skids, stalls, crashes, missed targets, dropped packets, etc. Similarly, most central nervous systems (CNS) activities are hidden from conscious awareness (4-6), implementing the sensing, decision making, and actuation necessary for robust control in complex environments.

Control theory makes strong predictions about how robust circuits must necessarily be implemented, largely independent of the device technology, all perfectly consistent with observations in neural circuits (1). Such claims are easily checked by experts in the math, but also hopefully intuitively plausible to neuroscientists generally. In particular, any natural parameterization of functional control circuits (e.g. lobster somatogastric ganglia (1)) is well-known to be large (high dimension), thin (even higher codimension), and nonconvex (24). If it were otherwise, engineering design would be much easier. As a simple analogy to explain these terms, consider a 2-dimensional piece of paper with lengths that are "large" by some measure and sitting in a 3-dimensional square box of comparable lengths. The larger these lengths the smaller the fraction of volume the paper will occupy in the box. So the paper can be both large, but "thin" as a fraction of the box volume. If the paper is bent or wrinkled, then it is also nonconvex within the box, since most straight lines between two different points on the paper will not remain in the paper.

An even simpler example is the set of words in most languages, which is large, but vanishingly thin as a fraction of all possible meaningless sequences of letters. There are 9!=362,880 different permuted sequences from just the 9 distinct letters "adeginorz," roughly the total number of English words, but only "organized" is a word. Humans would have some difficulty checking this claim, but computers do so easily (and make formidable Scrabble opponents). The set of English words is thus large and thin. The set of functional parameter values of any circuit will also typically be large but vanishingly thin and nonconvex in the set of all possible (mostly nonfunctional) circuits. This fact is largely independent of the notions of function, circuit, or parameter, provided they are sufficiently complex and realistic. Much of engineering theory is devoted to constructing special (higher dimensional, nonphysical, abstract) parameter embeddings that are convex, and thus algorithmically searchable for robust and functional design values. This idea that robust systems are large but thin and nonconvex in the space of all systems is a theme discussed below.

Another general feature of control systems is hard limits on robustness and efficiency (20). If the brain as controller has evolved to control *action*, then conscious thought may be in some sense a late evolutionary addition or byproduct (6-12). Both robust control theory and experimental evidence suggest that complex internal dynamic models are needed to resolve ambiguities in noisy sensations, as well as for use in planning for uncertain action, all in an uncertain, perhaps hostile, environment. When these models are implemented in slow neural hardware, management of the resulting delays almost certainly requires a heavily layered organization, a concept central to network architecture emphasized here. (Delay has little ill effect if there is truly no uncertainty, since then open loop control is adequate, but this ideal is never seen in practice.)

Pain and reflexes illustrate the sophisticated interplay of central and peripheral control, and that fast action receives priority. Fast, thick, myelinated, general purpose sensory fibers initiate withdrawal from painful stimuli while slow, thin, specialized fibers provide delayed, but detailed, information about the source of pain. This is a pattern seen throughout the organization of the nervous system, and throughout behavior we see this mix of reflex (fast, automatic, hidden, expensive) and reflect (slow, conscious), with reflex receiving priority in resources. The acquisition of skill (playing instruments, ball sports, chess, reading, etc) involves shifting "down" into fast reflex processes that start "high" and slow. Indeed, the more expert we are in an activity, the less we necessarily rely on conscious processes to perform, and such evidence for layering is found everywhere (4-6).

If brains are doing robust control using internal models, then illusions may be intrinsic. What reaches conscious awareness is the state of a simulation, not a direct perception of the world (9-12). Yet "seeing is believing" since what we "see" is not only a remarkably robust, integrated, dynamic state estimate of the external world blending multiple senses, but one that automatically focuses attention on information we need to take robust actions. That we can be almost arbitrarily fooled is one of the many unavoidable tradeoffs of our physiology, evolution, and brain architecture. So it is equally true that "seeing is dreaming" as known from a variety of well-studied illusions (3), dreams, and hallucinations. Functional losses due to CNS lesions (4,6) are often highly specific and reproducible, making us aware of myriad unconscious processes that were previously taken for granted and showing that our internal simulations use a distributed and parallel implementation to mitigate the

effects of hardware delays. The extreme gain and loss of capabilities in savants also suggest powerful but constrained simulation capabilities.

Computer and control technology

Modern robust control systems are typically implemented using digital hardware and software, and most computers are "embedded" in this way and thus permanently hidden. Examples are ubiquitous, from antilock brakes to automated collision avoidance to GPS in cars, to fly-by-wire aircraft. Networks and cloud computing that connect the relatively fewer (but still billions worldwide) PCs and smartphones have hidden routers and servers that control the flow of packets and files. The internal mechanisms are again manifest largely in the rarity of crashes, losses, errors, failures, etc., and in the catastrophic nature of rare crashes. Yet despite enormous progress, robots struggle to navigate the real world as effectively as rodents or even insects, and computers continue to fail in "Turing tests," though in fascinating ways that reveal much about both humans and computers (42). This enormous, hidden, cryptic complexity, driven by robustness, is both the greatest initial obstacle in using advanced information and control technologies as metaphors for biology, but also ultimately the key to important insights and theories (16,19).

