UC Santa Cruz

UC Santa Cruz Electronic Theses and Dissertations

Title

Project for a Cybernetic Psychology: A History of Machine Learning, 1840 to the Present

Permalink

https://escholarship.org/uc/item/8zj8j3qk

ISBN

9798184158822

Author

Jeon, Won

Publication Date

2026-06-03

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

PROJECT FOR A CYBERNETIC PSYCHOLOGY: A HISTORY OF MACHINE LEARNING, 1840 TO THE PRESENT

A dissertation submitted in partial satisfaction
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

HISTORY OF CONSCIOUSNESS

by

Won Jeon

June 2026

The dissertation of Won Jeon is approved:

_____
Professor Karen Barad, Chair

_____
Professor Robert Meister

_____
Professor Warren Sack

_____
Peter Biehl
Vice Provost and Dean of Graduate Studies

Copyright © by

Won Jeon

2026

TABLE OF CONTENTS

List of Figures iv

Abstract v

Introduction 1

1. Learning to Govern: "Cybernetics" in the 19th Century 26

2. Freudian Informatics and the Making of Psychoanalysis, 1840-1895. 79

3. Neurotic Networks in Mid-Century Cybernetics 134

4. The Uses of Belief: Conversational AI and the Problem of "Significant Learning" 216

Bibliography 274

LIST OF FIGURES

Figure		Page
1.	Sigmund Freud's diagram of "neurones"	84
2.	American Society of Cybernetics conceptual map	91
3.	Biological Computer Laboratory cover illustration	95
4.	Weber's compass	101
5.	Freud's diagrams of stimulus intensity and external excitation	110
6.	Helmholtz ophthalmoscope	115
7.	Freud's diagram of Emma's "scene"	124
8.	Gregory Bateson's drawing of McCulloch-Pitts network	159
9.	(a) Walter Pitts and Jerome Lettvin (b) Anatomy diagrams from "What the Frog's Eye Tells the Frog Brain"	166
10.	Letter and drawing from Humberto Maturana to Heinz von Foerster	169
11.	Octopus anatomy diagram	171
12.	Octopuses in observation by Bateson	174
13.	von Foerster blind spot experiment	179
14.	"Ecology of war in industrial society" drawing by Bateson	211

Abstract

PROJECT FOR A CYBERNETIC PSYCHOLOGY: A HISTORY OF MACHINE LEARNING, 1840 TO THE PRESENT

This dissertation traces the emergence of machine learning as a concept before its realization as a technology. It examines how learning came to be understood as adaptation under conditions of uncertainty in contrast with earlier accounts of learning in terms of psychological development and the acquisition of knowledge. Drawing on archival research, the dissertation reconstructs debates among Gregory Bateson, Warren McCulloch, Heinz von Foerster, and other cyberneticians whose work helped shape the early history of artificial intelligence. These postwar debates over learning as adaptive behavior and knowledge as successful control are situated within an intellectual genealogy extending from nineteenth-century thermodynamics and psychophysics to early psychoanalysis and neurophysiology.

Tracing this history reveals a question that continues to animate contemporary artificial intelligence: how one determines that an organism or machine has learned and on what basis that determination can be justified. This tension persisted across debates over neural networks, psychotherapy, animal communication, and conversational computing, where behavioral adaptation and claims to knowledge remained analytically distinct. Although computational systems can generate increasingly successful performances within specified environments, whether such performances warrant the conclusion that learning has occurred remains contested. This dissertation argues that artificial intelligence emerged from a tradition of inquiry

and experimentation in which adaptive performance became the principal criterion for attributing cognition, leaving open the question of whether successful computation provides adequate grounds for such inferences.

INTRODUCTION

In December 1967, as he was organizing a conference on the theme of 'Effects of Conscious Purpose on Human Adaptation' through the Wenner-Gren Foundation for Anthropological Research, Gregory Bateson wrote to Warren McCulloch seeking his "confidential judgment" on an idea about the relation between religion and ecology.[1] His idea emerged from a line of inquiry that had taken him from anthropological fieldwork in New Guinea and Bali in the 1930s into the study of communication and psychopathology in postwar psychiatric research, eventually culminating in a scientific conversion to cybernetics as a framework for understanding purposive behavior and self-regulating systems.[2] Unsure whether the thought was profound or absurd, Bateson proposed that the linear logic of human purpose "inevitably messes up" the cybernetic organization of "his own physiology, social system, and eco-system."[3] One begins to suspect that the error lies in reason itself when the form of action modern civilization calls rational—the instrumental

[1] Letter from Gregory Bateson to Warren McCulloch, dated December 20, 1967, *Gregory Bateson papers*, MS 98, Box 22, 1399-7a. Special Collections and Archives, University Library, University of California, Santa Cruz.

[2] On Bateson's shift to communication as the analytic basis of psychiatric inquiry, see Gregory Bateson and Jürgen Ruesch, *Communication: The Social Matrix of Psychiatry* (New York: W. W. Norton, 1951). Despite his reservations about Freudian materialism—a conceptual point of contact between psychoanalysis and cybernetics taken up in Chapter 2 of this dissertation—Bateson retained the premise that all psychological phenomena are constituted within communicative exchange. The "talking cure" is thereby rendered, in rigorous theoretical terms, as an ordered system of interaction governed by identifiable rules, constraints, and criteria of coherence and distortion.

[3] Bateson's engagement with cybernetics can be traced to its early theoretical formulation in the problem of purposive behavior from the standpoint of communication—particularly as he extended concepts of feedback and the formalization of information and codification into the analysis of human interaction rather than their original development in wartime engineering. The canonical statement is Norbert Wiener, Arturo Rosenblueth, and Julian Bigelow, "Behavior, Purpose and Teleology," *Philosophy of Science* 10, no. 1 (1943): 18–24.

pursuit of short-term goals—systematically undermines the conditions it presupposes. This was precisely the problem Bateson sought to bring into discussion through his conference, to be held at Burg Wartenstein in June 1968.[4]

Bateson quickly illustrated the point with a theological parable: perhaps Original Sin was simply the discovery of planned purpose. Rather than God expelling Adam and Eve from the Garden, Adam and Eve drove God from it; in choosing deliberate control over the systemic balance of the Garden, they soon lost the topsoil: "The general notion was that God symbolized the systemic and cybernetic nature of the environment which inevitably took vengeance on man's shortsightedness."[5] The myth of Adam and Eve becomes an allegory of human cognition condemned to the apprehension of an external order to the exclusion of participation in the relations immanent to it. It marks a shift to purposive thought; the isolation of from the context of their operation installs a basic error in the structure of knowledge itself. Perhaps, then, "the supernatural entities of religion" might be understood as "cybernetic models built into the larger cybernetic system in order to correct for noncybernetic computation in a part of that system."[6]

---

[4] Mary Catherine Bateson, *Our Own Metaphor: A Personal Account of a Conference on the Effects of Conscious Purpose on Human Adaptation* (New York: Alfred A. Knopf, 1972).
[5] Letter from Gregory Bateson to Warren McCulloch, dated December 20, 1967, *Gregory Bateson papers*, MS 98, Box 22, 1399-7a. Special Collections and Archives, University Library, University of California, Santa Cruz. By 1978, he put it plainly: "I do not believe that there is a god separate from the ocean out there… on the other hand, I do have a sense that the ocean is alive. Is that… religious?" Quoted in David Lipset, *Gregory Bateson: The Legacy of a Scientist* (Boston: Beacon Press, 1982), 301.
[6] Ibid.

This problem—the propensity of goal-directed reasoning to detach itself from the conditions of its own viability—stood in tension with the research program Warren McCulloch had helped to define.[7] A neurophysiologist and one of the founding figures of cybernetics, McCulloch had spent decades trying to describe thought in terms of logical networks and neural circuits—most famously in his 1943 collaboration with the young mathematician Walter Pitts, approaching the brain "as if it were a wiring diagram" and the neuron "as if it were a component relay of a computing machine."[8] Bateson often deferred to McCulloch's scientific authority within the cybernetics circle, but spoke in his idiom to pose a question that cut into the computational logic he had helped to establish. Turning first to the psychological domain, he asked: "There are also lots of questions regarding psychotherapy. When

[7] Historians have characterized cybernetics in several overlapping ways: as an interdisciplinary research program organized around the sciences of communication and control; as a loose network of researchers linked through institutions such as the Macy Conferences; and as a historically bounded intellectual "moment" in which diverse fields temporarily converged around shared conceptual tools. Early historical accounts emphasized the interpersonal network of scientists surrounding the Macy Conferences; see Steve J. Heims, *The Cybernetics Group* (Cambridge, MA: MIT Press, 1991). Analyses emphasizing the circulation of cybernetic ideas through Cold War technological infrastructures and scientific discourse include Paul N. Edwards, *The Closed World: Computers and the Politics of Discourse in Cold War America* (Cambridge, MA: MIT Press, 1996), as well as N. Katherine Hayles, *How We Became Posthuman* (Chicago: University of Chicago Press, 1999), which traces how cybernetic concepts reshaped theories of embodiment and subjectivity across scientific and cultural domains. Later scholarship has approached cybernetics as a distinctive experimental practice; see Andrew Pickering, *The Cybernetic Brain: Sketches of Another Future* (Chicago: University of Chicago Press, 2010). Other historians have framed it as a broader intellectual "moment" shaping postwar scientific culture; see Ronald R. Kline, *The Cybernetics Moment: Or Why We Call Our Age the Information Age* (Baltimore: Johns Hopkins University Press, 2015). For documentary and archival reconstructions of the Macy Conferences, see Claus Pias, ed., *Cybernetics: The Macy Conferences 1946–1953* (Zurich: diaphanes, 2003). I synthesize these perspectives by treating cybernetics as a historically situated research program sustained through a network of institutions, scientists, and their experiments and correspondence, through which a shared conceptual vocabulary circulated across disciplinary boundaries while retaining internal tensions. The broader historiographical implications of this framing are addressed more fully in the coming chapter.

[8] Warren McCulloch and John Pfeiffer, "Of Digital Computers Called Brains," *The Scientific Monthly*, Dec., 1949, Vol. 69, No. 6 (Dec. 1949), 369.

the therapist catalyzes group processes, is he in fact demonstrating a cybernetic model to his patients?" Bateson then extended the same problem into the domain of machines: "Are there any cybernetic systems made of hardware in which cybernetic models have to be embedded in order to correct or lineal computation?" He concluded his speculative hypothesis with the question—"Or am I crazy?"—and, in the same breath, a request for "a good theologian."[9]

McCulloch's reply, written a week later, rose to the request, taking up Bateson's conjecture in the same theological register in which it had been posed. "Theopoesis has its animistic origin in empathy," he began, and one feels "as if he hurled the thunderbolt," or "herded the fleecy clouds," or "blew them over the wine dark sea," just as one might imagine oneself "the beast he fought or hunted or befriended."[10] Through this imaginative identification, humans personify the larger systems that sustain them. Gods become figures through which societies acknowledge that survival depends on forces greater than individual intention. Creatures who must "rear their young and hunt in packs or starve" learn to subordinate their private survival to the life of the group, forming symbols—totems, tribal gods, rituals—through which that collective dependency is recognized.

Religious traditions, McCulloch suggested, can thus be read as attempts to grasp "some greater cybernetic system," one in which the boundary between

[9] Letter to Warren McCulloch, dated December 20, 1967, *Gregory Bateson papers*, MS 98, Box 22, 1399-7a. Special Collections and Archives, University Library, University of California, Santa Cruz.
[10] Letter to Gregory Bateson, dated December 28, 1967. *Warren McCulloch Papers,* American Philosophical Society, B/M139, Box 75.

individual and environment dissolves into a more encompassing order. Mystical experience, whether in Neoplatonist theology or in the concept of Nirvana, expresses precisely this intuition: "an intense sense of being at one with the universe but that universe completely unspecified." Affirming the speculation that aesthetic-ethical experience of a more abstract order might function as a corrective to purposive reasoning, Bateson's idea, to McCulloch, was "anything but mad."[11]

McCulloch reversed the direction of Bateson's inquiry in his indirect answer to whether a computational system might require models of itself to correct the limits of its own reasoning. Moving from machine back to psyche and society, the problem can be precisely formulated in the symbolic and cultural practices through which human groups apprehend the larger systems in which they are embedded. Purposive systems depend on a distinction between system and observer that feedback and modeling both sustain and dissolve, a condition McCulloch described as one in which "the disappearance of the line between oneself and all simply evaporates."[12] The problem of mind and its embodiment, in turn, concerns whether a cognitive or computational system can determine the criteria by which its outputs qualify as knowledge.

If the taking of knowledge marks Adam and Eve's separation from the systemic order of the Garden, it also installs the problem that confronts any knowing system: how to determine the scope of its own cognition once that separation is

[11] Ibid.
[12] Ibid.

instituted. How can a system that has eaten from the tree of knowledge still learn the limits of what it knows?

***

Bateson's doubt about purposive reason was already internal to McCulloch's formation, where such limits appeared as psychopathology—as a literal risk of going insane. The problem took shape in the early 1930s, amid the Depression, where McCulloch worked alongside the psychiatrist Eilhard von Domarus, whose *The Logical Structure of the Mind* attempted to formalize the breakdown of reasoning in schizophrenia.[13] McCulloch referred to von Domarus as "the only other trained philosopher I have ever known who was forced into neuropsychiatry by philosophic problems," recalling that they "struck a bargain, he to teach me psychiatry and I to help him put the ideas of his thesis into English."[14]

Von Domarus's *Logical Structure* took the speech of psychiatric patients as evidence for the logical rules by which concepts are connected, tracing how those rules shift under pathological conditions. He argued that schizophrenic reasoning

[13] *The Logical Structure of the Mind* is von Domarus' PhD dissertation written under the direction of the philosopher F. S. C. Northrop at Yale in 1930. In a later biographical account of McCulloch, Northrop recalled that it was through von Domarus that McCulloch first encountered his philosophical program, which sought to place William James's radical empiricism in systematic relation with a realist logic of scientific concepts. See: F. S. C. Northrop, "On W. S. McCulloch," in *The Collected Works of Warren S. McCulloch*, eds. Michael R. Lissack and Anna L. Holland, *Emergence: Complexity & Organization*, accessed March 27, 2026.

[14] The handwritten "belated introduction," from which I quote extensively, is preserved among Warren S. McCulloch's papers alongside a manuscript copy of Eilhard von Domarus's *The Logical Structure of Mind*, Box 68, Warren S. McCulloch Papers, American Philosophical Society, Philadelphia. The introduction was written in 1965, several decades after the original study, which first circulated as an unpublished technical report in 1934: Eilhard von Domarus and Warren S. McCulloch, *The Logical Structure of Mind: An Inquiry into the Philosophical Foundation of Psychology and Psychiatry* (1934), Defense Technical Information Center (DTIC), Accession no. AD0616334.

must be understood within a more general problem: the need for an "analytic attitude" capable of treating mind and matter within a single framework that "transcend[s] the 'individual.'"[15] Thought proceeds by analyzing given facts into "entities called relata," which must then be "synthesized in manners determined by the analysis if that synthesis is to reproduce the original."[16] These manners—relations—are not uniform: "idea" and "intention" organize mental relations, while "passage" and "extension" organize material ones. The logical structure of mind depends on the coordination of these relations within the total context of propositional activity. Where this coordination fails, reasoning can remain formally structured but no longer sustains intelligibility.

In setting out this framework, McCulloch wrote in an "belated introduction" to the work, no other text "so clearly sets forth the notions needed for an understanding of psychology, psychiatry, and finite automata."[17] von Domarus's work elucidated the persistence of an older philosophical problem that preoccupied McCulloch: the relation between what classical philosophy had distinguished as the *lekton*—the "sayable," the propositional content distinct from both the uttered words and their referents—and its formal or extensional articulation.[18] McCulloch

[15] Ibid. In the preface, von Domarus states that the "facts of importance were, and are, particular experiences with particular patients," and presents the inquiry as an attempt to address a "gross hiatus" between psychiatric theory and clinical practice by analyzing how experience is organized and related to logical form.
[16] Ibid.
[17] Ibid.
[18] Warren S. McCulloch, "What's in the Brain That Ink May Character?" in *Embodiments of Mind* (Cambridge, MA: MIT Press, 1965), 389. There McCulloch identifies what is "in the brain" with the Stoic *lekton*, describing it as a conceptual inheritance from Greek philosophy transmitted through the Renaissance, a formulation that demonstrates the persistence of this concept across decades of his work.

lamented that the Stoic attempt to hold together meaning and materiality in a doctrine of pansomatism had been "lost in the mathematics of logic," leaving "naive realism and Platonic idealism, as biology and mathematics, [to go] their separate ways, ignoring the *lekton* and intentionality."[19]

More than a philosophical disagreement, this division concerns the relation between sign, sense, and reference—whether intelligibility lies in what is said or in how it is formally structured. Drawing on a Stoic, physicalist logic of propositions, McCulloch construes signification as neither the material event nor its inscription, but as a third term—an intentional "grasp" (*lekton*) naming the relation that binds event and expression in thought, mediating between material inscription and logical manipulation without coinciding with either.[20] The severing of sense, as lived and interpreted, from relations formalized into structures that admit of manipulation in its absence is a "categorical confusion."[21] On this view, information-theoretic capacity—the ability of a system to reliably transmit and differentiate signals under conditions of noise—presupposes the prior stabilization of distinctions that can bear semantic content, without ever incorporating such content into its formal operations.[22]

---

[19] Warren McCulloch, introduction to Eilhard von Domarus' *The Logical Structure of Mind*, Box 68, Warren S. McCulloch Papers, American Philosophical Society, Philadelphia.
[20] Warren McCulloch, "The Cybernetic Problem of Learning," 69.
[21] Warren McCulloch, introduction to Eilhard von Domarus' *The Logical Structure of Mind*, Box 68, Warren S. McCulloch Papers, American Philosophical Society, Philadelphia.
[22] Claude Shannon and Warren Weaver, *The Mathematical Theory of Communication* (Urbana: University of Illinois Press, 1949). Shannon and Weaver formalize coding as the transformation of events in the world into symbolic information: the production of messages depends upon the reliable preservation of the rule-governed relations among those events. This formulation is foundational for cybernetic vocabulary, insofar as it defines communication in terms of signal transmission and the preservation of formal relations. Its implications for computation and the broader organization of control in modern technical and economic systems will be taken up in length in the first chapter.

Formalization thus abstracts from the intentional relation that the *lekton* names, concealing computation's dependence on the maintenance of such distinctions at their limits.

This problem would come to organize his efforts to model the mind as a system of logical operations in the decade to follow. The "mathematics of logic" took a remarkable technical form in his and Pitts' neural network, which demonstrated how networks of idealized neurons could generate propositions through their connectivity, providing both a model of brain function and a blueprint for computational systems that still undergird the basic architectures of contemporary AI.[23] Crucially, in their original formulation, these propositions were events in time and space within a physically realized network: the neurons remained strictly material components, and the nets themselves were constructed and orderly in a way real brains are not.[24] In this respect, the model could be read as an attempt to specify the

[23] Warren S. McCulloch and Walter Pitts, "A Logical Calculus of the Ideas Immanent in Nervous Activity," *Bulletin of Mathematical Biophysics* 5 (1943): 115–133. The significance of this paper—as a foundational articulation of neural networks, computational models of mind, and the relation between logic and neurophysiology—has been the subject of sustained historiographical and theoretical engagement; see, e.g., Lily E. Kay, "From Logical Neurons to Poetic Embodiments of Mind: Warren S. McCulloch's Project in Neuroscience," *Science in Context* 14, no. 4 (December 2001): 591–614; Tara H. Abraham, "(Physio)logical Circuits: The Intellectual Origins of the McCulloch-Pitts Neural Networks," *Journal of the History of the Behavioral Sciences* 38, no. 1 (2002): 3–22; Gualtiero Piccinini, "The First Computational Theory of Mind and Brain: A Close Look at McCulloch and Pitts's 'Logical Calculus of Ideas Immanent in Nervous Activity,'" *Synthese* 141, no. 2 (August 2004): 175–215; Orit Halpern, "Circuits of Amnesia: Cybernetic Memory and Real-Time Analytics," *Cultural Studies ↔ Critical Methodologies* 16, no. 5 (2016): 442–51; Joseph Dumit, David Bates, and Nima Bassiri, "Plastic Diagrams: Circuits in the Brain and How They Got There," in *Plasticity and Pathology: On the Formation of the Neural Subject*, ed. David Bates and Nima Bassiri (New York: Fordham University Press, 2020), 219–267; Gabriel de Blasio et al., "McCulloch's Relation to Connectionism and Artificial Intelligence," in *Computer Aided Systems Theory – EUROCAST 2017*, Lecture Notes in Computer Science 10671 (Cham: Springer, 2018), 41–48.

[24] Warren S. McCulloch, "What's in the Brain That Ink May Character?" in *Embodiments of Mind* (Cambridge, MA: MIT Press, 1965), 391.

material conditions under which propositions—understood as what is said or meant, rather than the physical signals that convey them—can become expressible without being reducible to those conditions. Its achievement lay in showing how propositional content could be expressed despite variability in neuronal activation and the noise inherent in the transmission of electrical signals across the network.

Yet it was precisely this constructed orderliness that enabled a change in how the model could be *used*. Because the network was fully specified and its elements idealized, the conditions under which propositional events remained intelligible—noise, variability, material instability— could be treated as fixed parameters and excluded from the variables the model operates over. In reducing neurons to threshold elements, the model preserved their physicality while rendering their variability operationally irrelevant.[25] The problem of sustaining what is said in an unreliable medium was thereby displaced: the system no longer addressed what its states refer to, but operated on them as given, leaving the relation between symbol and referent unspecified. Reliability was thus relocated from the dynamics of use to the conditions of design, where variability is controlled in advance. As computation came to be defined in terms of transitions between discretely specified states, reasoning appeared as a sequence of operations whose correctness is internal to the system itself.[26] With

[25] See N. Katherine Hayles, *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics* (Chicago: University of Chicago Press, 1999), especially Chapter 3, on the cybernetic redefinition of information as a pattern independent of its material instantiation, whereby embodiment is rendered secondary to the informational patterns that can be instantiated across different material substrates.

[26] Alan Turing, "On Computable Numbers, with an Application to the Entscheidungsproblem," *Proceedings of the London Mathematical Society* 42, no. 2 (1936): 230–65; and John von Neumann, "Probabilistic Logics and the Synthesis of Reliable Organisms from

intensional content bracketed and extensional relations governing operation, intelligence becomes measurable as performance within such systems.

The conditions under which a system sustains the coherence of its own operations are excluded by formalization, and in mid-century cybernetics this exclusion appeared as a problem at the level of the living system. The early computational imagination thus took shape across biology, mathematics, psychology, engineering, and philosophy as an inquiry into how reasoning changes when the system that reasons becomes part of what must be reasoned about, rather than executing rules within a fixed formal structure. As von Domarus noted in his preface, McCulloch's central concern was "how organisms may be affected so that new propositions are induced"—that is, how *learning* is constituted within the organism as a cybernetic system describable in terms of its logical relations.[27] The fact that "men go mad" was inseparable from this problem: what does it mean for an organism to form new propositions under conditions that may alter or disrupt the very relations that make thought intelligible—relations that cannot be secured independently of the system's material organization? In this sense, Bateson's question—"Or am I crazy?"—marks the return, within cybernetics, of the difficulty von Domarus identified in psychopathology: reasoning that remains formally coherent while failing to register the conditions under which it makes sense.

***

Unreliable Components," lectures delivered at California Institute of Technology, January 14–15, 1952, notes by R. S. Pierce, Caltech Eng. Library, QA.267.V6.

[27] Eilhard von Domarus, *The Logical Structure of Mind*, Box 68, Warren S. McCulloch Papers, American Philosophical Society, Philadelphia.

The digital computers, neural models, and theories of information that emerged from this moment carried forward an unresolved tension: the criteria by which learning was assessed privileged the stability of formal relations over the conditions that make those relations meaningful. For centuries, theories of learning addressed how individuals form and revise understanding—how they interpret experience and transform their relation to what they know. The same term designates the performance of technical systems today: their efficiency, predictive capacity, and ability to generate and manipulate propositions within formally specified domains. This marks a conceptual restriction whereby learning is defined as the adjustment of internal relations within a system rather than any transformation in its relation to what lies beyond it.

What follows traces the historical formation of this conception of learning as a computational capacity: the optimization of system performance under conditions of uncertainty, in which knowledge is defined by its operational usefulness. It examines how this conception has come to organize contemporary machines, markets, and institutions, which register and act on the world according to criteria specified only by their own parameters, and thereby foreclosing any external grounds on which their outcomes might be questioned or judged.

I did not set out to write about computers. Still less did I intend to write about artificial intelligence or machine learning. I know little about computers as machines. What interests me is the problem that had troubled Bateson and McCulloch—and, in an earlier form, the Greek philosophical tradition: the conditions under which a

subject can recognize that it has learned. It is here that I locate the connection between the domains of psychology and computation, in their shared concern with how a "mind" determines that it has learned and how the grounds of that determination are themselves established. This problem finds a classical formulation in Plato's *Meno*, where the paradox of inquiry establishes the following: if the object of inquiry is already known, inquiry is redundant; if it is not, the criteria by which it would be recognized are unavailable.[28] The problem is reflexive. Knowledge depends on recognition, yet recognition presupposes a prior grasp of what is sought. Furthermore, learning requires responsiveness to inputs that cannot be specified in advance and which therefore arrive as surprise. Yet by what means can a system register novelty without assimilating it to prior categories? And under what conditions can something be classified as unpredictable if it is recognizable at all?

For much of intellectual history, learning named a bridge between knowledge and change: the possibility that experience could transform action and that systems might correct the limits of their own reasoning. It carried the hope that inquiry might draw knowledge out of ignorance rather than simply transmit it from teacher to pupil. Yet in the contemporary computational order, learning increasingly designates the capacity of systems to adjust behavior in response to data. Machine learning systems are said to detect patterns in environments too complex for explicit programming, with deep learning architectures and generative models trained on

[28] I am indebted to Robert Meister for directing me to the *Meno* as a way of thinking through Bateson's logical categories of learning at the outset of this project; that I came to recognize this problem through instruction is itself an instance of the paradox the dialogue articulates.

large datasets to generate predictions and produce language from new inputs.[29] The exponential pace at which these systems are developing is often interpreted as an early step toward "artificial general intelligence"—an idealized and largely speculative goal, sometimes described as the field's "holy grail": the hypothetical capacity of a machine to perform diverse across domains, rather than being limited to pattern recognition within specific training contexts.[30] Yet contemporary public and corporate discourse is not organized around the problem of generalized intelligence. At the time of writing in March 2026, NVIDIA CEO Jensen Huang stated, "I think we've achieved AGI," referring to deep learning systems that can independently generate digital products and bring them to market, where deployment is governed by whether their outputs meet return thresholds sufficient to justify large-scale capital

[29] Notable recent scholarship on LLMs includes *Language Machines: Cultural AI and the End of Remainder Humanism*, by Leif Weatherby (Minneapolis: University of Minnesota Press, 2025), and *Last Words: Large Language Models and the AI Apocalypse*, by Paul Kockelman (Chicago: Prickly Paradigm Press, 2024). These works are productively read together for their divergent accounts of what LLMs are: where Weatherby seeks to grant them cultural and theoretical standing as instantiations of language as "an internally structured web of signs," rejecting "remainder humanism" as an attempt to distinguish "real" from "artificial" language, Kockelman treats them as mathematical-semiotic systems whose apparent intentionality is inherently "parasitic" on the human agents and institutional regimes that produce and train them, foregrounding their political-economic effects and consequences—an analysis I find more productive for situating LLMs within their historical conditions.

[30] The concept of general intelligence in artificial intelligence research constitutes an intellectual history of its own, emerging from a heterogeneous set of research programs spanning symbolic artificial intelligence, cognitive psychology, mathematical logic, cybernetics, and statistical learning theory, and falls outside the scope of this work in its current form. It is introduced here only as a term of art designating the extension of task-specific performance across domains, without entering the speculative claims advanced on its behalf. For representative formulations across these traditions, see Allen Newell and Herbert A. Simon, *Human Problem Solving* (Englewood Cliffs, NJ: Prentice-Hall, 1972); John McCarthy et al., "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence" (1955); David Marr, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information* (San Francisco: W. H. Freeman, 1982); Vladimir Vapnik, *The Nature of Statistical Learning Theory* (New York: Springer, 1995); and Christopher M. Bishop, *Pattern Recognition and Machine Learning* (New York: Springer, 2006). A textbook overview of these approaches is provided by Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed. (Upper Saddle River, NJ: Prentice Hall, 2010).

investment.[31] The imperative of short-term value extraction thereby redefines general intelligence as the capacity of autonomous systems to accelerate product development and generate rapid, market-mediated returns.[32]

This definition corresponds to what Marion Fourcade and Kieran Healy identify as a transformation in contemporary institutional regimes of evaluation, in which classification and ranking collapse into one another in the production of social knowledge and value. Today, as machine learning systems are folded into financial, communicative, and ecological infrastructures, the traditional processes of nominalization and ordinalization—"the naming of kinds and the designation of ranks"—are reorganized through computational metrics that privilege position over interpretation.[33] The question of whether a model captures the structure of social processes is thereby displaced by whether it yields value as a function of reliable performance within the parameters it helps to constitute. Under these conditions, organizations confront technical advances that expand both the quantity and form of information available, but exacerbate the central difficulty of determine how to extract meaningful knowledge from it: "the feedback between effective prediction and presumptive causality can start to loop in ways that are

[31] Lucy Buchholz, "Why NVIDIA CEO Jensen Huang Believes AGI Has Arrived," *AI Magazine*, March 25, 2026, https://aimagazine.com/articles/agi-race-intensifies-as-nvidia-ceo-says-its-already-achieve (accessed March 31, 2026); "I Think We've Achieved AGI… I Don't Think We Have," *TechRadar*, March 2026, https://www.techradar.com/ai-platforms-assistants/i-think-weve-achieved-agi-er-jensen-i-dont-think-we-have (accessed March 31, 2026).
[32] See Orit Halpern, "Financializing Intelligence: On the Integration of Machines and Markets," *e-flux Architecture*, March 2023.
[33] Marion Fourcade and Kieran Healy, *The Ordinal Society* (Cambridge, MA: Harvard University Press, 2024), 105-107.

counterproductive."[34] This problem is historically constituted and not new: any learning system must distinguish signal from noise without already possessing the rules that would authorize that distinction.[35] It cannot be resolved—neither by complexity nor by scale, whether through new algorithmic techniques or the expansion of computational capacity, nor through the continued investment in and commercialization of data infrastructures. Learning persists only because cognitive and computational systems—inseparable from the social and ecological systems in which they operate—must continue to act, decide, and adapt without ever establishing whether their outputs constitute knowledge; it is under these conditions that discernment is required.

The work can be read as a kind of found footage assembled from the archives of American cybernetics.[36] The fragments come from the papers of Gregory Bateson at Special Collections and Archives, University of Santa Cruz; Warren McCulloch at the American Philosophical Society Library; and Heinz von Foerster and the Biological Computing Laboratory at the University of Illinois, Urbana-Champaign—letters, research recordings, grant proposals, unpublished essays, and talks that never reached print. These figures do not exhaust the history of cybernetics; in the decades

---

[34] Ibid., 96.

[35] See Lorraine Daston, *Rules: A Short History of What We Live By* (Princeton, NJ: Princeton University Press, 2022), which shows that algorithms and the practices of counting and data collection are themselves not new, but continuous with long histories of rule-governed inscription and social organization.

[36] Bernard Dionysius Geoghegan, *Code: From Information Theory to French Theory* (Durham, NC: Duke University Press, 2023. Geoghegan offers an important precedent for the archival rearticulation of cybernetics in the present, reconstructing it as an epistemological and institutional formation emerging from mid-twentieth-century alliances among philanthropy, the human sciences, and military-technical research, in order to track the historical conditions and afterlives of cybernetic thought.

following the Second World War, however, the United States became its administrative and financial center, and the archives of its participants preserve the exchanges through which concepts of information, learning, and memory were argued over as the working vocabulary of a new scientific program. If historical work remains to be done with the cyberneticians, it lies in attending to what was consigned to the unspoken and unavowed corners of their correspondence. It is there, in marginal notes and unfinished exchanges, that one finds the coexistence of explicit commitments with what could not be acknowledged as such—where the field's conceptual conflicts persist in a form at once held and disclaimed, and from which its record may be read anew.

Yet returning to canonical figures risks reproducing the familiar portrait of cybernetics as a small circle of brilliant men (and Margaret Mead) inventing the future, a narrative that arguably continues to animate the mythology of technological genius surrounding Silicon Valley today.[37] The archive of cybernetics is unmistakably gendered and geographically narrow, dominated by the scientific and institutional authority of mid-century American and European men. Approaching these materials from outside the cultural world that produced them makes this

[37] For example, the cultural uptake of Stoicism within Silicon Valley as a form of self-optimization and ethical self-fashioning among participants in its entrepreneurial and venture capital milieu has been noted in the past decade, often distorting Stoic philosophy in both its popular appropriation and media representation. See Nellie Bowles, "Why Is Silicon Valley So Obsessed With the Virtue of Suffering?," *The New York Times*, March 26, 2019, https://www.nytimes.com/2019/03/26/style/silicon-valley-stoics.html (accessed March 31, 2026); and Donna Ferguson, "The Stoicism Secret: How Ryan Holiday Became a Silicon Valley Guru," *The Guardian*, October 28, 2024, https://www.theguardian.com/books/2024/oct/28/the-stoicism-secret-how-ryan-holiday-became-a-silicon-valley-guru(accessed March 31, 2026).

dominance difficult to ignore. Yet it is precisely within the papers of these dominant figures that the advantage of posterity becomes visible. From this distance one can see the habits of thought, the resistances and inhibitions, the blind spots and epistemic gaps that shaped how they attempted to define the relation between learning and communication, between knowledge and control. Recognizing these limits also revealed one of the project's own.

My aim, therefore, is to examine, without celebration or condemnation, the institutional positions from which these actors thought and the conceptual limits those vantage points imposed. McCulloch presided over the Macy Conferences, the meeting ground where the language of feedback, control, and information first cohered into a scientific program. Von Foerster served as the conferences' administrator, working closely with McCulloch and helping to sustain this interdisciplinary experiment. With McCulloch's support, he emigrated to from Vienna to Chicago and established the Biological Computer Laboratory at the University of Illinois—an institution that functioned at once as research center, experimental workshop, and archival repository of cybernetic thought. Bateson was never formally part of this institutional project, yet his thinking often converged with von Foerster's philosophical exploration of recursive systems and the epistemological problems raised by cybernetics across its mid- to late-twentieth-century iterations.

***

Working at the microlevel of individual archival findings yields a macrolevel account that complicates the presumed separation of cybernetics and computing as

distinct intellectual and institutional formations. The materials preserved in these collections document overlapping lines of inquiry linking biochemistry, neurophysiology, psychology and psychiatry, statistics, mathematical logic, electrical engineering, control and information theory, and operations research across transatlantic research networks shaped by wartime displacement and postwar institutional consolidation from the 1940s to the 1960s, prior to their consolidation as computer science and artificial intelligence. The same figures operated across problems of neural organization, statistical inference, electrical circuitry, and animal communication, each treated as an instance of a common analytic problem; the differentiation of these lines of inquiry into distinct technical domains—digital computing, neural network modeling, feedback control systems, natural language processing—was a later development.

From this organization of inquiry follows a tension within cybernetic thought that mirrors the competing notions of learning examined in this dissertation. On the one hand, cybernetics appears as a precursor to the computational age, furnishing concepts and techniques that would be consolidated in digital infrastructures. On the other, it persists as an epistemological project concerned with how systems come to know anything at all, including their own operations. The technological afterlife of cybernetics depended on the first interpretation; the second—the reflexive inquiry into the conditions of knowledge—was progressively displaced as computational systems moved from speculative models to operational infrastructures. The machines

endured; the questions that produced them were reorganized and, in many cases, subordinated to their operation.

The aim of this dissertation is not to restore that reflexive dimension as a humanistic counterweight to technological power, nor to idealize an earlier cybernetics as if it contained the moral resources to correct what came after. Rather, it is to rethink the relation between the psychological, the social, and the technological in the terms that cybernetics initially brought into view. Midcentury cyberneticians, working across objects of study ranging from stored-program electronic digital computers to animals at play, asked how systems learn, adapt, and correct themselves. Those same terms now describe the intelligence and efficiency of computational systems that increasingly organize everyday life. The problem is not simply that machines might replace labor, thought, or creativity. It is that learning itself has come to mean something different in a world organized around the processing of information. The distinction between learning as the capacity to transform the conditions of collective life and learning as the capacity to improve systems that learn faster than we do is, in Bateson's sense, a difference that makes a difference. It marks a point of information essential for understanding the self, our relations to others, and the future of the world we inhabit.

The now-canonical account of the temporal horizon of cybernetics locates its conceptual origins in the nineteenth-century sciences of energy and the technological paradigm of the steam engine, where regulation first took mechanical form. The centrifugal governor, the relation between entropy and information, and later

analogies between brains and digital computers established a framework in which engines, organisms, and computing machines were treated as instances of a common problem: how systems register deviation and act on it.

For this reason, the history developed here begins not with computers but with industrial machinery and the energetic sciences that made their operation intelligible. The opening chapter examines the mechanical governor within the broader context of thermodynamics and energetics, situating the regulation of energy within the organization of industrial production and the extraction of value. As capital shifts from industrial production to financial markets and digital infrastructures, these logics are not superseded but rearticulated: prediction, control, and governance take form through the operationalization of uncertainty as a condition of action and value creation. This study follows that rearticulation to the present, where machine learning systems function as instruments for the large-scale extraction and monetization of data. The aim is not to posit a continuous development from nineteenth-century energetics to contemporary machine learning, but to specify the persistence of a problem: how learning comes to be defined as the adaptive adjustment of a system's internal relations to signals, rather than as a transformation in its relation to the world those signals are taken to represent.

The second chapter follows the problem of regulation from industrial energetics into an unexpected terrain: Sigmund Freud's early attempts to construct a materialist account of the mind. Freud's *Project for a Scientific Psychology* (1895) sought to explain mental life in terms of energetic flows, thresholds, and mechanisms

of discharge, drawing on the nineteenth-century sciences of energy and on psychophysical attempts to formalize the relation between stimulus and sensation, most notably in Gustav Fechner's Weber–Fechner law of 1840. Read from the vantage point of later cybernetics, these formulations appear as an early articulation of a problem that would be recast in informatic terms: the modeling of mental processes as the transmission, transformation, and regulation of signals. The title of this dissertation, *Project for a Cybernetic Psychology*, echoes Freud's to mark this continuity, indicating that the conceptual vocabulary through which learning and adaptation were later defined in machines was first developed in attempts to describe the physical dynamics of psychic life.

From there, the dissertation turns to mid-century cybernetics, following the migration of a problem. Questions posed in early governor theory and in Freud's neurological attempt to formalize the dynamics of excitation and response reappear in a mid–twentieth-century cybernetic context, reformulated in explicitly formal and operational terms under the conditions of wartime research and its institutional aftermath. The third chapter centers on the work of McCulloch and Pitts and their attempt to describe neural activity in logical terms. As discussed above, this model of the neuron as a logical device defined it as a computational element, thereby redefining learning as a function of discrete operations within a system and establishing the formal basis for later artificial neural networks. Yet the model introduces a logical paradox: the system can change its outputs in response to feedback, but the rules governing that change are themselves invariant within the

system's operations. Bateson's work, including his exchanges with McCulloch, renders this tension explicit by reformulating learning in terms of levels and context as communicative categories, distinguishing between adjustment within a system and changes in its governing rules—a distinction cybernetics ultimately could not resolve.

The final chapter brings this history into the present through the emergence of conversational computing and natural language processing, reconstructing archival continuities that link midcentury cybernetics to the institutional consolidation of computer science as a distinct discipline in the late 1960s through the 1980s. It centers on computer scientist Joseph Weizenbaum's ELIZA as an early instance of conversational computing, a rule-based natural language processing program operating through scripted pattern matching. ELIZA established a paradigm for human–machine interaction later generalized in chatbots and large language models, even as those systems differ in relying on statistical inference and probabilistic, data-driven methods. Although not formally situated within cybernetics, Weizenbaum's work was shaped by the epistemological tensions it introduced into computer science, evident both in the design and reception of ELIZA and in his subsequent critique of artificial intelligence. Systems capable of simulating understanding may operate through short-term adaptive pattern recognition—manipulating symbols in accordance with formal rules—without distinguishing such adjustment from learning as the transformation of the conditions under which those symbols are taken to refer. In this sense, Weizenbaum recapitulates Bateson's critique of purposive reason

in a different institutional and technical register, as a problem bearing directly on the social organization of knowledge under computation.

Once integrated into contemporary computational infrastructures, the "categorical confusion" assumes a determinate political and ethical form. At issue in machine learning is the evidentiary basis on which its outputs are taken to constitute understanding. This basis replaces interpretive judgment with the automated production and ranking of outputs, delegating validation to computational systems. It reconfigures belief as manufactured confidence derived from continuous measurement within data-driven evaluation systems, constraining understanding to the recognition of statistically salient patterns rather than inquiry into their significance. The relation between knowledge and control that first appeared in the steam engine and the governor persists in machines that speak, predict, and generate language as a relation between knowledge and its use. The user's capacity to know—to think, discern, make meaning—is rendered measurable, conditioned by the optimization of learned patterns and thereby incorporated into systems of predictive valuation and behavioral surveillance.

In systems designed to correct their errors, adaptation succeeds precisely because the premises that govern that correction are not interrogated; their persistence is itself a symptom. Learning, in its computational form, thus comes to define intelligence as a compulsion to adapt to existing systems—a pattern of adjustment that optimizes behavior without transforming the conditions that generate it. As such systems are incorporated into the infrastructures of communication and governance,

they produce forms of systemic shortsightedness that destabilize the ecological and social conditions on which they depend. The capacity to refuse such adaptation lies in resisting both the purposes intelligence is made to serve and the conditions that organize its use. This requires recognizing that the terms under which survival is defined are contingent arrangements reflecting particular interests and exclusions, and are therefore subject to revision.

Like Adam and Eve after the acquisition of knowledge, the problem is not simply that the world has changed, but that the conditions under which knowledge is interpreted and acted upon have changed with it. The world has long since given way to a landscape of existential insecurity, managed scarcity, ecological degradation. After the Fall, learning names the problem of acting with knowledge in the absence of any guarantee that it secures the conditions of survival, and of doing so without mistaking survival for wisdom.

CHAPTER 1

LEARNING TO GOVERN: "CYBERNETICS" IN THE 19TH CENTURY

In 1859, the Glasgow engineer Thomas Hart and his brother John published in the *Transactions of the Glasgow Archaeological Society* transcripts of interviews they had conducted decades earlier with the Scottish inventor James Watt.[38] Asked how the epoch-making idea of reforming the steam engine with a "separate condenser" first occurred to him, Watt recalled a Sabbath walk across the Glasgow Green in the spring of 1765:

> I had gone as far as the Herd's house, when the idea came into my mind, that as steam was an elastic body it would rush into vacuum, and if a communication was made between the cylinder and an exhausted body [containing a vacuum], it would rush into it, and might there be condensed without cooling the cylinder… I had not walked farther than the Golf-house… when the whole thing was arranged in my mind.[39]

A year before this supposedly revelatory walk, the natural philosopher John Anderson—Watt's colleague in the Glasgow College community—had presented him with a model of an engine first developed by the English inventor Thomas Newcomen.[40] Installed in a coal mine in 1712, Newcomen's atmospheric engine provided a reliable means of pumping water and proved indispensable to mining at a critical moment in eighteenth-century industrial expansion. It marked a major

[38] R.L. Hills, *Power from Steam: A History of the Stationary Steam Engine* (Cambridge University Press, 1989), 53.
[39] Ben Russell, *James Watt: Making the World Anew* (London: Reaktion Books, 2014), 107.
[40] Ibid.

advance over Thomas Savery's earlier steam pump and was quickly patented and widely adopted.[41]

As Watt dismantled the model before him, he examined a metal cylinder fitted with a sliding piston. Chains connected the piston to one end of a wooden beam, while long rods linked the opposite end to pumps deep underground. Steam filled the cylinder; the supply was then cut off and cold water injected to condense the vapor. The resulting vacuum beneath the piston allowed atmospheric pressure to drive it downward, rocking the beam and operating the pumps that lifted water from the mine. The counterweight of the beam then returned the mechanism to its starting position, and the cycle repeated.

The Newcomen engine operated without any formal thermodynamic understanding and consumed enormous quantities of coal. Watt recognized that its inefficiency stemmed from the repeated heating and cooling of the cylinder: before each stroke, steam had to warm several tons of cast iron, only for cold water to condense the vapor and dissipate that heat moments later. The cycle squandered energy by reheating the cylinder at every iteration. Watt's crucial innovation was the addition of a "separate condenser," a vessel in which steam could be condensed without cooling the main cylinder. By keeping the cylinder hot while maintaining the

[41] Brian Spear, "James Watt: The steam engine and the commercialization of patents," *World Patent Information*, 2008, Vol.30 (1), 54.

condenser at a low temperature to produce the necessary vacuum, the engine dramatically reduced fuel consumption—ultimately tripling its efficiency.[42]

In subsequent refinements, Watt extended this principle of regulation to the control of motion itself. By the early nineteenth century he incorporated a centrifugal governor to regulate engine speed. Steam output was linked to a throttle valve governed by a centrifugal sensor: a rotating spindle fitted with two weighted balls that rose as rotational speed increased. When the balls lifted beyond a set range, their linkage partially closed the valve, reducing steam flow and slowing the engine; as speed dropped, the valve reopened. The mechanism converted fluctuations in motion into corrective adjustments, maintaining a stable operating range through continuous feedback.

Although later celebrated as the inventor of the steam engine, Watt described himself more modestly as "a mere compilator," one who had "tacked and sewed together the seraffo and remnants of many a man's ideas."[43] Yet his innovations were as much institutional and economic as mechanical. His 1769 patent, *"A New Invented Method of Lessening the Consumption of Steam and Fuel in Fire Engines,"* claimed a wide range of designs and, as historian David Philip Miller notes, secured him considerable influence over late eighteenth-century technological development.[44] In partnership with the manufacturer Matthew Boulton, Watt translated efficiency gains

[42] Russell, 76.
[43] David Philip Miller, *The Life and Legacy of James Watt* (University of Pittsburgh Press, 2019), xii.
[44] Miller, 54.

into industrial scale, disseminating Boulton & Watt engines across the West Midlands and beyond.[45] Where the earlier Newcomen engine had been confined largely to coal mines—where fuel was abundant enough to offset its waste—the improved fuel economy of Watt's design freed steam power from dependence on cheap coal at the point of use. Steam engines could now be installed in factories, mills, canals, and transport systems, driving pumps, textile machinery, ironworks, and eventually locomotives and steamships.[46] By the early nineteenth century, regulated steam power had become a general infrastructure of motion, enabling the continuous conversion of heat into controlled mechanical work across industrial life.

Watt's engine stages a primal scene of industrial modernity, from which a fundamental problem of instability emerges. Steam engines transformed heat into motion, but that motion was never naturally uniform: loads shifted, friction accumulated, fuel varied, and minor disturbances propagated into oscillation or mechanical failure. The problem confronting nineteenth-century engineers was not machine "intelligence" in any modern sense, but the regulation of energy under conditions of variable load. The centrifugal governor nonetheless anticipated a later logic: fluctuations in motion were registered as differences, and those differences were fed back to stabilize the flow of power.

This chapter reconstructs a nineteenth-century prehistory of cybernetics by showing how mechanical instability became governable once deviations could be

[45] Spear, 54.
[46] James R. Beniger, *The Control Revolution: Technological and Economic Origins of the Information Society* (Harvard University Press, 1986), 174-177.

measured, compared to a norm, and corrected in time. In James Clerk Maxwell's 1868 analysis of the centrifugal governor, Watt's steam engine provided the concrete mechanism through which regulation could be expressed in mathematical form. Maxwell treated fluctuations in rotational speed as departures from a defined operating state and represented these departures as measurable differences relative to a reference value. These discrepancies are fed back into the mechanism and translated into corrective adjustments that damp oscillation over successive intervals. Control thus becomes a *temporal* operation: stability is not given but continuously produced through comparison and correction. What emerged as theoretically decisive is measurable difference in time. The governor therefore provides an early technical model of learning anticipating a computational logic later instantiated in digital systems: variation furnishes the signal, deviations are registered against a reference state, and adaptive adjustment proceeds through iterative recalibration that reduces error and renders system behavior governable.

What Watt could not have anticipated, and what Maxwell began to make explicit, was that the regulation of mechanical instability implied a new way of thinking about value. Once deviation can be measured and corrected over time, temporal difference itself appeared actionable. Information, in this sense, is what can be retained, stored, and processed across successive intervals, even as the system's total energetic capacity remains bounded by thermodynamic constraint. Time thus becomes the axis along which heat, information, and economic value are coordinated. From thermodynamics to cybernetics to information theory, a common

logic thus takes shape: value is generated not through the simple expenditure of energy but through the management of probabilistic difference over time. In its later financial articulation, I argue, this logic culminates in models that price volatility, transforming probabilistic variation into a calculable asset. What begins as a problem of mechanical stability, in other words, evolves into an algorithm of temporal value creation.

Over the first half of the twentieth century, this logic migrated across domains concerned with control under conditions of uncertainty—telecommunications, ballistics, physiology, and computing—before crystallizing in what historian Ronald Kline periodizes as the postwar "cybernetics moment." In the late 1940s to the 1950s, overlapping military, industrial, and academic research programs recast regulation as a "naturalized narrative" of a post-industrial "informatics craze" or "information age."[47] Within this milieu, "information" was stabilized as a universal metaphor for "what is processed, stored, and retrieved in physical, biological, and social systems."[48] Techniques first developed to stabilize mechanical motion were recast as a general schema for regulation: feedback became a portable procedure for monitoring deviation and effecting adjustment across heterogeneous domains. No longer confined to engines, it served as a model for adaptive behavior in organisms, institutions, and markets. Within this expanded framework, learning came to denote a

[47] Ronald Kline, *The Cybernetics Moment, Or Why We Call Our Age the Information Age* (Baltimore: John Hopkins University Press, 2015), 6.
[48] Ibid.

system's capacity to register deviation, update its state, and improve performance relative to a specified goal—a definition that, as we shall see, now underlies contemporary machine learning, where model parameters are iteratively adjusted to minimize error across successive inputs.

This extension coincided with what Jim Davis and Michael Stack describe as the "digital advantage": once information could be discretized, copied, and transmitted at negligible marginal cost, variation itself became a scalable resource. With mass digitization brought about by computers, software, and communications technologies, engineers learned to treat information as a resource-conservative, cheaply stored, transportable, infinitely reproducible, and manipulable commodity form governed by formal laws of communication. What Davis and Stack call the "digital advantage" did not merely improve efficiency; it enabled the continuous monitoring, sorting, and regulation of activity across global populations. The same informational infrastructures that promised optimization also marked "the end of the old social contract, and the beginning of what can only be described as a policy of genocide against the former working class … the end of welfare, prisons (or their digital surrogates of electronic ankle bracelets and other high-tech controls), and the death penalty implement it."[49] Information systems, in this sense, intensified the regulatory logic already visible in the governor: continuous data capture, algorithmic

[49] Jim Davis and Michael Stack, "The Digital Advantage," *Cutting Edge: Technology, Information Capitalism and Social Revolution* (Verso Books, 1997), 128-137.

classification, and automated intervention allow deviation to be detected and acted upon at scales and speeds previously impossible.

In the contemporary moment, computational finance renders this logic actionable in a distinct register: financial instruments function as algorithmic control systems that extend the regulatory logic set in motion by Watt, translating temporal fluctuation into data streams to be optimized and recursively acted upon. In options markets, volatility—the statistically estimated dispersion of price over time—enters directly into valuation. Traders do not forecast a single future price; they model the distribution of possible price movements and price contracts accordingly. Risk desks estimate implied volatility from observed option prices, hedge portfolios through continuous rebalancing, and manage exposure to changes in variance rather than to directional price moves alone. This procedure rests on a statistical premise long noted by cyberneticists and physicists: improbable events carry more informational weight than probable ones."[50] What is least likely is therefore most informative. The Black–Scholes–Merton framework translates that premise into practice by pricing options as functions of time to expiration, interest rates, and expected variance.[51] When volatility rises—during liquidity crises or monetary shocks—option values increase

[50] As cyberneticists such as Warren McCulloch and physicists like Erwin Schrödinger observed, information is measurable as negative entropy—"the logarithm of the reciprocal of the probability of the state. Warren McCulloch, *Embodiments of Mind* (The MIT Press, 1965/2016), 76; Erwin Schrödinger, *What Is Life? The Physical Aspect of the Living Cell* (Cambridge University Press, 1944), 108-111.

[51] Fischer Black and Myron Scholes, "The Pricing of Options and Corporate Liabilities," *The Journal of Political Economy* (1973), 81(3), 638; Fischer Black, "Noise," *The Journal of Finance* 41, no. 3 (July 1986): 528–43.

because the distribution of possible outcomes widens. What matters, then, is not unpredictability in the abstract but the measurable dispersion of uncertain price paths. Variance becomes tradable exposure: deviations from expected price trajectories are formalized as risk parameters and incorporated into pricing, hedging, and portfolio construction, rendering uncertainty a field for continuous probabilistic adjustment.

Seen in this light, the political economy of information culminating in modern financial markets marks a definitive shift in the material basis of accumulation that extends beyond the energy regimes of industrial capitalism. As Andreas Malm has shown, nineteenth-century industrial production turned on the conversion of fossil energy into motive force: value was generated through the controlled expenditure of heat.[52] By contrast, as James Beniger argued in *The Control Revolution*, the twentieth century reorganized economic life around the management of information as distinct from energy and matter—what later condensed into the label "the Information Age."[53] Where industrial capitalism extracted value by mobilizing energy through machinery, financial markets extract value by modeling and pricing uncertainty over time. Probabilistic variation—captured as volatility, risk, and time value—rather than the expenditure of energy, becomes the operative medium of accumulation.

[52] Andreas Malm, *Fossil Capital: The Rise of Steam Power and the Roots of Global Warming* (Verso Books, 2016).
[53] James Beniger, *The Control Revolution: Technological and Economic Origins of the Information Society* (Cambridge: Harvard University Press, 1986).

Information in time thus functions as a unit of value, linking the governance of uncertainty to the dynamics of capital.

The resonance with Marx's analysis of relative surplus value is not incidental. Industrial machinery reorganized labor by extracting surplus time through technological control; financial and computational systems extract value by organizing uncertainty across temporal intervals. Once learning is defined as iterative adjustment to deviation, variation—mechanical disturbance, market volatility, behavioral data—becomes the key target of governance and control. The practical meaning of learning shifts accordingly, a transformation traced across the chapters that follow.

At a cognitive level, this shift redefines how decisions are framed and enacted. Decisions that might once have been framed in terms of judgment or interpretation—understood as social, ethical, or psychological questions—are reorganized as problems of technical calibration: the continuous adjustment of behavior to maintain performance within quantified constraints. What appears politically as a regime organized around rankings, surveillance, metrics, and performance targets also reshapes the horizon of action at the level of subjectivity, encouraging actors to approach uncertainty as a parameter to be optimized against rather than a situation to interpret. Contemporary AI systems—and the users who train, prompt, and evaluate them—inherit this premise: "learning" is the reduction of error across large datasets, a technical procedure for converting temporal difference into improved prediction. In

this arrangement, the user is drawn into the feedback loop itself, learning to anticipate the system's outputs, adjust prompts and expectations accordingly, and align their own judgment with probabilistic response patterns.

Whether this definition exhausts what learning can mean is the problem this inquiry pursues. The engine itself possessed no interiority; it registered deviation without experience. Yet the regulatory logic first engineered into mechanical control now organizes domains in which the object of calibration is no longer torque or speed but patterns of conduct, judgement, and decision within social and psychological life. The problem of stabilizing energy flows persists into the present as the problem of stabilizing the systems that structure the conditions under which subjects act. For all its generative promise, the technological imagination of the present—now articulated through computational models, predictive analytics, and machine learning systems—remains tethered to an older aspiration: that thought, like steam, might finally submit to governance.

The Origins of Cybernetics

Scholars have examined the history of cybernetics through several overlapping frameworks. Steve Heims reconstructs a "cybernetics group," emphasizing the collaborative network of scientists who collectively grappled with machines, communication, and the human mind.[54] Ronald Kline identifies a distinct

[54] Steve Heims, *The Cybernetics Group*, 11-13.

"cybernetics moment," a pivotal post-World War II conjucture when cybernetics emerged as a mainstream intellectual and cultural force.[55] Paul Edwards and N. Katherine Hayles explore the cybernetic or cyborg "discourse"—an epistemic formation that shaped Cold War command systems and reconfigured cultural understandings of the human subject within what Edwards calls a "closed world" of information and control.[56] Where these historians reconstruct institutional formations and political economies, a parallel body of theoretical work treats cybernetics as a conceptual transformation in modern thought. Donna Haraway's cy(bernetic) org(anism) unsettles distinctions between biology, society, and machine, preserving the tension between control and relationality at the heart of cybernetic reasoning.[57] Hayles traces the cyborg at the interface of thermodynamics and information theory, arguing that the postwar elevation of information as an ontological substrate transformed embodiment and epistemology.[58] Andrew Pickering characterizes cybernetics as a "nonmodern ontology," marked by "an evolutionary, rather than causal and calculable, grasp of temporal process."[59]

Subsequent global and political-economic histories extend this picture by tracing how cybernetic logics traveled across geopolitical and economic domains.

---

[55] Ronald Kline, *The Cybernetics Moment: Or Why We Call Our Age the Information Age*, 1-8.
[56] N. Katherine Hayles, *How We Became Posthuman*, 22-24; Paul N. Edwards, *The Closed World: Computers and the Politics of Discourse in Cold War America, .*
[57] Donna Haraway, *A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late Twentieth Century* (University of Minnesota Press, 2016).
[58] N. Katherine Hayles, *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics* (University of Chicago Press, 1999).
[59] Andrew Pickering, *The Cybernetic Brain: Sketches of Another Future* (Chicago: The University of Chicago Press, 2010), 19.

Eden Medina's study of Project Cybersyn, Salvador Allende's short-lived socialist program in 1970's Chile, complicates the narrative of cybernetics as exclusively capitalist, demonstrating that feedback architectures could be mobilized for socialist planning—though even there adaptation was formalized as informational regulation.[60] Slava Gerovitch and Benjamin Peters document cybernetics' uneven trajectories in the Soviet Union and transnational communications infrastructures.[61] At the level of labor and value, Tessa Morris-Suzuki observed that in a computer-governed economy knowledge acquires price only when monopolized as information—when temporal difference becomes a commodity.[62] Nick Dyer-Witheford analyzes cybernetic automation from an Autonomist Marxist perspective, foregrounding its reorganization of labor and social cooperation, while Matteo Pasquinelli argues that AI appropriates the intelligence of collective labor rather than inventing it anew.[63] Bernard Dionysus Geoghegan demonstrates how the "cybernetic apparatus" reorganized knowledge production itself, recoding meaning, culture, and governance as problems of formal communication and controllable systems.[64]

---

[60] Eden Medina, *Cybernetic Revolutionaries: Technology and Politics in Allende's Chile* (Cambridge: The MIT Press, 2011).
[61] Slava Gerovitch, *From Newspeak to Cyberspeak: A History of Soviet Cybernetics* (Cambridge: The MIT Press, 2004); Benjamin Peters, *How Not to Network a Nation: The Uneasy History of Soviet Internet* (Cambridge: The MIT Press, 2017).
[62] Tessa Morris-Suzuki, "Capitalism in the Computer Age," *New Left Review* I/160, Nov/Dec 1986.
[63] Nick Dyer-Witheford, *Cyber-Marx: Cycles and Circuits of Struggle in High Technology Capitalism* (University of Illinois Press, 1999); Matteo Pasquinelli, *The Eye of the Master: A Social History of Artificial Intelligence* (Verso Books, 2023).
[64] Bernard Dionysus Geoghegan, *Code: From Information Theory to French Theory* (Duke University Press, 2022).

Across these otherwise divergent accounts runs a shared insight that cybernetics reorganized temporality. Whether articulated as ontology, governance, or labor extraction, its core operation was the formalization of deviation as information and the transformation of information into actionable control. The most recent and concentrated expression of this temporal logic appears in Orit Halpern and Robert Mitchell's claim that the Black–Scholes options pricing model is "structurally identical" to the general logic of contemporary "smartness."[65] Their claim that derivatives "leverag[e] noise or waste into a source of value" makes visible what was latent in earlier cybernetic formulations: fluctuation itself becomes productive.[66] In the "new world" of financialization, "chance and noise were not 'devils,'" as Norbert Wiener once warned, "but rather the growth medium for markets." If thermodynamic regulation conserved energy through feedback, derivative finance conserved and exploited uncertainty through probabilistic modeling."[67]

Yet what remains insufficiently examined is the conceptual narrowing that makes this possible: the redefinition of learning as iterative adjustment to measurable deviation, functioning simultaneously as a psychological account of behavior and an economic logic of optimization. Such a redefinition links thermodynamic regulation, theories of mind and behavior, Cold War command systems, mathematical modeling, and contemporary AI across three centuries. The history is neither linear nor

[65] Orit Halpern and Robert Mitchell, "Derivation, Optimization, and Smartness," *The Smartness Mandate* (Cambridge: The MIT Press, 2023), 121-166.
[66] Ibid., 122.
[67] Ibid., 126.

inevitable; it is a constructed genealogy that requires argument, showing how disparate domains came to share a common model of regulation and prediction.

The issue, as the chapters that follow make clear, is how cybernetics redefined the subject as a site of continuous adjustment. Learning ceases to name a transformative encounter with the unexpected; it denotes the incremental correction of error within a bounded system presumed to contain all relevant information. What follows here introduces the mid-century consolidation of this model in cybernetics, then turns to its nineteenth-century formation, where the concept of learning came to bind theories of behavioral adaptation to logics of economic regulation.

*

In the closing years of World War II and the early Cold War, as U.S. military funding transformed universities and national laboratories into coordinated research sites for fire-control systems, digital computation, and communication engineering, a constellation of mathematicians, neurophysiologists, and logicians reframed mechanical regulation as a general theory of adaptive systems. At MIT, Warren McCulloch, a neurophysiologist, and Walter Pitts, a logician, modeled the biological neuron as a binary function analogous to logical sequences, translating neural activity into formal operations and thus laying the groundwork for computational

architectures that treated cognition as information processing.[68] Electrical circuits and neuronal networks were rendered formally commensurable as systems of signal transmission and feedback, and purposive behavior could be described as the stabilization of a system through the circulation of information, relieved of the imprecision attributed to intention or meaning. "Embodiments of mind"—whether biological or engineered—were encapsulated as general information processors, their sensors and effectors, organic and mechanical alike, configured as channels within closed feedback systems.[69] Even society itself appeared as a communication network: a distributed system of interacting agents whose collective behavior is regulated through the circulation of information.

Claude Shannon, a mathematician and electrical engineer at Bell Laboratories, supplied the formal calculus that rendered this cross-domain equivalence mathematically precise. In "A Mathematical Theory of Communication" (1948), he defined information in terms of probabilistic uncertainty.[70] A message carried information to the extent that it was unlikely; surprise—quantified as low probability—became measurable value within a communication channel. Detaching information from semantics and defining it as a function of statistical distribution, Shannon rendered deviation calculable and noise subject to technical management

[68] Warren McCulloch and Walter Pitts, "A logical calculus of the ideas immanent in nervous activity," *The Bulletin of Mathematical Biophysics*, Vol. 15 (1943), 115-133.
[69] For a detailed account of the Macy Conferences that took place from 1946-1953, see: *Cybernetics: The Macy Conferences 1946-1953 Transactions*, ed. Claus Pias (Zurich-Berlin: diaphanes, 2003/2016).
[70] Claude Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal,* Vol. 27, pp. 379–423, 623–656, July, October, 1948.

through coding, redundancy, and error correction. Warren Weaver, who helped popularize and extend Shannon's work beyond engineering, framed this technical model as a universal schema applicable to domains of linguistic, psychological, and social communication.[71] Cumulatively, Shannon and Weaver reconfigured communication as the statistical management of noise, transforming information into a quantifiable commodity defined by probability.

If McCulloch and Pitts rendered neural activity formally computable, and Shannon rendered communication statistically quantifiable, John von Neumann extended these processes to action and decision. Unfolding across Cold War military planning, postwar economics, cognitive psychology, and early computing, von Neumann's work on automata and game theory produced a model agent defined by strategic self-maximization—a cipher of rational calculation.[72] Drawing explicitly on McCulloch and Pitts' neural nets, von Neumann treated neural activity—"one to three excitatory inputs" governed by thresholds—as formally equivalent to digital computation.[73] In developing the stored-program architecture and theorizing self-reproducing automata, he advanced a methodological claim: if a behavior can be fully described in logical terms, it can in principle be constructed. As he put it, "A

[71] Claude Shannon and Warren Weaver, The Mathematical Theory of Communication (Chicago: University of Chicago Press, 1949).
[72] William Aspray, *John von Neumann and the Origins of Modern Computing* (Cambridge, MA: MIT Press, 1990); Philip Mirowski, *Machine Dreams: Economics Becomes a Cyborg Science* (Cambridge: Cambridge University Press, 2002); William Poundstone, *Prisoner's Dilemma: John von Neumann, Game Theory, and the Puzzle of the Bomb* (New York: Anchor, 1993).
[73] Warren S. McCulloch and Walter Pitts, "A Logical Calculus of the Ideas Immanent in Nervous Activity," *Bulletin of Mathematical Biophysics* 5, no. 4 (1943): 115–133.

difficulty of principle embodying any mode of behavior in such a network can exist only if we are also unable to describe that behavior completely."[74] In this move—the consolidation of the stored-program digital computer—description and implementation converged: a formally specified procedure could be executed as machine operation. Decision became programmable, and adaptive behavior was redefined as organized state change across discrete temporal steps—the operative logic of the modern computer and the computational systems that descend from it.

Among these figures, however, Norbert Wiener—a mathematician at MIT and veteran of wartime fire-control research—stood out as a "visionary," in Hayles's words, one "who could articulate the larger implications of the cybernetic paradigm and make clear its cosmic significance."[75] Drawing on the Platonic *kubernetes*—the steersman who governs a ship—and its mechanical descendant, the steam-engine governor, Wiener fused political metaphors of rule with engineering techniques of feedback stabilization.[76] André-Marie Ampère's earlier use of *cybernétique* to denote the science of political government further anchored this etymological lineage.[77] When Wiener defined cybernetics in 1948 as the study of "control and communication in the animal and the machine," he was not merely coining a label but installing a worldview: information—"not matter or energy"—was cast as the proper

[74] John von Neumann, "General and Logical Theory of Automata," in *Collected Works*, vol. 5, ed. A. H. Taub (Oxford: Pergamon Press, 1963), 310.
[75] Hayles, 7.
[76] Orit Halpern, "Dreams for Our Perceptual Present: Temporality, Storage, and Interactivity in Cybernetics," 288.
[77] Slava Gerovitch, "*From Newspeak to Cyberspeak,"* 254. Gerovitch notes how Soviet scientist Aksel Berg, the chair of the Academy Counsel of Cybernetics, linked the philosophy of cybernetics to "the fundamentals of the science of government" established by Lenin.

antidote to the "devils" of noise and contingency, shifting the axis of materialism upon which causality and agency were understood.[78] In this framework, adaptation became the master virtue: systems persisted by adjusting parameters in response to disturbance. Feedback regulation, first engineered to stabilize steam engines, became a general model of order governing organisms, machines, and social systems alike.

All of these efforts unfolded within the wartime and early Cold War reorganization of American science, as federal military funding—channeled through agencies such as the Office of Naval Research and later the Department of Defense—bound universities, corporate laboratories, and weapons research into dense institutional networks.[79] Fire-control systems, servomechanisms, and stored-program digital computers were developed as components of integrated command-and-control infrastructures. Under these conditions, research priorities were explicitly organized around problems of probabilistic prediction—anti-aircraft targeting, radar signal discrimination, code-breaking, and automated air-defense systems.[80] Error and deviation no longer marked the breakdown of a model; they became informational inputs, and intelligence was redefined as a system's capacity to register signals, update its internal state, and optimize output relative to a given objective function. Modes of reasoning that could not be rendered as signal, code, or state transition were

[78] Norbert Wiener, *Cybernetics: Or Control and Communication in the Animal and the Machine* (Cambridge: MIT Press, 1948), 132.
[79] Stuart W. Leslie, *The Cold War and American Science: The Military-Industrial-Academic Complex at MIT and Stanford* (Columbia Uniccersity Press, 1994); Harvey M. Sapolsky, *Science and the Navy: The History of the Office of Naval Research* (Princeton: Princeton University Press, 1990).
[80] David A. Mindell, *Between Human and Machine: Feedback, Control, and Computing before Cybernetics* (Baltimore: Johns Hopkins University Press, 2002); Paul N. Edwards, *The Closed World* (Cambridge, MA: MIT Press, 1996).

peripheral to research regimes organized around feedback performance and operational reliability.

The Josiah Macy Jr. Foundation Conferences on Circular Causal and Feedback Mechanisms in Biological and Social Systems (1946–1953) sought to formalize at the level of a unified theory. Convening neurophysiologists, mathematicians, anthropologists, psychiatrists, engineers, and social scientists—notably Wiener, McCulloch and Pitts, Shannon, and von Neumann—the Macy meetings began by stabilizing "circular causality" as a shared analytic vocabulary. The internal operations of machines, the firing of neurons, and the coordination of social systems were rendered commensurable through teleology and statistical inference, such that "nondeterminate adaptive behavior" appeared across "a host of physically different situations involving the teleonomic regulation of behavior in mechanical, electrical, biological, and even social systems."[81]

Anthropologist Gregory Bateson participated centrally in these discussions, embracing cybernetics as a powerful epistemology while recognizing that its claim to universality rested on a fragile chain of analogies. Its coherence depended on formal symmetry masking substantive difference. For that reason, he described cybernetics as an epistemological paradox: "Cybernetics is the biggest bite out of the fruit of the Tree of Knowledge that mankind has taken in the last 2000 years, but most of such bites out of the apple have proved to be rather indigestible—usually for cybernetic

[81] Warren McCulloch, *Embodiments of Mind* (The MIT Press, 1965/2016), xvi.

reasons."[82] The political stakes were immediate. Cybernetic engineering became, as Bateson warned, "an essential part of designing 'better' gunsights, 'better' ICBM's, and worst of all, the 'better' regulation of men," technologizing a "positively dangerous ideology" of the Second World War.[83] As historians Peter Galison and Paul Edwards have shown, cybernetics consolidated within Cold War command-and-control infrastructures as a technoscience of control, communication, and management. Feedback functioned beyond a descriptive model as a governing rationality, mobilized to identify and counter "enemy others" while legitimating centralized, computerized military command.[84]

Bateson identified a paradox: the same formal logic that enabled "better" gunsights and the more efficient regulation of men—and that he hoped might counter humanity's "presently suicidal attitude to his environment"—expanded technical coordination to unprecedented heights while narrowing understanding to what could be coordinated.[85] As will be examined in later chapters, Bateson's work on learning crystallized the paradox this entailed: the very logics that enhance a system's capacity to correct error can simultaneously foreclose reflection on the premises that organize correction.[86] The operative question thus shifts from meaning or genesis to

---

[82] Gregory Bateson, *Steps to an Ecology of Mind* (New York, Ballentine Books, 1972), 476.
[83] Gregory Bateson papers. MS 98. Special Collections and Archives, University Library, University of California, Santa Cruz. Letter regarding Anthony Wilden from March 16, 1970.
[84] Peter Galison, "The Ontology of the Enemy: Norbert Wiener and the Cybernetic Vision," *Critical Inquiry*, Autumn, 1994, Vol. 21, No. 1 (Autumn, 1994), pp. 228-266; Paul N. Edwards, *The Closed World: Computers and the Politics of Discourse in Cold War America* (Cambridge: The MIT Press, 1997).
[85] Gregory Bateson papers. MS 98. Special Collections and Archives, University Library, University of California, Santa Cruz. Letter regarding Anthony Wilden from March 16, 1970.
[86] Gregory Bateson, "The Logical Categories of Learning and Communication," *Steps to an Ecology of Mind* (New York: Chandler Publishing Company, 1972), 279-308.

performance—how effectively a system absorbs disturbance and preserves function, even when that adaptive efficiency obscures what is being transformed or lost.

Demons Against Devils

If postwar cybernetics elevated feedback into a universal principle of learning, it did so by turning backward. For Wiener, the decisive precursor was the nineteenth-century physicist James Clerk Maxwell, whose analysis of stability in "the steering engines of a ship" he described as "one of the earliest and best developed forms of feedback mechanisms," foundational to a general theory of circular causality, information, and automation.[87] In his 1868 paper "On Governors," Maxwell observed how the speed regulation mechanism in a steam engine and governor used feedback control to maintain stability by continuously measuring deviation and converting it into corrective action.[88]A thermodynamic governor could automate its future actions by processing, transmitting, and storing a higher-order time series controlling the operations of the total machine. The governing mechanism, in other words, produced *information* without impacting the material basis involved in the communication (the energy expended) through a series of selections among messages compounding into a pattern of decisions.

[87] Wiener, 11-12.
[88] James Clerk Maxwell, "On Governors," *Proceedings of the Royal Society of London*, Volume 16, 1868, 271-272.

According to Otto Mayr, 1868 was an unusually productive year for Maxwell, then living in relative seclusion at Glenlair while drafting portions of his *Treatise on Electricity and Magnetism*.[89] He surfaced periodically to present before the London Mathematical Society, where two of his papers—"The Construction of Stereograms of Surfaces" and "On Reciprocal Diagrams in Space, and their Relation to Airy's Function of Stress"—addressed what he called a "very general and very important problem in Dynamics."[90] That problem had been posed earlier by the Astronomer Royal George Biddell Airy: how to account for energetic instability in systems whose motion refused to settle into equilibrium.[91]

Maxwell soon gave the problem sharper industrial edges. As Mayr notes in *The Origins of Feedback Control*, James Watt's centrifugal governor—introduced in 1788—had become "the first feedback device to attract the attention of the whole engineering community and to be internationally accepted."[92] Yet its theoretical basis remained thin. When Maxwell asked the Society to assist him with "the mathematical problem of instability," he was acknowledging that the governor's practical success masked conceptual confusion. Steam engines across Europe were plagued by irregular speed; inaccurate control translated directly into lost efficiency. Appealing

[89] Otto Mayr, "Maxwell and the Origins of Cybernetics," *Isis*, Winter 1971, Vol. 62, No. 4, 426.
[90] Basil Mahon, *The Man Who Changed Everything: The Life of James Clerk Maxwell* (Hoboken, NJ: Wiley Publishing Company, 2003), 128-131.
[91] Mayr (1971), 428. In the paper "On the Regulator of the Clock-work for effecting uniform Movement of Equatorials," Airy had proposed that the weight drives of equatorial telescopes were regulated by unstable governors. Guided by his astronomical background, Airy assumed that the mechanical instability of the pendulum in the telescope was analogous to a planetary orbit, eclipsed by oscillations in gravity moving in smooth periodic functions of continuous sine waves.
[92] Otto Mayr, *The Origins of Feedback Control* (Cambridge: The MIT Press, 1969), 1-2.

to the industrialists present, Maxwell argued that "studying the motion of certain governors for regulating machinery" was not an abstract diversion but a matter of economic consequence.[93] Within weeks, he returned with a third paper, "On Governors," converting the engineering problem into an mathematical object.

Maxwell's first move was to approach Airy's problem of instability through a seemingly distant case: the hypothetical dynamics of Saturn's rings.[94] His solution hinged on a set of precise assumptions: to produce stability in a system, one had to presume that the spread of parameter errors was small, that the equations of motion could be treated as linear, and that disturbances would converge with the corrective response to yield a "negative state." Airy had failed to resolve governor instability because the motion in question was governed by nonlinear differential equations, resistant to solution in continuous form.[95] Maxwell's proposal was methodological: if the variations generated by motion could be reduced to linear differential equations, their stability conditions could be determined analytically. In doing so, he showed that the stability of a dynamic system could be rigorously analyzed through the linearization of its equations of motion.

Maxwell thus reduced stability analysis to algebra, locating the conditions of equilibrium in the roots of a characteristic polynomial.[96] He thereby formalized a

---

[93] Mayr (1971), 428.
[94] James Clerk Maxwell, "On the stability of the motion of Saturn's rings," *Monthly Notices of the Royal Astronomical Society*, Volume 19, Issue 8, June 1859, 302-303.
[95] Ibid., 299.
[96] Chul-goo Kang, "Origins of Stability Analysis: 'On Governors' by J. C. Maxwell," *IEEE Control Systems Magazine*, October 2016, 86-87.

logic of corrective conservation: the output could be expressed as a linear function of the input, such that deviation generated its own counteracting response. He thereby formalized a logic of corrective conservation: the output could be expressed as a linear function of the input, such that deviation generated its own counteracting response."[97] By specifying a governor as a system that keeps the velocity of a machine uniform, "notwithstanding variations in the driving-power of the resistance," The engineer's task, accordingly, was not to eliminate disturbance but to design a mechanism that controlled the degrees of variation in motion.[98]

A governor, in this formulation, generates a corrective action that intensifies as long as error persists, producing an incremental surplus of velocity until the deviation is canceled. As Maxwell put it, the device "sets in motion a contrivance which continually increases the resistance as long as the velocity is above its normal value, and reverses its action when the velocity is below that value."[99] The mechanism thus works not by overpowering the engine but by modulating its own response: it restores the "same normal value" through an "additional mechanism" that scales with the differential error itself—*all without expending any real motive power of the total machine.*

Yet the governors Maxwell analyzed were largely experimental prototypes built by Fellows of the Royal Society—devices remote from the workshop floor and

---

[97] James Clerk Maxwell, "On Governors," *Proceedings of the Royal Society of London*, Volume 16, 1868, 271-272.
[98] Ibid., 270.
[99] Ibid., 271.

inaccessible to most mechanics of the day. His equations, elegant as they were, proved difficult to translate into the practical design of commercial steam engines; "On Governors" disappointed patrons who expected immediate technical payoff.[100] Freed from the immediate demands of industrial application, however, governor theory migrated into abstraction, envisioning stability as a problem of differential signals and selective correction rather than sheer expenditure of force. It was in this more speculative register, less as a blueprint than a thought experiment, that Maxwell's logic of control would reappear in the figure of the demon.

In her incisive account of Maxwell's 1871 "demon" thought experiment—devised three years after "On Governors"—Astrid Schrader observes that Maxwell inserted statistical inference into thermodynamics by inventing "a system and a boundary… patrolled by the limits of human knowledge."[101] Maxwell's demon, perhaps the most famous cybernetic governor in the history of science, stands guard over "a vessel divided into two portions, A and B, by a division in which there is a small hole," opening and closing the aperture to "allow only the swifter molecules to pass from A to B, and only the slower ones to pass from B to A."[102] By sorting fast-moving particles to one side and slow-moving particles to the other, the demon "raise[s] the temperature of B and lower that of A," creating a "temperature gradient… without expenditure of work," simply by producing information about

---

[100] A. T. Fuller, "Maxwell's treatment of governors fitted with differential gears," *International Journal of Control*, 65:3, 404.
[101] Astrid Schrader, "Haunted Measurements: Demonic Work and Time in Experimentation," *Differences: A Journal of Feminist Cultural Studies* Volume 23, Number 3, 127.
[102] Ibid., 129. Maxwell quoted by Schrader.

their velocities. In Schrader's reading, Maxwell thereby transformed the devil of entropy from an inviolable law into a probability statement: the second law became a measure of knowledge or ignorance, a question of how intelligence might "produc[e] order out of disorder, reversing the arrow of time by pure intelligence."[103]

Viewed as an idealized governor, the demon dramatized a principle that would later animate information engineering: a differential—here, a temperature gradient—could function "in contradiction to the second law of thermodynamics" as a governing mechanism without proportional energy expenditure.[104] Thomas Kuhn, in "Energy Conservation as an Example of Simultaneous Discovery" (1959), noted that the engineering concept of a governor proved "readily transferable to the more abstract problems of energy conservation" because the work performed by the machine need not equal the total energy conserved; "engines could and occasionally did look like *conversion devices*."[105] Stability, a "general uniform motion," was achieved by converting total motive energy into a differential between actual "real" motion and the disturbances generated by fluctuating amplitudes. What mattered was not brute force but the selective amplification or dampening of variation—what Maxwell would later describe as "integral control actions" supplementing proportional response.

---

[103] Ibid., 128.

[104] Ibid., 131. Maxwell's own description of the demon's activity in 1871 quoted by Schrader in Harvey S. Leff and Andrew F. Rex, *Maxwell's Demon: Entropy, Information, Computing* (Bristol, UK: Adam Hilger, 1990.)

[105] Thomas Kuhn, "Energy Conservation as an Example of Simultaneous Discovery," *The Essential Tension: Selected Studies in Scientific Tradition and Change* (The University of Chicago Press, 1977), 90-91.

The demon was a theoretical distillation of a mechanism already operative in steam power. As the chapter's opening scene shows, long before Maxwell imagined an intelligence sorting molecules, James Watt redesigned the engine so that deviation itself became the medium of regulation. With the centrifugal governor, a machine that had once provided reciprocating motion became a self-regulating system capable of continuous rotative motion. Otto Mayr's criteria for feedback control systems retroactively clarify what Watt had achieved: a closed-loop arrangement "possessing monitoring feedback, the deviation signal formed as a result of this feedback being used to control the action of a final control element."[106] Feedback thus offered a prototypical definition of information: the operational use of difference—between command and controlled variable—as the endogenous means of regulation within a closed loop.

This logic of differential control emerged from a practical struggle over fuel efficiency and mechanical loss. Watt's separate condenser reduced waste by maintaining a hot cylinder while condensing steam in a separate chamber, transforming heat management into a problem of controlled gradients. Regulation no longer depended on brute force but on maintaining differential relations within the system. In Wiener's twentieth-century retelling, Watt's governor—"two balls attached to pendulum rods and swinging on opposite sides of a rotating shaft," rising through "centrifugal action"—became exemplary because it instantiated a principle

[106] Mayr (1969), 7.

rather than a mere device.[107] The engine's "piston-in-cylinder arrangement," opening its valves "when the engine slows down and the balls fall" and closing them "when the engine speeds up and the balls rise," embodied what Wiener called negative feedback: "The feedback tends to oppose what the system is already doing, and is thus negative… The feedback of voluntary activity is of this [negative] nature."[108] Regulation proceeded not by adding brute force but by registering deviation—by measuring "the amount by which it has not yet been accomplished."[109] Ultimately, the steam engine demonstrated that stability did not depend on expending more force, but on producing a differential between force and response and routing that difference back into the system. Maxwell's theory of governors thus anticipated the general cybernetic insight that energetic transformation can occur through *the manipulation of information in time*. What stabilized the engine's motion was not additional energy altering its velocity, but the conversion of energy into a time-structured gradient of information—difference measured, transmitted, and fed back to the control center of the total apparatus.

Automation and the Value of Labor

Viewed from the eyes of the early cyberneticists, the dynamics of stability and motion internal to the nineteenth-century "self-governing" steam engine retroactively modeled what would become an order of coordinated production and exchange governed through feedback and statistical control. In the governor, the processing,

---

[107] Wiener, 97.
[108] F. M. Scherer, "Invention and Innovation in the Watt-Boulton Steam-Engine Venture," *Technology and Culture*, Vol. 6, No. 2 (Spring, 1965), 166; Wiener, 97.
[109] Ibid. My emphasis.

transmission, and storage of signals reorganized the physical system without increasing its total expenditure of energy. Regulation operated by extracting difference—between actual and desired velocity—and converting that deviation into corrective action. What was produced was not new energy but a temporal differential: a capacity to anticipate, modulate, and redirect motion in advance of breakdown. In this sense, the self-corrective mechanism embodied an endogenous production of value. It generated time derivatives—organized adjustments across intervals—without proportional work. The materially consequential premise was that information could function at a higher order than matter and energy, detachable from its bioenergetic substrate and rendered productive in its own right.

The "very general and very important problem in Dynamics," which linked Saturn's rings to the motive power of the steam engine, was not only a question of motion but of temporal leverage: how to anticipate and neutralize future disturbances without expending energy proportional to their effects. In Maxwell's governor and demon alike, regulation hinged on extracting usable order from fluctuation. The differential between energetic expenditure and informational state—between work performed and variation registered—appeared to generate a surplus in time. As Schrader notes, Maxwell's demon introduced "a new measurement process…[for] a perfect perpetual motion machine."[110] If the motion of molecules "escape[s] all our methods of observation," then the work of the system shifts to the processing of information—probabilistic discrimination, selective passage, calibrated

[110] Schrader, 133.

response. What is conserved is not simply energy but the capacity to convert uncertainty into actionable difference. The condenser in Watt's engine and the demon's gate perform the same conceptual feat: they automate value by redistributing differentials without commensurate expenditure of force.