As a starting point, human memory layering appears to be very different from computers, as shown by various syndromes, lesions, and lab studies (6), but also competitions pushing the extremes of human memory (14). A standard technique among competitors memorizing sequences of meaningless symbols is to embed them in previously "prepared" complex and vivid 3D dynamic simulations (called "palaces") that can then be replayed to retrieve the symbols. For example, such methods allow experts to memorize a single pack of 52 shuffled playing cards with no errors in less than 22 seconds. Palaces are reused after the memories are actively purged. This illustrates not only that humans can repurpose innate (dynamic, modal, grounded, etc) simulation capability in lower layers for purely symbolic higher layer memory, and (it is claimed) the lack of real alternatives for rapidly storing and retrieving large amounts of symbolic data (14). That it works so poorly is perhaps less remarkable than that it works at all.

Computers have opposite memory capabilities from humans, in that massive amounts of purely symbolic, meaningless data are nearly instantaneously found, stored, searched, and retrieved, and as a result "Google" is now a verb. This is possible because computers and networks have arguably the canonical layered architecture in engineering, from VLSI chip design to the TCP/IP protocol stack (16,19), and a brief look at such architecture is a rich source of insights. Near "the bottom" is analog circuitry that is exquisitely organized (extremely large/thin/nonconvex) to create digital behavior when interconnected appropriately, at the expense of speed and efficiency. These analog and digital hardware layers are functionally distinct but physically coincident. Importantly, these hidden layers and interfaces are fundamental to the more obvious plug-and-play modularity they enable.

Many devices can be built purely out of hardware, but a software layer gives much greater flexibility at even more expense of speed and efficiency. As with the digital layer and its analog substrate, software only exists when embodied in hardware, but because software can be moved across hardware platforms it also has an existence that transcends any individual physical instantiation. Software is a "very special organization" of hardware, and similarly digital hardware of analog circuitry, but no simple terminology captures the full richness of this layering. Nevertheless, there are no mysteries here, just an impoverished language for description. Software too is richly layered. An operating system (OS) often has a "kernel" layer that manages and virtualizes the various hardware resources for higher layer application programs. For example, the hardware memory typically has its own separate layering of memories from small, fast, and expensive to large, slow, and cheap. This layering is within the hardware so is orthogonal to that of analog to digital to software.

By managing the use of layered memory cleverly, the OS kernel can provide applications programs with a virtual memory that has nearly the speed of the fastest hardware, with the cost and size of the cheapest. Such virtualization is a familiar and essential element of layering. So applications can use abstract named variables and higher level languages, the kernel then translates these names into virtual addresses, and ultimately into physical addresses, but the name-to-address translation process is hidden from the applications. This OS architecture provides a variety of robustness features, from scalability of the name and virtual address spaces, to resource sharing between applications, to security of the physical memory from application failures or attacks.

At the most basic level, the Internet TCP/IP protocol stack extends the functionality of the OS kernel across the network to multiple machines, allowing much broader resource sharing, and creating the illusion to

users of near infinite resources. Unfortunately, TCP/IP was designed decades ago not as a general purpose platform but primarily to be robust to physical attacks on hardware in relatively small networks with trusted users running minimal applications. It did this brilliantly, especially compared with the alternatives at the time, but modern usage is largely the opposite: Hardware is more reliable than software which is more trustworthy than users, and the network is large and supports a bewildering range of applications. In essence, TCP/IP is not strongly layered enough. It lacks a modern naming and virtual addressing mechanism, leading to problems with security, performance, scalability, multihoming, mobility, etc. for which resolution is hotly debated even among experts (43). That TCP/IP is in some ways inadequate is less surprising than that it works at all given the astonishing change it has enabled.

TCP/IP is an example of how architectures that are well-designed for extreme robustness can create evolvability as a side benefit, perhaps the essential benefit of good architectures and the focus of the rest of this paper. Networked and embedded control computers may ultimately be a good source of metaphor and theory for neuroscience because we know exactly how the system behavior depends on the technical details, and a rich and growing body of mathematics formalizes the insights (19). Unfortunately, these details are by design largely hidden from users, so while experts will find the previous discussion trivial and obvious, many readers may not. Also, despite abundant relevant tutorial material on computer architecture (less on networks), there is little discussion on which of its features arose from fundamental designs versus historical accidents of rapid evolution. For these reasons, we explore some additional case studies that are transparent and familiar, yet illustrate the fundamental concepts of complexity, architecture, layering, and robustness.

Layered architectures simplified

Clothing and textiles represent a simple case study in network architecture and the role of robustness and layering, with paper a special case. Though clothing may seem a frivolous illustrative example, it is based on many levels of complex technologies and reveals universal organizational principles, with details that are easily accessible to nonexperts. On the surface each clothing module or garment (coat, socks) looks fairly similar, hiding chemical and physical differences in weave, elasticity, water-resistance, breathability, UV protection, even insect repulsion. The constraints imposed by fashion trends on the success of clothing as a technology illustrate important points but will be de-emphasized, and our consideration of the architecture of an outfit focuses more on essential function and robustness in harsh environments.   The basic function of clothing is protection, providing comfort over a wide variety of external (weather, temperature) and internal perturbations (physical activity). Like other complex systems, complexity in clothing is driven by robustness to extremes more than by need to provide minimal function. Human skin appears optimized by evolution for dissipating heat during endurance running in the tropics (44-46), and offers little protection compared to heavy fur. Clothing provides that protection when needed.

Four fairly universal layers exist within textile architecture: 1) fibers that are spun into 2) yarn or thread which are woven or knitted into 3) cloth which is sewn into 4) garments. Cotton fibers are about 12-20 μm (microns) in width and several centimeters in length, roughly comparable to large neurons but with very different morphologies. A kilogram of cotton has less than a billion fibers, large but still much less than the number of neurons in a kilogram of brain, and the way in which fibers are interconnected is much simpler than neurons. Thus this is a simple but easily understood example of layering, in which the properties of textiles are not obvious from those of fibers. Tens to hundreds of fibers are spun into yarn or thread of essentially arbitrary length, which are woven or knitted into cloth that is nearly two dimensional, then sewn into garments, also of arbitrary size.

This layered construction is much simpler but parallels analog to digital hardware to software, and the polymerization of metabolic building blocks to macromolecules which assemble into networks and cells.   In all of these examples, the layered architecture illustrates universal principles of organization and protocols for construction. Each layer has the large/thin property for which functional alternatives are almost unaccountably numerous, but are nevertheless a vanishingly small fraction of all possible (e.g. random) configurations. Each layer must be exquisitely organized to produce the layer above it, which is not necessarily physically distinct. Garments are functionally distinct from the fibers from which they are physically composed, as are cloth and yarn.

The complexity of the textile architecture is driven by robustness tradeoffs, as all the layers from fiber to cloth can be completely collapsed to make paper, a nearly random connection of fibers with no intermediate layers. Paper is an extreme example of a degenerate special case of a layered architecture. Here "degenerate"

simply means that the constraints that define the architecture are relaxed or removed entirely. And with minimal additional complexity, paper can be sewn into specialized but unavoidably fragile garments. This makes clear that the complex internal, hidden layering is only for robustness and is not needed for minimal functionality, as paper can easily stand in for cloth in idealized environments. Similarly, small bio-inspired networks of metabolites and enzymes can be used to manufacture valuable chemicals, but lack the robustness and evolvability of whole organisms. The overall textile architecture has persisted for many thousands of years (the bacterial cell, billions) while technologies within layers evolved rapidly.

One feature of the fiber to garment (i.e. garment/sew/cloth/weave/thread/spin/fiber) layered architecture is that it is robust enough that we can temporarily defer its study and focus on something even simpler, which is how individual garments are further layered to make outfits. Outfits for harsh environments typically have three kinds of layers. The outer shell layer protects from wind and water, the middle or insulation layer provides warmth, and the inner or base layer is comfortable next to the skin and keeps it dry. These three layers each are composed of garments, which have within them the fiber to garment layers. So the garment to outfit layering, and the fiber to garment layering are in some sense "orthogonal," though there is no standard terminology. This illustrates an almost trivial but nevertheless crucial feature of organized complexity. Because this overall architecture is intrinsically so robust, we can temporarily take one dimension (fiber to cloth layering) for granted and view it as a platform for another simpler dimension of clothing, but one that also connects with more popular views of modularity, while still introducing some essential elements of architecture.

Layering of garments to make outfits is one obvious architectural feature of clothing providing robustness to environments. These modular layers are physically distinct (unlike fiber to cloth) and can be shed or reincorporated as needed. This has obvious parallels with software. Good programming practice includes breaking large algorithms into smaller subroutines with simple interfaces; this modularity is more familiar but less fundamental than the layering of analog to digital hardware to software that makes it possible in the first place. In the absence of robustness requirements, the necessary engineering aspects of architecture can recede and clothing can become considerably simplified or elaborate (as dictated by fashion). In perfect environments, little or no clothing is required.