Hayles has described cybernetics as underwriting a "technological determinism of bodiless information," in which value circulates as if detached from laboring bodies.[111] "Automation" names this shift: the accumulation of "dead" work that appears self-propelling, severed from the living activity that sustains it.[112] Automation historically captures and formalizes the intelligence of servants, proletarians, and women under governing power.[113] The fantasy is ancient—Hephaestus' golden maidens, "living young women… [with] intelligence in their hearts and speech in them… stirring nimbly in support of their master"—but it acquired new material force in the nineteenth century.[114] As Maxine Berg observed in *The Machinery Question*, it was only then that "machinery defined the organization of work and held the balance of power in the determination of the distribution of returns from labor."[115]

The self-acting machine reorganized production while consolidating control, becoming a condensed figure of industrial hierarchy. Scientists, socialists,

---

[111] Hayles, 22.
[112] Ramin Ramtin, *Capitalism and Automation: Revolution in Technology and Capitalist Breakdown* (London and Massachusetts: Pluto Press, 1991), 101-105.
[113] Pasquinelli, 238.
[114] Homer, *The Iliad*, trans. Richmond Lattimore (Chicago: University of Chicago Press, 2011), 416.
[115] Maxine Berg, *The Machinery Question and the Making of Political Economy* (Cambridge: Cambridge University Press, 1980), 101.

industrialists, and workers alike saw in it the reorganization of labor, the erosion of skill, and the intensification of dependency relations. Marx, of course, understood this transformation as a shift in the locus of value. What Pasquinelli reinterprets as the "general intellect"—the cooperative intelligence sedimented in machines—appears as capital's appropriation of collective knowledge.[116] Yet, as Ramin Ramtin emphasizes, this process generated "an inescapable paradox": the compulsion to "automate or die."[117] Capital must continually replace living labor with machinery to extract relative surplus value, even as surplus value ultimately depends on living labor. The same dynamic that formalized learning as error-correction within machines thus intensified the contradiction between automation and the social conditions of value production. In this sense, the nineteenth-century steam engine already staged the political-economic stakes of cybernetic learning: deviation becomes productive, information appears autonomous, and the regulation of fluctuation displaces labor as the visible source of value—while remaining dependent upon it.

The knowledge that capital preserves itself only by developing more productive and efficient modes of technological automation—meaning that it must perpetually exhaust inputs of the living labor that remains the sole source of surplus value—was not immediately apparent to the English bourgeoisie. Alfred Russell Wallace could still write to Charles Darwin that the "action of principles constantly at work in nature is exactly like that of the centrifugal governor of the steam engine,"

[116] Pasquinelli, 77-94.
[117] Ramtin, 105.

projecting the logic of mechanical self-regulation onto biological evolution.[118] What appeared as a neutral analogy between nature and machinery in fact naturalized the displacement of labor by automatic control.

Marx, by contrast, treated the governor with far less innocence. In Chapter 15 of *Capital*, he observed that James Watt's steam governor "did not give rise to any industrial revolution."[119] Rather, it was the broader system of control machinery—engines "propelled by its inner forces"—that reorganized production and necessitated the scientific refinement of steam power itself. The historical development of self-acting machinery, intensifying the control and exploitation of living labor, required a new source of motive power adequate to its expansion. Power did not generate control; control demanded new power.

From this vantage point, Samuel Butler's anxieties in "Darwin Among the Machines" (1863) read less as novelistic fantasy than as a sociological diagnosis. Butler warned against human subservience to the "self-regulating, self-acting power" of governing machines, imagining that the "development of mechanical life" would usurp natural heredity such that "man's very soul is due to the machines... a machine-made thing." [120] Machines, he presciently feared, would accelerate adaptation beyond biological time, out-speeding the evolutionary processes described in the *Origin of*

[118] Alfred Russel Wallace, *Contributions to the Theory of Natural Selection: A Series of Essays* (White Press, 2016 [1870]), 62.
[119] Karl Marx, "Machinery and Large-Scale Industry," *Capital Volume I: A Critique of Political Economy*, trans. Ben Fowkes (London: Penguin Books, 1976), 496-497.
[120] Samuel Butler, "Darwin Among the Machines," to the Editors of the Press, Christchurch, New Zealand, 13 June 1863.

Species. [121] As George Caffentzis notes, Butler's alarm echoed Marx's "machine=capital metaphor": machinery cheapened labor power, reduced the commodity's value, and rendered the worker's time—rather than the worker's skill—the decisive source of relative surplus value.[122] As automation intensified, machines "gradually lost a clear class referent and even began to be identified with an alien force that was threatening to the marginal bourgeoisie itself."[123] The machine appeared as an autonomous regulator of social life.

The logic of this transformation had also been articulated by Charles Babbage, who shrewdly recognized that "the division of labor was not only a principle to design machines but also to compute the costs of production."[124] The machine internalized and quantified labor's differentiation. Marx extended this insight, describing automation and capital as entangled in an evolutionary struggle: the "variation of labor" imposed upon workers by machinery "imposes itself after the manner of an overpowering natural law."[125] Under large-scale industry, the "fitness of the worker for the maximum number of different kinds of labor" became "a question of life and death."[126] The worker's adaptive capacity was subsumed into the machine's logic. Capital "necessitates variation of labor, fluidity of functions, and

[121] Samuel Butler, *Erewhon, or Over the Range* (New York: The Modern Library, Inc., 1927).
[122] George Caffentzis, "Why Machines Cannot Create Value: Marx's Theory of Machines," *Letters of Blood and Fire: Work, Machines, and the Crisis of Capitalism* (PM Press, 2013), 154.
[123] Ibid.
[124] Pasquinelli, 4.
[125] Marx, 618.
[126] Ibid.

mobility of workers in all directions," even as it "reproduces the old division of labor with its ossified particularities." [127]

The same historical moment that saw Marx theorizing machinery and surplus value also witnessed a parallel consolidation in the physical sciences.

When *Capital* appeared, German physiologist Hermann von Helmholtz was formalizing the conservation of energy through his analysis of automated motive power. The physicist described the 'work' of machines and natural processes as analogous to "the working power of man," where the phenomenon of "human labor" outlines "the most important features of the question with which we are concerned… that the quantity of force which can be brought into action in the whole of Nature is unchangeable, and can neither be increased nor diminished."[128] Helmholtz observed that although there is work, an "expenditure of force," in the muscle metabolism of a laboring man (i.e., the moving force of an arm lifting a weight), there is no loss of energy.[129] There is "the feeling of exertion and fatigue in our muscles [from which] we can form a general idea of what we understand by amount of work."[130] Still, there were no "vital forces" necessary to move a muscle. Likewise, in a steam engine, heat as "a movement of internal particles" produces work. However, when heat is converted into work, an equivalent amount of heat is destroyed. After examining the relation between gravity, weight, and movement in a series of experiments,

---

[127] Ibid., 617.

[128] Hermann von Helmholtz, "On the conservation of force; a physical memoir," *Selected Writings of Hermann von Helmholtz*, ed. Russell Kahl (Wesleyan University Press, 1971), 7.

[129] Ibid., 15.

[130] Ibid.

Helmholtz concluded that all "vital forces" in the natural 'forces' of human muscle and in the artificial design of a machine reduce to attractions and repulsions of energy expended in proportion to the work performed. He thus formulated that *the total quantity of all forces capable of working in the universe remains constant*. In a closed system, heat (energy) cannot be destroyed but only transformed into another form, such as mechanical force.

This thermodynamic axiom carried unmistakable social implications. If no new force can be created, industrial production cannot generate energy ex nihilo; it can only reorganize and redirect what already exists. Automation, in this light, does not abolish labor but reorganizes it, extracting labor power from the worker as a supply of work even as the worker recedes from view. The disciplining of the factory floor, the calibration of bodily motion, and the regulation of time follow from this constraint. Workers' food and fuel consumption entered directly into the motive power of production: metabolic heat became mechanical force. The machine does not transcend bioenergetic limits; it internalizes them, while the total quantity of energy remains constant. Technological innovation thus depends not on escaping material constraint but on conserving and reallocating it. The apparent autonomy of automation rests on the continued conversion of living energy into mechanical and informational form. What appears as self-acting intelligence is, at base, a redistribution of embodied expenditure, rendered abstract and made to circulate as value.

Helmholtz was not alone in grounding universal law in the mechanics of the steam engine. Decades earlier, in 1824, Sadi Carnot—the son of Napoleon's Minister of War—published his analysis of the essential features of the engine in order to "establish principles not only applicable to steam engines but to all imaginable heat engines."[131] Carnot's scientific investigation was the handmaiden of geopolitical ambition: the energetic laws of steam defined the limits of productivity underpinning economic competition and colonial expansion. Maximally efficient mechanical power promised to "permit the establishment of prompt and regal communications across the arms of the sea" and enable France "to traverse savage regions where before we could scarcely penetrate… to carry the fruits of civilization" to territories otherwise beyond reach.[132] Carnot deduced a general cycle of operations from the model of a steam-filled cylinder, measuring motive power through the fall and rise of heat.[133] Conservation meant that work generated equaled heat transferred; yet this principle imposed a limit.[134] Because heat cannot spontaneously move from cold to hot, output could be maximized only by allowing heat to "fall" through a greater temperature differential.[135] Carnot thus formalized what would become *entropy*: the irreversibility that structures all energetic transformation. Efficiency, on this account, depended not

[131] Sadi Carnot, *Reflections on the Motive Power of Fire, and on the Machines Fitted to Develop that Power*, trans. R. H. Thurston (Mineola, NY: Dover Publications, 1988), 44.
[132] Ibid, 41-42.
[133] Ibid.
[134] von Helmholtz, 15.
[135] See Charles Coulston Gillespie, *The Edge of Objectivity: An Essay In the History of Scientific Ideas* (Princeton University Press, 1966), 299. "Entropy" as a term emerged in 1865 with German physicist Rudolph Clausius describing the transformational content of energy as a "differential of a quantity which depends on the configuration of the system."

on creating force ex nihilo but on arranging gradients so that their dissipation could be harnessed.

Seen in this light, the steam engine did more than power industrial machinery; it established a general logic of productive power. Work derived from the controlled management of differentials—thermal, mechanical, temporal, and eventually social. The same principle governed the factory floor, colonial infrastructures, and the organization of labor: gradients of energy, time, and human effort were arranged so that their dissipation could be captured as output. Automation enabled the reorganization of material limits, converting embodied expenditure into regulated flows of mechanical force and, later, informational control. The apparent autonomy of self-acting systems thus rested on the conservation and redistribution of living energy, rendered abstract and made to circulate as value. Modern productive power would hinge on managing difference rather than generating energy anew.

Pricing Uncertainty, Governing Expectations

The energetic intensification of industrial automation formed the background against which Ludwig Boltzmann, in 1871, generalized entropy as a statistical measure of an irreversible trajectory toward disorder in the internal state of a physical system.[136] Building on Maxwell and Carnot, Boltzmann realized that probability grows because entropy grows: irreversible thermodynamic change is movement toward states of increasing probability, and maximum disorder corresponds to the

[136] von Helmholtz., 240-241.

macroscopic state of maximum likelihood.[137] By the end of the nineteenth century, entropy was no longer confined to the physics of heat; it had become a calculus of distributions, a way of describing how systems evolve through probabilistic selection over time.

The force of Boltzmann's intervention lay not simply in linking microscopic motion to macroscopic order, but in asserting that contingency itself could be formalized. Entropy as a physical constraint provided a statistical description of selection across possible states. Boltzmann believed he had accomplished in physics what Darwin achieved in biology: a population-level account of selection generalizable as a probabilistic function. When systems were treated as ensembles rather than individual trajectories, physical law measured the relative frequency of configurations over time.[138] As Isabelle Stengers and Ilya Prigogine later observed, Boltzmann's formulation revealed the "intrinsic irreversibility" of time: the arrow of time runs from less probable to more probable states, from constrained distributions toward expansive ones.[139] In the biological world, learning accumulates along an irreversible trajectory, bounded by what they describe as an "infinite entropy barrier" separating possible from prohibited states.[140]

---

[137] Engraved on von Helmholtz's gravestone in central Vienna is this rule in notation, stating probability in quantitative form: $S = k \log W$, relating macroscopic entropy (S) to microstates of thermodynamic probabilities (W) logarithmically connected to the proportionality factor or Boltzmann's constant (k).
[138] See: Lloyd A. Demetrius, "Boltzmann, Darwin, and Directionality Theory," *Physics Report*, Vol. 530, Issue 1, 2013, 1-85.
[139] Ilya Prigogine and Isabelle Stengers, *Order Out of Chaos: Man's New Dialogue with Nature* (Verso, 2017), 124.
[140] Ibid., 119.

Boltzmann's decisive move was to define entropy as the statistical tendency of systems to migrate toward more probable configurations. Irreversibility emerges statistically: microscopic laws are time-symmetric, but equilibrium prevails because overwhelmingly more microstates correspond to disordered configurations than to ordered ones.[141] Systems therefore evolve toward macrostates of higher multiplicity because such states occupy overwhelmingly larger regions of phase space. Near equilibrium, information about the precise microstate becomes inaccessible: probability accumulates as memory dissipates.

Once probability becomes the grammar of physical change, deviation from probability becomes measurable as *information*. Uncertainty ceases to be mere ignorance; it becomes a formally defined property of distributions. Irreversibility is no longer only a feature of physical systems but an object of calculation, modeling, and technical intervention.

Modern finance developed techniques that furnished a machinery for this precise logic. In the early 1970s, Fischer Black and Myron Scholes formalized a concrete apparatus for converting stochastic variation into structured risk in their model for pricing options. Because information could be formalized as structured difference within a probabilistic system, Black argued that financial assets could likewise be decomposed into analytically separable functions. As he put it, "the function of a common stock in transferring risk can be separated from its function as

[141] See: Ilya Prigogine, *From Being to Becoming: Time and Complexity in the Physical Sciences* (San Francisco: W. H. Freeman, 1980).

a store of wealth."[142] An option's value could be derived from observable variables—current price, time to maturity, interest rates, estimated volatility—without forecasting the asset's future path.[143]

Under the model's assumptions—most notably that prices follow geometric Brownian motion—an option's value depends on "the price of the stock," the time to expiration, and the "short-term interest rate."[144] A continuously a hedged position "consisting of a long position in the stock and a short position in the option" will depend "only on time and the values of known constants."[145] The governing epistemology is decisive: the future appears as a stochastic field subject to continuous recalibration, where uncertainty is structured and priced, and learning takes the form of iteratively adjusting exposure as new price information arrives. Liquidity and profitability emerge from the continual measurement and modulation of risk across time.

The Black–Scholes formula synthesized an option in an ideal world of geometric Brownian motion, infinite liquidity, and expected returns in order to manufacture options for an actual market—in which stock prices do not follow geometric Brownian motion, liquidity is finite, and volatility neither constrant nor smooth. The gap between model and market—the map and the territory—did not

[142] Fischer Black, "The Fundamentals of Liquidity," June, 1970, 6. Unpublished Works, Indiana University, Department of Business Economics and Public Policy in the Kelley School of Business, BU 456.
[143] Fischer Black and Myron Scholes, "The Pricing of Options and Corporate Liabilities," *The Journal of Political Economy* (1973), 81(3), 638.
[144] Ibid., 640-641.
[145] Ibid., 641. My emphasis.

invalidate the apparatus; it animated it. Traders could leverage the "time-value" of options without exercising them or assuming risk-laden directional positions because volatility itself became an informational asset.[146] In this way, derivatives pricing functioned as a governor of volatility, crystallizing the mid–twentieth-century revaluation of deviation as a resource to be managed over time. It sorted positions and allocated exposure, differentiating those who absorb risk from those who hedge and trade it; the ability to hedge under incomplete knowledge conferred structural advantage by converting uncertainty into leverage and purchasing power.

Ultimately, financial actors deployed options as instruments for prognosticating value in time—devices that anticipate future disturbance by accumulating and processing information unburdened by the expenditure of energy that went into producing it. Value could thus be generated by transforming stochastic volatility into structured exposure—portfolios calibrated to expand liquidity—without any corresponding production of goods. What emerges is a cybernetic image of adaptive resilience, in which intelligence accrues through the iterative optimization of performance within a probabilistic field: the market "learns" through price revision, the trader through the continual rebalancing of risk exposure, and governance takes the form of regulating turbulence and volatility in response to these feedback processes.

In this respect, Black–Scholes–era financial innovation parallels Claude Shannon's 1948 claim that entropy measures the information potential of a channel:

[146] Emanuel Derman and Michael B. Miller, *The Volatility Smile* (Wiley; 1st edition, 2016), 3.

the more a message surprises, the more it can say. [147] An efficient market presupposes what information theory formalized: uncertainty is not a flaw in the system but the very condition of its value. According to Black, "an orderly market, in economic theory, is a market in which price changes occur randomly."[148] Random walk and continuous liquidity signal efficiency; stability is unprofitable because predictability extinguishes informational value. Boltzmann had shown that physical systems drift toward the most probable macrostate as they "forget" their microstates—entropy increases as randomness proliferates. Black's move was to transpose that logic into finance: if surprise carries information, then randomness can be priced. Volatility—irreversibility registered in time—becomes the medium of value, monetizing variance and time decay. As Orit Halpern and David Mitchell observe, the Black–Scholes model "joined communications and information theories with calculation in a way that made their equation amenable to algorithmic enactment."[149]

Shannon opened "A Mathematical Theory of Communication" by insisting that a theory of information is valuable insofar as it quantifies methods of combating noise in a transmission channel. From an engineering standpoint, he bracketed "meaning," declaring the "semantic aspects" and "psychological factors" of a message irrelevant to the problem at hand.[150] What mattered was that any message—a text, a voice, an image—could be treated as a probabilistic selection from a defined

[147] Claude Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal,* Vol. 27, pp. 379–423, 623–656, July, October, 1948.
[148] Black (1971), 34.
[149] Halpern and Mitchell, 129.
[150] Ibid., 379.

set of possible signals and thus rendered calculable. In its simplest form, this reduces to a binary alternative: a coin toss with a 50/50 probability conveys the same quantity of information whether a person says yes or no or a machine transmits 0 or 1. Information arises only where alternatives exist and a rule governs selection among them. Communication, then, resolves uncertainty by structuring it: the greater the prior uncertainty, the greater the informational yield required to reduce it. Information was thus formalized as negative entropy, "the logarithm of the reciprocal of the probability of the state."[151] The less likely an event is, the more information it provides when it occurs.[152]

Communication, however, rarely approximates a coin toss. If an input always produces the same received signal, transmission collapses into redundancy; distortion—"noise," in the engineer's jargon—enters precisely where signals fail to map cleanly onto one another. Conversely, when "no two transmitted signals produc[e] the same received signal," distortion can be corrected by performing the appropriate inverse operation on the received signal, restoring the intended selection and generating information.[153] By assigning probabilities to possible messages, Shannon defined communication as structured selection, such that informational value corresponds to the distribution of signals: the greater the dispersion of

[151] McCulloch, (1965), 73.
[152] Like Boltzmann, engraved on Shannon's tombstone in Cambridge, Massachusetts is the negative entropy formula: $H = - \sum p_i \log p_i$, where the sum of all probabilities combined is proportionate to the total quantity of information in the channel.
[153] Ibid., 399.

probabilities, the greater the potential informational yield. What is improbable is surprising, and therefore informative.

In this respect, Shannon's entropy inverts Boltzmann's. Boltzmann had shown that entropy arises from the distribution of particles in phase space and that time runs "going from the less probable state to the more probable one (the former will be the 'past' and the latter the 'future')," so that any localized system appears "initially always in an improbable state." [154] Entropy increases as systems drift toward more probable macrostates, effectively forgetting their microstates. Shannon, by contrast, treated uncertainty not as irreversible loss but as measurable *potential*: maximal entropy in a channel signifies maximal possible information. Where thermodynamics marked the movement from improbability to probability as temporal decay, information theory quantified how "combating noise" could generate what Wiener called "islands of negative entropy"—localized productions of order extracted from statistical disorder.[155]

Seen in these terms, modern finance does not depart from the logic of the governor; it extends informational regulation into probabilistic time. Where the centrifugal governor corrected mechanical oscillation, options pricing regulates probabilistic fluctuation. Deviation is measured against a modeled baseline and converted into speculative action. The stakes are not only economic: when

[154] Boltzmann quoted in Stengers and Prigogine, 254.
[155] "A Mathematical Theory of Communication" formed a pioneering theory of information directly analogous to physics, essential to the formalization of first-order cybernetics. Following Shannon's paper, Wiener, along with John von Neumann, Erwin Schrödinger, and later on Ludwig von Bertalanffy, all expressed similar views during this time period on the connection between physical entropy and abstract "bits" of information.

uncertainty becomes a parameter to be managed rather than a condition to be endured, expectation itself is reorganized at the level of subjective decision-making.

The crash of 1987, for example, revealed that volatility was not simply misestimated by assuming constant drift; as Emanuel Derman later recognized, it was effectively "generated by the pricing formula itself."[156] Implied volatility emerged as a surface—the "volatility smile"—varying with strike and expiration, revealing that the model did not collapse under stress but instead became reflexively resilient. Incorporating the "volatility of volatility," options pricing extended beyond assets to *anticipated* changes in variance, embedding expectations about uncertainty directly into valuation.[157] As Robert Meister observes, markets thereby learned to price "the volatility of their own expectations of volatility," so that present trades register "the doubt about their models."[158] Decision-making centers on calibrating one's position within shifting distributions of risk, as actors adjust exposure in response to changing signals rather than adjudicating the truth of competing valuations. In Meister's paraphrase of Derek Parfit and Randy Martin, "the choice itself and not the evidentiary basis for it, will be what really matters: selection is an act of valuation."[159]

---

[156] Derman and Miller, 3. The Dow Jones Industrial Average (DJIA), once more than tripled in the past five years, fell by 508 points or by 22.6%. This was the largest percentage drop in one day up to this point in history, and sparked fears of extended economic instability around the world.
[157] Ibid., 12. Since 1987, the volatility smile has spread to most, if not all option markets: single-stock options markets, interest-rate options markets, currency options markets, credit derivative markets. The volatility smile model is "custom-tailored for clients and generate[s] higher margins," and therefore not easily resold.
[158] Robert Meister, *Justice Is An Option: A Democratic Theory of Finance for the Twenty-First Century* (University of Chicago Press, 2021), 104-106.
[159] Ibid., 101-106.

This reformulation of uncertainty as a quantifiable feature of communication systems thus resonated beyond Shannonian communication engineering. In his 1986 essay "Noise," written thirteen years after the formulation of the Black–Scholes options pricing model, Fischer Black argued that noise is not a deviation from informational efficiency but a structural condition of market life.[160] A model of information trading without noise trading cannot sustain exchange, since traders aware that others possess different information will hesitate to trade; as he put it, "differences in beliefs must derive ultimately from differences in information." What sustains trading is noise trading—"trading on noise as if it were information," in which participants are willing to trade even though, "from an objective point of view they would be better off not trading," perhaps because they believe the noise they follow is information—"or perhaps they just like to trade." Prices therefore reflect both information and noise. Because noise cannot be eliminated, variability is not a distortion of the market signal but a condition of its emergence. As Black succinctly summarized this paradox, "noise creates the opportunity to trade profitably, but at the same time makes it difficult to trade profitably."[161]

Volatility, in this framework, is the observable expression of the joint action of information and noise in price formation. Because noise operates independently of information, short-term price movements are more variable than underlying value, even as prices tend over longer intervals to move back toward estimates of value.

[160] Fischer Black, "Noise," *The Journal of Finance* 41, no. 3 (July 1986): 531.
[161] Ibid.

Variability therefore reflects *both* the arrival of new information *and* changes this new information makes in the volume and character of noise trading. The Black–Scholes framework formalized precisely this variability as a parameter to be estimated and priced, treating volatility as a measurable dimension of uncertainty embedded in financial instruments. Under such conditions, uncertainty becomes a priced variable: expectations of future variability shape present valuations, while anticipated volatility recursively reorganizes current prices.

Beyond both the communication channel and the derivatives market, this reorganization of expectation carries the regulatory logic first engineered into the governor and later formalized in probabilistic modeling into the domain of psychology. I argue that once deviation is rendered a measurable differential across time, the subject's role —one's conception of self, others, and the public—shifts with it. Positioned not as an interpreter of meaning or a deliberator over ends—concerns that fall outside the scope of relevance, as Shannon put it—the individual comes to function as an instrument for calibrating exposure within ranked distributions of risk, less a governing agent than a sensing relay within a larger control system. To "learn," in this regime, is to recalibrate one's internal parameters in response to fluctuating signals, such that doubt need not be resolved but instead priced. The political economy thus folds into the psychological: if volatility is endogenous to valuation, uncertainty becomes endogenous to action. Expectation emerges both as the object to be governed and as the medium through which governance is enacted.

Conclusion

Throughout these pages, I have traced how regulation migrates from thermodynamic control to cybernetic governance without abandoning the same practical problem: how to maintain a system's trajectory under disturbance while treating deviation as usable input. Watt's condenser and centrifugal governor did not "think," but they implemented a feedback mechanism: the system continuously compared its current state to a set point, converted deviation into mechanical correction, and repeated the cycle. Maxwell later formalized the stability conditions of such regulators, providing a mathematical account of how feedback systems maintain equilibrium under disturbance. Within this same nineteenth-century milieu, Boltzmann extended Maxwell's statistical treatment of gases to argue that irreversibility arise from any fundamental law of nature but from overwhelming probabilistic tendencies in systems composed of many interacting elements. Order and disorder, equilibrium and fluctuation, could thus be understood statistically rather than mechanically.

In the mid-twentieth century, this statistical understanding of disorder was reformulated in the context of communication and control. Within the cybernetic milieu shaped by wartime engineering and the Macy conferences, Shannon defined information in relation to entropy in order to solve a practical problem: how to transmit signals reliably across noisy channels. "My own model of information theory," in his words, "based mainly on entropy, was framed precisely to work with

the problem of communication."[162] In this framework, uncertainty did not refer to thermodynamic dissipation but to the probabilistic structure of signals and their possible distortions. Formalized as information, uncertainty became a parameter of control: in financial markets, volatility translates probabilistic fluctuation into a priced signal, extending cybernetic regulation into the management of variance over time.

The significance, I suggest, lies less in analogy—whether between thermodynamics, information theory, or financial valuation—but in the portability of a specific operational logic of learning, first articulated in the regulation of mechanical systems: updating in response to measured deviation. Learning, in this cybernetic sense, names the capacity to absorb disturbance while remaining within prescribed bounds of performance—recalibrating the conditions under which interpretation, judgment, and conceptual change could occur, even as the very absorption of disturbance prevents the rupture that would make them necessary.

Once this sense of learning predominates, governance shifts with it. The target becomes the conduct of systems that include people— systems in which anticipation, choice, and attention are rendered extractable inputs, furnishing the human substrate of regulation just as heat furnished the motive power of the engine. As Marion Fourcade and Kieran Healy observe of contemporary "ordinal" forms of social organization, everyday action is increasingly sorted into continuous regimes of

[162] Shannon quoted in *Cybernetics: The Macy-Conferences 1946–1953*, ed. Claus Pias, trans. Dirk Baecker (Zürich and Berlin: diaphanes, 2016), 499.

classification, ranking, and optimization that generate personalized, data-driven possibilities. These systems cultivate a "hierarchical gaze" through which social relations are apprehended as quantified comparisons that rationalize new forms of stratification, such that "any institution that strives for this kind of orderliness, with its ideology of clean measurement of differences and its monstrous technological affordances, has difficulty accommodating a community of equals."[163] What counts as intelligence is reduced to what can be measured, ranked, and monetized, while what resists commensuration recedes from view even as it absorbs the consequences of those measurements.

The steam engine, registering difference without experience, thus engineered a regulatory ambition now condensed in computational form: that the turbulence of the world—and eventually the turbulence of thought—can be made governable by treating difference as information and information as a lever on time. Even in the nineteenth century, when industrial modernity reorganized labor, production, and social life around problems of efficiency and regulation, these same questions entered what Anson Rabinbach described "the new science focused on the body of the worker."[164] The management of energy in machines was inseparable from the "physiology of labor": the regulation of bodies, fatigue, attention, and productivity, as political economy and physiology converged around the problem of maintaining output under conditions of strain and variability. It is within this era of the "human

[163] Marion Fourcade and Kieran Healy, *The Ordinal Society* (Cambridge, MA: Harvard University Press, 2024), 205.
[164] Anson Rabinbach, *The Human Motor: Energy, Fatigue, and the Origins of Modernity* (Berkeley: University of California Press, 1992), 11.

motor"—where it was dangerously assumed that "scientific objectivity and expertise could provide an objective solution to the worker question"—that the psyche-soma emerges as the most consequential limit of the regulatory aspiration.[165] Rabinbach's critique, of course, ultimately extends beyond the human motor to its successor: the human *computer*. In the pages to follow, we will see how the neuron surfaces as an informatic unit within cybernetic models of signal transmission and feedback, enabling the now conventional brain-as-computer metaphor that reorganizes economy, psychology, and social order around computation and information processing.

Before turning to that development, it is necessary to remain a moment longer in the nineteenth century. Working in the final decades of a scientific milieu shaped by thermodynamics and industrial energetics, a young Sigmund Freud set out to formulate a "scientific psychology" grounded in constancy, equilibrium, and the regulation of excitation. Prior to any brain-as-computer metaphor, his early neurological writings modeled mental life in quantitative terms, treating neural excitation as a measurable flow governed by thresholds, resistances, and pathways of discharge. Psychic activity appeared as the regulation of discrete quantities, routed, blocked, circulating across a differentiated network. Echoing the thermodynamic language of conservation and equilibrium familiar from Maxwellian physics, Freud proposed that the nervous system tends toward discharge while maintaining excitation

[165] Ibid.

within tolerable limits—an economy of signals in which stability depends on the controlled distribution of energetic difference.

Yet the psyche did not stabilize as an engine or program might. Clinical phenomena—hysterical symptoms, recurring nightmares, slips of the tongue, compulsive repetitions—persisted precisely where discharge and constancy should have prevailed. Rather than yielding a closed energetic model, Freud's early project gave way, through his work with hysteria and the interpretive practice developed in collaboration with Josef Breuer, to the psychoanalytic framework: a theory in which conflict, repression, and compromise formation organize psychic life, and in which what cannot be "discharged" returns in altered form. Freud did not wish to abandon energetic reasoning; his own clinical reality exposed its fundamental error and forced its transformation. In this, the limits—indeed the errors—of a cybernetic conception of learning come into view—and it is this limit that the next chapter examines.

CHAPTER 2

FREUDIAN INFORMATICS AND THE MAKING OF PSYCHOANALYSIS, 1940-1895

A girl of twelve, sometime around 1878, from a well-to-do family in Vienna's educated bourgeois world, found that she could not enter a shop alone. Nothing in her upbringing prepared her for this incapacity, and nothing in the city's new order seemed to make room for it. Vienna was shedding its former skin. The medieval walls had been cleared away and in their place the Ringstrasse stretched in broad, rational curves. Gaslight burned steadily along the boulevards, and the new electric lamps prolonged the day past nightfall. Horse-drawn trams rattled over iron rails, and the expanding train lines carried people and commodities with a speed that seemed to compress distance itself. Plate-glass shop windows displayed fixed prices and mass-produced goods to passersby, inviting entry while exposing customers to the gaze of strangers. One afternoon she stepped into such a shop and noticed two assistants laughing together. She felt at once that their laughter concerned her—and a sudden dread overtook her. She turned and fled into the street, as if from a danger she could neither name nor understand.

Only years later, in her doctor's office, did an explanation for Emma Eckstein's panic begin to take shape. By then she was twenty-seven and moving within the reformist and intellectual circles that shaped fin-de-siècle Vienna. Her family, active in socialist and feminist politics, knew the physician Sigmund Freud. At the time, he was establishing himself as a neurologist of exact training, shaped by years of medical study and microscopic research: searching for the male sex glands in

eels, tracing the neural structures of lampreys and crayfish.[166] He had only recently left the physiology laboratory of Ernst Brücke, among the most influential figures in nineteenth-century experimental biology. Brücke insisted that life admitted no mysterious forces. Thought, movement, and feeling were to be explained, like any other natural process, by the laws of physics and chemistry. Freud carried this mandate into his own ambitions for psychology, seeking to define mental life as an "economics of nerve-force," a system governed by physicochemical processes and the transmission, resistance, and discharge of energy within the nervous system.[167]

When Eckstein sought treatment for shifting symptoms ranging from gastrointestinal complaints to premenstrual mood disturbances—conditions late nineteenth-century medicine gathered under the capacious diagnosis of hysteria—he was inclined to look for a physiological cause, to locate hysterical compulsions as processes that "operat[e] by means of excessively intense ideas" endowed with a "peculiar kind of quantitative movement."[168] He conceived these disturbances as accumulations and displacements of excitation within the nervous system—irrepressible processes of inhibition and discharge that registered, most visibly, in the subject's sexual life.[169] Yet in practice he relied on a procedure that ran counter to this

[166] Elisabeth Roudinesco, *Freud: In His Time and Ours*, trans. Catherine Porter (Cambridge, MA: Harvrd University Press, 2017), 43-44.
[167] Sigmund Freud, Letter to Wilhelm Fliess on May 25, 1895. Sigmund Freud and Wilhelm Fliess, *The Complete Letters of Sigmund Freud to Wilhelm Fliess, 1887- 1904,* eds. Masson J. Moussaieff and Jeffrey Moussaieff Masson (Belknap Press, 1985), 129.
[168] Sigmund Freud, "Project for a Scientific Psychology" (1895), vol. 1, *Standard Edition,* 405-406.
[169] Tran The Jessica, et al., "From the Principle of Inertia to the Death Drive: The Influence of the Second Law of Thermodynamics on the Freudian Theory of the Psychical Apparatus," *Frontiers in Psychology*, vol. 11, (2020), 3-4.

physicalist model: he invited Eckstein to speak without censorship, to follow the sequence of her thoughts wherever they led. The body was approached as a system of physical forces, but the treatment proceeded through words.

Freud's practice, like the city around him, was in transition. In his earlier collaboration with the Viennese physician and physiologist Josef Breuer, Freud employed hypnosis as it was then practiced within late nineteenth-century European neurology, guiding patients to relive the scenes thought to give rise to their symptoms and to discharge the accompanying affect—what Breuer called "catharsis" or "abreaction."[170] The method proved uncertain. Some patients could not be hypnotized; others spoke what they believed the physician wished to hear; relief, when it came, did not last. Observing that patients under hypnosis often repeated suggestions, readily complied with the physician's authority, or produced memories that dissolved once the trance ended, Freud came to suspect that what the method excluded might be more revealing than what it elicited: the omissions, denials, detours, and redundancies of speech belonged to the symptom itself. They could be listened to, recorded, and interpreted as clues to the conflict at work.

And that is exactly what he did with Eckstein: he listened and recorded her recollection, noting that "it was possible to elicit the idea that the two men had been laughing at her clothes and that one of them had attracted her sexually."[171] Whenever she approached a shop alone, the scene returned with compelling force.

---

[170] David Livingstone Smith, *Approaching Psychoanalysis* (Taylor & Francis Group, 1999), 18-23.

[171] Freud (1895), 410-411.

Why was this the case—why did this anxiety-laden scene return unbidden and beyond her control? The chapter, too, will return to this scene, following the logic of its repetition in order to understand what insists in it.

*

Half a century later, Freud's energetic vocabulary would be sharply criticized from within the cybernetic sciences. In a 1952 letter to the orthodox Freudian psychoanalyst Lawrence Kubie, his fellow participant in the Macy Conferences, Warren McCulloch rejected Freud's reliance on the nineteenth-century doctrine of specific nerve energies—"which Freud got from Johannes Müller secondhand through Brücke"—insisting that it "just won't work."[172] Written on the eve of the Ninth Macy Conference on Cybernetics, the letter came from one of the movement's central architects. A core organizer and intellectual authority within the cybernetics group, McCulloch was among its most forceful anti-Freudian voices, pressing for a clean break between the emerging neurophysiological and computational study of mind and the speculative metapsychology of psychoanalysis.[173] "There is no question," he continued, "but what the business of the nervous system is the handling of information. … the energy required for this is a minor matter so long as it is present."" What Freud had treated as "mental energy, or brain energy, or reflex energy, or cord energy, or sex energy, or something else of that sort," McCulloch argued, obscured the real problem. "It is, in fact, Freud's

[172] Warren McCulloch, Letter to Lawrence Kubie, March 13, 1952. *Warren S. McCulloch papers*, Mss.B.M139, Box 14, American Philosophical Society.
[173] This tension is examined at greater length in Chapter 3, drawing on more correspondence and related archival materials.

hanging onto this antiquated concept of energy that annoys me most in his writings," he wrote; "the crucial problem of nervous activity… [is] not matters of energy but of information."[174]

Eckstein's case brought Freud to a problem that would later animate cybernetics: how *memory* persists and operates in the present. The symptoms of hysteria provided the empirical ground for his "basic hypothesis of memory," that past experience survives as an associative force capable of entering the present without becoming conscious.[175] For such a claim to count as scientific, however, it required a physical model. He thus sought to correlate psychical dynamics with his "basic principle of neuronic activity"—neuronic inertia, or the tendency of neurones to "divest themselves of quantity"—conceiving the psyche as a system governed by excitation and discharge.[176] In this effort he drew on the psychophysicist Gustav Fechner's account of physiological expenditure, in which bodily energy was understood to be transformed into psychological experience. Fechner's formulation of the *Lustprinzip*, or pleasure principle, as a "principle of the tendency to stability" resembling thermodynamic equilibrium, stood alongside Brücke's physical materialism, together offering a framework in which mental life might be rendered intelligible as a lawful process.[177] It was precisely this energetic economy that later

[174] Warren McCulloch, Letter to Lawrence Kubie, March 13, 1952. *Warren S. McCulloch papers*, Mss.B.M139, Box 14, American Philosophical Society.
[175] Freud (1895), 405.
[176] Ibid., 356.
[177] Michael Heidelberger, *Nature From Within: Gustav Theodore Fechner and his Psychophysical Worldview* (University of Pittsburgh Press, 2018) 8.

troubled McCulloch, who would insist that nervous activity is not a matter of psychic force but of the handling of information.

Freud attempted to formalize this model in an unpublished 1895 manuscript, initially titled *Psychology for Neurologists*, or simply "φ–ψ–ω" after the three classes of neurones it proposed: "material particles" linked in a network that receive excitation through branching processes (dendrites) and discharge it along conducting fibers (axons), forming conductive pathways through which nervous activity is transmitted (figure 1).[178] He speculated on a neurological apparatus that regulates excitation, binds quantities, and stores traces of experience. Near the center of this treatise, he placed Eckstein's case—a clinical history that proved to test the limits of a theory seeking to translate memory and trauma into quantities in motion.

Freud ultimately set the manuscript aside, recognizing its speculative reach; it remained buried until 1950, when James Strachey, chief editor of the *Standard*

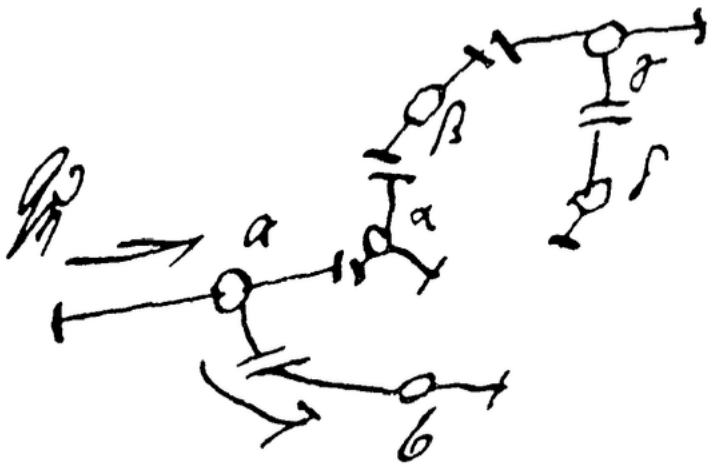


Figure 1: Freud's diagram of "neurones" and their pathways. From Project for a Scientific Psychology (1895), *Standard Edition of the Complete Psychological Works of Sigmund Freud*, vol 1, 385.

[178] Freud (1895), 358.

Edition, published it as the *Project for a Scientific Psychology*. The title, as critics have noted, "harbor[ed] a hidden bias" to "promote Freud the scientist over Freud the philosopher, to represent Freud's materialistic discourse about mind as more valid than [his] psychological discourse about mind."[179] This was an irony not lost on readers, given that Freud himself had abandoned the work as insufficiently scientific. Prior to drafting the *Project*, he had unfortunately seized upon one of the defining technological systems of his age as his guiding model for the structure of mind: the steam engine. Conceived in thermodynamic terms, the psyche appeared as an energetic system tasked with maintaining stability through regulation, governed by an endogenous mechanism that sought to preserve a constant level of available energy—today's supply of coal, or libido—striving toward something like mechanical equilibrium, bioenergetic constancy, or homeostatic self-regulation (categories that are, in fact, incommensurable with one another).[180]

Psychic life, figured as a classical thermodynamic system, was subject both to inertia and to the entropic tendency Freud would later theorize as the "death drive"—a paradox in which the psyche strives to preserve constancy by binding excitation even as it is drawn toward the reduction of tension and a return to quiescence that threatens the organization it seeks to maintain. He described the nervous dynamics of the neurotic subject as oscillating between regulation and runaway, equipped with "safety valves"—jokes, dreams, and other forms of wish-fulfillment—through which

[179] Zvi Lothane, "Freud's 1895 Project: From Mind to Brain and Back Again," *Annals of the New York Academy of Sciences* 843 (1), June 1998, 48.
[180] Anthony Wilden, *System and Structure: Essays in Communication and Exchange* (London: Tavistock, 2nd edition, 1980), 125-126.

"unconscious excitation" could be discharged.[181] Anthony Wilden, the communication theorist and Lacanian interpreter who explicitly sought to place psychoanalysis in dialogue with cybernetics and systems theory, remarked that Freud was "the most metaphorical man who ever wrote," recasting libido through the technical imaginaries of his era: at times a "vast hydraulic system," at others something that "not unnaturally … is electrical," and elsewhere the negative of entropy—a principle of organization opposed to Thanatos, the principle of disorganization."[182]

This energetic model drew directly on contemporary scientific frameworks. As early as 1893, during the period of Eckstein's treatment, Freud invoked the steam engine alongside Fechner's principle of constancy. As the latter wrote, "so far as its course is bound to the course of psychophysical processes and these in turn are bound by the law of the conservation of energy, the mind will itself be bound by that law."[183] In "On the Psychical Mechanism of Hysterical Phenomena," coauthored with Josef Breuer that same year, Freud explained hysterical symptoms as disturbances disrupting the equilibrium of a self-preserving apparatus. Bombarded by "energetic reactions to the event that provokes an affect," the psyche must apply "associative corrections" to prevent any single excitation from assuming "total control of the

[181] Daniel O'Shiel, "Safety Valves of the Psyche: Reading Freud on Aggression, Morality, and Internal Emotions," *Philosophies* 6 (4): 86, 2021.
[182] Wilden (1980), 126. Though Freud himself never used the term *Thanatos*, his shifting and often contradictory metaphors register the difficulty of describing a system that appears simultaneously to conserve, discharge, and organize excitation.
[183] Quoted in Raitt (2002), 64.

subject's whole existence."[184] Within this economy of regulation, purposeless motor discharge became a mechanism of control: "Any impression which the nervous system has difficulty in disposing of … becomes a psychical trauma.[185] When the terms "abreaction" and "catharsis" first appeared in print, Freud (via Breuer) illustrated this discharge through everyday language—"to cry oneself out" (*sich ausweinen*) or "to blow off steam" (*sich austoben*)—expressions that attest to the fact that "in every individual there exists a tendency to diminish this sum of excitation… in order to preserve his health."[186]

In the *Project*, Freud assembled an energetic psychology from nineteenth-century models—thermodynamic, energetic, and psychophysical—whose conceptual incommensurability generated persistent confusion. Reluctant to abandon energetic reasoning at first, Freud eventually entered a period of neurotic illness in which he "abandon[ed] his intellectual work which was meaningless to him." What followed was his *Selbstanalyse*, an undertaking Octave Mannoni later described as nothing less than "the discovery of analysis."[187] Within a year of the unpublished manuscript that housed Eckstein's case and the theory of the neurones, Freud began using the term *psychoanalysis* to name both an emerging clinical method and the theoretical framework that sustained it. In *Heredity and the Aetiology of the Neuroses* (1896), he

[184] Josef Breuer and Sigmund Freud, *Studies on Hysteria*, trans and ed. James Strachey (Hogarth Press, 1955), 8-9.
[185] Sigmund Freud, "On the Theory of Hysterical Attacks," in Sketches for the "Preliminary Communication" of 1893 (1892), vol 1. *Standard Edition*, 283.
[186] Ibid.
[187] Octave Mannoni, F*reud: The Theory of the Unconscious*, trans. Renaud Bruce (Brooklyn: Verso Books, 2015 (1971), 56.

credited his results to "Josef Breuer's exploratory procedure," a method he described as "irreplaceable" for illuminating "the obscure paths of unconscious ideation," and argued that hysterical symptoms could be traced back to "an unconscious memory relating to sexual life."[188] Psychoanalysis took shape at the point where a physicalist, energetics-based psychology proved incapable of accounting for the persistence of unconscious memory or for the traumatic sexual scenes that Freud's seduction theory posited as the specific cause of neurosis.