Similarly, even the most complex architectures allow for much simpler degenerate special cases (analogous to paper) under idealized circumstances. Tradeoffs abound within each layer in fabric, weight, cost, durability, fasteners, etc. With changes in technology or conditions, garments can become obsolete or evolve (e.g. body armor, Velcro, etc). Highly optimized, robust, and efficient garments that are finely specialized for a specific layer, body part, and individual are typically fragile to other uses (other layers, body positions, or wearers of different size and shape) and may be costly. Simple wraps or rags are very versatile but at best yield outfits that are fragile to environment and movement. Most garments fall between these extremes.

If we knew nothing about the layering of garments we might learn little from observing intact outfits, but could begin reverse engineering the architecture via lesions or knockouts in controlled experiments. These might require harsh experimental conditions, and perhaps appropriately instrumented "crash dummies". Damage or loss of a garment layer can cause very specific loss of robustness: outer layer to wind and/or water, middle layer to cold, and inner layer to comfort. Changes to fiber types, yarn, or sewing could be "lethal" at different levels, revealing their functional role. Most informative would be small changes with large consequences, such as unraveling a seam to reveal the role of sewing in garment construction, or disruption of a weave or knit to reveal its role in cloth integrity.

Architecture as constraints that deconstrain

The view of architecture as "constraints that deconstrain" (17,18) originated in biology, but is consistent with engineering (16) and illustrated by clothing: A robust architecture is constrained by protocols, but the resulting plug-and-play modularity that these shared constraints enable "deconstrain" (i.e. make flexible) systems designed using this architecture. "Constraints" give a convenient starting language to formalize and quantify architecture, and ultimately a mathematical foundation (19). Concretely, consider a given wardrobe which is a collection of garments, and the problem of assembling an outfit that provides suitable robustness to the wearer's environment. Three distinct but interrelated types of constraints are universal in clothing, as in all architecture (16): 1) component (garment) constraints, 2) system (outfit) constraints, and 3) protocol constraints. So in combination, diverse, heterogeneous components (garments) that are constrained by materials and construction

combine "synergistically" (via protocols) to yield outfits that satisfy system constraints not directly provided by any single component. We will use "outfit" to describe a functional, robust set of garments, and "heap" (Craver uses "aggregates" (47)) to describe a random collection not required to have any other system features.

The protocols that constrain how garments make outfits are simple and familiar, and a minimal view is that each of g garment categories (e.g. socks, sweaters, coats, boots, hats, etc...) is constrained to a specific layer and body position, and thus to a specific and essentially unique location within an outfit. Suppose for simplicity a wardrobe has n garments of each type for a total of ng garments. If any of the n garments of a specific type can be part of an outfit, then there are a total of $F = n^g$ possible outfits. For example, for $n = g = 10$ there are $ng = 100$ total garments but 10^{10} (10 billion) distinct outfits that obey the layered architecture. On the other hand, there are 2^{ng} subsets or heaps of ng garments, so that if protocols are ignored, the number of *un*constrained garment heaps is vastly larger. For $n = g = 10$ there are $2^{ng} > 10^{30}$ such heaps, so heaps chosen without regard to protocols have a vanishingly small chance of being outfits, another example of large/thin.

The discussion of clothing so far provides only a static view of architecture. In reality the core of good architecture is ability to facilitate change over many timescales, including overall architecture (millennia), manufacturing technology (centuries), garments (decades), and outfits (daily). This example can be expanded to hint at the dynamic and control dimensions of both us and our clothing. Well- constructed outfits respond so automatically to movement that wearers can normally ignore the hidden internal complexity that makes this possible, just as we do the control of movement itself.

The roughly minute-to-hour timescales needed to assemble outfits also illustrates the role of dynamic control within an architecture. The most obvious control is the actual "forward" assembly of a specific choice of g garments into a layered outfit. Most of the protocols which govern this process are readily learned by children, though specialized garments may require complex control (e.g. bowties, shoe laces). Ideally, the protocols are complete enough that any outfit made that obeys them will automatically satisfy system constraints (this is rare in engineering, as it is hard to design protocols with such guarantees). Humans easily visualize (simulate) what an outfit will look like from seeing the separate garments, but often still need to "try them on" to be sure of details of fit and appearance. Our simulators are robust but imperfect.

More subtle and complex (and less easily learned) is the "backward" process of choosing these garments to match the day's specific systems constraints, which are most dependent on weather and the wearer's activities. This takes simulation to another level. Since layering allows dynamic reconfiguring of an outfit in real time, this backward selection control or management process potentially interacts with forward assembly control on all time scales. The backward process of choosing components is typically much more complex (and less obvious) than the forward assembly process, and this can lead to confusion about the role of control and feedback in architectural design. Hopefully in this concrete example the processes are obvious, even if the language for describing it is inadequate.

On longer time scales (days to decades), the user might assemble a wardrobe of garments, again guided by the overall architecture by which garments make outfits. Manufacturing technologies can change on year to century timescales but must both reflect the architecture and only slowly change aspects of it. New technologies (such as spandex and Velcro) relax component constraints and allow new systems without fundamentally changing the architecture of clothing, which has persisted for millennia. Here again, in the engineering context, robustness largely drives complexity, since without changing system and component constraints, the protocols and control processes for connecting and reconciling them could be vastly simpler (e.g. standardized uniforms).

Bowties, hourglasses, pathways, flows, and control

Another aspect of constraints that deconstrain is the relatively small diversity in the protocols and processes that connect layers and enable vastly greater diversity in the materials that constitute a layer. This aspect of architecture can be visualized as a "bowtie" or "hourglass" (depending on whether layers are visualized horizontally or vertically) (27). For example, the fairly universal, homogeneous process of sewing (the bowtie knot or hourglass waist) takes an almost unaccountably greater and extremely heterogeneous diversity of cloth (fanning in to the knot) into an even greater diversity of garments (fanning out of the bowtie). Similarly, the reactions and metabolites of core metabolism are largely universal, connecting extremely diverse layers of catabolism to

moderately diverse biosynthesis. The basic processes and codes underlying transcription and translation are highly conserved, but the specific genes and gene products are extremely diverse.

Once the diverse garments sewn from cloth become components in the layered outfit architecture, they are categorized into the much less diverse *types* of garments and assembly protocols that in turn make a hugely diverse set of outfits. Threads and yarns are more diverse than the few canonical weaves, knits, or knots that create diverse textiles. The least diversity is in the fibers (from plant, animal, mineral, and synthetic origins) and the spinning processes that make yarn and thread. (What is vastly diverse is the geographic origins of these fibers.) Thus the great diversity and heterogeneity *within* each layer also varies *among* layers, and even depends on the categories used to define diversity. Garments (or cells) are unaccountably diverse and deconstrained when viewed in detail as the result of the garment/cloth/yarn/fiber (or DNA/RNA/protein) architecture, but much less so when viewed as satisfying system constraints of that architecture, the component constraints of the outfit architecture, or the constraints on cells or cell types. The protocols between layers are typically more fixed and much less diverse by any measure than the layers they connect.

Since they are the fixed points in robust architecture, when protocols are subject to attack, the system can fail catastrophically. Seaming is the protocol that sews fabric into garments, providing structure and function. Seams make the important difference between wraps and clothes, but are the main source of clothing's fragile robustness. If the seam connecting shoulder to sleeve unravels, a coat is useless. The greatest fragility of universal knots/waists is that they facilitate hijacking and attack by parasites and predators. Viruses hijack cellular transcription/translation machinery, and predators exploit the fact that they share with their prey universal and essential metabolic building blocks. In neuroscience, dopamine's role in a robust and flexible reward system (knot of the bowtie) is fragile to hijacking by addiction (5). The stitches used in seams are not as tight as fabric weave and so seams are internalized and often protected by lining, illustrating how good architectures allow hiding of necessary fragilities as much as possible. Our skulls, cardiovascular system, blood-brain barrier, and immune systems similarly protect fragilities of our brains to trauma, intense activity, and infection, at the expense of the overhead to maintain them.

In both textiles and biology, obvious natural pathways and "flows" of materials and "information" assemble systems from components. Indeed, depicting these architectures in terms of pathways rather than layers has been the dominant view in science (and until recently in engineering as well). While not inconsistent with layering, the emphasis on a pathway view has limited our understanding of control, complexity, and robustness. While the interplay between computational complexity, constrained optimization, and robust control has been deeply explored in the last decade, with broad applications including the Internet (19), power grids, and systems biology, no universal and accessible taxonomy for describing these various "flows" and their various "complexities," has emerged, even within engineering.

We already saw that the backward process of deciding on an outfit that satisfies the constraints of a given activity and environment is vastly more complex than the forward process that assembles the outfit from a set of garments. Similarly the feedback control (e.g. of looms and sewing machines) within each layer of textile manufacturing is vastly more complex than the forward flow of materials from fibers to textiles. Additional complexity comes from the backward flow of textile design that turns constraints on textiles into specifications on the manufacturing processes, as well as the supply chain management of the resulting process in response to customer demand.

Biology has similarly complex feedbacks. There are ten times as many fibers feeding back from the primary visual cortex to the visual thalamus as there are in the forward "flow" (6). Wiring diagrams that include both autocatalytic (e.g. of ATP, NADH, etc) and control feedback in metabolism are so much more complex than the usual depictions of relatively simple tree-like flows of metabolites, that they are rarely drawn in detail except at the level of small circuits. Even complete wiring diagrams do not reflect the true complexity of control (20). The control of transcription and translation is vastly more complex than the basic forward polymerization processes themselves. While we have no difficulty understanding the basic nature of these feedbacks and their roles in these specific architectures, the lack of an adequate language to generalize and/or formalize is still a roadblock, especially since engineering jargon is domain specific and heavily mathematical.