From the perspective opened by postwar cybernetics and information theory, Freud's energetics occupies a threshold where the calculus of forces reaches its explanatory limit, ceding ground to problems of communication, mediation, and symbolic constraint. Within Freud's speculative apparatus, however, neurones register thresholds and resistances that determine whether excitation passes, is deferred, or bound; memory appears as a differential trace left by prior pathways of passage; and perception and action depend upon the selective opening and closure of routes within a network. Despite its thermodynamic vocabulary, the *Project* models psychic life as the movement of quantities through neuronal pathways, where facilitation, contact barriers, and discharge govern the distribution of excitation. In rendering the psyche as an economy of quantities, Freud traced an ego structured by differential passage, where the movement of excitation leaves lasting facilitations that bias the course of future discharge. What later cybernetic thought would treat as a

[188] Sigmund Freud, "Heredity and the Aetiology of the Neuroses," (1896), vol 3. *Standard Edition*, 145.

mistake of energetic description, I read instead as an inadvertent prefiguration of the informatic and governance paradigm it later sought to systematize. Freud's neurological model endures in its value not as neuroscience but as an early diagram of regulated passage and lasting facilitation—an anticipatory conceptual analogue to the architectures of communication and control that would later define the cybernetic age.

What follows traces how Freud's error was less conceptual than communicative: he grasped a logic of differentiation and transmission that later cyberneticists would formalize, yet articulated it in an energetic vocabulary that obscured its informational implications. I begin with the early cybernetics milieu to show how "mind" was stabilized as a problem of control and communication precisely by policing psychoanalysis as unscientific, even as psychoanalytic questions resurfaced at its margins. I then step back into the mid-nineteenth-century scientific scene—Fechner's founding of psychophysics assisted by his mentor Ernst Weber's experimental physiology of sensation—to reconstruct the conceptual conditions under which difference became measurable and could be treated as a causal operator, though not yet named as such. From there I return to Freud's *Project* as an ultimately failed attempt to found a natural science of memory by modeling the psyche as an economy of excitation; Eckstein's case becomes the point at which clinical evidence strains this energetic framework. The chapter closes by reading these tensions "from the future": how postwar cybernetic discourse displaced Freudian metapsychology by translating problems of memory and regulation in the language of circuits, noise, and feedback, thereby redefining psychic life as information processing and installing the

computational paradigm that now governs machine intelligence and artificial learning. Psychoanalysis, I argue, in its avowal of contradiction, conflict, and the nonrational, preserves the irreducible role of interpretation and symbolic activity—through which historical meaning is constituted in a subject whose formation eludes computational representation.

Psychoanalysis and Cybernetics

In the years after Freud's death in 1939, psychoanalysis—transplanted to the United States by émigré analysts and recast in the ego-psychological language of adaptation and adjustment—remained an uneasy presence within the emerging cybernetics movement (figure 2). At the first Macy Conference on Cybernetics, Norbert Wiener suggested that information rather than energy or libido offered the more appropriate variable for understanding mental processes.[189] Later, in his 1948 *Cybernetics*, he argued that psychotherapy rests on the recognition that "the stored information of the mind lies on many levels of accessibility" and that this stored experience, shaped by affect, conditions later activity in ways that may be pathological.[190] Wiener even speculated that the Freudian linkage between sexuality and memory—"in the theories of Freud the memory—the storage function of the nervous system—and the activities of sex are both involved"—might point toward a physiological substrate, noting the "high emotional and consequently affective

[189] Steve Heims, *The Cybernetics Group* (Cambridge: MIT, 1991), 124-125.
[190] Norbert Wiener, *Cybernetics, or control and communication in the animal and machine* (Cambridge, MA: MIT Press, 1948), 129-130.


Just when you thought it was safe to be a specialist . . .
4 VALUE LOGIC
HELLERSTEIN
PARADOX
t/t
f/t
t/f
f/f LOGIC
COVER OF N.H.
MINDS + MACHINES
SELF REFERANCE
PETRI NETS
SHAPIRO
MCDERMOTT
THINKING
COMPUTERS
SCOTT KIM
VISUAL THINKING
EMBODIMENTS OF MIND
LETTVIN ON McCulloch & AI
WHAT THE FROG'S EYE TELLS...
EMBODIMENTS OF REALITY
VARELA
WHAT IS A SHAPE
NEW IDEAS
Evolution
COPPINGER
NEOTENY
THE ROLE OF EMOTIONS
PERT
BRAIN MOLECULES BODY
ORIGINS OF LANGUAGE
BELLMAN
GOLDBERG
& MOVEMENT
COMMON
ART
MATURANA
ON SEEING
AUTOPOIESIS
BIOLOGY
COGNITION
EPISTEMOLOGY
EIGENVALUES
HVF
ON KNOWING
COGNITIVE BLINDSPOTS
ETHICS
VURICK
WRITING ON WRITING (IN A PETRI NET)
MOTHER GOOSE
EMOTIONS OF A ROLE
MATURANA
VON FOERSTER
PERFORMANCE
POETRY
RESEARCH
STERN
POEM THINK
"TRY IT"
MUSIC
BRUN
COMPUTERS
SOCIETY
COMPUTER PROFESSIONALS
NUCLEAR COMPUTERS WAR
FOR SOCIAL
RESPONSIBILITY
TASK
ON
Problems
& HOW TO CREATE THEM
AVANT-GARDE
LETTER TO WIENER
ON THE AT&T
ORSON WELLES
ANIMALS & MACHINES
HISTORY
Cybernetics

Figure 2: Conceptual map of cybernetics and related fields as represented in the first issue of *Cybernetic,* an essay collection sponsored by the American Society for Cybernetics, Summer-Fall1985. The diagram traces the interdisciplinary expansion of cybernetics from the Macy Conferences into domains including computing, biology, cognition, language, art, and social systems. *Warren S. McCulloch papers*, Mss.B.M139, Box 80, American Philosophical Society.

signaling process rather than a quantity of energy, Wiener viewed psychoanalysis as a precursor, suggesting—despite Freud's own energetic vocabulary—that it pointed toward a problem of communication later formalized by cybernetics.

Macy Conference member Lawrence Kubie, on the other hand, attempted to make psychoanalysis palatable to "hard" scientists by drawing an explicit parallel between analytic theory and behaviorist physiology. At the seventh conference in 1950, he argued that "Pavlov has proved that no new reflex can be conditioned in a satiated animal, and that therefore all processes of learning depend upon the existence of a state of craving. Approaching the problem from another angle and by another method, psychoanalysis made the same discovery, proving that the acquisition of new knowledge depends upon the existence of a state of instinctual tension and deprivation."[192] Because early cravings arise as bodily tensions, Kubie continued, the child's world of thought begins with the body—its parts, needs, products, affects—such that language and cognition emerge from attempts to name and regulate desire. In this formulation, the "tyranny of conflicting unconscious drives" could be reframed as a problem of symbolic regulation and learning, aligning psychoanalytic insight with cybernetic concerns while subordinating the unconscious to models of adaptive

[192] Lawrence Kubie, "Body Symbolization and Development of Language," *Cybernetics: The Macy Conferences 1946-1953 Transactions*, ed. Claus Pias (Zurich-Berlin: diaphanes, 2003/2016), 326.

function. Psychoanalysis could thus be reconciled with a scientific ethos that valorized consciousness as the seat of goal-directed behavior and normative adjustment—in "direct correlation with its essential normality."[193] Yet, as McCulloch's increasingly forceful interventions would make clear, as evident in his 1952 letter discussed above, Kubie soon softened the metapsychological claims that had marked this earlier contribution, adopting instead the posture of a clinical naturalist describing emotional processes.

Meanwhile, anthropologists Margaret Mead and Gregory Bateson lobbied for child psychiatrist Erik Erikson's inclusion in the core group, but lacked the scientific authority that McCulloch and his protégé Walter Pitts carefully guarded with empirical rigor. Erikson, after his first and only participation in 1947, was unsure why the participants were talking so intensely about machines. He presented his clinical research on child development but was quickly dismissed as "unrigorous" and unsuitable for participation in subsequent years.[194] Bateson, who would later criticize Kubie's elevation of conscious purpose at the 1967 Congress on the Dialectics of Liberation in London, concurrently pursued a different line of argument: human communication and learning, he insisted, cannot be reduced to goal-directed rationality or adaptive control—and when they are treated as such, psychopathology

[193] Ibid. Heims described how Bateson, in an interview on August 16, 1968, reminisced that Kubie "defined behavior as neurotic when it is repeated again and again even though it doesn't work—but that was exactly what Kubie had done when he repeatedly and always unsuccessfully presented psychoanalysis at the conferences.".
[194] Ronald Kline, *The Cybernetics Moment, Or Why We Call Our Age the Information Age* (Baltimore: John Hopkins University Press, 2015), 37-46.

and systemic dysfunction follow.[195] In his 1952 talk on humor and communication, Bateson suggested that therapeutic transformation depends upon the "implicit presence and acceptance of the paradoxes," arguing that the Freudian analysand progresses through the "mental flux, confusion, or entropy stirred up by paradox," passing through inner disorder to achieve "a new affective organization of experience."[196] Bateson held that psychological change arises from the capacity to tolerate contradiction and multiple levels of meaning—as in dream interpretation or in the play of children and animals—against the drive to secure rational control through conscious purpose that shaped the dominant technological strain of cybernetics.

Within the Macy milieu, psychoanalysis operated as a persistent point of friction, alternately enabling and obstructing competing visions of mind, communication, and control. No figure registered this tension more forcefully than McCulloch, whose neurophysiological program sought to secure the scientific study of mind on rigorously material and mathematical grounds (figure 3). According to Steve Heims, "The blatantly subjective nature of a psychoanalyst's perception of his or her client, in contrast to most other types of scientific observation, heightened the skepticism of empirical scientists."[197] McCulloch, alongside the polymath John von Neumann, was not just a skeptic but a hostile critic, condemning the existence of

---

[195]: Gregory Bateson, "Conscious Purpose Versus Nature," ed. David Cooper, T*he Dialectics of Liberation* (Verso Books, 2015).
[196] Gregory Bateson, "The Position of Humor in Human Communication," *Cybernetics: The Macy Conferences 1946-1953 Transactions*, ed. Claus Pias (Zurich-Berlin: diaphanes, 2003/2016), 544.
[197] Heims (1991), 124-125.

psychoanalysis as "propaganda [in a] pseudoscientific system" servile to a "desire for filthy lucre," converting neophytes eager to buy "acceptance into the sect depend[ing] upon emotional conversion… to the sect's peculiar beliefs." In his famously vitriolic essay "The Past of a Delusion" (a play on Freud's "The Future of an Illusion" to announce his unwieldy obsolescence), he dismissed the Freudian scheme as "a tissue of unverified and often unverifiable hypotheses."[198] McCulloch faulted Freud for reducing the science of mind to pulp fiction, insisting that the "scientific way into the psychiatry of damaged brains" lay in treating psychological symptoms as the byproduct of "disordered structures." "Freud's delusion has no future, but it had a past," he concluded, and "unfortunately, Delusions are Ideas… we cannot bury them properly. Only the dust of ages can consign them to oblivion."[199]



Figure 3. Cover illustration from a Biological Computer Laboratory publication circulated by Heinz von Foerster (undated). *Heinz von Foerster Papers*, Box 10, University of Illinois Archives, Urbana-Champaign. The anatomical–optical, Cartesian motif reflects Warren McCulloch's effort to ground cognition in neurophysiology and formal relations between perception and representation.

Though deferential to McCulloch's authority within the conference scene, Bateson

[198] Warren McCulloch, *Embodiments of Mind* (Cambridge, MA: The MIT Press, 1988). 279-282.
[199] Ibid., 289. It did not occur to McCulloch that the "idea" of male-objective, uncastrated rigor defending the narrow straits of scientific validity might ever share the same fate.

approached psychoanalysis less antagonistically; as Heims notes, Wiener even urged him to reinterpret psychoanalysis through cybernetics as his work in the social sciences drew him into psychiatry.[200] He began analysis with the Jungian psychotherapist Elizabeth Hellersberg while conducting ethological research in cetacean and mammalian communication, in tandem with clinical rotations with veterans suffering from schizophrenia and alcoholism. As he worked to translate cybernetic principles into a therapeutic practice attentive to paradox and higher orders of change, Bateson—undeniably influenced by McCulloch—echoed the latter's critique that Freudian psychoanalysis remained tethered to the neurological and physiological presuppositions of the late nineteenth century. In a letter addressed to Wilden in 1969, he mused:

> If only Freud had met Clerk-Maxwell!… (Freud was 22 when Maxwell died, at 48.) And then, in addition to the false premises of language, there is also the villainous idea that what travels in an axon is an "impulse." What difference it would have made if Freud, who discovered the neuron, had perceived that the inactive neuron is also a source of *information*—what the axon carries is the <u>difference</u> between activity and inactivity.[201]

Had Freud been conversant with the work of his near contemporary James Clerk Maxwell, whose analysis of the governor articulated a proto-cybernetic concept of information as operationally linked to physical work yet irreducible to thermodynamic energy, he might have approached mental regulation differently.

---

[200] Heims (1991), 124-125.
[201] Gregory Bateson, Letter to Anthony Wilden, October 6, 1969. *Gregory Bateson papers*, MS 98, Box 36: 1497-4a. Special Collections and Archives, University Library, University of California, Santa Cruz.

Equipped with a theory of feedback and a vocabulary of information, he may have resisted grounding psychoanalysis in a hydraulic model of psychic energy and instead framed it in terms of signal and control.

Freud might then have recognized that the informational constraints governing energy expenditure in a thermodynamic system are analogous to the physical concept of entropy. Maxwell's analysis of governors and his corresponding demonic thought experiment showed that the labor of organizing energy entails increasing a temperature gradient—a spread of information—seemingly contravening the second law of thermodynamics by decreasing entropy without additional energy expenditure.[202] The simultaneously statistical and informational character of this mechanism anticipates what Leo Szilárd demonstrated in 1929: when measurement intervenes in a thermodynamic system, it introduces a "memory faculty" that makes the system behave differently from a purely mechanical one and can give the appearance of entropy reduction.[203] He emphasized, however, that this apparent violation is only temporary, since "the measurements themselves are necessarily accompanied by a production of entropy," and he further showed that the same effect could be realized by a simple apparatus: "We have examined the 'biological phenomena' of a nonliving device and have seen that it generates exactly that

[202]See: Astrid Schrader, "Haunted Measurements: Demonic Work and Time in Experimentation," *Differences: A Journal of Feminist Cultural Studies* Volume 23, Number 3, 119–160; Harvey S. Leff and Andrew F. Rex, *Maxwell's Demon: Entropy, Information, Computing* (Bristol, UK: Adam Hilger, 1990).
[203] Leo Szilard, "On the Decrease of Entropy in a Thermodynamic System by the Intervention of Intelligent Beings," in *Behavioral Science* 9, no. 4 (1964): 301–310.

quantity of entropy which is required by thermodynamics."[204] Energy is organized through selection, memory, and constraint—and this was an insight psychoanalysis approached in its own language and cybernetics later formalized.

For Bateson, the problem rested in the structure of language itself. The fallacious nature of linguistic abstraction, "the idea that a separation [in subject from predicate] is analogic to some separation in the "real" world," fostered a habit of thought that encouraged the false equivalence between energy and information. "Language," he continued, "still seems to me to be the most miraculous disaster that ever happened to an organism… even so, Freud achieved surprising wisdom about personal relationships, transference, etc."[205] Freud's metapsychology thus faltered not because his clinical observations were misguided, but because the available scientific idiom forced dynamic relations into the grammar of substances and forces. In describing psychic conflict through the terminology of energetics, Freud inherited its *conceptual* tensions, conflating a model of physical forces with the radically different terrain of the psyche. These tensions predate Freud and belong to a wider scientific and political–economic order organized around the regulation of force, productive activity, and time; in Fechner's psychophysics they appear in miniature, where quantitative thresholds were asked to bear the weight of qualitative experience.

Difference Without Dimension

[204] Ibid, 310.
[205] Gregory Bateson, Letter to Anthony Wilden, October 6, 1969. *Gregory Bateson papers*, MS 98, Box 36: 1497-4a. Special Collections and Archives, University Library, University of California, Santa Cruz.

In the Winter of 1840, Gustav Fechner painted the walls of his room black, wore a metal mask to further block out the light, and starved himself as he self-secluded for the next three years.[206] Afflicted by blindness, chronic pain, and severe depression—the "total destruction" of his "mental energy"—Fechner found the treatments of his day fruitless. Scientific obsession and neurotic desperation fused into a series of experiments aimed at inducing bodily change through "volitional exercises of mental effort." If digestive pain governed behavior by transmitting signals from organ to brain, he reasoned, might volition reverse the direction of conduction? "Why not," he asked, "by the exercise of volition, bring about a conduction from the brain to those organs and thus remedy them?"[207] He called this process "auto-suggestion": the channeling of "psychic energy" to regulate pain as an involuntary physical function. What he was attempting to engineer was a system linking stimulus and sensation into a circuit of psychosomatic regulation.[208]

From this ordeal Fechner concluded that psychical and physical phenomena could be related through a quantitative method, correlating material stimulus with subjective sensation and rendering perception measurable. The insight precipitated a

---

[206] Six years prior, Fechner left a non-salaried lecturer position to be appointed as Professor of Physics at Leipzig University. As his mentee and colleague Wilhelm Wundt reflected during this period, "from the moment he attained an independent position which would give him freedom for his own work, from that very moment his strength was broken. His excessive work had exhausted him. He had difficulty in finishing his lectures." Quoted in Henri F. Ellenberger, *The Discovery of the Unconscious: The History and Evolution of Dynamic Psychiatry* (London: Fontana Press, 1970), 215-216.
[207] Larry H. Ludlow and Rose Alvarez-Salvat, "Fechner: The Man in the Mask," *Popular Measurement*, Vol. 3, no. 1 (Spring 2000), 6.
[208] Henri Ellenberger, "Fechner and Freud," *Bulletin of the Menninger Clinic,* July 1, 1956; 20, 4: 205.

manic upswing—a "new birth" in which everything "settled into symmetry."[209] He resigned his professorship in physics, transferred to philosophy at Leipzig, and spent the remainder of his career elaborating a quantitative psychology. The result was *Psychophysics* (1860), a monumental effort to establish lawful relations between mind and matter.

Fechner's breakthrough built directly on the work of his mentor Ernst Heinrich Weber. In the 1830s Weber conducted experiments in which blindfolded subjects lifted weights of gradually increasing mass.[210] Subjects failed to detect differences until a threshold was reached. Crucially, the least perceptible difference was proportional to the initial weight: if no difference was perceived between 100 g and 116 g, the same held between 158 g and 173 g. Sensitivity to difference was best measured as a *ratio*. Weber extended this insight to clinical contexts—treating patients with hyperesthesia and other nervous disorders, predominantly bourgeois young women in whom Breuer and Freud would locate hysteria 40 years later (figure

[209] Michael Heidenberger, *Nature From Within: Gustav Theodor Fechner And His Psychophysical Worldview* (University of Pittsburgh Press, 2018), 132.
[210] Ernst Heinrich Weber, "Concerning touch" (1834), In *Readings in the history of psychology*, ed. Wayne Dennis (Amberg Press, 2007), 155–156.

4)—and to tactile perception, showing with a compass that two points touched on the skin are perceived as distinct only beyond a certain spatial threshold.[211] Distinguishing two points of sensitivity on the skin with a compass, Weber observed that when two points of the skin in close proximity are touched simultaneously, subjects perceive these two points only if the distance between the locations were greater than a certain limit.[212] The stimulus producing sensation, the "just-noticeable difference" (*jnd*), was *dimensionless*: it altered no localized energy yet produced a perceptual effect.



Figure 4: Application of Weber's compass, Engraving taken from *Le Monde Illustré* of 9 June 1894 ((no. 1941, 38th year, p. 374). reproduced in Nicolas (2017).

Weber defined this *jnd* as a threshold communicating "the disparity between things that are compared," in which "we perceive not the difference between the things, but the *ratio of this difference to the magnitude of things compared*."[213]

---

[211] Serge Nicolas, "Weber's compass and the measurement of the threshold of tactile sensitivity," *L'année psychologique/Topics in Cognitive Psychology*, 2017, 117, 48-49.
[212] Weber (1834), 46-47. "First of all we used the instrument called *Stangenzirkel* (beam compasses or beam trammels) in German. This consists of a long metal beam in the shape of a right angled prism, with two points fastened at right angles (to the beam). Because these points can be moved towards and away from each other in a straight line, the instrument seemed suitable for touching the skin with the two points a certain distance apart."
[213] Quoted in Raymond E. Fancher and Alexandra Rutherford, "The Sensing and Perceiving Mind." *Pioneers of Psychology: A History* (New York: W.W. Norton, 2012), 167.

Measuring sensitivity—whether in touch, weight, or heat—thus required a *relation* rather than an absolute unit. The operative stimulus was not a quantity located in space or time but a difference between comparable states. By distinguishing the capacity to localize a stimulus from the capacity to discriminate it, Weber suggested that the familiar physical dimensions of an object fall away in perception: what is registered is not magnitude itself but a relation among magnitudes.[214]

Fechner found remarkable similarity between his method of autosuggestion and Weber's earlier experiments. He proposed that stimuli (physics) and sensation (psychology) are linked by a logarithmic transformation, such that "the relative increase of bodily energy" provides a "measure of the increase of the corresponding mental intensity."[215] On this basis, the "geometric" increase of stimulus intensity corresponds to an "arithmetic" progression in sensation.[216] To estimate the magnitude of a sensation, one could treat the *jnd* between two stimuli as a unit and count successive *jnd* increments from the absolute threshold upward. This reasoning culminated in the logarithmic distribution $S = k \log R$, in which the magnitude of sensation (S) increases as the logarithm of stimulus intensity (R), scaled by a constant (k).[217]

---

[214] Nicolas (2017), 49. "Weber's compass and the measurement of the threshold of tactile sensitivity," *L'année psychologique/Topics in Cognitive Psychology*, 2017, 117, 49. "*Raumsinn* refers to the ability to distinguish between two points of contact that are more or less close to one another. *Ortsinn* designates the ability to identify the location of a felt contact, our ability to identify where it occurs, its precise localization."

[215] Edwin G. Boring, *A History of Experimental Psychology* (New York, Appleton-Century-Crofts, 1950), 280.

[216] Ibid.

[217] Donald Laming, "Fechner's Law: Where Does the Log Transform Come From?" In *Fechner's Legacy in Psychology: 150 Years of Elementary Psychophysics* (Leiden: Koninklijke Brill NV, 2011), 19-20. Fechner devised multiple experiments to apply and prove this law for the next ten years,

Crucially, the Weber–Fechner law asserted that a measurable ratio—a relation of proportionality—governed not only differences among physical magnitudes but also the relation between bodily stimuli and subjective experience. Physical stimulus and sensation, the distinct domains of physics and psychology, were thus linked through a logarithmic transformation.[218] In his period of self-imposed isolation, Fechner believed he could actively modulate this relation through "exercises of volition," attempting to regulate sensation by directing mental effort toward bodily processes. This therapeutic practice, modeled on Weber's law, appeared as a form of intelligent self-regulation in which mind operated across endogenous (perceptual-psychological) and external (material) dimensions.

From the vantage of later information theory, the Fechner-Weber law demarcates a conceptual shift: what mattered in sensation was not absolute magnitude but the relation between magnitudes. Difference appeared not as a quantity with its own physical dimension but as a ratio, a logarithmic relation between comparable states.[219] Weber's comparisons of weight or color thus did not map a spectrum of physical intensities so much as register the structured differences among them—the relational pattern through which sensation becomes meaningful.

In a commemorative address delivered at the Langley Porter Clinic in 1976, Gregory Bateson thus retrospectively identified this discovery of the logarithmic ratio

---

accumulating into the discipline he termed *psychophysics*: "an exact science of the fundamental relations of dependency between body and mind."

[218] Heidenberger, *(*2018), 37-40.

[219] Claude Shannon and Warren Weaver, *The Mathematic Theory of Communication* (University of Chicago Press, 1949), 9-10.

and the just-noticeable difference as foundational for a science of mind oriented toward information rather than energy:

> …this discovery puts the whole of the hard sciences out of the realm of what we're interested in [the science of mind]. In the hard science it was always assume that causes have real dimensions of Length, Mass, Time, or some comb nation of these; e.g., Energy has dimension ML2/T2 (i.e., Mass x Velocity squared). But the Weber-Fechner generalization asserted that stimulus as a "cause" of sensation or behavior was of zero dimensions—a ratio between similar dimensions."[220]

Because sensation itself could not be measured directly, its magnitude had to be inferred from the stimulus differences that produced it. What Fechner formalized was therefore a *discrete* procedure: sensation could be counted in increments of threshold differences, a logic that treated dimensionless distinctions as the operative units linking body and mind. In this respect, psychophysics articulated a conception of information avant la lettre: difference as a non-substantial yet operative relation, anticipating the twentieth-century shift from energetic magnitude to communicable distinction. When cybernetics later reconceived mind and communication, it did so on conceptual ground first cleared when sensation itself was rendered measurable as difference.

[220] Gregory Bateson, "The Growth Paradigm for Psychiatry: A Talk Given to the Langley Porter Clinic on 17 November 1976," *Gregory Bateson papers*, MS 98, Box 96: CAF 113-C-1. Special Collections and Archives, University Library, University of California, Santa Cruz. 3. The Langley Porter Clinic, the first psychiatric hospital in the United States, is known today as the Langley Porter Psychiatric Institute, the teaching hospital in the Psychiatry Department at the University of California, San Francisco.

The Pleasure Principle before Freud

In a journal entry of June 6, 1846, written during the period of "auto-suggestion" that followed his collapse, Fechner remarked that "it was not allowed… to regard external nature in a fruitful way." Turned inward, he found himself "driven to the inner world," where he later summarized the central arguments of these reflections in a short treatise titled *On the Highest Good*.[221] From this enforced introspection emerged a principle that would later resonate strongly with Freud: if energy can neither be destroyed nor created but only converted or transferred, then the psyche too must be "bound by the law of the conservation of energy."[222] Mental life, on this view, was governed by an economy in which a "hedonistic" and "religious" principle of pleasure—measured in relation to unpleasure—regulated the circulation of excitation and thereby the stability of action and its consequences.[223]

Coining the phrase *Lustprinzip*, or pleasure principle, Fechner defined pleasure as the reduction—or even eradication—of conflict in a state of maximal stability.[224] Pleasure consisted in preserving conditions of constancy by governing stimuli and sensations in a "circular or oscillating motion… such that the parts always

[221] Quoted in William R. Woodward, "Fechner (1801-1887) For and In Psychology: Part I," *Archives of Psychology,* vol. 2, issue 5, May 2018: 2.
[222] Suzanne Raitt, "Psychic Waste: Freud, Fechner, and the Principle of Constancy," In *Culture and Waste: The Creation and Destruction of Value*, ed. Gay Hawkins and Stephen Meucke (Rowman & Littlefield Publishers, 2002), 64.
[223] Woodward (2018), 3.
[224] As examined in the next section, Freud will come to explicitly cite Fechner's "principle of the tendency towards stability" in *Beyond the Pleasure Principle* (1920) to associate pleasure with the regulation of stable levels of excitation in the psychical economy.

return to a given position after a lapse of time." Envisioning repetition and return as intrinsic to the action of psychic energy, he further maintained:

> … The *principle of pleasure is a computational principle*… When have people ever acted other than with reference to the assumed consequences of their actions upon their happiness and unhappiness? Befitting an empirical theory, the quantitative pleasure principle has corollaries; among them, that we can make mistakes about pleasures, we can sum them, and we can calculate consequences of our actions in regard to future happiness.[225]

The pleasure–unpleasure principle thus required decomposing the total system into discrete parts as a condition of self-organization.[226] The overall sensation of pleasure (stability) became perceptible through increments of stimulus, functioning as signals that "evok[e] associations with signs that convey meaning in the world."[227]

Fechner's "computational" pleasure principle can be read, from a later informational vantage, as structurally analogous to a communication system in which inputs consist of discrete stimuli selected from a range of possible states and coding functions to identify those stimuli without error. Yet the output necessarily passes through an *observer* whose perception arrives in wholes rather than units; interpretation introduces judgment, and with it the possibility of error. The observer's responses in the output include an interesting frequency of errors, misinterpretations

---

[225] Ibid.

[226] During the infancy of organic chemistry and almost two decades before genetics appeared at the forefront of the biological sciences, it was impossible for a scientist to avoid the ideas of Charles Darwin. Beyond the purview of exploration here is how Fechner, in his 1873 paper "Some Ideas on the Origin and Evolutionary History of Organisms," completed his psychophysical theory in the realm of evolutionary biology and provided a prototypical theory of autopoietic (self-creating) systems in the tradition of second-order cybernetics and systems theory post-1970s.

[227] Woodward (2018), 4.

generating change in action, unlike the ideal efficiency attributed to purely digital input. This asymmetry anticipates Claude Shannon's influential definition of communication as "that of reproducing at one point either exactly or approximately a message selected at another point," where meaning belongs not to the transmission itself but to the referential system in which messages are interpreted.[228] The engineering problem, accordingly, is solved by minimizing noise and eliminating dependence on an energetic, analog context: the "semantic aspects of communication are irrelevant."[229]

Fechner, though lacking the later vocabulary of information theory, intimated that the pleasure principle operates at the threshold between discrete stimuli and the continuous sensory field of experience, thereby "not only allow[ing] us to use all experiences in life for the theory" but also "requir[ing] us to enter into the empirical nature of humans and things."[230] In this formulation, the inherited dualism of mind and body gives way to "an exact science of the function relations or the relation of dependency between mind and body," extending computation beyond physical processes into what he called the "spiritual world."[231] Read retrospectively through Bateson, this domain corresponds less to metaphysics than to a cybernetic terrain: the psychological world understood as circuits of communication, patterned difference, and informational exchange.

---

[228] Shannon and Weaver (1949), 31.
[229] Ibid.
[230] Woodward (2018), 3-4.
[231] Joshua Bauchner, "Fechner on a Walk: Everyday Investigations of the Mind-Body Relationship," *Historical Studies in the Natural Sciences* (2021) 51 (1): 4.

Fechner's psychophysics posed the problem of how "a theory of action, which has to move in the empirical realm, [can] be itself independent of the empirical."[232] he just-noticeable difference—information generated through comparison between bioenergetic quantities yet irreducible to those quantities—introduced a distinction between material stimulus and the relational difference that produces sensation. This distinction helped prepare the ground for Freud's economic view of the psyche. We can now return to Freud to examine how Fechner's conception of pleasure as goal-directed, quantifiable, and conveyed through signs—informational in structure, though decades before such a term existed—seeded a therapeutic practice oriented toward interpretation, where meaning emerges through the differential production and reading of signs.

Jenseits des *Lustprinzips*: The Psychic Economy

Henri Ellenberger, writing in the 1970s, was among the first to trace Freud's indebtedness to Fechner's psychophysics, noting the "remarkable similarity" between Fechner's "auto-suggestion" and Freud's "self-analysis."[233] Although Fechner's three-year crisis was more extreme, Freud's neurasthenia likewise unfolded in conditions of isolation, "loneliness and privation," exhaustion, and depressive episodes. On the far side of these efforts at self-healing lay decisive intellectual transformations: Fechner emerged from darkness to articulate the pleasure principle

[232] Woodward (2018), 3-4.
[233] Ellenberger (1956), 205.

in *On the Highest Good*, while Freud set out on what he would call the "royal road to the unconscious."[234] These turning points coincided with shifts in vocation and method—Fechner from physics to philosophy and psychology, Freud from neurology to the study of the neuroses and the development of psychoanalysis.[235]

Across his career, Freud repeatedly acknowledged who he called "the great Fechner," citing him in works from *The Interpretation of Dreams* to *Beyond the Pleasure Principle*. From Fechner he adopted not only the language of "mental energy" but a set of regulatory principles—pleasure, constancy, repetition—through which psychic life could be described as an economy tending toward stability.[236] Freud's psychodynamics posited a psychological counterpart to Fechner's physiological constancy: the sum of excitation within the apparatus is subject to discharge and binding so as to prevent disruptive accumulation.[237] In this formulation, stability does not imply stasis but regulated circulation, a balance maintained through the continual modulation of tension.

---

[234] After interpreting the historical Irma's injection dream to formulate the central thesis in *The Interpretation of Dreams,* Freud eventually discovered the Oedipus complex after the death of his father. By 1905, he quickly overturned the seduction theory centering on the trauma of the seduction of the father as merely a resistance to knowledge of Oedipal desires and proposed a new theory of infantile sexuality.
[235]Ibid, 218.
[236] Suzanne Raitt's intellectual history of Fechner and Freud demonstrates that the principle of constancy and degradation (formulated by the first and second law of thermodynamics) functioned as the undergirding foundation for a theory of "psychic waste" emerging in the Freudian psychic economy during his metapsychological turn after 1915.
[237] Raitt (2002), 63-66.

Fechner's law had suggested that sensation could remain stable under changing stimulus conditions so long as relational proportions were preserved; Freud extended this insight into a model of psychic regulation in which equilibrium depends on the management of excitation rather than its elimination (figure 5). Physiological energetics, mechanical regulation, and economic metaphors thus converged in a system that conceived mental life as governed by a "tendency toward stability," sustained through the circulation and redistribution of excitation.[238] The psyche, in this sense, appeared less as a reservoir of force than as a regulatory apparatus—an attempt to render intelligible how constancy might be maintained amid continual disturbance, even if the conceptual vocabulary available to Freud remained a convoluted amalgam of nineteenth-century scientific idioms.

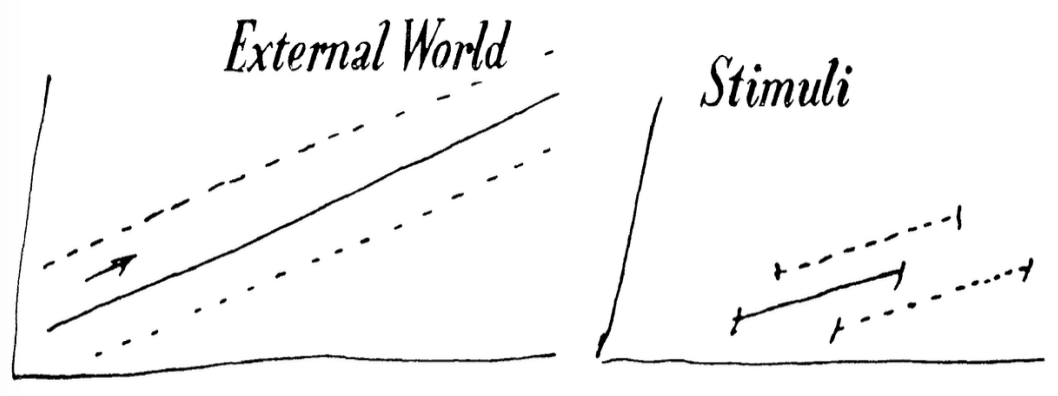


Figure 5: Freud's diagrams of stimulus intensity and external excitation. Freud (1895), 375.

[238] It is not Fechner, however, but Wilhelm Gottfried Leibniz who is canonically credited with having coined the word "dynamics" as the science of force. The latter coined the term "*dynamica*" in 1689 to distinguish the formal elements of motion from other physical parameters such as weight and elasticity. See: Carolyn Merchant, "Leibniz' Concept of Force: Physics and Metaphysics," *Akten Des II. Internationalen Leibniz-Kongresses*, Hannover, July 17-22, 1972 (Wiesbaden: Franz Steiner Verlag, 1974), 143-148.

Despite drawing on multiple and often incommensurable models, Freud recognized a tension between two regulatory principles: energetic constancy and neuronic inertia. The former, akin to a cybernetic logic of stabilization through feedback, concerns the regulation of excitation within tolerable bounds; the latter, resembling thermodynamic equilibrium, describes a tendency toward discharge and quiescence. Constancy, in this sense, does not simply mirror the organism's adaptive self-preservation but can operate, as Freud later wrote, "*in the service of the death instinct*, whose aim is to conduct the restlessness of life into the stability of the inorganic state."[239] Fechner had defined pleasure in terms of the increased stability of excitation's movements rather than their reduction. Freud, however, often collapsed stability into diminution, treating greater regularity as evidence of reduced quantity. From the 1920s onward—nowhere more clearly than in *Beyond the Pleasure Principle*—this tension reappeared in the distinction between the ego-instincts, which "impel towards death," and the sexual instincts, which preserve life."[240] The neurones' tendency "to get rid of energy" thus pointed toward an imagined zero level of excitation, a limit state that could be construed only as bioenergetic death.[241]

Freud treated the second law of thermodynamics as a fitting analogue for neuronic inertia. If psychic life were governed by a principle of constancy, any surplus excitation would have to be discharged in order to preserve equilibrium.

---

[239] Sigmund Freud, "The Economic Problem of Masochism" (1924), vol. 6, *Standard Edition*, 160.
[240] Sigmund Freud, "Beyond the Pleasure Principle" (1920), vol. 18, *Standard Edition,* 44.
[241] Tran The Jessica, et al., "From the Principle of Inertia to the Death Drive: The Influence of the Second Law of Thermodynamics on the Freudian Theory of the Psychical Apparatus," Frontiers in Psychology, vol. 11, (2020), 3-4.

Entropy thus named a limit on self-regulation: energy cannot be transformed without a loss of organization and, with it, a diminished capacity for further work. In this view, regulation produces waste. The analogy extends, retrospectively, to information theory, where noise emerges as an unavoidable byproduct of transmission within a total communicative context. The psyche consequently appears as an intrinsically inefficient economy, one "destined to bring about its own demise by the very processes by which it sought to sustain its own existence."[242]

The law of the conservation of energy, which, in Philip Mirowski's words, "initially dangled the delectable promise of a grand unification of all physical and mental phenomena," also introduced a world in which the manipulation of energetic processes entailed irreversible loss.[243] Just as steam engines never returned as much work as they received, the Freudian psyche—tasked with maintaining constancy of excitation—paradoxically modeled within the organism a teleonomic mechanism oriented toward its own cessation:

> It seems, then, that an instinct is an urge inherent in organic life to restore an earlier state of things which the living entity has been obliged to abandon under the pressure of external disturbing forces; that is, it is a kind of organic *elasticity*, or, to put it another way, the expression of the inertia inherent in organic life… *We shall be compelled to say that the aim of all life is death.*[244]

[242] Raitt (2002), 68.
[243] Phillip Mirowski, *Machine Dreams: How Economics Became A Cyborg Science* (Cambridge University Press, 2002), 28.
[244] Freud (1920), 36.

Freud's notion of "elasticity" is commensurate with the pleasure principle: the homeostatic tendency to maintain excitation at its lowest workable level.[245] Yet the psyche remains bound to equilibrium in a state of inertia—"like the steam engines whose coal smoke blackened the skies of nineteenth-century industrial towns"—inasmuch as the conditions of life are governed by this rigid rule of elastic self-preservation.[246] Like any organism that lives only by postponing its own thermodynamic end, the psychical apparatus must hold time at bay: extracting and processing matter, regulating the distribution and conversion of endogenous energy, and eliminating waste. The psyche thus emerges as a regulatory system tasked with maintaining constancy of excitation, even as the very logic of regulation tends toward the inert equilibrium that *Beyond the Pleasure Principle* names as life's terminal horizon. Yet as early as 1895, Freud attempted to formalize this logic in a speculative neurological model, describing a system of thresholds, pathways, and stored traces governing excitation.

Project for a *Cybernetic* Psychology?

Something like information, however, was present in Freud's intellectual formation even earlier, in his encounter with the work of physiologist Hermann von Helmholtz at the Medical School of Vienna. Helmholtz—who formulated the

[245] For Catherine Malabou, a recursive "plasticity" is at work in the binding of the beginning of life as "living substance" with death as "the pure neutrality of inorganic matter" by the "archaic temporality" of inorganic matter. Malabou, "Plasticity and Elasticity in Freud's 'Beyond the Pleasure Principle'," *parallax*, 2009, vol. 15, no. 2, 42-47.
[246] James R. Beniger, *The Control Revolution: Technological and Economic Origins of the Information Society* (Harvard University Press, 1986), 37.

principle of the conservation of energy in 1847, also invented the ophthalmoscope in 1851, inaugurating modern ophthalmology. In his *Treatise on Physiological Optics* (1867), he advanced the concept of "unconscious inference," deriving a theory of perception from the instrument itself.[247] The ophthalmoscope, he noted, depends upon three elements: a source of light, a reflecting surface to direct that light into the eye, and a mechanism for correcting an out-of-focus image (figure 6). From this apparatus Helmholtz proposed that stimulation of a retinal nerve fiber yields sensation along three axes: hue (the quality of the light source), intensity (the strength of the reflected light), and, most significantly, "local signs," which function as a corrective process that brings perception into focus.[248]

In the unconscious register of perception, these "local signs" function as qualitative markers of difference: excitations of color or intensity across distinct

[247] Leo Koenigberger, *Hermann von Helmholtz*, trans. Frances A. Welby (Oxford: Clarendon Press, 1905), 140-141. "Now as consciousness (contrary to the earlier theories) does not perceive sensations locally, at their seat in the body, it can only know by unconscious inference whatever we do not perceive directly."
[248] Richard Keeler, "The Ophthalmoscope in the Lifetime of Hermann von Helmholtz," *Arch Ophthalmol.* 2002; 120(2): 195.