Hidden complexity, illusions, and errors

Related but important research themes can be mentioned only briefly while recapping the main points. For example, the emphasis on dynamic and mechanistic explanations in the philosophy of neuroscience (47-49) is compatible, complementary, and hopefully can help lead to a more coherent and consistent shared language, which is desperately needed. The dangerous illusions and errors that plague individuals are often amplified by institutions, and this is relevant to engineering as well, since policy and politics often trump technology (39). Arguably the most dangerous and pervasive of popular illusions is that our actions are unconstrained by hard tradeoffs, a problem increasingly acute in everything from teaching evolution to dealing with global warming. A unique case study in human error because it is entirely within science is the genre of research that has dominated mainstream literature for decades under the rubric of "new sciences" of "complexity" and "networks" (NSCN). NSCN is relatively new in neuroscience but already has an appealing narrative (50) and extensive and accessible reviews (51-52). Claims that NSCN has been a success in other fields (50-52) is supported by impact factor measurements, but NSCN has led to persistent and systematic errors and confusion that show no sign of abating (see (16,30-32) and the many references therein). While the goals of NSCN research are somewhat consistent with ours, particularly in neuroscience, the methodology is not. Most concepts and terminology in NSCN do not overlap with control theory, but those that do can have opposite meaning. A thorough discussion is beyond the scope of this paper, but a simple instance is illustrative.

Engineered and biological systems necessarily make ubiquitous use of nonlinearity, recursion, feedback and dynamics, which in NSCN are almost synonymous with unpredictability, fractals, self-similarity, or chaos (16). In engineering quite the opposite is true. For example, the amplifiers in the sensors and actuators that enable robust controllers are necessarily extremely nonlinear. Digital computers are recursive and also extremely nonlinear (i.e. switching using transistor amplifiers in hard saturation) but are the most repeatable complex systems we build. Ironically, to a naïve observer the analog behavior of the billions of transistors (each much smaller and simpler than a neuron) per chip in a computer might appear bewilderingly noisy, chaotic, and unpredictable, with no hint of the almost perfectly robust, repeatable digital behavior that results at the interfaces and that engineers use to build systems. Of course, this is exactly the purpose of the "very special large/thin/nonconvex organizations" that constitute digital/analog and other forms of layering. This finely tuned, hidden diversity and complexity underlying robust systems is also the opposite of "complexity" in NSCN, which emphasizes minimally tuned, mostly random interactions of large numbers of homogeneous components that yield surprising "emergent self-organization" and "order for free" (50-52).

Recall that in the clothing example, for a fixed number of garment types g the number of outfits $F = n^g$ is constrained to polynomial growth in the number n of each garment type, whereas all possible "heaps" grow exponentially in n. A related source of confusion in biology is whether biology is "fine-tuned vs. robust," as if these were mutually exclusive. There is robustness in the large (polynomial growth) sets of structured and functional networks, and fine tuning that makes these a thinly small subset of the vastly larger (exponential growth) set of random nonfunctional networks. Highly evolved biological systems are large/thin and both fine-tuned (obey strict and far from random protocol constraints) and robust. Indeed, this is the essence of constraints that deconstrain, and a necessity, not a paradox. The connection between the large/thin feature of neural circuits (1) and robust control is even deeper, as sketched above.

In a completely different direction, research on human evolution has recently exploded in both depth and accessible exposition (44-46), and the picture emerging is also complementary and compatible to ours. Compared with other great apes and top predators, humans are physically weak and slow with thin skin, no protective fur, and small guts that digest raw food poorly, all possibly fragile side effects of evolved robustness to running long distances in hot weather (46). When paired with even minimal technologies of weapons (e.g. simple sticks and stones), fire (for cooking, protection, and warmth), and teamwork, we go from helpless prey to invincible top predators (45), whose main threat then comes from other, similarly equipped humans. Our layered biological architecture of brain to mind is now augmented by layered technological architectures such as fiber to garment to outfit, metal to tool, weapon, machine, analog to digital, hardware to software, all of which expand the "cognitive niches" (44) in which we can robustly function. We have now eliminated many fragilities to our environment but replaced them with new and potentially catastrophic fragilities of our own making.

The tradeoffs we see throughout these architectures and systems between efficiency and robustness, as well as between robustness to various different perturbations, are necessities and not accidents, though choices are still abundant within the resulting constraints. Versions of such tradeoffs can be formalized and made

mathematically precise (20), and the SI presents a simplified tutorial on the mathematics and model systems. More speculative but plausible is the claim that layered architectures are also necessary to effectively balance these tradeoffs, as evidenced by their ubiquity in biology, physiology, and technology. That is, inner/lower layers are large/thin/nonconvex and *must* remain hidden within the system for robustness. For example, inner and middle garments must remain hidden behind the outer shell to provide comfort and warmth in a harsh environment, whereas the middle and outer layers must be segregated from skin for comfort. Since each garment maintains a separate identity that can be easily recovered by disassembling an outfit (or heap), this is perhaps the paradigm for "modularity" (51), but is a very special case and on its own quite misleading without considering the protocol and component constraints.

The component constraints on garments within the shell/insulation/base layering of outfits depend on material properties that derive from the "orthogonal" garment/cloth/yarn/fiber layering of textiles. In these orthogonal layers, garments are "very special large/thin/nonconvex organizations" of fibers/threads which must lose their individual identity. "Loose" threads (disobeying protocols, even minimally) make garments fragile. Similarly cells are a special organization of macromolecules, digital hardware of analog circuitry, software of hardware, brains of cells, and perhaps minds of brains. If the brain is layered at the highest modular level (6,15), roughly analogous to outfits/garments then there is an orthogonal layering of tissues down to cells down to macromolecules that occurs within each of the macro brain layers, analogous to the garment to fiber layering.

In all these layered architectures, the virtualization of the lower layer resources is an illusion that can be maintained almost perfectly in normal operations of minds, outfits, machines, and computers. The hidden complexity is primarily needed to create this remarkable robustness and evolvability, not minimal function, and is only revealed by pushing systems to their extremes by perturbing the environment or components in lower layers outside the constraints the systems evolved to handle (3-6). Once the layered architectures are in place, both our minds and software are free to rapidly evolve independently given the right environment, though the large differences between digital and brain hardware imply very different constraints (6). This plasticity is one of the main benefits "deconstrained" by the "constraints" of the bowtie/hourglass protocols that create the layering (17,18,27). The mind/brain layering is much more complex than the most complex current technology of embedded and or networked software/hardware/digital/analog, but the latter would be utterly incomprehensible without the right conceptual framework, mathematics, and tools, most of which are still unknown or relatively new to neuroscience (7). We hope that the connections between neural and technological architectures will help to demystify some aspects of this complex continuum.

References

1. Marder E (2011) Beyond the mean: Variability, compensation and modulation in neurons and circuits, *Proc Natl Acad Sci USA*
2. Purves D, Wojtach WT, Lotto RB (2011) Understanding vision in wholly empirical terms. *Proc Natl Acad Sci USA*
3. Chabris C, Simons D (2010) *The Invisible Gorilla: And Other Ways Our Intuitions Deceive Us*. (Crown The Crown Publishing Group, New York).
4. Gazzaniga M (2011) *Who Is in Charge: Free Will and the Science of the Brain* (Harper Collins (Ecco), New York).
5. Linden, D.J. (2011) *The Compass of Pleasure: How Our Brains Make Fatty Foods, Orgasm, Exercise, Marijuana, Generosity, Vodka, Learning, and Gambling Feel So Good*, Viking Press, New York
6. Eagleman D (2011*). Incognito: The Secret Lives of the Brain*. Pantheon.
7. Nagengast AJ, Braun DA, Wolpert DM (2010) Risk-sensitive optimal feedback control accounts for sensorimotor behavior under uncertainty. *PLOS Comput Biol* 6(7):e1000857.
8. Grafton ST (2010) The cognitive neuroscience of prehension: recent developments. *Exp Brain Res* 204: 475-491.
9. Cisek P, Kalaska JF (2010) Neural mechanisms for interacting with a world full of action choices. *Ann Rev Neurosci* 33: 269-298.
10. Konrad K (2007) Decision theory: What "should" the nervous system do? *Science* 318:606.
11. Shadmehr R, Smith MA, Krakauer JW (2010) Error correction, sensory prediction, and adaptation in motor control. *Ann Rev Neurosci* 33:89-108.
12. Barsalou LW (2008) Grounded cognition. *Ann Rev Psych* 59: 617-645.