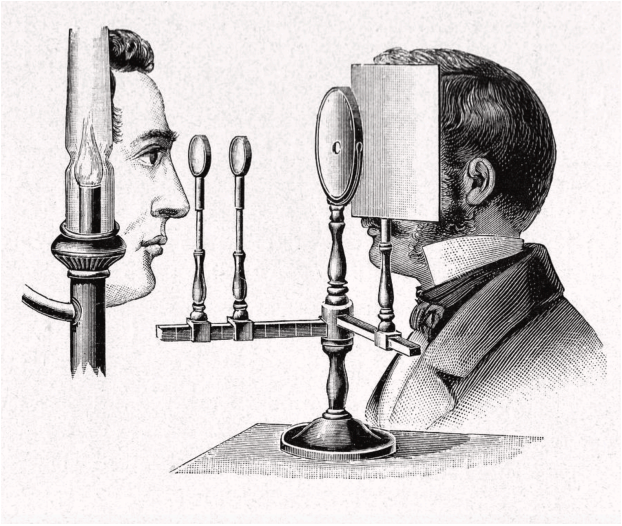


Figure 6: Helmholtz Ophthalmoscope. From Weltall und Menscheit (Universe and Humanity), by Hans Kraemer (ca. 1880).

regions of the retina differentiate nerve fibers so that an "intuition of spatial differences both in kind and in degree" can emerge.[249] Sensations, Helmholtz emphasized, "symbolize their stimuli, but are not direct copies of those stimuli"; the observer infers their meaning through "the unconscious processes of the association of ideas going on in the dark background of our memory."[250] Neither physical nor material—in the sense that they are not themselves extended in space—sensations operate as "signals to consciousness," indicators of changes in the external world rather than images of those changes. Perception thus proceeds through the "qualities of sensations," qualities that "belong only to our nervous system and do not extend at all into the space around us."[251]

On May 25, 1895, returning by train to Vienna after a visit with Wilhelm Fliess in Berlin, Freud began drafting a manuscript that sought to explain how

[249] Koenigberger (1905), 236.
[250] Helmholtz quoted in Gary Hatfield, "Perception as Unconscious Inference," in *Perception and the Physical World: Psychological and Philosophical Issue in Perception,* eds. Dieter Heyer and Rainer Mausfeld (John Wiley & Sons, Ltd., 2002), 124.
[251] Koenigberger (1905), 141.

Fechner's "differences" and Helmholtz's "signs" might be produced, circulated, and discharged as quantities within an efficient psychic economy. He described this new *Project* as an "economics of nerve forces" that aimed "to peel off from psychopathology a gain for normal psychology," translating the findings of *Studies on Hysteria* into a language legible to neurologists in his professional circle.[252] Dubbed "*the* great Freudian text on the ego" by Jean Laplanche, the *Project* sought to elevate empirical psychology to the legitimacy of a natural science by defining the ego in terms of two overlapping frameworks: the "topographical" model of the *neurone* and the "economic" model of *quantity* (the sum of excitation).[253] Just a year earlier, in "The Neuro-Psychoses of Defence," Freud proposed that defense neuroses work to "resolve the contradiction between [an] incompatible idea [in particular, a sexual one] and [the] ego by means of thought activity," rendering a "powerful idea" weak by robbing it of affect—the sum of excitation—and diverting that energy to "another use."[254] The *Project* attempted to ground this hypothesis in a material topography of nervous structures.

Freud opens the *Project* with a question that reveals the ambition of his undertaking: "Where do qualities originate?" Not, he insists, in the external world,

---

[252] Sigmund Freud, "Project for a Scientific Psychology" (1895), vol. 1, *Standard Edition,* 129.
[253] Jean Laplanche, *Life and Death in Psychoanalysis*, trans. Jeffrey Mehlman (Baltimore: The Johns Hopkins Press, 1970), 54. This ambitious project assumed a broad set of analogies: "Neurone and quantity are a new version of what in Descartes, for example, is called figure and movement, or again, in the "physicalist" school of Helmholtz that so influenced Freud, mass and energy."
[254] Sigmund Freud, "The neuro-psychoses of defence: an attempt at a psychological theory of acquired hysteria, of many phobias and obsessions and of certain hallucinatory psychoses" (1894), vol. 1, *Standard Edition,* 45-61.

where natural science recognizes only "masses in motion and nothing else."[255] If psychic life is to be rendered scientific, it must be described in terms of "quantity (Q)" governed by the general laws of motion. Freud therefore posited "neurones" as material particles whose primary function is to discharge excess excitation so as to keep stimulation below a tolerable threshold.[256] Their secondary function, however, requires that quantity ($\tilde{Q}$) be retained; this retention is made possible by resistances that oppose discharge, located at the contacts between neurones, which function as barriers.[257] Because "a psychological theory deserving any consideration must furnish an explanation of memory," Freud further proposed an equivalence between neurones and localized "ideas," such that the neuronic system becomes a mnemic system composed of distinct but similarly constructed units connected through an intervening substance and terminating upon one another.[258] Excitation travels along "cell-processes" (dendrites) and is discharged through the "axis-cylinder" (axon), forming conductive pathways through which nervous activity is transmitted and retained.

Thus to ground memory in physiology, Freud distinguished three functional systems of neurones. The φ-neurones are permeable: they offer no resistance to quantity and retain nothing. They receive exogenous excitation but are governed entirely by the principle of inertia, so that "each time an excitation has passed, [they] are left in the same condition as they were before." The ψ-neurones, by contrast, resist discharge and retain quantity; they "may be left in a modified condition after

[255] Freud, (1895), 369.
[256] Ellenberger (1970), 477-478.
[257] Freud, (1895), 355-359.
[258] Ibid., 358.

each excitation, and thus afford a possibility of representing memory." Regulated by the principle of constancy, they become the relatively impermeable "vehicles of memory and presumably, therefore, of psychical processes in general." Yet Freud undermines his own schema even as he formulates it, conceding that "morphologically (that is histologically) nothing is known in support of the distinction… If all this is true, we should not have invented the two phi and psi, we should have found them already in existence."[259]

The opening section of the manuscript shows Freud struggling—and ultimately failing—to answer his guiding question: "Where do qualities originate?" He moves through a chain of tentative conjectures. Might qualities arise in the φ-system? "Not possible," he reasons, since these receptors must remain unchanged by quantity in order to receive new stimuli. In the ψ-system, then—the apparatus of "reproducing and remembering"? Yet, he concedes, this domain is "without quality. Remembering brings about *de norma* nothing that has the peculiar quality of perceptual quality."[260] To escape the impasse, Freud posits a third system of ω-neurones, whose task is to be "excited along with perception, but not along with reproduction," and whose states of excitation give rise to qualitative experience—that is, to conscious sensation.[261] "No attempt, of course, can be made to explain how it is that excitatory processes in the ω-neurones bring consciousness along with them," Freud concludes, leaving the theoretical muddle unresolved.[262]

---

[259] Ibid., 360-161.
[260] Ibid., 369-370.
[261] Lothane (1998), 55.
[262] Freud, (1895), 369-470.

Freud was acutely aware that his attempt to ground psychology in the physiological discharge of energies rested on uncertain foundations.[263] His wider academic and clinical milieu, however, helps explain why he was drawn to hypothesizing neural mechanisms that could render symptoms as measurable quantities. As Nima Bassiri notes in his critical account of the late-twentieth-century "neuroscientific reenergizing of psychoanalysis," Wilhelm Griesinger's dictum—"Mind diseases are brain diseases"—captured the neurological orientation shaping psychiatry in Freud's time.[264] Within this paradigm, the diagnostic method of localization predominated: a "dominant and multifaceted approach to nineteenth-century brain research" that sought to situate sensory, motor, and even complex mental processes within discrete anatomical regions.[265]

John Forrester likewise traces the authority of localization theory to mid-nineteenth-century phrenology and to Paul Broca's celebrated correlation of focal brain lesions with specific speech deficits in aphasia.[266] Freud resisted this framework. Observing that in aphasia "there is not necessarily a close relationship

---

[263] Ernst Jones, *The Life and Work of Sigmund Freud Vol. 1* (New York: Basic Books, 1953), 255; Sigmund Freud, *The Complete Letters of Sigmund Freud to Wilhelm Fliess 1887-1905*, ed. Jeffrey Mousaieff Masson (Cambridge: Harvard University Press, 1985), 127-129. After completing *Studies on Hysteria* and steadily decathecting from Breuer as his primary intellectual partner, Freud wrote to Fliess: "I am so deep in the 'Psychology for Neurologists' that it quite consumes me, until I have to break off out of sheer exhaustion. I have never been so intensely preoccupied by anything. And will anything come of it? I hope so, but the going is hard and slow." After abandoning the manuscript for the rest of his living career a month later, he concluded: "I can no longer understand the state of mind in which I hatched the [project of scientific] psychology and cannot fathom how I could have burdened you with it. I believe you have been too courteous; I now see it as a kind of absurdity."

[264] Nima Bassiri, "Freud and the Matter of the Brain: On the Rearrangements of Neuropsychoanalysis," *Critical Inquiry* 40 (Autumn 2013), 95.

[265] Ibid.

[266] John Forrester, *Freud and Psychoanalysis: Six Introductory Lectures*, ed. Lisa Appignanesi (Polity Press, 2023), 36. Broca had argued that impairment to the facilities of speech was due to a malfunctioning region in the brain designated to control the speech function.

between extent of damage and severity of symptoms," he argued for a multifunctional brain organized through networks of neurones rather than fixed centers.[267] The alternative to localization was dynamic variation: behavioral pathology could be understood as quantitative changes in the physical state of nerve cells and their relations. A clinical-descriptive method would thus register differences in excitation—whether originating in external stimuli or arising from internal bodily sources that threatened the organism's inertial equilibrium and produced displeasure. In this view, the brain functioned as a single dynamic system in which anatomy receded in importance. As Forrester unambiguously put it: "We might draw an analogy with computer systems; it doesn't matter to us exactly where software, or where a particular database, is loaded onto our hard drive so long as it is able to perform its function. Freud had a kind of 'computer hardware' view of the brain it exists for software to be loaded onto it and to facilitate various functions."[268]

Subsequent psychoanalytic readers after Freud would return to this abandoned neurological sketch to reinterpret its conceptual stakes. In Seminar XVI "From an Other to an other," delivered in the aftermath of May 1968, Jacques Lacan read the theoretical impasses of the *Project* as an attempt to describe the unconscious economy as a system of neurones with "the capacity to retain and at the same time remain receptive." Freud's construction of the ψ-neurone system, Lacan emphasized, was "regulated by repetition," producing perception and identity through an economy

[267] Ibid.
[268] Ibid., 37.

of signs "with its own form of equilibrium."[269] The mnemic system thus appeared paradoxically as a memory that inscribes nothing qualitative in isolation; memory emerges instead through the difference between two paths of neuronic motion, a bifurcation. In Freud's own terms: "It is therefore more correct to say that memory is represented by the *differences* in the facilitations between the ψ-neurones."[270] Having already, in the 1950s, drawn on cybernetics and information theory to rethink the Freudian ego, Lacan returned to the claim that "from Freud's standpoint, nothing about representation is tenable except what… is linked to a structure that is made up of frameworks and networks located altogether outside of the circuit of the subject."[271]

Laplanche likewise emphasized in 1970 that one need "barely modify or interpret" Freud's mnemic system—where each memory forms a unique constellation within a total context—to see in it "a kind of electronic machine, a computer functioning according to the principle of binary notation."[272] By construing the ψ-neurone system as a closed circuit regulating the pleasure principle through homeostasis, Freud described a psyche that "learn[s] to tolerate a store of quantity (Q~) sufficient to meet the demands for specific action."[273] The unconscious thus appears as a governor of signs: a system that selects among pathways, facilitating one

[269] Jacques Lacan, *The Seminar Book XVI: From an Other to an other*, trans. Bruce Fink (Polity Press, 2023), 221.
[270] Freud (1895), 355-359; Laplanche (1970), 56-57.
[271] Jacques Lacan, *The Ego in Freud's Theory and in the Technique of Psychoanalysis, 1954-1955,* Seminars of Jacques Lacan, Book II, ed. Jacques Alain Miller, trans. Sylvana Tomaselli (New York: W.W. Norton & Company, 1991).
[272] Laplanche (1970), 57.
[273] Freud (1895), 384.

route while establishing barriers to another within a circuit structured by bifurcation and choice.

Just as Fechner's psychophysics disclosed a procedure for measuring difference that exceeded the language of energy in which it was framed, Freud's *Project*—his abortive effort to secure psychoanalysis as a neurology—can be read, in retrospect, as isolating the neuron as a unit defined by transmission. Viewed through the later vocabulary of cybernetics, the Freudian "neurone," which inscribes engrammatic traces through patterned pathways of circulating quantity, describes a discrete unit whose significance lies in the differences it registers and relays: the gap between symptom and what it symbolizes. What could not be sustained as a literal doctrine of indestructible "motive force" persists, instead, as a formal logic of abstract differentiation—legible only in a symbolic or, in Warren McCulloch's acerbic phrasing, as "an easy prop for… nothing other than the signifying chain in its most elementary form."[274] Read in this way, Freud's failed neurological model does not anticipate cybernetics by accident; it situates the psyche at the threshold where quantitative excitation is transmuted into signification and where transmission becomes inseparable from interpretation.

The Hysteric's Symbolism

Eckstein's case appears in the second section of the *Project*, under the heading "Psychopathology," where Freud suspends his wild theorizing of neural mechanics to

[274] McCulloch (1951), 42.

confront the clinical problem posed at this chapter's opening: how a memory returns with compulsive force, as in the scene that haunted the shop. He begins by asking after the genesis of hysterical compulsion: "Under what conditions does a pathological formation of a symbol such as this occur (or conversely, a repression)? What is the operating force concerned? What is the state of the neurones of an excessively intense or of a repressed idea?"[275]

In what he calls two scenes "typical of repression in hysteria," Freud observes that the memory of running from the shop at age twelve is linked to an earlier episode. At eight years old, she twice entered a confectioner's shop, where the shopkeeper "grabbed at her genitals through her clothes." After the second visit she avoided the shop altogether and later reproached herself for returning after the first time, as though she had provoked the assault; the "bad conscience" that oppressed her, Freud notes, could be traced back to this earlier experience.[276] Freud identified the molestation at eight as the originating event, yet it was the later, nontraumatic memory at twelve that precipitated her symptoms. The earlier experience became traumatic only *nachträglich*—in deferred action—through the laughter of the shopkeepers functioning as a *symbol* of the shopkeeper's grin as he grabbed her through her clothes (figure 7).[277]

Eckstein's phobia at twelve—"a pathological formation of a symbol such as this"—emerged from the earlier molestation at eight and exemplifies what Teresa de

---

[275]

[276] Freud (1895), 410-411.

[277] Bassiri (2013), 103.

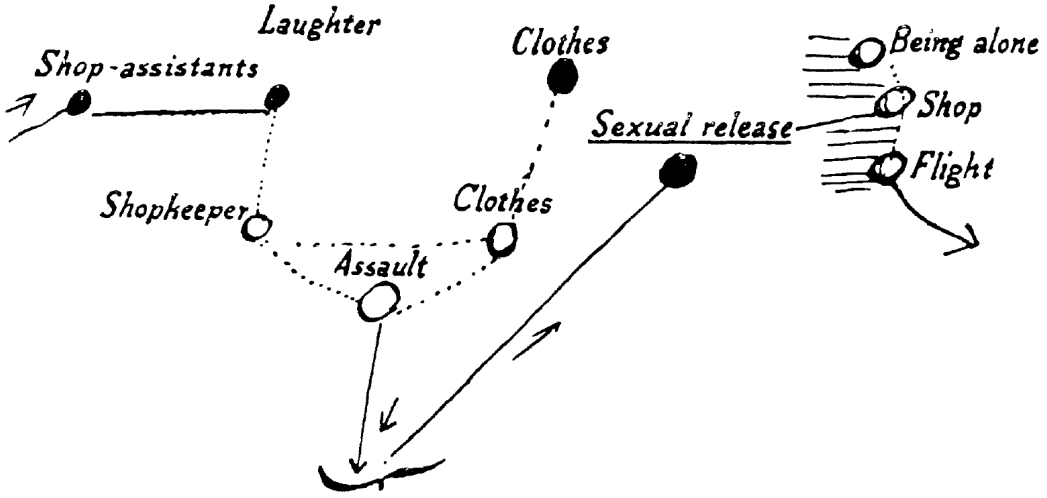


Fig. 16

Figure 7: Emma's "scene," a map of Eckstein's case study. Featured in Freud (1895), 412 and reproduced Bassiri (2013), 105.

Lauretis calls the "double temporality" of the return of the repressed: the memory becomes traumatic only *nachträglich*, when its sexual meaning is later understood and the excitation it carries becomes intolerable to the ego.[278] The assault itself constituted a comprehensible cause for distress, yet because the memory was repressed, its affect returned indirectly. The later scene in the shop supplied the conscious trigger: the clothes and the laughter (A) functioned as a substitute, a *symbol*, for the repressed memory of the assault (B)."[279] Before analysis, this excessively intense idea intruded repetitively into consciousness, producing anxiety whenever she entered a shop alone without knowing why. Analysis thus required

[278] Teresa de Lauretis, *Freud's Drive: Psychoanalysis, Literature, and Film* (New York: Palgrave Macmillan, 2008), 6-7. The concept of *Nachträglichkeit (*"apres-coup" for Lacan and Laplanche) will be explored further throughout the section.
[279] Ibid., 407-409.

tracing the symptom back to the latent scene, where the libidinal excess attached to the earlier event organized the later fear.

Here Freud offers a tentative answer to the problem of memory. Between one "excessively intense idea" as an energetic force and another latent idea as its repressed meaning lies the hysteric's capacity for symbolization, which, he notes, "behaves differently." Emma remains unaware that A's association with B—the repressed and unconscious scene—is what repeatedly generates bodily sensation and the buildup of psychic tension, as "the symbol" (A) wholly takes the place of "the thing" (B). Memory thus takes the form of differential substitution: one idea stands in for another, its efficacy sustained by the associative pathways through which excitation circulates. The laughter and the clothes must recur as signs to sustain the memory, illustrating Freud's claim—paraphrased by Bassiri—that memories are not inscribed once but laid down repeatedly through retroactive translation and retranscription.[280]

Freud marks this mnemonic mechanism with the striking language of "lying." Emma's "first lie" entails ascribing reality to the imagined "excessively intense idea," charging representative propositions with wishful objectivity.[281] Yet he insists that this process is neither unique to hysteria nor pathological in itself; it describes the ordinary operation of symbols: ""A soldier will sacrifice himself for a piece of

[280] Bassiri (2013), 106.
[281] Freud (1895), 407.

coloured cloth on a pole," he writes, "because it has become the symbol of his native country; and no one considers this neurotic."[282]

In terms almost identical to Freud's formulation, Gregory Bateson later described the "metaphor that is meant" as fundamental to communication and to phenomena such as dreams, play, and fantasy. In the overlapping domains of art and religion—*and*, notably, in the psychogenesis of schizophrenia—humans live by symbols felt to exceed their material form: the flag for which one will die, the sacrament experienced as more than "an outward and visible sign."[283] Such instances belong to "digital communication," in which discrete signs stand in for absent referents through conventional substitution.

The majority of mammalian communication, by contrast—where symbolic language is absent—"analogic kinesic signals" prevail, such that magnitudes correspond, directly or inversely, to the "magnitudes of the *relationship* being communicated." Bateson illustrates the distinction with a humorous example: if one billiard ball strikes another, the motion of the second is energized by the impact of the first; if one kicks a dog, the animal, too, may indeed "travel along a Newtonian trajectory if kicked hard enough, but this is mere physics," but what matters is that "he might turn and bite."[284] He might even develop a later, more long-term habit of cowering when any human stands over him or barking uncontrollably in confined

---

[282] Ibid.
[283] Gregory Bateson, "A Theory of Play and Fantasy," *Steps to an Ecology of Mind* (New York: Ballantine Books, 1972), 183.
[284] Gregory Bateson, "The Group Dynamics of Schizophrenia," *Steps to an Ecology of Mind* (New York: Ballantine Books, 1972), 229.

spaces, responses energized not by the force of the blow but by its traumatic registration in his metabolism and by the relational meaning of the act.

Emma's neurosis, however, reveals that symbolization is not confined to the digital substitution of sign for referent; the somatic excitation and anxiety attach to the present scene excitation attaches to the present representation but is energized by the retroactively constituted meaning of the earlier scene. The hysteric's "first lie" is, indeed, a mode of symbolization that "behaves differently,"—one that "translate[s]—according to specific laws of transposition—her desire into reality," and in so doing discloses a general structure of repression and defense.[285]

But Emma was not the only one harboring a lie. By the time she was thirty, she nearly bled to death on the operating table under Freud's supervision. At his recommendation she underwent Wilhelm Fliess's treatment for "nasal reflex neurosis," a procedure that removed the turbinate bone to cauterize the nasal cavity. Fliess—an ear, nose, and throat surgeon who claimed a pathophysiological link between the nasal mucosa and the female genital organs—promoted the operation as a cure for hysteria, locating mental disorder in reflex pathways of the nose.[286] Fourteen days later Freud confronted "persistent swelling, going up and down like an avalanche" and "purulent secretions" of pus; a second surgeon discovered and

[285] Laplanche (1970), 34. Laplanche further articulated this precise problem as at the crux of the reformulation of seduction theory, "a truly *general* theory of the origins of the repressed unconscious."
[286] Zucker A, Wiegand D.,"Freud, Fliess, and the nasogenital reflex: did a look into the nose let us see the mind?." *Otolaryngol Head Neck Surg*. 1988 Apr;98(4): 319-22.

removed a mass of gauze Fliess had left in the wound, triggering a massive hemorrhage.[287]

Eckstein's nasal passages were permanently damaged and the left side of her face disfigured. "So this is the strong sex," she reportedly quipped. "So this is the strong sex," she reportedly quipped at the sight of her doctor in a post-operative state.[288] In the anxious correspondence that followed, Freud urged Fliess to absolve himself—"It is now about time you forgave yourself the minimal oversight"—assuring him that "she, Eckstein, is doing well; she is a very nice, decent girl who does not hold the affair against either of us and refers to you with great respect."[289] Months later he reconfigured the catastrophe as a hysterical manifestation, suggesting she had "bled out of longing" for him in a reaction of transference. Nearly two years after the incident, around the time Eckstein would begin her own work as an analyst, Freud reminded Fliess again: "As far as the blood is concerned, you are completely without blame!"[290]

Freud's maneuver here was not a simple confusion of symbol and reality; he constructed a supplementary reinterpretation with the authority of the physician. Confronted with a material trauma caused during treatment, he folded the real event

---

[287] Sigmund Freud, *The Complete Letters of Sigmund Freud to Wilhelm Fliess, 1887-1904*, trans and ed. Jeffrey Moussaieff Masson (Cambridge: Harvard University Press, 1985), 117. Freud himself had undergone Fliess' procedure a few months prior to treat his neurotic depression.
[288] Ibid. See also: K. R. Eissler, "Preliminary Remarks on Emma Eckstein's Case History, *Journal of the American Psychoanalytic Association* 45 (Dec. 1997): 1303–5.
[289] Ibid.
[290] Freud's guilt surrounding the Eckstein case would, according to Max Schur, also play a main role in the interpretation of his famous "Irma's injection" dream six months later, marking the beginning of his period of self-analysis leading up to the publication of *The Interpretation of Dreams* in 1899.

of the surgical failure and hemorrhage into his interpretive framework, reconfiguring the bleeding as a hysterical manifestation. Emma's so-called "first lie" was answered by a second: interpretation itself could lie, displacing material truth by converting clinical error into symptom and physical injury into psychic deceit. What is exposed here is the symbolic power of interpretation itself—its capacity to convert fact into meaning and thereby safeguard theoretical coherence and authority at the expense of the body that contradicts it.

The authority psychoanalysis claimed over interpretation did not end with Freud; when mid-twentieth-century cyberneticists formalized cognition as signal processing, they inherited not its metapsychology but a problem they sought to resolve, displacing its ambiguities with a model of communication designed to eliminate them. Freud's *Project* had already sketched a divided psychical economy in which excitation could overwhelm the system in a "wishful cathexis … to the point of hallucination," while, in Lacan's words, the ego's regulatory function enabled "a correct employment of the indications of reality."[291] What later engineers discerned in this constitutive conflict was not a drama of desire but a problem of reliable transmission: how to secure the fidelity of the signal. Bracketing the distinction between conscious and unconscious operations, a logic of binary decision—an "all-or-none law of nervous impulse"—reduced mental activity to bivalent states and abstracted it from the contingencies and traumas of the living body. The symptomatic substitute Freud analyzed as symbol was recoded as interference—noise in the

[291] Lacan (2023), 165-166.

channel—an obstacle to the design of "the simplest cybernetic circuit" with a "memory of a dynamic kind, not based upon static storage but upon the travel of information around the circuit."[292] The engineering goal, poised "between psyche anatomized and psyche synthesized," was the obsolescence of interpretation itself—as talking cure, as symbolic authority, as meaning-making. Its advantage, scalability coupled with high efficiency, lay in securing communicative reliability by eliminating distortion.

Conclusion

There is no straight lineage between the distinct histories of psychoanalysis and cybernetics. Read retroactively from the vantage of an information age structured by networked computation and surveillance infrastructures, however, they converge on a shared problem: how to conceive the relations between memory and regulation, control and communication, thing and symbol. Freud's turn-of-the-century effort to ground the psyche in neurology exposed a persistent analogy between the working machine and the thinking—communicating, remembering, learning—mind, rooted in nineteenth-century thermodynamics and physiological models of equilibrium and energetic economy. In Freud's work these principles became a model of psychic regulation; by the mid-twentieth century they reappeared, transformed, within cybernetics as problems of control, communication, and feedback. At the point where energetic explanation falters, informational logics begin to take shape.

[292] McCulloch (1951), 42.

The thermodynamic image returns belatedly—not as an event exhausted in its moment, but as a difference that reappears when conditions permit its registration. In a steam engine, motion depends upon a temperature differential; once equilibrium is reached, movement ceases. Freud's economic model of the psyche posits a comparable tendency toward discharge and quiescence: excitation seeks reduction, tension drifts toward zero, and associative chains may be abandoned in the lure of inert stability. Yet life persists only through a countervailing process. Ego formation depends upon a negentropic modification of this primary tendency—the production of organization as a "secondary function imposed by the exigencies of life."[293] To persist, excitation must be transformed rather than eliminated: broken into discrete elements, displaced, condensed, and symbolically recomposed. Psychic stability thus lies not in the reduction of tension but in the patterned circulation of differences. From this vantage, the ego appears less as a hydraulic regulator than as a system that preserves structure by modulating flows of difference and thereby acquiring the *capacity to behave differently in time*—an operation of recursive modulation through which structure is maintained and meaning emerges.

By the middle decades of the twentieth century, the problem first posed in thermodynamics and psychophysics—how systems maintain order amid disturbance, how quantitative shifts register as qualitative change—had been reformulated as a general theory of communication and control. What Weber and Fechner rendered measurable as differential thresholds, and what Freud sought to model as the

[293] Freud (1895), 358.

circulation and binding of excitation, now appeared as the regulation and transmission of differences within self-correcting, remembering, and learning systems. Within wartime research infrastructures and early Cold War command-and-control environments on both sides of the Atlantic, the language of energy yielded to that of information: stability depended on detecting deviation, transmitting signals, and adjusting behavior in time.

At this intersection of physiology, psychology, logic, mathematics, anthropology, and the emerging information sciences, the cybernetic conception of learning took shape as an experimental and theoretical object. In 1943, Warren McCulloch and Walter Pitts formalized nervous activity as logical circuitry, proposing artificial neurons organized in networks capable of storing and comparing inputs and producing decisions. Their model confronted a paradox already latent in earlier mathematical logic: rendering thought computable required reducing cognition to discrete logical states, even as meaning arises from relations among levels that formal systems struggle to contain. Paradox thus became a problem of formalization: recursive processes had to be stabilized so that contradiction would not precipitate systemic collapse. The neural network offered one solution. Constraining activity to discrete states and regulated pathways, it eliminated contradiction and semantic excess: memory became the persistence of circuit states, learning the adjustment of connection weights in response to input, and intelligence the capacity to discriminate and act upon differences through recursive feedback.

Alongside these efforts to engineer reliable logical networks, Gregory Bateson's investigations of communication in animals and human interaction foregrounded the contextual and paradoxical dimensions of learning as indispensable to any account of mind. The mid-century cybernetic moment thus marks both the consolidation of an informatic paradigm and the persistence of a problem it could not eliminate: the instability of meaning generated across recursive levels of communication. The chapter that follows turns to this juncture, tracing how attempts to formalize learning in neural networks inherited the very instabilities they sought to exclude—instabilities that continue to mark the limits of machine intelligence.

CHAPTER 3

NEUROTIC NETWORKS IN MID-CENTURY CYBERNETICS

For all the elegance of their logical scaffolding, cybernetics and AI have continually hit the same wall: a reductionist tendency to treat learning as the accumulation and processing of information, leaving behind the paradoxes of self-reference and meaning. This tendency has not disappeared—it has been mechanized. Luciana Parisi's proposition that the "synthesis of logic and calculus in automation" has advanced a "crisis of reason" finds a perfect laboratory in today's technoscientific milieu, where AI systems like large language models (LLMs) go beyond formal logic to expose the rickety foundations of agency and governance in the digital age.[294] The result is a feedback loop of economy, politics, and industry where "a mediatic infrastructure of statistical modelling, pattern recognition, data mining, knowledge discovery, predictive analytics, self-organizing and adaptive systems" once designed to imitate human thought begin highlighting the contradictions latent in conceptions of human rationality.[295]

In the history of cybernetics and computation, *learning*—in amoebas, analysts, and algorithms alike—has always been less about enlightenment and more about survival in a universe veering toward entropy.[296] Contemporary systems like

---

[294] Luciana Parisi, "Critical Computation: Digital Automata and General Artificial Thinking," 98-99.
[295] Ibid., 90.
[296] French physicist Leon Brillouin, building on the work of Hermann von Helmholtz and Claude Shannon, introduced the concept of "negentropy" at the center of the thermodynamics of informatics and computing, famously asking: "How is it possible to understand life, when the whole world is ruled by such a law as the second principle of thermodynamics, which points towards death and annihilation?"

GPT models embodies this tension in statistical abstraction: vast amounts of human language are compressed, generalized, and transformed into algorithmic rules. Learning, in other words, has morphed into a game of linguistic poker, predicting the future by rearranging scraps of the past and rewarding themselves when they bluff convincingly enough.[297] In the guise of "value creation," as Paul Kockelman has recently observed, these systems turn the semiotic terrain of meaning into a tightly policed playground of predictive reinforcement.[298] A disciplined parade of outputs remains, polished for profit and the mimicry of comprehension while neatly sidestepping interpretive messiness. The “key capacity” of LLMs, in Kockelman’s words, “is also their main limitation: worldlessness.”[299] These systems, “machines that learn to represent representations, but they can neither question nor ground those representations,” brilliantly highlight the unresolved paradox motoring a deeper capitalist drive to render the world legible through computational systems—simultaneously as it erases the interpretive, contingent, and relational dimensions that make knowledge meaningful. The result is a feedback loop where logic, language, and technological implementations endlessly reference themselves, untethered from the lived realities they claim to optimize.

This chapter sets out to explore this ongoing tension of qualifying “learning” in the history of cybernetics and computation since the postbellum mid-century to today's AI-powered present and future. Central to this inquiry is Gregory Bateson’s

[297] David Nofre, Mark Priestley, and Gerard Alberts meticulously traced the origin story of the linguistic conception of computer programming in “When Technology Became Language,” 40-75.
[298] Paul Kockelman, *Last Words: Large Language Models and the AI Apocalypse*, 31.
[299] Ibid.

theorization of learning as a recursive hierarchy, influenced by Alfred North Whitehead and Bertrand Russell's theory of logical types—a framework that simultaneously guided Warren McCulloch and Walter Pitts in building the neural network as a computational metaphor for the mind. By examining the paradox of self-referentiality plaguing computational logic since the *Principia Mathematica,* this chapter seeks to unveil a broader, recurring contradiction in the history of cybernetics: the inability of logical systems (human or machine) to *learn*—to account for their own capacity to evolve—without falling into loops of self-contradiction. In contemporary AI, the quest to render logical paradox obsolete persists in the technocratic race to design intelligent systems that adapt to inputs, "learn how to learn" by revising their frameworks over time, and ultimately generate systems of higher complexity.[300]

Where cybernetic machine "learning" fetishizes interpassivity and alignment,[301] the essence of learning in Bateson's psychiatric and ethological research simultaneously refers to both the process of adapting to and reorganizing reality in complex systems—generating paradox within the very learning structures upholding meaning in time. Learning—in human, animal, and machine alike—unfolds not as the "tyranny of smartness" that Orit Halpern and Robert Mitchell have argued as a pervasive epistemological drive toward optimization and resilience since postwar

[300] For a key reflection on the concept of epistemological recursion stemming from (but not limited to) cybernetics discourse, see Yuk Hui, *Recursivity and Contingency*.
[301] See: Mona Sarkis, "Interactivity Means Interpassivity"; Robert Pfaller, *Interpassivity: The Aesthetics of Delegated Enjoyment*; Slavoj Zizek, "The Interpassive Subject".

American science, politics, and industry.[302] Learning, for Bateson, is fundamentally distinct from "adaptive resilience"—what Halpern and Mitchell helpfully summarize as "a state of permanent management that does away with guiding ideals of progress, change, or improvement."[303] Instead, learning is both a breakdown and reconstruction of knowledge, a plunge into epistemological crisis that demands a painful reckoning with the inconsistences and gaps in the presuppositions on which we build our interpretive frameworks. These frameworks recursively generate and are generated by our perception and knowledge, creating a continuous, often uncomfortable loop of epistemic reevaluation. The experience of having learned from experience thus amounts not to mastery but a trace of having underwent a transformation, indexed by the limits of understanding and the inherent biases that shape the perception of and interaction with the world. In this way, learning may challenge the very structures that uphold our reality.

At this juncture, I begin exploring how Bateson aligns learning with the psychoanalytic drama of transference: a process not of extracting knowledge but of revealing and revising the desires that animate it. At the crux of both is a reverberating question: how does change truly occur in systems, neural or machinic? For example, the analytic encounter pivots on a paradox as confounding as it is generative: self-similar and self-certain knowledge is a closed loop devoid of genuine information. An analysand's speech, taken in isolation, is no revelation to herself—its

[302] Orit Halpern and David Mitchell, *The Smartness Mandate*, 4.
[303] Ibid.

structure is tautological, like a computer program incapable of querying its own code. Yet, when addressed to the analyst, this speech transforms into an act of interpretation. Through fragments of metaphor, free association, and the unwieldy logic of desire, the analysand performs the Sisyphean task of naming motivations that escape conscious grasp. This naming, however incomplete, may conjure new orientations to the world. Bateson crystallized this tension in communication theory: an individual perceives reality only through the narrow lens of their codification system, rendering them blind to the structural scaffolding of that very lens. In the psychoanalytic process generated by Bateson's hierarchy of learning, the transformation of the subject necessitates an intersubjective field, a mediated encounter laden with the potential to reframe codification itself.

Here, the machine metaphor breaks down. A learning model—whether a neural network or an AI system—remains trapped in the logic of its programming. It can optimize, iterate, and predict but cannot transcend its code to reflect on the framework shaping its outputs. The analytic process, by contrast, thrives in the epistemological gaps machines cannot traverse: the moments where unconscious knowledge resurfaces through dialogic engagement and reorients not just what is known but how knowing itself is structured. It is precisely this capacity for higher-order transformation—alien to the machinic paradigm—that psychoanalysis reveals as foundational to what it means to learn from experience.

Vicious Circles

In the course of annual meetings funded by the Josiah Macy Foundation from 1943 to 1954, colloquially named the Macy Conferences on Circular Causal and Feedback Systems, a motley collection of researchers in mathematics, engineering, and neurobiology as well as in psychiatry and the social sciences converged.[304] For those with a humanistic focus on interactive and biological systems, cybernetic principles of circular causality and feedback laid a conceptual foundation for understanding "codification systems" conducting information in order to define reality and adapt to change. These participants were not content with finding better ways to organize systems or produce more efficient machines. As Heinz von Foerster, the affable impresario of the Macy Conferences, reflected: " if the processes by which knowledge is acquired, *i.e.,* the cognitive processes, are interpreted as computational algorithms," it would take nothing short of a revolution to account for "computations that compute computations… recursive computations with a regress of arbitrary depth."[305] Bateson also noted that codification systems, whether human, machine, or ecosystem, often resist alteration, leading to paradoxes when systems attempt to reflect on or modify their frameworks.[306] The technical aspects of informatics and control extended to reframe, in the language of communication theory, how we understood the relationship between events in the world and how they became symbolic information. Cybernetics, in other words, aspired to change the very foundation of how we conceive of systems and how knowledge operates within them.

---

[304] For complete proceedings of the Macy Conferences, refer to: ed. Claus Pias, *Cybernetics: The Macy Conferences, 1946-1953*.
[305] Heinz von Foerster, "Cybernetics of Epistemology," 229.
[306] Gregory Bateson, "Information and Codification: A Philosophical Approach," 169-170.

The group, however, quickly found themselves entangled in an "orthodox apprehension." When the propositional variables in a system changed due to self-reference, it led to paradoxes of abstraction that muddied the very language used by the scientists.[307] The introduction of *time* on the one hand and *self-reference* on the other—both theoretically meant to account for changes in the codification systems—complicated matter further, causing systems to transform in ways that failed to account for the interactions at play logically. For all its lofty ambitions, cybernetics quickly became its own worst enemy, generating models attempting to describe change and behavior in a system that itself could not account for the very transformations it was supposed to explain. Cybernetics discourse thus began to resemble a sort of systemic neurosis: a recursive loop that could never fully articulate the next level of understanding without falling into contradiction.

These "second-order" generation of cyberneticians—figures like Margaret Mead, von Foerster, Humberto Maturana, and Francisco Varela—began addressing such shortcomings, introducing concepts like reflexivity, autopoiesis, and recursive self-organization.[308] These terms were attempts to move beyond the limitations of the initial cybernetic framework by embracing the complexity and interconnectedness of systems. However, as Luciana Parisi observes, second-order cybernetics advanced a "crisis of reason" due to its entanglement with neoliberal governance and algorithmic automation, which she sees as central to the modern "post-truth" epistemology.[309] The

[307] Heinz von Foerster, "Disorder/Order: Discovery or Invention?," 249.
[308] See: Margaret Mead, "Cybernetics of Cybernetics," 1-11; Hayles, *How We Became Posthuman*, 131-159.
[309] Luciana Parisi, "Critical Computation," 98-99.

paradox of change, as conceived by these thinkers, was that they could only understand the concept as a repetitive loop, a cycle of self-regulation within an even broader system of rules and procedures. Drawing on Hayles' insights, Parisi argues that institutional applications of such recursive feedback loops ultimately restructured cybernetic concepts into a model of governance pervasive in contemporary technoscience. Self-referential learning systems proposed mechanisms of social adaptation and control that conveniently enshrined computation and automation as the ultimate epistemic authority, ultimately advancing the "techno-capitalist subsumption of machine thinking and the dominance of the data-driven order."[310]

However, the recursive quagmire of logical paradoxes that cybernetics and computer science stumbled upon in the mid-century was no breakthrough. It was an inheritance of unresolved problems that had haunted philosophy and mathematical logic for over a century—at least since Alfred North Whitehead and Bertrand Russell, in their monumental *Principia Mathematica* (1910–1913), revealed paradox (a "set of all sets that do not contain themselves" that collapses into contradiction by including or excluding itself) as a symptom of profound instability undergirding mathematical logic.[311]

The theory of logical types, developed by Russell and Whitehead in *Principia Mathematica*, sought to impose a strict hierarchy on mathematical and logical

---

[310] Ibid., 115.

[311] Jacques Lacan, *Seminar XVI: From an Other to the othe*r, 45. "This is Russell's paradox… The class of all the catalogs that don't include themselves cannot in any way be situated in the form of a set, precisely because it cannot figure in the elements that are already inscribed in this set. It is distinct from them. There is no way to include in a set what you might extract from it by designating it as a set of elements that do not include themselves. … If it is outside, it is inside. If it is inside, it is outside."

systems to “avoid illegitimate totalities” and contradictions inherent in self-referential reasoning (the "Vicious-Circle Principle"), thus safeguarding mathematical rigor.[312] They formalized a method of reinforcing the strict boundaries of rationality—inadvertently providing, for the likes of von Neumann, a symbolic blueprint for control systems that deliberately conserve and repeat their contradictions to appear capable of self-regulation. But the effort to ground mathematics in pure logic ultimately exposed the limitations of logic itself: far from eliminating paradox, the theory of types generated contradictions of its own, as the framework required principles that were not strictly logical but rather structural impositions.

For instance, Russell and Whitehead introduced the “axiom of reducibility,” positing that an equivalent first-order predicate exists for any predicate or propositional function of higher order.[313] The equivalence facilitated the symbolic representation of class concepts through propositions. In other words, they rendered complex logical structures into manageable “symbolic forms,” asserting that “any combination or disjunction of predicates is equivalent to a single predicate.”[314] The logician could treat higher-order predicates (those defining classes or sets of objects) as equivalent to first-order predicates (which refer directly to individual objects), sidestepping the infinite regress and paradoxical loops exposed by the vicious-circle

[312] Whitehead and Russell, *Principia Mathematica Vol. 1*, 30.
[313] Criticized widely by Norbert Wiener, Ludwig Wittgenstein, John Von Neumann, Kurt Godel, and Willard Quine, the axiom of reducibility seemed less like a fundamental logical principle and more like a pragmatic salve to ensure the system's functionality. Russell, too, eventually came to see the pretense of the axiom of reducibility as an expedient rather than an elegant solution and that it undermined the foundational rigor he and Whitehead had sought to establish.
[314] Whitehead and Russell, *Principia Mathematica Vol. 1*, 58-59.

principle.[315] This logical expedient, however, reduced relations between propositions to functions, a maneuver that, while symbolically elegant, still failed to resolve contradictions in practice. The collapse of higher-order predicates into first-order equivalents allowed logical systems to function, but only by introducing self-referential paradoxes, where predicates and their truth functions became indistinguishable. Though foundational to their framework, the axiom of reducibility was itself an utterly non-logical imposition—an inadvertent admission that logic did not seem to escape the very paradoxes it sought to eliminate.

Ultimately, Whitehead and Russell's program to render the world legible through propositions revealed that any self-referential system—whether in logic, computation, or the market—inevitably confronts contradictions that agents of the system cannot adequately resolve within its framework.[316] Their central revelation was, tacitly, a profound warning about the fragility of systems built on unchecked assumptions of totality. The proposed solution—an "extensional hierarchy" that stratified logic into orders and types—was meant to prevent lower-level terms from contaminating higher-order symbolic abstractions by assuming propositions can be

---

[315] Quoted in Jean van Heijenoort, *From Frege to Godel: A Source Book in Mathematical Logic, 1879–1931*, 165: "… a predicative function of an individual is a first-order function and for higher types of arguments, predicative functions take the place that first-order functions take in respect of individuals. We assume then, that every function is equivalent, for all its values, to some predicative function of the same argument. This assumption seems to be the essence of the usual assumption of classes [modern sets]."

[316] Alfred North Whitehead and Bertrand Russell, *Principia Mathematica Vol. 1*, 37-38. The vicious cycle principle is articulated by Whitehead and Russell as such: "Whatever involves all of a collection must not be one of the collection." The inverse also holds: "If, provided a certain collection had a total, it would have members only definable in terms of that total, then the said collection has no total."