13. Pinker S (2007) *The Stuff of Thought*, Viking Adult
14. Foer J (2011) *Moonwalking With Einstein: The Art and Science of Remembering Everything*. (Penguin Press, New York).
15. Hawkins J, Blakeslee S (2004) *On Intelligence: How a New Understanding of the Brain will Lead to the Creation of Truly Intelligent Machines*. (Times Books, New York).
16. Alderson DL, Doyle JC (2010) Contrasting views of complexity and their implications for network-centric infrastructures. *IEEE Trans Systems Man Cybernetics—Part A: Syst Humans* 40:839-852.
17. Kirschner M, Gerhart J (1998) Evolvability. *Proc Natl Acad Sci USA* 95: 8420-8427.
18. Kirschner M, Gerhart J (2005) *The Plausibility of Life*. (Yale University Press, New Haven).
19. Chiang M, Low SH, Calderbank AR, Doyle JC (2007) Layering as optimization decomposition: A mathematical theory of architecture. *Proc IEEE* 95: 52-56.
20. Chandra F, Buzi G, Doyle JC (2011) Glycolytic oscillations and limits on robust efficiency. *Science*. In press.
21. Doyle JC (1978) Guaranteed margins for LQG regulators. *IEEE Trans Autom Contr* 23: 756-757.
22. Glover K, Doyle JC (1988) State-space formulas for all stabilizing controllers that satisfy an H-infinity-norm bound and relations to risk sensitivity. *Syst Contr Lett* 11: 167-172.
23. Doyle JC, Francis BA, Tannenbaum A (1992) *Feedback Control Theory*. (Macmillan Press, New York).
24. Zhou K, Doyle JC, Glover K (1996) *Robust and Optimal Control*. (Prentice Hall, Upper Saddle River).
25. Yi TM, Huang Y, Simon MI, Doyle J (2000) Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proc Natl Acad Sci USA* 97:4649-4653.
26. Csete ME, Doyle JC (2002) Reverse engineering of biological complexity. *Science* 295:1664-1669.
27. Csete ME, Doyle JC (2004) Bow ties, metabolism, and disease. *Trends Biotechnol* 22: 446-450.
28. Doyle JC, Csete ME (2005) Motifs, stability, and control. *PLOS Biol* 3(11):e392.
29. Doyle JC, Alderson DL, Li L, Low S, Roughan M, Shalunov S, Tanaka R, Willinger W (2005) The "robust yet fragile" nature of the internet. *Proc Natl Acad Sci USA* 102:14497-14502.
30. Keller EF (2005) Revisiting "scale-free" networks. *Bioessays* 27: 1060-1068.
31. Lima-Mendez G, van Helden J (2009) The powerful law of the power law and other myths in network biology. *Mol Biosyst* 5:1482-1493.
32. Willinger W, Alderson D, and Doyle JC (2009) Mathematics and the internet: A source of enormous confusion and great potential. *Notices Amer Math Soc* 56:586-599.
33. Schulz K (2010) *Being Wrong: Adventures in the Margin of Error*. Ecco
34. Heffernan M (2011) *WILLFUL BLINDNESS: Why We Ignore the Obvious at Our Peril*. Walker Books. New York
35. Freedman DH (2010) *WRONG: Why experts keep failing us--and how to know when not to trust them*, Little, Brown and Company;
36. Ioannidis JPA (2005) Why most published research findings are false. *PLoS Medicine* 2005, 2(8):e124.
37. Trikalinos NA, Evangelou E, Ioannidis JP (2008) Falsified papers in high-impact journals were slow to retract and indistinguishable from nonfraudulent papers. *J Clin Epidemiol*. 61(5):464-70.
38. Stiglitz J (2010) *Freefall: America, Free Markets, and the Sinking of the World Economy*. W. W. Norton & Company.
39. N Oreskes, EM Conway, *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming*, Bloomsbury Press
40. Buffett HG (2009) *FRAGILE: The Human Condition*, National Geographic
41. Kelly K (2010) *What Technology Wants*. Viking Adult
42. Christian B (2011) *The Most Human Human: What Talking with Computers Teaches Us About What It Means to Be Alive*. Doubleday
43. Day J (2008) *Patterns in Network Architecture: A Return to Fundamentals*, Prentice-Hall
44. Pinker S (2010) The cognitive niche: Coevolution of intelligence, sociality, and language, *PNAS*, vol. 107, suppl. 2, 8993–8999
45. Wrangham R (2009) *Catching Fire: How Cooking Made Us Human*. Basic Books
46. Bramble DM, Lieberman DE (2004) Endurance running and the Evolution of Homo. *Nature*. 432: 345-352.
47. Carl F. Craver (2007) *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*, Oxford University Press, USA
48. Waskan JA (2006) *Models and Cognition: Prediction and Explanation in Everyday Life and in Science*, MIT Press

49. Kaplan DM, Bechtel W (2011) Dynamical models: an alternative or complement to mechanistic explanations? *Topics in Cognitive Science*, 3, 438-444

50. Chialvo DR (2010) Emergent complex neural dynamics, *Nature Physics*, Vol 6, Pgs:744–750

51. Bassett DS, Gazzaniga MS (2011) Understanding complexity in the human brain, *Trends in Cognitive Sciences*, May 2011, Vol. 15, No. 5

52. Bullmore E, Barnes A, Bassett DS, Fornito A, Kitzbichler M, Meunier D, Suckling J (2009), Generic aspects of complexity in brain imaging data and other biological systems, *NeuroImage*, 47 (2009) 1125–1134

53. Hasty J, Hoffmann A, Golden S (2010) Systems biology of cellular rhythms: from cacophony to symphony. *Curr Opin Gen Dev* 20: 571-573.