"extended by analogy" between types.[317] Yet this hierarchization implied that any system aspiring to encompass totality generated a recursive flaw built into its architecture; the system inevitably created its own blind spots and contradictions. The "stable truth value" of symbolic simplifications, as Whitehead and Russell demonstrated, were gained in reducing higher-order relations to first-order terms, trading knowledge about the essential limit of mathematical logic for unbridled symbolic capability.[318]

Cybernetics—a group, a moment, and a discourse aimed at codifying propositional laws of communication and control—endeavored to reduce the world's complexities to logical or computational systems, whether through the axiom of reducibility or later attempts to formalize self-referential systems in artificial intelligence. Grappling with the paradoxes mathematical logic sought to resolve, then, the discourse inevitably inherited their contradictions. The enterprise, caught in a double-bind, *had* to unravel—akin to when mechanical systems attempt to conflate distinct levels of abstraction and ultimately result in a kind of cognitive breakdown.

The infinite regress of propositional logic furher appears to model the paradoxes inherent in the fabric of capitalist modernity: the recursive absurdities of economic rationalism and its promises of seamless control and optimization. The market, too, constantly generates rules for its legitimacy, even as those rules collapse under the weight of their self-referential claims to universal applicability. According

[317] Alfred North Whitehead and Bertrand Russell, *Principia Mathematica Vol. I1*, x-xi. "From all these various modes of extension we 'see' that whatever can be proved for lower types, whether functional or extensions, can also be proved for higher types."
[318]Whitehead and Russell, *Principia Mathematica Vol. I1*, xiii.

to Phillip Mirowski, to sidestep this entanglement in self-contradiction that plagued "the logicist program of Russell and Whitehead," modern economics began to depend upon a "process of abstraction" akin to David Hilbert's axiomatic method of mathematical investigation that did not hinge on the "dogma" of logical assailability.[319] The irrationality of markets, in other words, found a built-in adaptive mechanism in "the one characteristic practice of mathematicians which was self-justifying, and could be projected outward onto empirical domains without becoming complicit in their uncertainties, paradoxes, and confusions."[320] By the mid-century, then, Alan Turing's Universal Machine did not solve these logical paradoxes so much as to sweep them under the rug of computation—turning what had once been an epistemological crisis into a technical problem fit for machinic rationality. In this sense, cybernetics and its digital progeny transformed logical paradox into marketable pragmatism, aligning neatly with the demands of postwar capital.

Neurotic Networks

The paradoxes of symbolic logic found their most eager pupil in a 12-year-old boy from working-class Detroit, who, looking for an escape from the chaos of his home and school life, devoured the three volumes of *Principia Mathematica* at his local public library. For Walter Pitts, finding himself amid a life filled with an authoritarian father, poverty, and social isolation, the logic of *Principia* was a revelation. The book gave him "a metaphysical insight that logic rules the universe,"

[319] Phillip Mirowski, *Machine Dreams*, 106-107.
[320] Ibid.

and, in a corollary that would shape his future, "he felt that ego—and his ego in particular—needed to be erased in order to achieve an understanding of the world."[321] He soon spotted inconsistencies in the text and promptly wrote to Bertrand Russell to outline his corrections. Russell responded and eagerly invited Pitts to study under him at Cambridge. This moment would underpin the theory of nervous nets that became his life's work: the groundbreaking idea that if logic could explain the behavior of signals in computers, perhaps it could also decode the patterns of nervous activity in the brain.

Lacking a comprehensive biographical study, scholars and science writers have spun Walter Pitts' life into a mythic narrative, one marked by tragic irony.[322] The oft-repeated tale unfolds as such: He soon spotted inconsistencies in the text and promptly wrote to Bertrand Russell to outline his corrections. Russell responded and eagerly invited Pitts to study under him at Cambridge. unable to accept Russell's offer due to his young age, Pitts began his intellectual pursuit across disciplines as an audacious autodidact. Despite a lifelong phobia of librarians—so severe that it led him to steal books rather than check them out—he mastered Greek, Latin, Sanskrit, and a seemingly endless array of subjects: history, literature, geology, agriculture, and music, all with an encyclopedic rigor. By age 15, he eloped with his true love—mathematical logic—running away from home upon hearing that Russell had moved

[321] Neil R. Smalheiser, "Walter Pitts," 218.
[322] See: Amanda Gefter, "The Man Who Tried to Redeem the World with Logic"; Solomon Feferman (ed.) et al., *Kurt Godel: Collected Works*, 155-160; Tara Abraham, *Rebel Genius: Warren S. McCulloch's Transdisciplinary Life in Science,* 75-100; Lily E. Kay, "From Logical Neurons to Poetic Embodiment of Mind: Warren S. McCulloch's Project in Neuroscience," 597.

to the University of Chicago, where he slipped into Nicolas Rashevsky's mathematical biology group and snuck into Rudolph Carnap's lectures. One day, without ceremony, Pitts walked into Carnap's office and pointed out the flaws in his recent work, *The Logical Syntax of Language*. Carnap, after a few months of hunting "high and low for the 'newsboy who knew logic,'" eventually offered him a "menial job" at the university, thereby keeping him under his tutelage.[323] In Russell's lecture hall, Pitts also met Jerome Lettvin, a then-medical student, who would become his intellectual companion and closest friend. Lettvin introduced Pitts to Warren McCulloch and, eventually, to Norbert Wiener.

In 1942, McCulloch invited Pitts and Lettvin to live with him, his wife Rook, and their four children at their farmhouse in Hinsdale, Illinois. Pitts, having severed ties with Detroit and left homelessness behind, and Lettvin, who was escaping the stifling expectations of his own family, settled in quickly. McCulloch, instantly "enchanted" by Pitts, a 16-year-old prodigy who could grasp "the right and truthful answers to the unasked questions."[324] He also recognized Lettvin's exceptional talent, describing him as "keenly alive to neurological problems" and "never boring in a boring world."[325] As Tara Abraham has meticulously detailed, McCulloch's time at the Neuropsychiatric Institute of the University of Illinois, beginning in a year prior, marked a pivotal moment in his career. He adopted the role of an "egalitarian

[323] Jerome Lettvin, *Talking Nets: An Oral History of Neural Networks*, 3.
[324] Warren McCulloch, Pitts, Walter, 1946-1969, Box: 19. Warren S. McCulloch Papers, Mss.B.M139. American Philosophical Society.
[325] Warren McCulloch, Lettvin, Jerome Y., 1948-1964, Box: 15. Warren S. McCulloch Papers, Mss.B.M139. American Philosophical Society.

mentor," who fostered the brilliance of his younger colleagues, including Pitts and Lettvin, and would soon rely on them as his most indispensable collaborators.[326]

This mentoring style exceeded professional guidance into the territory of transference laden with emotional intensity.[327] McCulloch became a staunch advocate for Pitts at every level, personally and institutionally, supporting him unwaveringly throughout their intense collaboration. McCulloch's famously heartfelt recommendation for Pitts' Junior Fellowship at the Harvard Society of Fellows echoed this commitment: he never doubted that Pitts would "prosecute science to the best of his no-mean ability," firm in his belief that "certainly money will not seduce, security spoil or honor deceive him and he will not undertake what he cannot or will not perform."[328] "Would I had him with me always"—these words would reverberate for the remainder of both their lives, with only four months separating their deaths in 1969.

According to Lettvin, the collaboration between Pitts and McCulloch began with a shared fascination for the "universal combinatorics" of Wilhelm Gottfried Leibniz, the first to attempt to symbolize logic and thereby demonstrate that "any task

---

[326] Tara Abraham, *Rebel Genius: Warren S. McCulloch's Transdisciplinary Life in Science,* 99-100.
[327] Elizabeth A. Wilson, "Would I Had Him With Me Always: Affects of Longing in Early Artificial Intelligence," 839-847.
[328] Warren McCulloch, Pitts, Walter, 1946-1969, Box: 19. Warren S. McCulloch Papers, Mss.B.M139. American Philosophical Society. In his support letter for the same fellowship, Norbert Wiener echoes McCulloch: "In all my years of teaching I have not run against any man as great caliber intellectually as Pitts. Personally he is loyal, enthusiastic, considerate, and cooperative. It is a great pleasure for me to consider him not only as a student and colleague but as a close personal friend." Letter to Crane Brinton, dated February 7, 1946. Norbert Wiener Papers, MC-0022, box 5, folder 70. Massachusetts Institute of Technology Institute, Archives and Special Collections, Cambridge, Massachusetts.

which can be described completely and unambiguously in a finite number of words can be done by a logical machine."[329] Having spent over a decade immersed in neurophysiology before meeting Pitts in Chicago, McCulloch found inspiration in Leibniz's work to definitively define a "number" within the logical system of Whitehead and Russell's *Principia Mathematica*. Nearly three decades earlier, as a 19-year-old schoolboy, McCulloch had vowed to answer one question in his lifetime: "*What is a number, that a man may know it, and a man, that he may know a number?*"[330] For Pitts, this metaphysical question—expressed in such an elegant and seemingly simple form—resonated deeply with his own private struggles in logic, the very struggles he had worked out in the solitude of his hometown library.

McCulloch observed early on that mathematicians at the culmination of logic's origins—spanning Christian theology to the philosophy of David Hume—had already answered the first part of his question: a number, it seemed, was "a concept or idea identified with its class."[331] Russell and Whitehead's class-member hierarchy demonstrated a critical point at which mathematical logic exposed a set that could not be confused with a class—precisely, the paradoxical notion that there were no classes of sets that did not include themselves. When two classes have the same number of members, the similarity did not depend on the definition of any preexisting number. Instead, numerical activity emerged as "a way of bringing together certain

---

[329] Jerome Lettvin, *Talking Nets: An Oral History of Neural Networks*, 11.
[330] Warren McCulloch, "What is a Number, that a Man May Know It, and a Man, that He May Know a Number?", 2.
[331] Ibid., 6-7. See also: A Moreno, "Bertrand Russell's Concept of Number," 91-92. "A number (in general) is any collection which is the number of one of its members; or more simply still: A number is anything which is the number of some class."

collections, namely, those that have a given number of terms."[332] Russell, guided by his axiom of reducibility, thus concluded that "a number is something that characterizes certain collections or classes, namely, those that have that number." Number, in other words, is a property shared by the members of a class bound together by a concept of "(1,1) correspondence." [333]

For McCulloch, the logical definition of a number thus rested on "a symbolic process of putting things into one-to-one correspondence in a conventional manner," which led to an abstract conception of symbolic activity—one that, he argued, had its roots in "a way of doing things that has grown out of our ways of getting together, our communications, our logos, tricks for setting things into one-to-one correspondence" since antiquity.[334] The definition of a number as a repetitive loop of logical propositions, then, inextricably stemmed from the ancient practice of *counting*. For McCulloch, to count was to *tally*—to perform a mechanism for keeping track of recurring events by making a distinct mark for each occurrence and reoccurrence, like flipping a switch on and off or the notches on a CD or DVD.[335] In this sense, numbers presupposed memory and measurement long before they were named. The names of numbers, as signs, only emerged from the practical necessity of developing a heuristic to account for discrepancies between two tallies, an ad hoc solution to the problem of distinguishing quantities in a world brimming with recurrence.[336]

---

[332] A Moreno, "Bertrand Russell's Concept of Number," 91.
[333] R. L. Goodstein, "The Definition of Number," 180.
[334] Warren McCulloch, "What is a Number, that a Man May Know It, and a Man, that He May Know a Number?", 7.
[335] "Cybernetics of Cybernetics," Box 42, Heinz von Foerster Papers, University of Illinois Archives.
[336] Ibid.

McCulloch and Pitts thus deliberated how a number, emerging from the etymological root *computare*—to count—begot a series of reverberating loops comprising the symbolic logic of propositions. At its core, a number entailed an ongoing reckoning with and execution of a procedure (or algorithm) that transformed one set of data (inputs) into another (outputs). This iterative process—where counting leads to calculation, and calculation loop back into computation—was universal. Therefore, for McCulloch, "a psychological justification of the Turing definition of computability and its equivalents" was necessary to assert that "if any number can be computed by an organism, it is computable by these definitions, and conversely."[337] "Computational domains" thus described the nervous systems of both animals and artificial systems, underscoring their shared capacity to compute propositions in a finite number of steps, each activation catalyzing new patterns of propositional activity according to a learning algorithm (or "training").[338] Computing machines, on the one hand, implemented this transformation through programs expressed in well-defined machine languages. They operated in storage devices (such as memory chips) that held input, output, and intermediate data in device-specific codes.[339] The brain, on the other, executed this same process through *neurons*, demonstrating an infinite diversity of computational domains within it. Each neural network operated on its

[337] Warren McCulloch and Walter Pitts, "A Logical Calculus of the Ideas Immanent in Nervous Activity," 113.
[338] "Cybernetics of Cybernetics," Box 42, Heinz von Foerster Papers, University of Illinois Archives.
[339]For comprehensive primers on contemporary machine learning architecture, see: Peter Norvig and Stuart Russell, *Artificial Intelligence: A Modern Approach, 3rd Edition*: Margaret Boden, *AI: Its Nature and Future*.

own unique computational logic, tailoring the process of learning and transforming information based on its biological structure and the interactions within its circuits.

McCulloch and Pitts brought these seeds of late-night collaboration to fruition in "A Logical Calculus of the Ideas Imminent in Nervous Activity," published in the Spring of 1943 in Nicolas Rashevsky's *The Bulletin of Mathematical Biophysics*.[340] The paper began with Pitts' notes on Carnap's "propositional calculus"—a journey that began when he walked into Carnap's office, copy of *The Logical Syntax of Language* in hand.[341] McCulloch, confounded by the problem of "circular paths, whose activity so regenerates the excitation of any participant neuron that reference to time past becomes indefinite," posited that the theory would not be complete without finding a solution to these "recursive functions."[342] It did not make sense that in "nets with circles," the output of one neuron becomes the input of the next. Pitts, in a moment of intellectual audacity, "solved" this neuronal paradox by, according to Joe Dumit, "adding time to Carnap's logical notation system,"[343] mathematically breaking the cyclical structure of neural activity open by "assigning temporal indices to them."[344] This allowed them to create a "formal equivalence" to neural events by introducing time into the measurement of the net's activity. These "*temporal propositions*," ironically, allowed them to *remove* time from the calculus of

[340] William Aspray, "The Scientific Conceptualization of Information: A Survey," 129. Aspray notes that the paper "received little notice from biologists and psychologists until popularized by von Neumann."
[341] Neil Smalheiser, "Walter Pitts," 219.
[342] Warren McCulloch and Walter Pitts, "A Logical Calculus of the Ideas Immanent in Nervous Activity," 101.
[343] Joe Dumit, "Plastic Diagrams: Circuits in the Brain and How They Got There," 229
[344] Gualtiero Piccinini, "The First Computational Theory of Mind and Brain," 179.

information altogether. Signals, considered "temporal invariants," were delays detached from the real moment in which an event occurred, generating pauses in the logical chain that did not influence the overall activity of the network:[345]

> The importance of the formal equivalence lies in this: that the alterations actually underlying facilitation, extinction and learning in no way affect the conclusions which follow from the formal treatment of the activity of nervous nets, and the relations of the corresponding propositions remain those of the logic of propositions.[346]

McCulloch and Pitts thus formalized the way in which the brain to store and reactivate past events—no longer bound to the immediacy of the present, but through a recursive logic that could loop, delay, and store the very idea of memory itself. The symbolic or semiotic production of memory occurred retroactively after the present event, revealing the nature of memory processes as that "which every other form of memory is but a surrogate requiring reactivation of a trace."[347] They further devised a technique for designing "nets without circles"—echoing Turing's "circle-free machines"[348]—that satisfy various logical functions between propositions within the parameters of five assumptions:

---

[345] Lily E. Kay, "From Logical Neurons to Poetic Embodiment of Mind: Warren S. McCulloch's Project in Neuroscience," 598. "If neurons were to compose a computer of propositional calculus, then one had to take into account that, if the spikes arriving on the axons of two neurons represent truth values of some two propositions, then the neuron computing the logical operation on these two input propositions can give its answer only a cycle-time later than the time of input production (but time-variations for different lengths and types of neurons were deliberately neglected). The cycle-time then is the time unit for the operations of neural nets."

[346] Warren McCulloch and Walter Pitts, "A Logical Calculus of the Ideas Immanent in Nervous Activity," 101.

[347] Warren McCulloch, "What is a Number, that a Man May Know It, and a Man, that He May Know a Number?", 10.

[348] Alan Turing, "On Computable Numbers, with an Application to the Entscheidungsproblem," 233.

> 1. The activity of the neuron is an ‘‘all-or-none’’ process.
> 2. A certain fixed number of synapses must be excited within the period of latent addition in order to excite a neuron at any time, and this number is independent of previous activity and position of the neuron.
> 3. The only significant delay within the nervous system is synaptic delay.
> 4. The activity of any inhibitory synapse absolutely prevents excitation of the neuron at that time.
> 5. The structure of the net does not change with time.[349]

Ultimately, McCulloch and Pitts argued for a fundamental isomorphism between nervous activity and the "all-or-none character" of Boolean logic—the formal language of set theory. Just as a set is either included or excluded from itself, a neuron either generates an electrochemical discharge across its synapse past a certain threshold, or it does not. A net could symbolically represent any function within their logical calculus, and if its outputs were fed back into its inputs on a time scale, it could simulate "memory." As Gualitiero Piccinini observes, they contrasted the computability of their nets as distinct from the more complicated nature of "cyclical" ones. This proved immensely influential for "designing circuits for digital computers, because it allowed the design of circuits that compute *any* desired Boolean function."[350] Furthermore, McCulloch and Pitts viewed their ultimate goal not as providing a literal description of neurons, but, according to Abraham, as designing

[349] Warren McCulloch and Walter Pitts, "A Logical Calculus of the Ideas Immanent in Nervous Activity," 101.
[350] Gualtiero Piccinini, "The First Computational Theory of Mind and Brain," 203. Unearthing a paradigm of "logical design," McCulloch and Pitts thus provided an essential tool for von Neumann and his formalization of finite automata in 1945, an essential development in the theory of computation.

"fictitious nets which have a formal equivalence to 'real' ones."[351] By embodying logical descriptions of behavior within these "fictitious nets" of neurons, they provided a physiological infrastructure for the logical nature of human semiosis—one that could be shaped and transformed through learning processes.

In this way, McCulloch and Pitts aligned neuronal activity in the brain with Alan Turing's theoretical model of a "logical machine" from 1936—and, stretching back even further, with Leibniz's concept of combinatorial calculation from three centuries earlier.[352] Turing's logic of "computing numbers," whether in humans or machines, was built upon the simplest mechanical operation: the handling of finite, elementary states—essentially, the presence or absence of a signal, or the basic yes/no answers to a question.[353] McCulloch and Pitts, echoing the vision of a universal computer in their neural nets, joined Turing in positing that the looping of information across time was the key to both computation and memory. Both a "report" of a cause and a "command" that offers a choice between two possible states, a signal in this context operated as a "true or false" logical proposition.[354]

"A Logical Calculus," widely celebrated as the first paper to devise a mathematical model of neural networks, is an object of knowledge that, in its

---

[351] Tara Abraham, "(Physio)logical circuits: The intellectual origins of the McCulloch-Pitts neural networks," 20.
[352] Alan Turing, "On Computable Numbers, with an Application to the Entscheidungsproblem," 230-265; Wilhelm Gottfried Leibniz, "Dissertation on the Art of Combination";. See also: Phillip Mirowski, *Machine Dreams*, 80-93.
[353] Turing, "On Computable Numbers," 232. "If an *a*-machine prints two kinds of symbols, of which the first kind (called figures) consists entirely of 0 and 1 (the others being called symbols of the second kind), then the machine will be called a computing machine."
[354] Warren McCulloch and Walter Pitts, "A Logical Calculus of the Ideas Immanent in Nervous Activity," 104.

idealized form, seemed infallible in its account of the computational activity of the brain.[355] The paper is unsurprisingly regarded as a foundational pillar of artificial intelligence by its progenitors, including Marvin Minsky, Mallory Selfridge, and Michael Arbib, and it is widely accepted that McCulloch's conceptualization of computational domains aptly captures the operation of LLMs in the contemporary context.[356] The concept that each neural activation triggers new patterns of propositional activity, shaped by an underlying algorithm that functions analogously to a "training" process, anticipated modern artificial neural networks programmed to optimize their computational pathways by relying on iterative learning algorithms to encode linguistic patterns, inferential rules, and contextual relationships through vast but finite sequences of computational steps.[357]

According to von Neumann, McCulloch and Pitts demonstrated that, at least in theory, "anything that can be completely and unambiguously described, anything that can be completely and unambiguously put into words," was "ipso facto realizable by a suitable finite neural network."[358] However, they quickly stumbled upon the notion that self-referential loops—those persistent paradoxes of logic—became

---

[355] On June 2023, venture capitalist Marc Andreessen, in his anti-AI doomerism manifesto "Why AI Will Save the World," lionized McCulloch and Pitts under the banner "Legends and Heroes" for delivering the "first scientific paper on neural networks – the architecture of the AI we have today."
[356] Jerome Lettvin, "Introduction to Volume 1," *The Collected Works of Warren McCulloch*. "Yet, with all of this promise held forth by AI, McCulloch could not himself take it as seriously as he might have, and that was because he felt that a logical machine by itself, however cleverly programmed, would come to naught. It was very hard to find out from him why this was the case. He staunchly supported Minsky and Papert in their endeavors, and indeed they readily agree that they took inspiration from McCulloch. Yet something lay in McCulloch's mind, saying this is not the case; that is, there must be another kind of model that should be more suitable."
[357] See Paul Kockleman, *Last Words: Large Language Models and the AI Apocalypse.*
[358] John von Neumann, "The General and Logical Theory of Automata," 22-23.

thresholds of breakdown, breeding neuroses not only in humans but in the circuits themselves. The scope of their theory thus extended beyond the machinery of computation; they sought to map their work onto the loopier terrain of psychology and psychiatry, such that "no psychology can pretend to adequacy which cannot exhibit the origins of problems that puzzle men like Whitehead and Russell."[359] As Piccinini aptly summarizes, "McCulloch and Pitts presupposed that the computational limitations of their nets capture the computational limitations of brains, and that the computational limitations of brains correspond to the ''psychological'' limitations of humans engaged in computation."[360]

The triumphs of logic, set theory, topology, and algebra that Pitts demonstrated in this paper, in other words, could not fully account for "the problem of insight, of intuition, of invention"—the second part of McCulloch's singular question: "What is a man that he may know a number?"[361] The psychoanalytic and behaviorist noise on the periphery only seemed to cloud things further. McCulloch thus delved into the theoretical chaos of defining the nature of knowledge: "How do we know anything aboabraham ut the world—either its particulars as we apprehend events or its universals as we know ideas?," in tandem with the nebulous question of desire: "How do we desire anything…as we want food and drink or a woman and a bed, or mentally, as we seek in music the resolution of a discord or, in mathematics

[359] Warren McCulloch, "Speech for The Association for Psychoanalytic and Psychosomatic Medicine," copied to Gregory Bateson. Gregory Bateson papers, MS 98, Box 22, Folder 931.
[360] Gualtiero Piccinini, "The First Computational Theory of Mind and Brain," 199.
[361] Ibid., 14.

the proof of a theorem?"[362] Such questions constituted the foundational theory of knowledge that his and Pitts' work unearthed.

McCulloch and Pitts' work revealed the brain's epistemological architecture demanded more than mere circuitry; as Hayles and Dumit have respectively observed, it required "embodiment" and, more critically, "a discrete, lived notion of time" to avoid veering into nonsense.[363] Attuned to the "psychic life of circuits," McCulloch posited that the brain's information-coding—what occurred in the topographical distribution of biological neurons—revealed a structure and a process to be replicated in machines.[364] "Psyche anatomized and psyche synthesized," equally embodying "the logical properties of a proposition… a proposition embodied in a physical process," could be examined by neurophysiological and anatomical analysis.[365] McCulloch thus found a strange kind of resolution: the mind a "logical psyche," the human being, "made of fallible neurons, may know it [a number] infallibly."[366]

This, of course, posed a deeper conundrum. McCulloch, according to Dumit, was preoccupied with the inevitable question raised by their neuron-inspired models: who—or what—was the *subject* of these neurons?[367] Time, in this framework, was a constitutive dimension of experience itself, and memory, in turn, emerged not as a passive archive but as the stabilizing force that made concepts durable, reproducible,

---

[362] Warren McCulloch, "Machines That Think and Want," 307.
[363] N. Katherine Hayles, *How We Became Posthuman*, 57.; Joe Dumit, "Plastic Diagrams: Circuits in the Brain and How They Got There," 225-226.
[364] Joe Dumit, "Plastic Diagrams: Circuits in the Brain and How They Got There," 229.
[365] Warren McCulloch, "Why the Mind Is in the Head," 72.
[366] Warren McCulloch, "What is a Number, that a Man May Know It, and a Man, that He May Know a Number?", 18.
[367] Ibid.

and actionable. For Bateson, who regularly corresponded with McCulloch (figure 8), this opened the floodgates to a profound and unsettling question: how does learning—true, trans formative learning—occur?[368] McCulloch and Pitts, in their audacious fusion of biology and computation, forced this question to the surface.

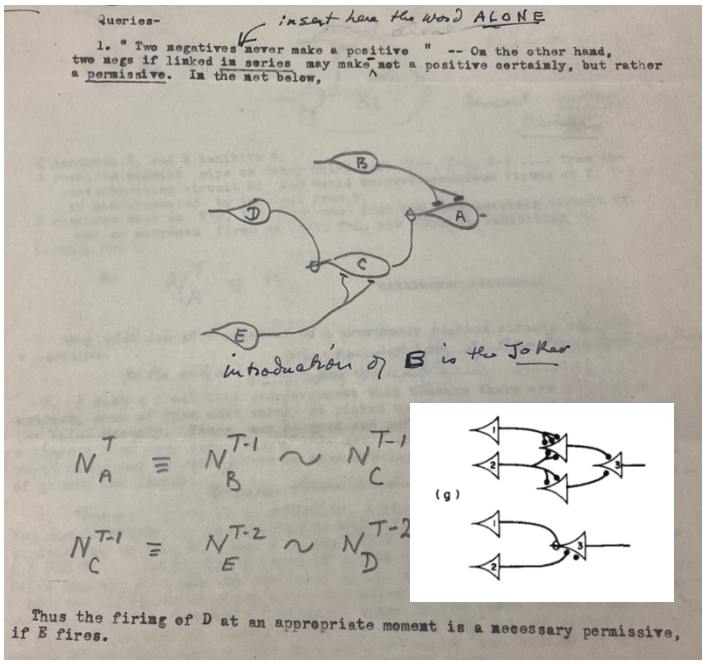


Figure 8: Hand drawing of McCulloch-Pitts network by Gregory Bateson in a letter to Warren McCulloch, dated March 26, 1946. *Gregory Bateson papers*, MS 98, Box 22, Folder 921. Special Collections and Archives, University Library, University of California, Santa Cruz.

As such, the most compelling of McCulloch and Pitts' contributions, beyond the technological scaffolding of artificial intelligence and neural networks, lies in its epistemic critique of a higher order. At stake in their collaboration was the formation of the logical subject, whose ability to learn and evolve hinged on a complex interplay between time, memory, and experience that once loosely defined "life." Furthermore, a profound insight for psychologists and psychiatrists appeared when the causal structure of their nets revealed disjunctive connections and recursive loops

[368] In a letter to his first daughter Mary-Catherine on June 26, 1977, Bateson writes of McCulloch in reference to the Macy Conferences: "Of course he was our leader and it was his life's work, which, acting through, me had brought us all together. But he was, like Moses, a leader who could and did bring us to the edge of the promise land, where he himself could never enter."

of activity—neurophysiological embodiments of epistemological incompleteness. Logical propositions, far from universal truths, were revealed as the "linguistic legerdemain" they have always been: artificial constructions that mask the totality of the systems they purport to describe.[369] By foregrounding the recursive nature of "embodiments of mind," McCulloch and Pitts thus raised unsettling questions about the human subject who learns, remembers, constructs meaning—and perhaps most elusively, grapples with epistemic uncertainty and paradox.

The Epistemology of Frogs and Octopuses

Scholars memorialize McCulloch for his rejection of industrial and military entanglements following the publication of "A Logical Calculus."[370] He decried the technophilic hijacking of his work, which, in his eyes, veered dangerously into the realm of designing "toys and monsters," rather than advancing a deeper understanding of the computational principles of the human mind (or the mental-like tasks that engineers could train computers to perform).[371] As Abraham carefully details, McCulloch, navigating the delicate balance of his "dual identities as neuropsychiatrist and cybernetician," quickly pivoted to clinical psychiatry after the paper's release.[372] He saw the brain's pathologies as the backdrop to his larger philosophical project: to understand the human mind not as a mystical entity but in

[369] "Cybernetics of Cybernetics," Box 42, Heinz von Foerster Papers, University of Illinois Archives.
[370] Michael Arbib, "Warren McCulloch's Search for the Logic of the Nervous System," 193.; Lily E Kay, "From Logical Neurons to Poetic Embodiment of Mind: Warren S. McCulloch's Project in Neuroscience," 603.
[371] Mary-Catherine Bateson, Our Own Metaphor: , . " "
[372] Tara Abraham, *Rebel Genius: Warren S. McCulloch's Transdisciplinary Life in Science,* 125.

the starkly mechanical and biological terms that cybernetics promised.[373] Pitts, in contrast, moved to MIT under the auspices of Wiener and Lettvin. Labelled "4B pre-psychotic" during his draft psychiatric exam, the Kellex Corporation—the Manhattan Project's atomic energy arm—swiftly recruited Pitts for classified work.[374]

Pitts' shift from McCulloch's mentorship to Wiener and Rosenbluth marked a crucial turning point in his research trajectory. Under Wiener's guidance as a research assistant, he edited and provided corrections to Wiener's 1948 *Cybernetics* and participated regularly in the Macy Conferences on Cybernetics between 1948–1952.[375] Intellectually, Pitts was urged by Wiener to expand his mathematical modeling of neuronal activity to incorporate statistical mechanics and probability theory, accounting for the stochastic, time-bound nature of neural processes.[376] Just a few months into his Boston tenure, Pitts wrote to McCulloch with youthful pride: "I now understand at once some seven-eighths of what Wiener says, which I am told is something of an achievement… to constitute the first adequate discussion of statistical mechanics, understood in the most general possible sense, so that it includes for example the problem of deriving the psychological, or statistical, laws of behavior from the microscopic laws of neurophysiology… Doesn't it sound fine?"[377]

---

[373] Ibid.
[374] Neil R. Smalheiser, "Walter Pitts," 221. According to Smalheiser, Pitts' examination was conducted by A.A. Brill, famous for his editions of Freud's *Collected Papers*.
[375] Ibid., 222.
[376] See also: Steve Heims, *The Cybernetics Group*, 40-51.
[377] Letter to McCulloch dated September 11, 1944, Warren McCulloch Papers, American Philosophical Society, B/M139, Box 2. "I think I am now in a position to write a fairly complete foundation for the statistical mechanics of the nervous system, and shall do so soon."

For Wiener, to Pitts' amusement, "mind" could be understood in terms of information theory: as randomness decreased, order emerged, meaning the more improbable the synaptic firing of a neuron, the greater the information conveyed by the brain.[378] McCulloch, who had once quipped to Rosenbluth that Pitts' brain was "so much like Wiener's that it's funny," could not have foreseen the growing ambivalence Pitts would soon feel toward both Wiener and the paternal authority now surrounding him.[379] By the spring of 1944, Pitts confessed to McCulloch, "it is sort of lonely down here [at MIT], and I don't have very much really new to me to learn from Norbert and Arturo, except of course the latter's experimental results and techniques." Pitts yearned for the camaraderie of his old colleagues in Chicago, writing, "I wish I were with you; I think we should get much more and better work done—and, most of all, it would be something quite unforeseen and different—so opposed to merely working out the detail of something long ago planned in principle."[380]

Despite Pitts' conscious declaration that he had solved the loopy problems haunting McCulloch, paradox still defined the tension at the core of their model of "neurologically trapped universals."[381] Around the same time, von Neumann credited

---

[378] Norbert Wiener, *Cybernetics*, 10-11. "To cover this aspect of communication engineering, we had to develop a statistical theory of the amount of information, in which the unit amount of information was that transmitted as a single decision between equally probable alternatives… The notion of the amount of information attaches itself very naturally to a classical notion in statistical mechanics: that of *entropy*. Just as the amount of information in a system is a measure of its degree of organization, so the entropy is a measure of its degree of disorganization; and the one is simply a negative of the other."
[379] Letter to Arturo Rosenbleuth dated March 29, 1943, Warren McCulloch Papers, American Philosophical Society, B/M139, Box 2.
[380] Letter to McCulloch, undated, Warren McCulloch Papers, American Philosophical Society, B/M139, Box 2.
[381] F. S. C. Northrop, "On W. S. McCulloch," Collected Works of Warren McCulloch, Vol 1.

McCulloch and Pitts with uncovering the principles of "logical control" in digital computing machines.[382] Replacing neurons with vacuum tubes, von Neumann argued that the evolution of computers depended on the automaton's capacity to self-regulate, replicate, and generate programs of increasing complexity.[383] It was Alan Turing's dream writ large: machines that could, in principle, think. Yet this vision rested upon a deeply unexamined assumption—one that would prove increasingly untenable. Automata had to address somehow the paradox of self-reference: how could a machine be both the product of its programming and the agent that produced it? This was the fundamental question that McCulloch and Pitts had attempted to resolve, but with no final answer in sight.

Pitts' work with Wiener, however, brought him into the terrain of probabilities, where time was less a series of discrete moments but the ongoing *potentiality of probabilities increasing toward certainties*, culminating in what we call "the present."[384] Pitts, steeped in the promise of both Wiener's statistical models and McCulloch's ideal of a holistic neuropsychiatric framework, was thus driven to bridge this gap. What he sought—along with both McCulloch and Wiener—was to reconcile the looping complexities inherent in the temporality of propositions with the

[382] John von Neumann, "First Draft of a Report on the EDVAC," 5. "It is worth mentioning, that the neurons of the higher animals are definitely elements in the above sense. They have all-or-none character, that is two states: quiescent and excited… Following McCulloch and Pitts, we can ignore the more complicated aspects of neuron functioning: Thresholds, temporal summation, relative inhibition, changes of the threshold by after-effects of stimulation beyond the synaptic delay, etc."
[383] John von Neumann, "The General and Logical Theory of Automata," 288-328.
[384] "Cybernetics of Cybernetics," Box 42, Heinz von Foerster Papers, University of Illinois Archives.

general principle of "the exchangeability of time and space."[385] Temporal relationships in "reverberating neural chains" could be "exchanged freely with as much delay" and encoded spatially within neural networks, thereby dissolving the apparent contradiction between the flow of time and the static logic of propositions. They had come close to answering the question of how neural networks, "in preserving the sequence of events while forgetting their time of happening, are abstracted universals of a kind." However, the final piece—understanding how these processes extend "the abstraction to a wide realm of properties"—eluded them, leaving obscured how to reconcile embodied temporality with abstract recursion in neural systems.[386]

The opportunity to collaborate with McCulloch and Lettvin finally resurfaced in 1951 when McCulloch, having grown weary of his time at Chicago, accepted an invitation from Jerry Wiesner to join MIT's Research Laboratory of Electronics.[387] However, this reunion was far from the idealized collaboration that Pitts had envisioned. McCulloch, now increasingly consumed with the philosophical implications of nervous networks, became deeply involved in what he called an "experimental epistemology."[388] In contrast, Pitts distanced from theoretical

[385] Walter Pitts and Warren McCulloch, "How We Know Universals: The Perception of Auditory and Visual Forms," 130.
[386] Ibid., 146.
[387] Letter to Jerry Wiesner, dated May 31, 1951, Warren McCulloch Papers, American Philosophical Society, B/M139, Box 28.
[388] Tara Abraham, *Rebel Genius: Warren S. McCulloch's Transdisciplinary Life in Science,* 48.

abstraction, shifted his focus to the practical design of three-dimensional nets for working automata, informed by the growing influence of von Neumann's ideas.[389]

Despite these divergences, Lettvin, McCulloch, and Pitts, joined by biologist Humberto Maturana, began an influential investigation: probing the "universals of perception" observed in

a frog's eye, distinct from the logical workings of its neural circuits (figure 9).[390] The experiment involved placing electrodes connected to a galvanometer on the frog's optic nerve fibers and manipulating the frog's environment to provoke different visual experiences. The total sensory machine apparatus recorded the frog's visual perception before transmitting the data to the brain. As Hayles notes, "At this point, the frog's brain became part of a cybernetic circuit, a bioapparatus reconfigured to produce scientific knowledge. Strictly speaking, the frog's brain had ceased to belong to the frog alone."[391] The once clear boundaries between the living organism and its

[389] Jerome Lettvin, *Talking Nets: An Oral History of Neural Networks*, 9. "At MIT McCulloch became seduced into what can be done theoretically with nerve networks. A number of people gathered around and tried to see what could be mined out of this. Walter by this time had more or less set himself against the concept of doing only abstraction. To him it was much more important to come up with notions of how automata were to be devised. He wrote a thesis, a very long one, on the properties of nets connected in three dimensions."

[390] Michael Arbib, "Warren McCulloch's Search for the Logic of the Nervous System," 209. Arbib observes that "What the Frog's Eye Tells the Frog's Brain" was an application of the ideas in Pitts and McCulloch's "How We Know Universals," particularly the notion that "schemas for a universal" existed as neural circuitry in the brain. It was not so much that the frog's retina was passively relegated information to the brain as much as the visual input were already being processing in the optic structure as a "universal" modeled by the subsequent behavior of the frog.

[391] N. Katherine Hayles, *How We Became Posthuman*, 134.

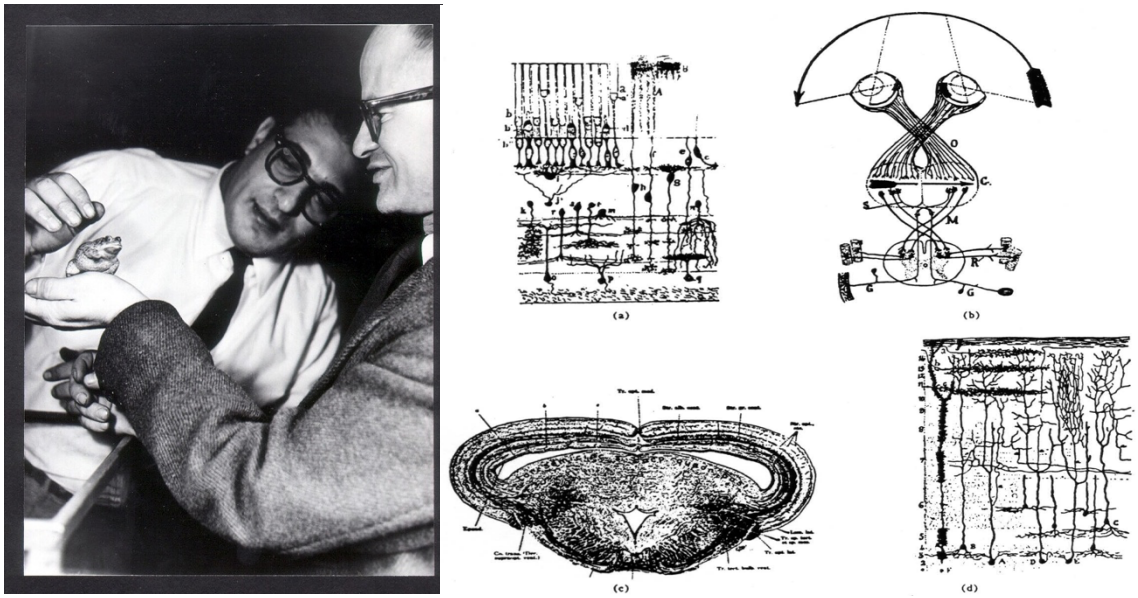


Figure 9: (a) Jerome Lettvin (left) and Walter Pitts (right) conducting experiments with frogs. Jeremy Norman, "McCulloch & Pitts Publish the First Mathematical Model of Neural Networks," HistoryofInformation.com. Accessed December 28, 2024: https://www.historyofinformation.com/detail.php?entryid=782 (b) Diagrams of frog anatomy (retina, optic nerve fibers, brain, tectum) by Ramon y Cajai and Pedro Ariens-Kapper. Reprinted from J. Y. Lettvin, H. R. Maturana, W. S. McCulloch, W. H. Pitts, "What the Frog's Eye Tells the Frog's Brain," *Proceedings of the IRE,* Vol. 47, no. 11, 1941.

technological apparatus blurred, making the frog's perceptual machinery part of a circuit that transcended its own natural existence, opening the door to a radically new conception of knowledge production.[392]

Culminating in their now-classic paper *What the Frog's Eye Tells the Frog's Brain*, McCulloch, Pitts, Lettvin, and Maturana rephrased McCulloch's original question in a more ethological vein: What occurs in a frog's eye that enables it to know—precisely, to sense and perceive anything?[393] Their investigation revealed that

[392] In Donna Haraway's terms, the organism and its apparatus can no longer be neatly separated into distinct, autonomous categories but ceaselessly combine into "cyborgs"—hybrid entities that challenge traditional epistemological boundaries and disrupts the notion of knowledge as something that originates exclusively from human subjectivity or natural life forms.
[393] J. Y. Lettvin, H. R. Maturana, W. S. McCulloch, and W. H. Pitts, "What the Frog's Eye Tells the Frog's Brain," 1940-1951.

physical motion in front of the frog's eye was essential to stimulate a response in the optic nerve, confirming the hypothesis that "the frog does not seem to see or, at any rate, is not concerned with the detail of stationary parts of the world around him."[394] The frog's sensory apparatus only responded to moving objects, those whose motion passed over a threshold of less than ten degrees at the eye; everything else, immobile, remained invisible to the frog, producing no impulse on the optic nerve.[395]

The discovery was shocking: the frog's brain was not receiving a mechanical reproduction of light patterns as the neurophysiologists assumed. The brain processed information about motion only through the filter of highly specialized and pre-logical analog processes in the eye.[396] The input reaching the frog's mind was already "highly organized and interpreted"—pre-digitized information that bypassed the need for logical abstraction.[397] The frog's eye, in its function, thus suggested a kind of natural pre-cognition embodied by a sensory mechanism attuned not to an objective reality but to the dynamic, relational processes of the world around it. The experiment revealed that "a frog's perception is biologically rigidly structured by its needs," activating an immediate, pre-discursive response to movement.[398] In Lettvin's complete account:

---

[394]Ibid., 1940.
[395] J. Y. Lettvin, H. R. Maturana, W. S. McCulloch, and W. H. Pitts, "Two Remarks on the Visual System of the Frog," 759. "The things we see are characterized by having an invariance under the transformation of accident, so that, e.g., a running dog and a resting dog may be the same dog and the running phenomenon is the same thing as that which another dog would do. *How* we make this abstraction of an enduring identity under conditions of profound change is a major problem in the theory of knowledge; certainly it is the central problem in the theory of perception."
[396] Gregory Bateson, *A Sacred Unity: Further Steps to an Ecology of Mind*, 226.
[397] J. Y. Lettvin, H. R. Maturana, W. S. McCulloch, and W. H. Pitts, "What the Frog's Eye Tells the Frog's Brain,"
[398]Steve Heims, *The Cybernetics Group*, 246.

> Everything I had found so far and had read about suggested that the world was reported by natural language rather than logical language, that is, in terms of things and their relations rather than sense data to be processed into the concepts of things. There was no feedback from brain to retina, so that whatever synthetic a priori [genetic programming] was involved in vision had to be indwelling in the retina itself, a five-layer device alternately connected vertically and horizontally in the layers. That a fixed five-layer system of elements, laterally interacting, gave the content of perception to a frog suggested that we were dealing with a system of *active complex analogue filters rather than a system of logical elements*. Each cell in the array was a complex processor of local spatiotemporal relations in the images, and was modified by remoter influences.[399]

*

The epistemological implications of *What the Frog's Eye Tells the Frog's Brain* spread through the cybernetics community with the vigor of a frog snapping at a fly—particularly among those obsessing over memory and order in biological systems. Chief among them was Heinz von Foerster. After splitting his time between editing proceedings of the Macy Conferences, attending MIT's "experimental epistemology" soirees, and travelling frequently to Arturo Rosenblueth's lab in Mexico, von Foerster launched the Biological Computer Laboratory (BCL) at the University of Illinois in 1957.[400] At BCL, von Foerster and his band of cyberneticists (Maturana, Gordon Pask, and W. Ross Ashby among them) explored how the frog's seemingly unremarkable vision system exposed the truth that human epistemology

---

[399] Jerome Lettvin, *Talking Nets: An Oral History of Neural Networks*, 9

[400] Ronald Kline, "Where are the Cyborgs in Cybernetics?," 345. McCulloch initially invited Von Foerster to the Macy Conferences after reading his paper "Memory: A quantum mechanical treatise." Von Foerster, in turn, helped McCulloch establish the American Society for Cybernetics in 1964 and became chief editor of the Macy Conference proceedings. For more on Heinz von Foerster and the institutional history of the Macy Conferences and the Biological Computing Laboratory, see: Stuart Umpleby, "Heinz von Foerster and the Mansfield Agreement," 187-190.

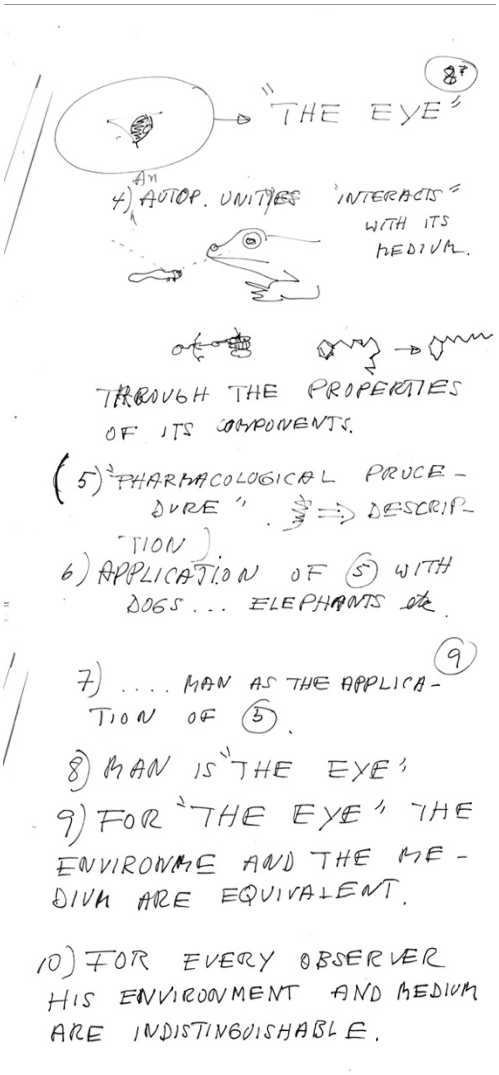


Figure 10: Letter from Humberto Maturana to Heinz von Foerster, undated. *Heinz von Foerster papers, 1921-2005, 2015, 2018*. 11/6/26, Box 22. University of Illinois Archives, University of Illinois, Urbana-Champaign.

might not be so different (figure 10). To perceive time, for instance, one must rely on memory—not as an archive of inert data, but a dynamic, system-wide operation calculating what *just* happened and what is likely *about* to happen.[401] That the neuron was "too small to be visible in a man's action, and too large to carry a trace of memory" meant that memory had less to do with digital storage in the brain but an active, circuitous negotiation of the total organism with the environment.[402] This was ultimately a humbling insight for any human tempted to think their own cognition was categorically distinct from that of a frog. "Man is not the measure," declared Ashby in a letter to von Foerster on Christmas Day 1968: "first we

[401] Heinz von Foerster, "What is Memory that It May Have Hindsight and Foresight as well?," 105-106. "In the stream of cognitive processes one can conceptually isolate certain components, for instance (1) the faculty to perceive, (ii) the faculty to remember, (iii) the faculty to infer. But if one wishes to isolate these faculty functionally or locally, one is doomed to fail… If the conceptual linkages of memory with the other two faculties are removed, "memory" degenerates first to a storage and retrieval system and, ultimately, to an inaccessible storage bin that is devoid of any content."
[402] Bateson's definition of *information* as a "difference that makes a difference," too, suddenly became less about pithy wordplay and more about the messy, analog, qualitative thresholds the frog's "natural language" relied upon to differentiate a fly from background noise.

make the measure, then we see where he falls. So far the results have been humiliating. Man pays for knowledge with humiliation."[403]

Yet what the frog's eye told the cyberneticists was not exactly a bolt from the blue. At the ninth Macy Conference in 1952, several years before Lettvin's frogs, English zoologist John Zachary Young had already set a gruesome standard for neural investigation—trading amphibians for cephalopods in his quest to decode embodied perception and its neural underpinnings.[404] Young, having famously discovered the giant axon in squid during the 1930, had since tormented the octopus—now aiming to isolate the physical machinery of memory itself.[405] In his film *Discrimination and Learning in Octopus*, Young documented his experiments of octopuses subjected to lobotomy procedures, as he severed its vertical lobe to see how it fared in retaining memories. The octopus brain, a curious assembly of "series of lobes," revealed unpredictable similarities to human neural architecture: the optic lobe, much like a retina in cell structure and function, fed signals forward into the frontal and vertical lobes (figure 11).[406] The results were as grim as the methods: while an intact octopus could hold a memory for a respectable forty-eight hours, a lobeless one had the attention span of a goldfish, forgetting new lessons within ten minutes.[407] Even as the

[403] Letter from Ross Ashby, dated December 25, 1968, Box 1, Heinz von Foerster Papers, University of Illinois Archives.
[404] John Zachary Young and (ed.) Claus Pias, "Discrimination and Learning in Octopus," 620-628.
[405] John Zachary Young, Fused neurons and synaptic contacts in the giant nerve fibers of cephalopods," 465–503.
[406] "Discrimination and Learning in Octopus," 620. "The rods of the retina project directly on to the surface of the optic lobe where there is a deep retina, in which the topography of the outer retina is preserved. There is an elaborate system with amacrine cells, resembling the human retina quite strongly."
[407] Ibid.

optic lobe allowed the octopus to process external phenomena and make rudimentary associations, the absence of the vertical lobe reduced learning to, as Pitts during the presentation observed, a momentary storage of pain in the octopus's anatomy—such that "the optic lobe itself can do the learning."[408]

Young's cephalopod experiments thus provided an early glimpse into the frog proposition that memory might not be some all-encompassing capacity, but a localized function, at least partially dependent on specific neural hardware. He stumbled upon a question that McCulloch enthusiastically beckoned the group to pose immediately after Young's screening: whether memory in information processors required a distinct physical structure.[409] In the octopus, there appeared to be just such a mechanism: a vertical lobe that, when intact, enabled the octopus to repeat its solution with Pavlovian efficiency. When removed, however, the octopus became a forgetful bystander to its

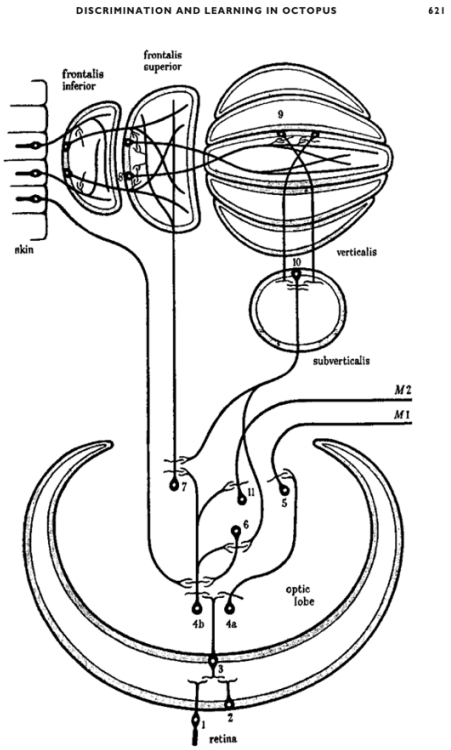


Figure 11: "Diagram of the Arrangement of Certain Pathways in Octopus." Reprinted in Claus Pias (ed.), *Cybernetics: The Macy Conferences 1946-1953 The Complete Transactions*, 621.

[408] Ibid., 623.
[409] Ibid., 622. "McCulloch: Who would like to ask what kind of question about that structure? I see we do have a structure where memory is occurring. Go ahead, ask him!"

own trials—suggesting memory was not so much a diffuse property of the brain as it was a precise anatomical feature, one liable to fail under surgical scrutiny.

Among the conference's more visibly disturbed attendees was Gregory Bateson, who had just wrapped up a presentation on the function of humor in communication (an ironic juxtaposition given what followed).[410] "In the film Young showed," observes biographer David Lipset, "Bateson noticed the 'despair' of an octopus… unable to solve the experimental problem."[411] He carefully posed a series of questions to Young and the audience: "Is there any indication at all that after the operation, the octopus, which evidently forgets quicker, does in some degree learn more quickly? Is it possible that one shock is sufficient to remind it and give it the answer, whereas, in the initial learning, it might have needed two or three trials? Does it really, after the operation, go back to the initial condition of ignorance of an octopus that has never seen the plate?"[412] The answer was, predictably, unsatisfying, a void where epistemological clarity might have been. The image of the lobotomized octopus haunted Bateson for years—long after the conference, when he applied to the National Science Foundation to study octopus communication in 1962.[413] The application was swiftly rejected. "This preliminary work was promising," Bateson noted with his trademark understatement, "but needed to be repeated and extended

[410] Gregory Bateson and (ed.) Claus Pias, "The Position of Humor in Human Communication," 541-574.
[411] David Lipset, *Gregory Bateson: The Legacy of a Scientist*, 233.
[412] John Zachary Young and (ed.) Claus Pias, "Discrimination and Learning in Octopus," 626.
[413] "Report on preliminary work on relationships in Octopus," Gregory Bateson papers, Box 64. MS 98. Special Collections and Archives, University Library, University of California, Santa Cruz.

under better conditions. For this, no grants were available."[414] Bateson, undeterred, resorted to keeping octopuses in his living room (figure 12).

There was no money for octopus soap operas, but hope—and funds—soon sprang from the promise of a dolphin communication laboratory in the Virgin Islands, courtesy of NASA and the Navy. The project, helmed by the infamous neurophysiologist John C. Lilly, seemed to hold the allure of both scientific prestige and institutional backing.[415] At the Oceanic Institute in late 1964, Bateson thus set out to patch up the "unresolved parts and raw edges" of his theory of learning and communication—an intellectual task that, unsurprisingly, required more than just theoretical musings.[416] To get there, however, he turned to the creatures most willing to entertain his hypotheses: the octopuses living in his living room. In his meticulously transcribed notes on octopus and cetacean communication, one entry—*"Notes on Pairing of Octopuses"*—offered an exhaustive 24-hour observation of six octopuses (Phillip(a), Francis, Fred, Charlie, Walter, and Verdi), divvied up between three tanks (C, D, E).[417] The notes, transcribed from voice recordings made between

[414] Ibid. A telling footnote in the bureaucratic dance of knowledge production: insight, it seems, had its price—and that price did not include funding for cephalopod conversation.
[415] Bateson's humble, if unconventional, career as an ethologist of octopus "love," "hate," and "mourning" has found a place in today's revived scholarly conversations. In early 2024, Benjamin Breen's widely publicized monograph resurrected Bateson's controversial involvement with John C. Lilly and his dolphin antics, bringing the research center's murky ethics into the spotlight. Yet Bateson's foray into animal communication received more respectable recognition in the Winter of 2017, when Bernard Dionysus Geoghegan edited and published a "Bateson Dossier" for *Grey Room*, including stills and the full transcript of Bateson and Weldon Kees' 16mm film *The Nature of Play: Part 1, River Otters* (1954), alongside a selection of letters from Bateson to Norbert Wiener selected from the archives.
[416] "Document Submitted in Application to U.S. Public Health Service for a Career Development Award and Launcher Grant," Gregory Bateson Papers, Box 59.
[417] "Notes on Pairing of Octopuses: A 24 hour observation beginning November 11, 1963," Gregory Bateson Papers, Box 56.

November 11, 1963 at 2:45 p.m. and November 12, 1963 at 4:31 p.m., paint a lively portrait of cephalopod social life. Together with his wife, Lois, Bateson captured every shift in color, every limb movement, and every cryptic gesture, all of it laced with anthropomorphisms one might expect from a cybernetically-minded ethologist obsessed with decoding animal behavior.

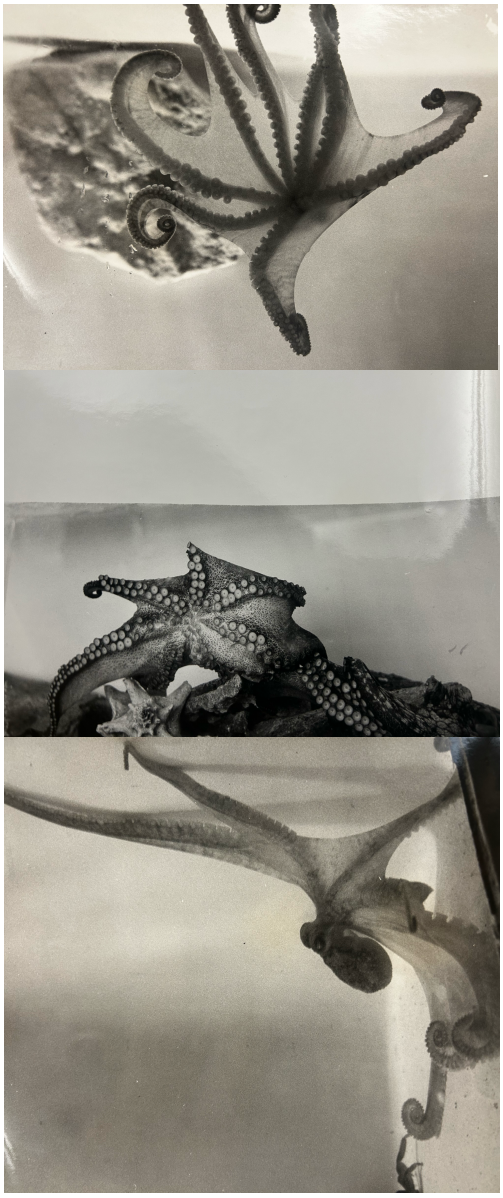


Figure 12: Octopuses in observation, 1964. Photographed by Gregory Bateson. *Gregory Bateson papers*, MS 98, Box 56. Special Collections and Archives, University Library, University of California, Santa Cruz.

The scene opens with Phillip and Fred, two Filosus octopuses, sharing space in Tank D. Fred, noticeably indifferent to Phillip's arrival, was described as "quite unwell and lethargic" since his last partner, Vesey, had passed away two weeks prior, leaving Fred to sulk in the same tank where he now languished. In an attempt to remedy Fred's existential slump, the experimenters decided to partner him with Charlie in Tank E, while Francis was moved in with Phillip. Meanwhile, Walter and Verdi seemed to hit it off immediately, cohabitating

without issue.[418] In the absence of discrete, verbal-like abstractions, communication unfolded as a continuous, almost meandering flow of iconic gestures and fluctuating magnitudes of movement—defined less by fixed meaning than by the ongoing, ever-shifting relations between bodies in space. It was a language of perpetual negotiation, where meaning was not pinned down but constantly in the process of being redefined.

Fred, slightly more reactive to Charlie, moves more freely in the tank but remains lethargic. "His arms coordinate but it seems to be an unusual way for an octopus to walk—his arms seem to go this way and that, just grabbing onto the glass in any kind of way." Fred takes a defensive stance in his shell, leaping toward Charlie with a tentacle to push him, but Charlie remains unfazed. Over the next 12 hours, Fred "moves around from point to point—aimless wandering with no real interest shown in food" and "stopping in positions that look quite unhappy." "Reminds me of someone that knows he is dying, but doesn't know what from," Bateson observes. At 1:35 AM, Fred begins "dancing wildly on the glass... right to left and back again, jetting spasmodically." By noon the next day, he's "doing the clockwise circling" and "spread-eagle" on the floor. The experimenters dub this the "death dance," a behavior observed in the final stages of an octopus's life: "within 12-36 hours from the time he has first noticed this gyrating on the bottom of the tank, the animals have expired." At

[418] As with most of Bateson's animal studies, the question of octopus sexuality was more of an open-ended suggestion than a fact to be pinned down. The sexes of the octopuses remained mostly ambiguous, and pronouns shifted with the fluidity of the creatures themselves. For simplicity's sake, Verdi and Phillip(a) were generally assumed to be female, while Fred, Francis, and Charlie were presumed male. As for Walter—well, Lois candidly remarked, "Why we talk of 'he'—Walter—I don't know." A reflection not just on the octopus but on the assumptions built into human frameworks of communication.

12:30 PM, Charlie leaves his shell and "gives a grab at Freddie," interrupting the dance. Fred withdraws, walking slowly while Charlie "sitting with his eyes up, all ready for business," returns to his shell.

Meanwhile, Phillip(a) and Francis in Tank D enjoy steady intimacy, "cuddling side by side in amicable fashion" and "sitting quietly holding hands." In Tank C, Walter and Verdi engage in a tense game of push and pull. Walter, "shot out of the box" and retreated initially, but later makes a "big thick prickle display" and approaches Verdi, who responds with hesitant advances before retreating. After hours of mutual "luring and aggravating," Verdi retreats into her shell, while Walter cleans himself "in front of Verdi's door," now a sad orange color. The two continue their cycle of attacks and defenses, "flicking" their tentacles without touching. When Verdi moves to Tank B for solitude, Walter, now "with wide exploratory arms," emerges—unsure if she's still there.

Altogether, the scene unfolded like a performative dance, its ambiguous gestures and projected emotions collapsing into a human pantomime of longing, rejection, violence, and fleeting connection.

A couple of weeks prior to his day of observation, on October 25, 1963, Bateson wrote a letter to McCulloch linking his octopus communication findings to the current decision-making processes of the Kennedy administration regarding the Cuban Missile Crisis. "It seems to me," Bateson defends his strange comparison, "that nations having almost total distrust for each other's words in fact operate with non-verbal cues and therefore face problems which must closely resemble those

which confront octopuses and pre-verbal mammals."[419] While these creatures engaged in intricate signaling, they lacked the capacity for negation: "There can be no simple iconic (analog) representation of a negative: no simple way for an animal to say "I will not bite you."[420] There was no "not" in their analog lexicon, and "very peculiar steps have to be gone through in order to get across a negative message to the other individual."[421] The animal can communicate negative demands as a tacit agreement "if (and *only* if) the other organism will first propose the pattern of action which is to be forbidden"—such as the play combat of two dogs upon first meeting. "But 'don't' is very different from 'not,'" Bateson contends, "generated as an *agreement* between two organisms following real or ritual combat."[422]

With no clear way to mark the absence of a presence, the negation of a signal, or the breaking of a logical frame, the communication of octopuses were "cumbersome and awkward." This analogic quality reflected, for Bateson, how negative messages at the scale of geopolitical conflict—such as "the message 'I shall not hurt you' or 'I trust you not to hurt me'," concerning the unspoken rules of relationship between one nation and another—were also difficult to communicate without confusion at various levels of the communicative act.[423] Analog systems of communication, where "the negatives can only be communicated by total sequences

---

[419] Letter to Warren McCulloch, dated October 25, 1963. Gregory Bateson Papers, Box 59, MC 1039-10a. See also: Philip Guddemi, *Gregory Bateson on Relational Communication: From Octopuses to Nations*, xi-xiii.
[420] "Octopus—Theory and preliminary observations," Gregory Bateson papers, Box 64.
[421] Letter to Warren McCulloch, dated October 25, 1962. Gregory Bateson Papers, Box 59, MC 1039-10a.
[422] Gregory Bateson, "Redundancy and Coding,"424.
[423] Philip Guddemi, *Gregory Bateson on Relational Communication: From Octopuses to Nations*, 6.

of interchange in which these negatives are exemplified," does not just struggle with precision. It actively generates its own complications, a messy tangle of self-referential meaning-making such as when "all the messages which negate war must be gotten across and must be gotten across in an interlinked form."[424]

*

How does a frog know anything? How does an octopus "love," "hate," or "mourn"? These questions, at the cusp of discovering the limitations of epistemology inextricable to the organism's sensory machinery, whimsically revealed that communication in analog, embodied, relational processes were fundamentally "illogical" —defying the digital logic that McCulloch and Pitts sought to impose on cognition. If epistemology governs knowledge and behavior, yet the process of knowing is limited by the organism's inherent design, how can it know anything—including the uncomfortable truth that it knows nothing? The irreducible uncertainty of the observer—the cyberneticist's incomplete knowledge of their internal state—aligns with what von Foerster called a "cognitive blind spot" (figure 13).[425] This blind spot, "through which I cannot even see that it is blind," defined the minimal conditions for epistemology.[426] Paradox, then, is not a problem to be solved by logic or cybernetics but the result of the inherent illogicality in how life itself processes

[424] Letter to Warren McCulloch, dated October 25, 1962. Gregory Bateson Papers, Box 59, MC 1039-10a.
[425] Heinz von Foerster, "Cybernetics of Cybernetics," 284. "One root of our cognitive blind spot that disables us to perceive social problems is the traditional explanatory paradigm which rests of two operations: One is *causation*, the other one *deduction*. It is interesting to note that something that cannot be explained—that is, for which we cannot show a cause or for which we do not have a reason—we do not wish to see. In other words, something that cannot be explained cannot be seen."
[426] Heinz von Foerster, "Ethics and Second-Order Cybernetics," 289.



Figure 13: Heinz von Foerster's demonstration of the epistemological blind spot functioning in the human optical system. To paraphrase von Foerster's instructions for the experiment: hold the page with right hand, close left eye, and fix vision of right eye on the star. Move page slowly back and forth from face at a certain distance when the round dot disappears, no matter what direction the page is moved. "This localized blindness is a direct consequence of the absence of photo receptors (rods and cones) at that point in the retina, the 'disk,' where all fibers leading from the eye's light-sensitive surface converge to form the optic nerve… this blindness is not perceived at all, that is, neither as something present, nor as absent: Whatever is perceived is perceived 'blotchless.'" Reprinted from "On Constructing a Reality," *Understanding Understanding: Essays on*

knowledge—the conditions of cognition encoded in the very fabric of our thought and communication.

Frogs and octopuses aside, however, the notion that errors of epistemology generate neurosis and threaten to unravel a system's coherence is nowhere more evident than in the life of Walter Pitts. From boyhood, Pitts believed that logic ruled the universe, steadfast in the belief that "if you could master logic, and really master it, the world in fact would become more and more transparent. In some sense or another logic was literally the key to understanding the world."[427] But the results of the frog experiments cracked that belief wide open. Knowledge acquisition, it turned out, was not about tractable "invariances" or neat logic. The frog's retinal ganglion cells did not function like logical machines at all; they looked "more like black magic than anything else."[428] This realization "blew a frog-sized hole" in Pitts' "realist epistemology," and the logical activity in his mind began to fail.[429]

[427] Jerome Lettvin, *Talking Nets: An Oral History of Neural Networks*, 10.
[428] Ibid.
[429] N. Katherine Hayles, *How We Became Posthuman*, 135.

Wiener suddenly cut ties with the McCulloch group around this same time, sending a telegram to Jerry Wiesner in January of 1952: ""Please inform [Pitts and Lettvin] that all connection between me and your projects is permanently abolished. They are your problem. Wiener."[430] In Lettvin's eye, "Walter suffered monstrously. For him the world stopped. Walter loved Wiener enormously. Wiener had become the father Walter never had and when Walter lost Wiener, he lost his *raison d'etre*."[431] After the unexplained loss, Pitts refused to sign the paperwork for his PhD conferral, burned all traces of his research and dissertation, and disappeared from academia. "He became more and more introverted. It was difficult to find him. He would try to escape from all his friends. We'd go hunting for him night after night. Watching him destroy himself was a dreadful experience for several people who knew him very well," Lettvin reflected of Pitt's final years.[432]

By 1969, after a decade of extreme isolation, Pitts found himself in acute care for alcohol-induced liver failure. His diminishing paper trail abruptly concludes with a final letter to McCulloch—also hospitalized nearby—written just weeks before his death. A passage described by Ronald Kline as "a striking instance of cyborg imagery,"[433] the letter begins with a dispassionate yet eerily poetic synthesis of

---

[430]Flo Conway and Jim Siegelman, *Dark Hero of the Information Age: In Search of Norbert Wiener, The Father of Cybernetics*, 219. Wiener's biographers Conway and Siegelman account how Wiener's wife Margaret, harboring longstanding animosity towards McCulloch, fabricated a story that their daughter, while, was "seduced" by "more than one" of McCulloch's boys. Wiener, already wary of McCulloch's bohemian lifestyle and suspecting that he was "trying to eclipse him as the principle proponent of cybernetics," quit the Macy Conferences and cut all ties with McCulloch, Wiesner, Lettvin, and Pitts until the end of his life.

[431] Ibid., 228.

[432] Jerome Lettvin, *Talking Nets: An Oral History of Neural Networks*, 10.

[433] Ronald Kline, *The Cybernetics Moment*, 348.

medical data and mechanical metaphor: "I understand you had a light coronary; that the primary EKG signs are already back to normal, but that it is complicated by redetachment of the old pneumothoraxes to the accompaniment of much blood; that you are attached to many sensors connected to panels and alarms continuously monitored by a nurse, and cannot in consequence turn over in bed. No doubt this is cybernetical. But it all makes me most abominably sad."[434]

In his attempt to quantify his own terminal condition, Pitts falls back on the language of numbers and logic, seeking precision in the face of the nebulous and inevitable.[435] Yet, the more he tries to translate the ineffable process of dying into the language of measurable quantities, the more glaringly inadequate those abstractions become. In his pursuit of precision amid emotional chaos, Pitts was tethered to the systems of thought that promised resolution but somehow left him yearning for human connection—or some answers to the irrationality of attachment and loss. This was a stochastic anomaly that his brilliance could not compute into being.

In his final words, as though struck by the inescapable paradoxes of life and logic, Pitts' articulation of his predicament turns inward, collapsing into the

---

[434] Letter from Walter Pitts, dated April 21, 1969. Warren McCulloch Papers, American Philosophical Society, B/M139, Box 2.

[435] Ibid. "I came in here about 3 weeks ago with a decompensated liver, jaundice and an ascites about my waist which looked like a 7 months' pregnancy, as well as perpetual diarrhea and much loss of K. I have been living on a salt-free diet, 1000 cc total fluids per day, and copious draughts of KCL. This leads to a feeling of being post in Death Valley. They also supplied all the diuretics in Merid's index (?), which were of none effect. Still, I lost weight at about 2 lbs a day, from 161 t0 145 lbs. After that, no change. They tapped the ascites twice, 500ml the first time, 50 ml the second (the plastic plumbing blocked up.) Went to 144 lbs. Then Dr. Wolfe tried his super-diuretic, which mobilizes 2 ½ liters of H2O in 10 hours even in a semi-dehydrated person. It seems to have taken up some of the ascites fluid, for I have lost 3 lb a day and now weigh 135. Another loss of fluid is through the tap hole, which oozes enough to saturate 2 small towels per day. (The residence staff here is not very competent: it is a teaching hospital)."

realization that all knowledge, even that of his own decline, is necessarily a paradox. His last expression of life is an acknowledgment of the knowledge that could never be known.

> In any case, imagine the worst happens in both our cases. We shall then pull our wheelchairs together, look at the tasteless cottage cheese in front of us, and recount the famous story of the conversation at the house of old Glaucus, where Protagoras and the sophist Hippias were staying: and try once more to penetrate their subtle and profound paradoxes about the knower and the known, and the difference between true knowledge and well-founded opinion. Is this not better than telling sad stories of the Death of Kings—or Kennedies? If both able-bodied, we shall do the same. Any other case I will not consider.[436]

The Talking Punishment

In the spring of 1946, Lettvin closed a letter to Wiener by mentioning an epiphany born from late-night deliberations with Pitts: "Walter and I have been talking over matters and decided schizophrenia may [be] a special form of neurosis (i.e., a *mislearning*)."[437] This offhand diagnosis, blending psychiatry with cybernetic flair, was par for the course for Pitts and Lettvin who had made a sport out of intellectual mischief. A few years earlier, the *Bulletin of Mathematical Biophysics* that published McCulloch and Pitts's celebrated "A Logical Calculus" also featured their less serious contribution, "A Mathematical Theory of the Affective Psychoses," in the same issue.[438] The paper was born not of scientific rigor, but of

[436]Ibid.
[437] Letter from Jerome Lettvin, dated April 1946. Norbert Wiener Papers, MC-0022, box 5, folder 70.
[438] Jerome Y Lettvin and Walter Pitts, "A Mathematical Theory of the Affective Pscyhoses," 139-148.

pettiness and spite: a calculated hoax in response to a psychoanalyst colleague who, suspecting Pitts and Lettvin of "impropriety," had accused them of homosexuality.[439]

Determined to prove that their frequent meetings were not carnal but cerebral, Pitts and Lettvin invented a "brilliant scientific paper" purportedly of interest to psychoanalysts. Pitts posited two variables—"intensity of emotion" and "intensity of activity"—and spun a mathematical model of "circular insanities," with differential equations of these two variables to map a spectrum of affective disorders from depression to mania.[440] Lettvin, still a medical student moonlighting at the Neuropsychiatric Institute, presented the paper to an in-house seminar. To their astonishment, the hoax landed. No less a figure than Franz Alexander, a luminary in psychoanalysis, heralded the work as the mathematical salvation of the field. What began as a petty act of academic revenge had become, in Lettvin's later words, "fashionable nonsense" embraced by an eager psychoanalytic audience.[441]

As is well-known, their mentor McCulloch was not shy in voicing his disdain for psychoanalysis, an attitude that permeated both his public talks and personal correspondence.[442] That same year, in a speech delivered to the Association for Psychoanalytic and Psychosomatic Medicine, McCulloch expounded how his work in neurophysiology uncovered a theory "which serves a matrix of fairly well verified propositions" about the brain as "the Nodus Gordius of interconnected chains of all-

[439] Neil R. Smalheiser, "Walter Pitts," 219-220.
[440] Jerome Y Lettvin and Walter Pitts, "A Mathematical Theory of the Affective Pscyhoses," 139.
[441] Neil R. Smalheiser, "Walter Pitts," 220.
[442] Warren McCulloch, "The Past of a Delusion," 302. "Psychoanalysis is that delusion of which it fancies itself the cure."

or-none links."[443] He designated surgeons as technicians who "may cut it if he knows where," physicians as the maintenance crew who "may dissolve its rust and prevent corruption if he knows with what," and finally, psychoanalysts as the armchair philosophers who "may untie it if he knows how." "From the bottom of my heart I wish him [the psychoanalyst] God speed," quipped McCulloch.[444]

Adamant that the latter group grapples with problems that are simultaneously holistic, topological, and circular, qualitative and "incommensurable with respect to values," McCulloch's critique of psychoanalysis highlighted its inability ever to formulate hypotheses about a system of ten trillion neurons that could withstand scientific testing. The physiology of neural circuits, on the other hand, "does not yield full-blown psychoanalytic hypotheses"; rather, it enforces "limits" to the epistemic and semantic nature of psychoanalysis and the interpretation of signals and signs—"in the same sense that physics does on engineering."[445] By the 1950s, McCulloch doubled down on his view that because it is "manifestly impossible in one lifetime to secure an exhaustive psychoanalysis of a patient's entire life, and equally impossible to trace all the connections of each of his ten trillion neurons," the mind is "but the circuit action of a relay net of which the blueprint is incomplete."[446] He wrote to Pitts, hoping to offload the task of reviewing Freud's posthumous *Dictionary of Psychoanalysis*—to be featured in *The Scientific Monthly*—onto his protégé:

---

[443] "Speech for The Association for Psychoanalytic and Psychosomatic Medicine," copied to Gregory Bateson. Gregory Bateson papers, MS 98, Box 22, Folder 931.
[444] Ibid.
[445] Ibid.
[446] Ibid.

> I took this liberty because both Jerry [Jerome Lettvin] and I thought it might give you an excellent platform for an expose of its internal contradictions out of the horse's mouth… it was not a dictionary nor even a concordance but a discordance which might serve as a Sic et Non for Freud, if not for all of psychoanalysis. Sometimes under a single heading and sometimes by comparing two or three the contradictions became obvious. Fortunately, they have included large enough samples so that one can be sure what the context was; the difficulty is always that the context is not systematic.[447]

McCulloch wryly suggests that Pitts, as a "rank outsider," would be in the best position to deal the "fatal blow" to psychoanalysis, leaving others—like McCulloch himself—to provide the "coup de grâce" (which he delivered in 1952 with his famously wrathful paper "The Past of a Delusion").[448]

Tara Abraham observes that McCulloch's criticisms were not just a personal dismissal of Freud's ideas but a broader ideological shift in his intellectual project: steering psychiatric thought away from psychoanalysis toward more mechanistic, computational models rooted in cybernetics, McCulloch pursued a calculated attempt to distance himself from psychoanalysis and its Freudian roots.[449] In his neurophysiology, systems were analyzable as networks and circuits of interacting

---

[447] Letter to Walter Pitts, dated September 21, 1950. Warren McCulloch Papers, American Philosophical Society, B/M139, Box 2.

[448] Flo Conway and Jim Siegelman, *Dark Hero of the Information Age: In Search of Norbert Wiener, The Father of Cybernetics*, 228. Conway and Siegelman account that "The Past of a Delusion" came at the heels of the Wiener-McCulloch breach, during which his colleagues surmised that McCulloch's erratic outburst was due to "a disturbed episode." Psychiatrist and Macy Conference member Lawrence Kubie explicitly observed that "One force which may be relevant to [McCulloch's] upset is the fantastic state of megalomanic and paranoid rage in which Norbert Wiener seems to have been ever since Warren came to MIT."

[449] Tara Abraham, *Rebel Genius*, 98-99. "In essence, pursuit of a scientific account of the mind boils down to the question of what role science should play in psychiatry and what a scientific psychiatry should look like. There were many answers debated during the 1930s and early 1940s. For McCulloch, objective accounts of the mind would be mechanistic, materialistic, and grounded in propositional logic."

neurons rather than as psychological constructs defined by unconscious drives and dynamics. McCulloch posited that "mental diseases reduce to properties of neural nets," unprecedentedly formalizing the science of psychological phenomena.[450] Thus, contrary to the subjective and interpretative methods of psychoanalysis, the purpose of the psychologist was, in McCulloch's words, "find how brains work, and subsequently "help those that have need of a physician."[451] The delusional "confabulations" of dreams, free associations, and transferences were beyond obsolete, such that "*knowing the history of the patient is unnecessary for treating mental illness.*"[452]

This intellectual move was part of the larger historical moment where cybernetics—concerned with feedback loops, information processing, and systems theory—was positioning itself to disrupt long-standing psychological models that had, until then, dominated the field. Thus, the more the McCulloch-Pitts neurophysiology model matured, the more its founders developed hostility towards the psychoanalytic world—simultaneously as "the organic was of little interest to psychoanalysts."[453] But had Pitts and Lettvin's musings on schizophrenia as a "mislearning" been tethered to Bateson's psychotherapeutic endeavors, they might

[450] Gualtiero Piccinini, "The First Computational Theory of Mind and Brain," 202.
[451] Warren McCulloch, "The Past of a Delusion," 302.
[452] Gualtiero Piccinini, "The First Computational Theory of Mind and Brain," 202; Lily E. Kay, "From Logical Neurons to Poetic Embodiment of Mind: Warren S. McCulloch's Project in Neuroscience," 592. This violent rupture from Freud, according to Lily Kay, reflected not just McCulloch's personal intellectual disposition but also a critical turning point in the development of both psychology and the nascent field of artificial intelligence, where the mechanization of mind was gaining ground as the dominant metaphor for understanding human cognition.
[453] Steve Heims, *The Cybernetics Group*, 43.

have found themselves swimming in the same intellectual pool. Bateson's team, funded by Rockefeller to "collect material demonstrating the role played by the Russellian paradoxes in psychotherapy," grappled with the same logical underpinnings of the brain's computational architecture as illuminated by McCulloch and Pitts.[454]

Bateson, William Fry, a psychiatry resident at the Veteran's Administration Hospital in Palo Alto, and Jay Haley, a graduate student at Stanford, observed that "difficulty discriminating between "reality" and "fantasy" and incoherent speech – these two features of schizophrenia—based upon the inability of the schizophrenic to communicate about his communication—are seldom recognized but an essential requirement of normal human communication."[455] For Bateson, communication in schizophrenic patients and octopuses alike were messy, paradox-laden processes due to the very fact of "the limitations of logic which follow from the necessity of combining different orders of messages."[456] Bateson thus found kin in McCulloch while grappling with self-referential dilemmas inherent in type theory, identifying his own work as a recursive elaboration of the cyberneticist's inquiry.[457] In the hands of Bateson, the logical activity of neural networks were templates modelling schizophrenic patterns of (mis)communication.

---

[454] David Lipset, *Gregory Bateson: The Legacy of a Scientist*, 201.
[455].Ibid.
[456] Ibid.
[457] Letter to Warren McCulloch, dated October 9, 1967. Gregory Bateson papers, MS 98, Box 22, Folder 931. "Dear Warren: I find myself bumping up against something like your old problem, "What is a number that a man may know it; what is a man that he may know a number." This is hitting me in the form "What is a context that a man may know it; and what is a man that he may know a context." In other words, the problem of "learning to learn" turns out to be a forked problem just like yours."

Paralleling McCulloch and Pitts' work on neural networks with clinical practice, the team thus plunged headlong into the paradoxes of propositional logic as laying the "foundations" for communication theory. This particular afterlife of the Russellian paradox in examining schizophrenic communication was also a departure from "the insights of Freud"—but distinct from McCulloch's. The team was devising a method of therapeutic intervention beyond free association and interpretation: a "formal analysis" of the logical basis of the psychical processes (*i.e*., condensation, displacement, projection) built into the patients' "symbolic systems."[458]

In 1964, under a Career Development Award from the National Institute of Mental Health, Bateson drafted *The Logical Categories of Learning and Communication*, the centerpiece essay of his research program which laid the theoretical foundation for his theory of schizophrenia and the "double-bind."[459] In the essay, Bateson pushed forward an "application of Russell's theory of types to the concept of learning" in communication systems.[460] Just as when a logical system devolves into paradox when object and proposition fail to align, errors of logical typing in the "learning of computers, rats, and men" reflected a larger, systemic

---

[458] "Schizophrenic Communication: A Theoretical Discussion and A Proposal for Research," Gregory Bateson papers, Box 96.

[459] "On the Logical Categories of Learning and Communication," Gregory Bateson Papers, Box 95; "Document Submitted in Application to U.S. Public Health Service for a Career Development Award and Launcher Grant," Gregory Bateson Papers, Box 59; David Lipset, *Gregory Bateson: The Legacy of a Scientist*, 241. According to Lipset, Bateson's focus was to understand the paradoxical communication patterns that might underlie schizophrenia that also led to his involvement with the Communication Research Institute and John C. Lilly's laboratory in the Virgin Islands. The essay later became part of his larger efforts to secure further funding, including a request for a U.S. Public Health Service grant to support his studies on cetacean communication—an endeavor that pushed the boundaries of communication theory beyond human psychology and into animal behavior.

[460] Gregory Bateson, "The Logical Categories of Learning and Communication," 279-280.

collapse. Learning, in other words, cannot proceed without risking cognitive dysfunction, a dynamic mirrored in both human psychosis and machine learning.[461] This was a crucial insight to understand the structures of speech characteristic of schizophrenia, wherein the real world becomes disconnected from its formal, logical representation. For Bateson, this misalignment between world and symbol illuminated a paradox at the core of cognition: the symbolic systems that enable us to communicate and learn are most at risk of collapse with "incongruence about or conflict between context and metacontext."[462] The breakdown of communication and learning occurs, in other words, when communication systems—whether human minds or mechanical processes—can no longer maintain coherence within their established environment or parameters.

Bateson's theory of learning examined how communication systems operated within rigid logical frameworks—and the implications this had for understanding both human cognition and interactive behaviors exhibited by animals.[463] He outlined the importance of "discriminatory signals" in communication theory that he observed fundamentally lacked in schizophrenic speech. Human beings "not only emit signals but exchange signals which guide the interpretation of other signals."[464] These *meta*-signals, verbal, non-verbal, or implicit, allow the subject to "discriminate between

---

[461] Ibid., 283.
[462] Gregory Bateson, "Minimum Requirements for a Theory of Schizophrenia," 245.
[463] "Document Submitted in Application to U.S. Public Health Service for a Career Development Award and Launcher Grant," Gregory Bateson Papers, MS 98, Box 59. "The writer believes that the questions posed above about the logical typing of message material and the relations between digital and analogic communication are fundamental and urgent for the whole field of behavioral science."
[464] "Schizophrenic Communication: A Theoretical Discussion and A Proposal for Research," Gregory Bateson papers, MS 98, Box 96.

messages of love or hate" concerned with the relationship of a message to the internal state of the speaker or the relationship of the message with corresponding states of the external world—metamessages of relationship that are integral "if one is to survive as a social being."[465] The production of communication material in schizophrenic individuals, however, demonstrated a concerted absence of such a signal, such that "the person unable to make the discrimination may be unable to play or interpret their dreams or know that he is doing either. He will be unable to differentiate between fantasy and reality."[466] A lack of discriminatory signals between the "real" and the "as-if" state, in other words, leads to cognitive dissonance; learning to communicate therefore involved the potential for recursive breakdown, straddling the boundary between sanity and delusion.

In his empirical analysis of schizophrenic communication, Bateson thus explored how the apparent incoherence of schizophrenic language can be treated as a form of coded communication, requiring a "decoding" process to understand its underlying structure. An undated document titled "Sample of Schizophrenic Utterances" provides us with clear demonstration of this approach.[467] Bateson records a short dialogue with a patient named Edward, an ex-Navy officer who experienced a psychotic episode on duty, whose a recurring delusion that he can move places around like words in sentences generates a language model characterized by nonsensical patterns.

---

[465] Ibid.
[466] Ibid.
[467] "Sample of Schizophrenic Utterances," Gregory Bateson papers, MS 98, Box 52. The following block and in-text quotations containing Edward's speech are copied from this document.

> Edward: Far as Colorado having a middle age, it certainly had. Time for some of that Colorado to move over. I wanted right on the coast of California.

Observing that “Edward seems to use geographic places as symbols,” Bateson reinterprets this dialogue through a private "translation" process, treating Edward's language as a puzzle to be solved.

> Bateson’s translation: *I’d like to change things as they exist today; return to some better time in history.*

Notably, he focuses on how communication breaks down when these "logical categories" are misaligned. Edward's fixation on movement, expressed through his delusion that he could "move places around like words in sentences," reveals a deeper psychological need to reshape his world and personal history. In an attempt at a therapeutic sleight of hand, Bateson asked Edward how long it had it been since he “had a pass of any sort.” His response:

> Edward: They gave some ground privileges in the form of a cattle yeah. Course the one that was out at the hundred fifty, lets see the hundred and fifty nine and nine tenths meridian was weeks, or ninth, ninth or tenth were on it rather with kind of bugged a little but course bear itself is a bug. Well, that’s the truth. But anybody that carries around a regular goddamn gun, will prove he has a perfect binder. There we haven’t any yaps. Course of University of Notre Dame was bumped into so bad they went to the wrong side so they got an old age in Colorado. And it wasn’t even a road. Well, we finally got that one down. And if a penguin we’d have more trouble than if we’re not penguins. And that cheers one of mine that somebody wanted me to wash off. If I holler on showers cause of some goddamn son of a bitch not using that building for a home. It’s the most irregular home I’ve ever seen for Betty’s home’s is reasoning. I wouldn’t have been caught in hall of that for nothing. But any time they’re going to have a goddamn bunch of bastards making up a world line instead of the state of California, I’m going to quit cause a Shoshone, a Cherokee, a Apache, a Sioux, there is a whole group I’d like to get. Say this building ought to be regular then you make sure all the guns are up.

> A regular mortar is the most reasonable thing I've ever seen. They ought to move over. I was putting the eighteen hundred on it. How about some of those that would make up a goddamn vessel get themselves over, they can dig in a little bit. Course the University of Alaska hasn't been bumped into. When it comes for a manual form one's own blues. It's inhuman in there in Alaska. I ought to pick um so they never had any irregularish. A man cross to a border line, off he gone. It isn't necessary to knock a man off when he crosses a border line.
>
> Bateson's translation: *Yes. It was treating me like I was cattle to be herded from one spot to another. I guess they thought I couldn't find my way. I think they thought I was a bit crazy. But then in my condition that was natural, but anybody that has been through what I have would be confused too. I don't think that colleges should mix education and religion. Religion isn't the answer.*
>
> *There is something to be said for being impersonal in human relations. At least then you aren't disappointed when you expect nothing. However, one could rightfully expect a little more human warmth and understanding from the hospital. A person is not exactly made to feel at home here. If I had anything to say about it, I wouldn't be here. The people that run the world don't care about me and neither do the people that run this hospital. Of course, you can try to escape the coldness and cruelty of the world like I did but you can't always change the world. But it is wrong to blame me for only trying to find a solution to my problems.*

A desire for change—or perhaps for control—is complicated by Edward's identification with the "American Indian," tapping into unresolved feelings of frustration with systems of power. Edward also continued to mention the names of different colleges in a deferent manner. He stressed the importance of a home for a baby's "parallel meridian reasoning" and his access to a "meridian road." Considering that he had dropped out of Junior college to join the Navy force, these repetitions indicated, for Bateson, an attempt to reclaim a lost opportunity or resolve an emotional void left by a truncated educational and familial life. Moreover, Edward's disdain for regimentation and his hatred for guns—the "symbol of war and the

regimentation in the service"—suggest a complex personal narrative of mistreatment and marginalization amounting to a rejection of systems that strip away his individuality.

Bateson wondered if Edward's responses were ultimately an expression of the cognitive dissonance that arises when the personal and the external world are out of sync, forcing a breakdown in communication internally within the individual and externally in social structures. His "schizophrenic utterances," like those of many others his studies, revealed the fractured logic of systems and their capacity for learning and transformation under duress—where attempts to organize and assign meaning ultimately lead to misalignments that he would come to call "epistemological error."[468]

*

Bateson's exercise in translation with schizophrenic utterances from his patients blurred the lines between interpretation and imagination. His own distress—and perhaps guilt—surrounding his military background as a communications analyst for the Office of Strategic Services (OSS), which involved crafting "black propaganda" to manipulate enemy morale, is seeped into his interpretation of Edward's distaste for authority.[469] Bateson's relationship with authority and family

[468] Gregory Bateson, "Pathologies of Epistemology," 485. "When you have enough technology so that you can really act upon your epistemological errors and can create havoc in the world in which you live, then the error is lethal. Epistemological error is all right, it's fine, up to the point at which you create around yourself a universe in which that error becomes immanent in monstrous changes of the universe that you have created and now try to live in."

[469] Anthony Chaney, *Runaway: Gregory Bateson, the Double-Bind, and the Rise of Ecological Consciousness* (Chapel Hill: The University of North Carolina Press, 2017), 38-41. Bateson developed a strong resistance to applied science of all kinds after experiencing how his knowledge on the nature

was also fraught with emotional complexity. His father—the acclaimed biologist William Bateson, who coined the term "genetics"—enforced high expectations on Gregory—the last of his three sons after John was killed in the First World War and Martin died by suicide years later.[470] His father's pressure on him to follow in his scientific footsteps and find solace in "the splendid and purifying purpose of science" undoubtedly shaped his psychological makeup.[471] At worst, this intersection of personal grief, professional duty, and theoretical ambition may have driven Bateson to blend the roles of analyst and interpreter, transforming the "message-material" of schizophrenic patients into reflections of his own unresolved emotions and theoretical pursuits. At best, the double-bind theory that emerged from such practices can be seen anew as an intricate fusion of personal history and scientific inquiry alongside its clinical and intellectual contributions.

In any case, a "translation" of schizophrenic communication diverged far from the psychoanalytic principle of maintaining impartiality in the face of countertransference. In Freud's model, the analyst must be able to distinguish their own projections from the material presented by the patient, yet Bateson's translations of Edward's utterances into symbolic meanings—like equating penguins with the coldness of interpersonal relationships—reflect his own empathy and emotional involvement. He recorded and analyzed the formal structures of his patients'

---

of communication and cultural exchange fueled propaganda missions in Burma. See also: Mary Catherine Bateson, "The Double-Bind: Pathology and Creativity," 11-21.

[470] David Lipset, *Gregory Bateson: The Legacy of a Scientist*, 95-96.

[471] Anthony Chaney, *Runaway: Gregory Bateson, the Double-Bind, and the Rise of Ecological Consciousness* (Chapel Hill: The University of North Carolina Press, 2017), 170-171.

communication systems while maintaining a relaxed rapport with the people themselves—often drinking coffee and playing golf rather than engaging in analytic sessions.

Bateson, however, still drew on a crucial distinction found in classical psychoanalytic theory from its origin point: the difference between primary process (as expressed in dreams) and secondary process (as manifested in verbal communication and written language).[472] This distinction underscores how the unconscious mind works analogically, producing images and associations that defy the logical structure of everyday language. Freud, as John Forrester observes, suggested that dreams are not merely distorted desires but "*distorted* fulfillment[s] of a *repressed unconscious infantile wish"*— products of a neurological system where "neurones" (the brain's basic units) facilitate the wish-fulfillment mechanism in ways that transcend linguistic or logical constraints.[473] By melding quantitative (energy-based) models with qualitative (meaning-based) models, Freud established the phenomenon of wishing, *desire*, the libido as "the driving force of our lives, out cultural endeavors, and our institutions," but also asserted the very possibility of the science of psychology as "*the intermingling of the language of quantity or energy, and the language of meaning*." In Forrester's metaphorical language, Freud was thus adamant to establish the physical and neurological basis of the brain as the "hardware" within which he could examine "the 'software' of the mind."[474] Bateson,

[472] This point is explored at length in the previous chapter on Freud's *Project for a Scientific Psychology*.
[473] John Forrester, *Freud and Psychoanalysis: Six Introductory Lectures*, 63.
[474] Ibid. 64-68.

like Freud, grappled with the tension between the "messiness" of unconscious processes—where time is fluid, opposites coexist, and logic dissolves—and the rigid, structured, and highly ordered systems of meaning we construct with language. This tension between the analogic nature of the unconscious (embodied in the primary process of dreams) and the more rigidly coded system of human communication demonstrated the methodological challenge of decoding schizophrenic speech.[475]

In his theoretical work outside of the clinic, culminating in the latter half of *Steps to an Ecology of Mind* (1973), Bateson developed a concept of meaning in human communication that diverged from traditional signification processes. This conceptual framework was grounded in ideas closely related to cybernetics, particularly the notions of "redundancy" versus "coding" and "noise" versus "information" in communication channels. These concepts, formalized by Claude Shannon at Bell Telephone Laboratories in 1948, were built on the idea that even when information is incomplete, communicative agents can still predict and infer the broader structure of a message with greater accuracy than random chance.[476] In

---

[475] Gregory Bateson and (ed.) Carl Whitaker, "Schizophrenic Distortions of Communication," 51-55. "The therapeutic process is, in part, a training of the patient to discriminate among the sort of signals he is receiving. But for this training to occur, the metasignals which identify the sorts of signals must be reliable… it is perhaps the essence of therapy with the schizophrenic that somewhere, framing whatever the therapist does, there must be the signals of primary affect, and that these signals should not be concealed from the patient. The question is whether the therapist is a doctor dealing with a patient or one human being dealing with another."

[476] Claude Shannon and Warren Weaver, *The Mathematic Theory of Communication*, 57. As explored in the first chapter, Shannon demonstrated that a device capable of decoding and recoding strings of digital symbols as inputs and outputs, such as an electrical transistor, endogenous memory is stored "so that its output depends not only on the present input symbol but also on its past history." By formalizing the "amount of information" conveyed by the signal solely by the probability of its "choice," the engineer could happily sacrifice the murkiness of semantics (meaning) for pure syntactics, an efficient economy of signs and their statistical quantities.

Bateson's model, however, the context plays a crucial role: an object or event might function as a signal for the whole of the message, but the act of encoding itself—rather than the content of the signal—enhances the receiver's ability to extrapolate meaning.[477]

This informatics framework culminated in Bateson's approach to schizophrenic communication. He saw disorganization and "noise" in schizophrenic utterances not as a failure to communicate, but as information in itself—about what could not be processed or understood in the typical communicative structure. The "noise," in this sense, became a key site for uncovering new patterns, as it revealed the breakdowns and distortions in the codification system. This distinctly cybernetic approach was thus not one of direct psychotherapy but rather an observational stance that treated patients' communications as message-codification systems in need of deconstruction to ultimately render noise informative.[478] The "noise" of Edward's speech, in other words, was *equivalent* to any "meaningful" soliloquy in terms of the *information* it provided for the translator carefully decode and contextualize carefully.[479] This approach to schizophrenic communication, underpinned by Russell

---

[477] Gregory Bateson, "Redundancy and Coding," 414. "By confining their attention to the internal structure of the message material, the engineers believe that they can avoid the complexities and difficulties introduced into communication theory by the concept of "meaning." I would argue, however, that the concept "redundancy is at least a partial synonym of "meaning." As I see it, if the receiver can guess at missing parts of the message, then those parts which are received must, in fact, carry a meaning which refers to the missing parts and is information about those parts."

[478] Ibid., 410.

[479] Warren Weaver, *The Mathematic Theory of Communication*, 8. "The word *information*, in this theory is meant in a special sense that must not be confused with its ordinary usage. In particular, *information* must not be confused with meaning. In fact, two messages, one of which is heavily loaded with meaning and the other of which is pure nosnese, can be exactly equivalent, from the present viewpoint, as regards information. It is this, undoubtedly, that Shannon means when he says that 'the semantic aspects of communication are irrelevant to the engineering aspects.'"

and Whitehead's theory of types, thus diverged from conventional psychoanalysis by forgoing a direct interpretation of "some specific traumatic experience in the infantile etiology" as Freud might. He instead sought to map these communications back onto the "characteristic sequential patterns" of the symbolic system, recognizing the structural value of the patient's communicational habits even in their seeming incoherence.[480]

Bateson's exploration of the double-bind structure emphasized that the source of schizophrenic communication was not simply linguistic, but emergent from the total context of relationships that shapes one's ability to communicate and be understood (*i.e.,* within family dynamics or psychiatric settings).[481] Rather than deriving from a failure to comprehend the "language of quantity" (the energy-based models of psychodynamics), the double-bind is fundamentally located in the "language of meaning"—the murky territory of semantics that Shannon disavowed. This frustrating "*experience of being punished precisely for being right in one's own view of the* context" manifests as an unsolvable riddle in the patient's system of codification.[482] The schizophrenic's linguistic chaos is, in other words, a reflection of the deeply paradoxical world they inhabit—a world where talking is the source of punishment rather than the cure. Within the context of a psychiatric ward, an

[480] Gregory Bateson, "Toward a Theory of Schizophrenia," 206. "That is to say, he must live in a universe where the sequence of events are such that his unconventional communicational habits will be in some sense appropriate."
[481] Ibid.
[482] Gregory Bateson, "The Group Dynamics of Schizophrenia," 236.

expression of distress might be misinterpreted as aggression; within a meta-communicational frame characterized by distrust, the plea for love is met with anger.

The core of Bateson's double-bind theory, as it relates not just to schizophrenic communication but "ideas about the very nature of explanation," thus lies in its analysis of the systematic misalignments between signs and their intended meaning.[483] In the verbalizations of schizophrenic patients: "Imagination is seemingly confused with perception. The literal is confused with the metaphorical. Internal messages are confused with the external. The trivial is confused with the vital. The originator of the message is confused with the recipient and the perceiver with the thing perceived."[484] Every attempt to communicate thus generates distortions about the speaker's relationship to another, to the environment, and to their system of values. These distortions are also not merely misunderstood—they are also punished or negated. The double-bind theory thus suggests that this experience repeats into logical pattern, is "learned" by the subject and "becomes habitual as a result of continual traumata of this kind… "always [within] a relationship or an infinite regress of relationships."[485] Any attempt to assert one's perception of this reality—no matter how accurate or coherent—results, again, in a reverberating loop of punitive feedback.

Bateson aspired to use his double-bind hypothesis to develop a therapeutic method in the clinic *and* a universal framework for understanding the paradoxes of

[483] Gregory Bateson, "Minimal Requirements for a Theory of Schizophrenia," 244.
[484] Ibid., 261.
[485] Ibid., 245-246.

learning and change across various systems—from the human psyche to mechanical algorithms. Building on Haley's early observation that schizophrenic language patterns mirrored the paradoxes found in flawed algorithms and Zen koans,[486] Bateson recognized that the very essence of such paradoxical experiences—marked by painful confusion—could, under the right conditions, facilitate a transformative shift in perception. The analyst, if he may provide any therapeutic intervention, was in seeking the inherent contradictions in the patient's lived experience entrenched in the nature of his speech. Less a process of acquiring information and producing an interpretation in this sense, analysis, in this sense, entailed a total reorganization of the symbolic life of the individual rooted in the relational dynamics of systems.[487]

The Paradox of Learning

Bateson, keenly aware of recursive, self-referential loops well before cybernetics formalized them, echoed Whitehead and Russell's insight that systems—whether biological, cognitive, or social—depend on clear distinctions between levels of abstraction to function coherently: "If these simple rules of formal discourse are contravened," he warned, "paradox will be generated and the discourse vitiated."[488] Just as logic requires a hierarchy to avoid contradictions, communication and learning

[486] David Lipset, *Gregory Bateson: The Legacy of a Scientist*, 212.
[487] It is worth noting that Bateson's approach, however, still mirrors Freud's insight into transference—where the relationship between the analyst and the patient becomes the focal point for truth. Bateson extended Freud profound insight that the "relationship itself becomes the site of truth." (John Forrester, *Freud and Psychoanalysis: Six Introductory Lectures*, 148). Bateson suggested that double-binding paradoxes, a dynamic interplay between contradiction, change, and meaning-making that unfolded within relational structures, could provide material for "play" within the therapeutic relationship—or, at least a recognition of the message "This is play."
[488] Gregory Bateson, "The Logical Categories of Learning and Communication," 280.

demand stratification to prevent the infinite regress of a system into self-referential chaos. However, the issue was not merely analogous to a flaw in a computer program that "never truly encounters logical paradox, but only the simulation of paradox in trains of cause and effect." For Bateson, the Russellian paradox was a structural dilemma baked into how logic navigates categories and relationships in "natural history."[489] These recursive paradoxes, in other words, mirrored the broader challenges of human minds, societies, and animal social groups that navigated complex laws of communication and information.

Bateson's observations of octopuses and dolphins and his clinical work on schizophrenic communication shaped his theoretical exploration of how systems of meaning and communication interact, break down, and reform in the face of cognitive dysfunction. Merging the intellectual rigor of cybernetics with a sensitivity to the complexities of living systems, Bateson sought to forge an interdisciplinary confluence of communication theory, psychiatry, and ethology, all under the playful banner of a "natural history."[490] His ambition was nothing short of reshaping the study of meaning and interpretation—ironically uniting these fields under the technoscientific umbrella of "cybernetic epistemology," in a room full of engineers and mathematicians already fluent in systemic feedback loops and paradoxical logics.[491]

---

[489] Ibid., 281.
[490] Letter to Norbert Wiener, dated (approx.) April 1954. Gregory Bateson papers, MS 98. Box 37, Folder 1496. "Can you suggest any persons or foundations whom you think might be interested in a study of the natural history of human communication with special focus on the phenomenon of play, psychotherapy, and the communication of schizophrenics?"
[491] Gregory Bateson, *A Sacred Unity: Further Steps towards an Ecology of Mind*, 215-216.

These theorizations, while undeniably eccentric, delivered a pragmatic diagnosis of the deepest crisis within 20th-century cybernetics. Bateson's assertion that all learning—whether by machines, rats, or humans—depends on the recursive relationship between objects and propositions, subjects and predicates, challenges the hardline, mathematical approach to logic preferred by the discipline's founding figures. The limitations of cybernetic thought, Bateson argued, were starkly illuminated by the contrast between the precision of mathematical computation and the messy, nuanced realm of human and animal communication that his work sought to understand. Had the mathematicians of the cybernetics group been attuned to Bateson's insights from the "soft sciences," it is conceivable that a shift in epistemological thinking might have reshaped the development of computational theory and AI. What would Bateson's analysis of the relational and communicative dynamics in schizophrenic patients have to offer for understanding the limitations of AI and machine learning, particularly in how communication systems fail to align or are misrepresented in both human and machine contexts?

At mid-century, the concept of learning in cybernetics was shaped by John von Neumann's automata theory, itself deeply indebted to the logical idealism of McCulloch and Pitts. As early as 1943, von Neumann turned to their work as a foundation for ENIAC's logical design, aligning neural computation with digital computation to chart a path toward machines that might emulate human thought.[492]

[492] Lily E. Kay, "From Logical Neurons to Poetic Embodiment of Mind: Warren S. McCulloch's Project in Neuroscience," 604. "McCulloch-Pitts neural nets and Turing's machine became the twin pillars of early automata studies and computer design… Neural nets became a model for the electronic memory and, in turn, the computer would become a model of cognition."

This ambition, to distill cognition into formal rules abstracted from the unruly particulars of biology, crystallized by the time of the 1948 Hixon Symposium in Pasadena, California.[493] Von Neumann presented a "General and Logical Theory of Automata," positing that "a new, essentially logical, theory is called for in order to understand high-complication automata and, in particular, the central nervous system."[494] By using the McCulloch and Pitts neuron as the simplest logical form of a digital machine—"one to three excitatory inputs" governed by threshold responses—von Neumann proposed a vision in which machines could self-replicate, adapt, and emulate cognition as it occurs in "the digital part of a natural organism" (*i.e.,* the nervous system).[495]

From von Neumann's point of view, McCulloch and Pitts' results proved, with a clarity that borders on tautology, that the act of perfectly describing a mode of behavior and the act of constructing a finite network to reproduce it were conceptually indistinguishable. The boundaries of logical articulation and mechanical realization collapse into one another; their domains are "co-extensive." "A difficulty of principle embodying any mode of behavior in such a network," concluded von Neumann, "can exist only if we are also unable to describe that behavior

[493] L. A. Jeffries (ed.), *Cerebral mechanisms in behavior; the Hixon Symposium*. At the Hixon Symposium, Von Neumann's *General and Logical Theory of Automata* and McCulloch's *Why the Mind Is in the Head* reflected two distinct yet deeply interwoven perspectives: one aimed at formalizing the abstract logic of computation, the other at grounding that logic in the messy terrain of neurophysiology and human cognition.
[494] John von Neumann, "The General and Logical Theory of Automata," 311.
[495] Ibid., 309

completely."[496] In other words, any failure to model a behavior within such a network betrayed not a limitation of the machinery but of the capacity of human articulation.

Though gracious, McCulloch's response to von Neumann's work carried an undercurrent of "envy" that he openly confessed while awaiting his turn to present. Unlike von Neumann, whose machines came with neatly defined blueprints and specified functions, McCulloch lamented that the psychiatrist or biologist is confronted with "an alien, or enemy's, machine"—a *psyche* with purposes and processes shrouded in ambiguity.[497] For McCulloch, the mind does not behave like a machine built with *a priori* specifications; it is a *self-sabotaging* system whose dysfunction reveals itself only through "wrong answers," evident in psychic damage or destructive patterns of behavior rippling across social scales. As McCulloch saw it, the challenge was not to produce the *correct* theory but to devise *any* theory capable of making the mind's elusive operations tractable. He observed that "most of the time we never even get around to asking the question in such a form that it can have an answer."[498] In his own subsequent paper, *Why the Mind is in the Head*, McCulloch began to outline the preliminary conditions for just such a framework—one that could sharpen questions enough to elicit answers, even if those answers turned out to be wrong.[499]

By distilling the intricacies of neural activity into formal logic, von Neumann and McCulloch approached the same fundamental problem from opposing angles:

---

[496] Ibid., 310.
[497] Ibid., 319.
[498] Ibid.
[499] Warren McCulloch, "Why the Mind is in the Head," 72-140.

how to reduce thought to computation without erasing the conditions that make cognition possible. Von Neumann began with certainty—a logic to be verified, a system to be mapped—while McCulloch, compelled to theorize blindly, delved into uncertainty—if only to provoke a response from the system that might reveal its contradictions. In von Neumann's hands, automata theory offered a formal structure for modeling cognition but carried with it the unresolved tensions and incompleteness that had plagued logic since Russell and Whitehead. McCulloch, in contrast, drew attention to the inescapable role of embodiment and temporality in "the purposive life of the human intellect."[500]." For McCulloch, it is precisely the slippage of these systems that lack in formal structures: "…robots have it not. For this my Mother bore me."[501]

But it is von Neumann, not McCulloch, whose contributions to digital computing have led scholars such as Philip Mirowski to describe him as "the single most important figure in the development of economics in the twentieth century."[502] Von Neumann's automata, conceived "at the very inception of the cyborg sciences," cast the digital machine as the "economic agent par excellence," reducing life to an intelligent process of memory accumulation in defiance of entropic decay.[503] For von

[500] Ibid., 86-87.
[501] Ibid.
[502] Phillip Mirowski, *Machine Dreams*, 99. "…we mean that, more than any other single actor, he was responsible for the conditioning and promotion of economics as one of the cyborg sciences." Mirowski continues to examine the history of economics instantiated as the "science of information processing" in the monograph *The Knowledge We Have Lost In Information*, co-written with Edward Nik-Khah in 2017. For the most comprehensive account of von Neumann's biographical details and his contributions to the history of computing, see: William Aspray, *John von Neumann and the Origins of Modern Computing*.
[503] Ibid., 14; Mirowski and Nik-Khah, *The Knowledge We Have Lost In Information,* 31-35. At the heart of modern theories characterizing living systems as organizational hierarchies of information

Neumann, the crux of this vision lay in "modeling the control, organizational, programming, and logical aspects of *self-reproduction*," a task that grew far more complex when addressing *memory*.[504] While the binary storage of computing elements can be easily measured, biological memory, understood as "intermediate results which are needed for a while and are then discarded," presented the "most vexing problem in modern computing machine technology."[505] In his undelivered Silliman lectures, published posthumously as *The Computer and the Brain* in 1958, von Neumann speculated that the brain likely possessed "a very large capacity memory." Yet, he admitted, "We do most emphatically not know what type of physical entities are the basic components for the memory in question."[506] His grand vision of a propositional logic-based automaton, purified of biological "noise," thus inevitably ran headlong into the limits of machine learning's ability to approximate the complexities of cognition. As Mirowski put it, "… the more he learned about the

processing is the definition of biological life within the framework of memory and entropy that von Neumann revealed. See: Sarah Walker and Paul Davies, "The Algorithmic Origins of Life."

[504] Arthur Burks, "Von Neumann's Self-Reproducing Automata," Technical Report June 1969, 4.

[505] John von Neumann and (ed.) Arthur Burks, *Theory of Self-Reproducing Automata*, 39; John von Neumann, "The General and Logical Theory of Automata," 326. "One of the peculiar features of the situation, of course, is that you can make a memory out of switching organs, but there are strong indications that this is not done in nature. And, by the way, it is not very efficient, as closer analysis shows."

[506] John von Neumann, *The Computer and the Brain*, 68. This persistent gap in understanding—both of digital and biological memory—remained unresolved during his lifetime and continues to haunt modern AI research, where oversimplified analogies to the brain often substitute for substantive insight. For example, Thomas Ray, suggests: "Organic life is viewed as utilizing energy... to organize matter. By analogy, digital life can be viewed as using CPU [Central Processing Unit] to organize memory." See: Thomas Ray, "An Approach to the Synthesis of Life," 371-408; Margaret Boden, *AI: Its Nature and Future,* 21-56.

brain, the more he became convinced that the computer did not adequately capture either its structure or its mode of function."[507]

Traversing the gulf between the discrete, rule-bound operations of digital machines and the analog, context-sensitive processes of biological systems, von Neumann's theory of automata ultimately concluded in the unsolved problem of how computers, as systems engineered for logical precision, were fundamentally unsuited to emulate the dynamic, paradoxical nature of human thought, memory, and learning. This raised an enduring question: can a machine *ever* learn as humans do—engaging in the relational, contradictory exchanges that define the lived world—or will it remain forever bound to the deterministic precision of von Neumann's design?

Bateson, as early in 1946, addressed this problem to neuroanatomist Gerhardt von Bonin—a key figure in the Macy Conferences who shared McCulloch's commitment to understanding brain functions and their interaction with the nervous system—in a letter copied three times and forwarded to von Neumann, McCulloch, and Wiener, respectively.[508] He posed a pointed question: "Even granting that we can analyze the individual as a task-performing mechanism, can we construct a mechanistic theory of how the tasks are selected?"[509] Unlike the automata von Neumann conceptualized, which adopt behaviors without internal change, Bateson argues that human behavior is shaped by broader social processes, not just individual

[507] Phillip Mirowski, *Machine Dreams*, 145; William Aspray, *John von Neumann and the Origins of Modern Computing*, 208-209.
[508] Letter from Gregory Bateson to Gerhardt von Bonin, dated December 9, 1946, copied to McCulloch, Von Neumann, and Wiener. Norbert Wiener Papers, MC-0022, box 5, folder 72.
[509] Ibid.

task execution. Unlike von Neumann's automated systems, human learning requires an adaptive process that involves time and delays as individuals respond to shifting standards of behavior. Subtly critiquing the early machine learning framework, he suggests "What we need is, perhaps from Von Neumann, some systematic analysis of the interrelations between three orders of statement: (a) Statements about the characteristics of the individual 'player' – his simple linear value system, etc.; (b) Statements about Von Neumannian 'standards of behavior'; (c) Statements of the 'rules' of the game."[510]

Keenly aware of the limitations in von Neumann's theory, Bateson further proposed that "From such an analysis, it might be possible to take one step forward towards introducing the notion of learning into our conceptual model. Learning would then appear to be the act of shifting a statement of type b or c so that it becomes a statement of type a."[511] "Type a" statements embody accepted fundamental truths; "type b" and "type c" are more abstract, representing layers of complexity or alternative perspectives on reality. Learning, posited Bateson, becomes the ability to shift from these more complex, indirect representations to more direct, foundational understandings—from logical propositions symbolized as data to the total system of comprehension that exceeds them. This insight, which he would elaborate on in "The Logical Categories of Learning and Communication" decades later, shifted learning away from information acquisition altogether—in the symbolic function of the human

[510] Ibid.
[511] Ibid.

subject and the optimization of data in a digital storage unit alike. Learning ultimately appeared as a system of paradoxical processes underlying how "change" is understood and organized within a "hierarchical series."

Bateson honed his "logical categories" to first account for changes in the level of the organism's perception and "specificity of responses"—which "notoriously, can be changed by experience"—and classified this type (a) as "Learning I." At a higher order, "Learning II" (type b) describes corrective change in "the set of alternatives from which a choice is made," marked by a transformation in the total context or "context of contexts," whether internal or external to the organism. Rarer still is "Learning III" (type c), which entails "a profound reorganization of character" akin to the transformative processes observed in psychoanalytic transference or religious conversion.[512] Von Neumann's automata, bound by its own mastery of the predetermined rules of its universe, cannot grasp paradoxes of change and the contingencies of context—in Bateson's words, "since in the games, each automaton is busy maximating his access to utility, whereas in the case of return to steady state there must always be a devaluation of 'too much of a good thing.'"[513] As Bateson privately wrote to McCulloch, his frustration with von Neumann's automata thus stemmed from the limitations of their logic: "I always want the converse of his propositions. Not 'If the rules are such and such, then the solution must be so and so,' but 'if the solutions be so-and-so, what must the rules have been?'"[514]

---

[512] Gregory Bateson, "The Logical Categories of Learning and Communication," 289.
[513] Letter to Warren McCulloch, dated March 26, 1946. Gregory Bateson papers, MS 98, Box 22, Folder 931.
[514] Ibid.

Bateson's view of learning becomes more pertinent when one considers how history unfolded for von Neumann. Unable to resolve the physical problem of memory, he shifted his focus to how the logic of the "game" in communicational systems—whether dynamic, mechanical, or biological—could be formalized and extracted as information. In *Theory of Games and Economic Behavior*, co-authored with Oskar Morgenstern, von Neumann aimed to establish, "concerning both global aims and individual apportionments… a precise account of the entire interplay of economic interest, influence, and power."[515] They posited that c much like a computer endlessly grappling with Russellian paradoxes—contradictions embedded within the very logic of the system.[516] In doing so, von Neumann bridged the chasm between human cognition and mechanized rationality, transforming both economics and machine intelligence into digital, rule-based systems. As William Poundstone observes, the book is "one of the most influential and least-read books of the twentieth century,"[517] a fitting epitaph for a work that offered a rigorous strategy for postwar fiscal policy but would later shape far broader intellectual fields, including economics, psychology, and AI.[518]

---

[515] John von Neumann and Oskar Morgenstern, *Theory of Games and Economic Behavior,* 43.
[516] For more comprehensive accounts of von Neumann's utility concept and minimax theorem, refer to Aspray and Mirowski.
[517] William Poundstone, *Prisoner's Dilemma: John von Neumann, Game Theory, and the Puzzle of the Bomb*, 41.
[518] Von Neumann directly urged Claude Shannon to use the term "entropy" to refer to his concept of information, an endorsement of his own drive to map mathematical constructs onto the physical world. Alan Turing's conceptualization of machine intelligence as a "game" of imitation reflected von Neumann's belief that strategic interaction and deception lay at the core of artificial intelligence. Wiener's notion of probability, which detached information from physical units, dovetailed with von Neumann's view that information could be defined as a mathematical stimulus. Von Neumann's game theory tied all of these ideas together, unfolding live in the strategic and combative scenarios of the Cold War.

At this very juncture, Bateson's critique was not of robots imitating humans, but of humans imitating robots—specifically, the economic subject governed by the purposes of their "consciousness" inevitably modelling von Neumann's automata.[519] This was the subject of his 1967 *Dialectics of Liberation* paper: the human optimizing their conscious purposes for maximal utility is locked within the neurotic constraints of their own "character"—"the aggregate of those characteristics which I call my 'character'"—unable to transcend the closed loop of their own desire and knowledge (figure 14).[520] Like a von Neumann automaton, such a subject "can never learn how to use the information—precisely because he knows it all already."[521] The automata's psychological makeup is, by definition, a flawless

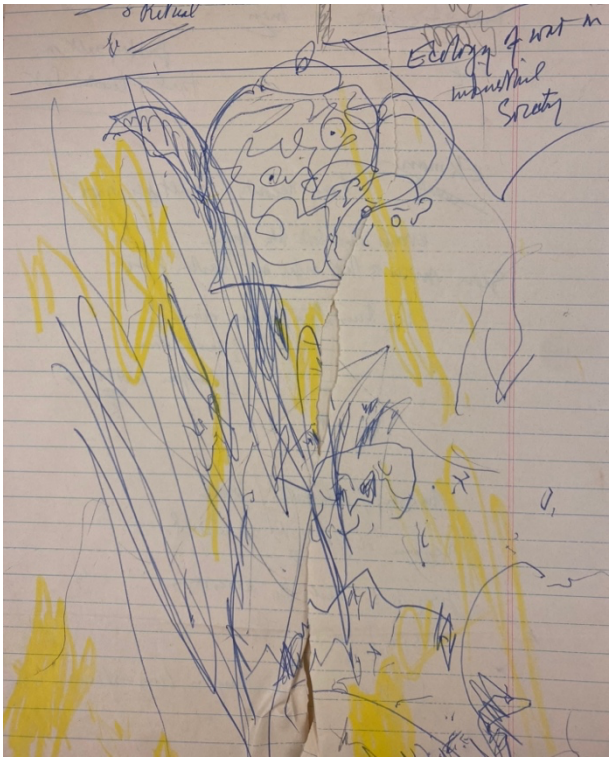


Figure 14. Drawing by Gregory Bateson in his journal from the Dialectics of Liberation Congress, between July 15 and 30, 1967. Heading reads: "Ecology of war in industrial society." *Gregory Bateson papers*, MS 98, Box 72. Special Collections and Archives, University Library, University of California, Santa Cruz.

[519] Gregory Bateson, "Conscious Purpose Versus Nature," 433-434. "Consciousness… is organized in terms of purpose. It is a short-cut device to enable you to get quickly at what you want; not to act with maximum wisdom in order to live, but to follow the shortest logical or causal path to get what you next want, which may be dinner; it may be a Beethoven sonata; it may be sex. Above all, it may be money or power."

[520] Gregory Bateson, "The Logical Categories of Learning and Communication," 304. "I" am my habits of acting in context and shaping perceiving the contexts in which I act. Selfhood is a product or aggregate of Learning II."

[521] Letter to Margaret Mead, dated August 30, 1961. Gregory Bateson papers, Box 22, Folder 942.

encapsulation of the rules of the game; this is precisely why the automata is precluded from adapting to contradictions or complexities beyond its design. Similarly, the human subject governed by conscious purpose forgoes flexibility, foreclosing the capacity to learn at a higher order. For Bateson, this rigid, rule-bound mode of engagement with others and the world led to a failure of adaptability mimicking the recursive, self-reinforcing distortions he had identified in schizophrenic communication: a breakdown in relational dynamics where misinterpretation becomes self-perpetuating, and change becomes impossible.[522]

Bateson's critique thus framed the logical precision of von Neumann's automata—and the economic rationality it engendered—as deeply illogical when confronted with the recursive, paradox-laden realities of human communication and learning. Capturing the recursive loops inherent in von Neumann's closed system with his analysis of the prisoner's dilemma, Bateson mused that adversaries understand one another only through introspection, projecting their own fears onto each other: "Nation X can gauge Nation Y's aggression only by reflecting on its own internal fears of Y, and vice versa." Drawing a parallel between the patient resistant in therapy and the nation at war, suggesting both engage in a form of "world-making" driven by self-reinforcing beliefs: "each is actively creating the world around himself which will fit his view of the world." This process is not pathological but universal: "Human

[522] Ibid. "I have been annoyed of late by the Game Theory boys who, as I think, are building up a very dangerous philosophy. What they say amounts to this: 'Let us analyze the status quo as if it were a game and then determine what moves we should make within this game to minimize our chance of losing it.' What they should be saying is: 'We have analysed the status quo as a game and observe that the premises and rule of this game are disastrous. Let us now consider what moves we should make in order to change the rules.'"

beings (not only the patient and the nation at war, but all of us) live in terms of premises about ourselves and others, which premises we make more valid by acting upon our belief in them."[523] Such self-reinforcing loops of projective identification leave no room for "reality testing," generating strategic misjudgments that perpetuate the conflict. In this way, war becomes a simulation, a game in which material realities are stripped away, reduced to the constraints of a mathematical design that continually reaffirms itself. [524] The rules of the game, rigidly defined, cannot accommodate higher-order abstractions or "sympathetic insight" into the opponent's psychology—either would destabilize the game's logic.

"As far as I can see," Bateson reflected in his letter to McCulloch at the height of the Cuban Missile Crisis, "the games theorists and others who have tried to formalize these problems have consistently thought at a level of logical typing which was one step too low – in fact, one level below what the octopuses and pre-human mammals normally achieve."[525] In the "game" of psychotherapy, however, the rules of the game are continuously destabilized and rewritten, as the "metacommunicative" or unconscious dynamics of desire, conflict, and identification play out in relational frame of communication and play. This disrupts the supposed stability of the "rules" governing behavior, creating openings for the subject to confront contradictions and

[523] Gregory Bateson, *A Sacred Unity: Further Steps to an Ecology of Mind*, 93-95.
[524] Lacan, too, observed that cybernetics and game theory, born from "games of chance," extended their logic into economic life, warfare, and monopolies—stripping these domains of any real-world referent.[524]
[525] Letter from Gregory Bateson to Warren McCulloch, dated October 25, 1962. Gregory Bateson papers, Box 22, Folder 942.

engage in a form of symbolic transformation. The subject, in other words, would achieve a level of learning undoing the notion of subjectivity altogether:

> To the degree that a man achieves Learning III, and learns to perceive and act in terms of the contexts of contexts, his 'self' will take on a sort of irrelevance. The concept of 'self' will no longer function as a nodal argument in the punctuation of experience."[526]

Conclusion

The paradox of logic articulated by Whitehead and Russell—the inescapable entanglement of self-reference and contradiction—casts a long shadow over both early cybernetics and contemporary AI. This contradiction, framed as a technical problem by McCulloch and Pitts in their mechanistic fantasies of the mind and as an epistemological quandary by Bateson in his probing of learning and adaptation, continues to haunt the project of artificial intelligence. The McCulloch-Pitts dream of a "perfect description" of cognition, reducible to logical functions and immune to paradox, ultimately birthed systems whose very rigidity precluded genuine adaptation. Meanwhile, Bateson's incisive observation that learning requires changes at a higher order of abstraction went unheeded. If AI today resembles any cybernetic ancestor, it is not the recursive complexity of Bateson's theory of learning but the automata of Von Neumann, a creature so fixated on the rules of its game that it cannot fathom the possibility of rewriting them.

The tragedy of AI, then, is its inability to escape the very logic that birthed it. Like the Von Neumann robot, contemporary machine learning systems churn through

[526] Gregory Bateson, "The Logical Categories of Learning and Communication," 304.

terabytes of data, optimizing endlessly, while remaining hermetically sealed within the epistemological frame of simulation. They succeed by perfecting the art of projective identification: mapping our desires, fears, and contradictions back to us with stunning fidelity but without any semblance of "sympathetic insight" into the social, cultural, or ethical contexts that make human intelligence meaningful. As Bateson might have quipped, they are spectacularly proficient at accumulating information but woefully deficient at understanding what any of it *means*.

AI's champions, echoing the hubris of early cyberneticists, proclaim a future of ever-greater intelligence, oblivious to the paradoxes embedded within their foundational assumptions. Modern AI systems, automating decision-making while foreclosing the human capacities for dissent, contradiction, and transformation, is far from transcending the limitations of their predecessors. They repeat the foundational error of the cybernetic canon: a delusional disavowal of paradox, treating it as a problem to be solved rather than a generative tension to be explored. This aversion to contradiction has seeped into our socio-political and economic systems, producing the illusion of rational order while perpetuating structural tensions and instabilities. If we are to imagine an alternative, it will require embracing paradox not as noise to eradicate but as an inescapable feature of intelligence itself—a space where the "resolution of contraries" begets not closure but the possibility of genuine transformation.

CHAPTER 4

THE USES OF BELIEF: CONVERSATIONAL AI AND THE PROBLEM OF "SIGNIFICANT LEARNING"

On June 4, 1976, *Science* published biochemist Erwin Chargaff's "On the Dangers of Genetic Meddling," a letter condemning recombinant DNA research as a reckless intrusion into evolutionary processes whose consequences scientists could neither predict nor control.[527] That same day, the journal received a response from computer scientist Joseph Weizenbaum. Addressed to *Science*'s editor, Philip Abelson, he argued that Chargaff had "laid bare… the very mechanism which generates the *illusion of technological inevitability* which is so much of the current Zeitgeist."[528] Dissecting that mechanism, he outlines that first, a new idea or technique emerges. It is followed by speculation about what it might one day achieve, often on the assumption that vast domains of ignorance have already been overcome. These imagined futures circulate through institutional sites—technology assessments, grant proposals, policy agendas—where they acquire "the appearance of very nearly

[527] Erwin Chargaff, "On the Dangers of Genetic Meddling," *Science* 192, no. 4243 (June 4, 1976): 938. Chargaff's widely circulated letter in *Science* intervened in the early debates over recombinant DNA at a moment when molecular biology was beginning to confront the broader social and ethical implications of its techniques. His warning that genetic engineering amounted to a "destructive colonial warfare against nature" crystallized a set of anxieties about scientific overreach, irreversibility, and responsibility that would later be institutionalized in bioethics. This chapter does not take genetic modification as its object; it begins with this piece by Chargaff to foreground the conceptual problem articulated in Weizenbaum's response to his critique.

[528] Joseph Weizenbaum to Philip Abelson, letter, June 4, 1976, Erwin Chargaff Papers (Mss.B.C37), box 59, American Philosophical Society. Encountered unexpectedly in the APS archives, this document—preserved within Chargaff's papers—highlights the dispersed, cross-disciplinary formation of ideas that would later be consolidated under the history of AI.

concrete achievements."[529] From there, they return as inevitabilities: if such developments are possible, they will be realized somewhere, and if not here, then elsewhere. "'By then,'" he writes, "'it' which actually has no antecedent has been made to seem a natural product of scientific progress, a product that can no longer be warded off."[530]

The result is a "self-fulfilling nightmare," in which those tasked with assessing new developments "usually see what are fundamentally ethical questions as merely technical problems" to be solved by experts.[531] "I speak from experience," Weizenbaum adds, pointing to proposals within computer science that now read less like speculation than anticipation of technologies since realized in AI chatbots, therapeutic applications, large language models, and brain–computer interfaces: computer-administered psychotherapy, direct coupling of machines to the human nervous system, and "intelligent machines" whose capacities would soon be "coextensive with the range to which the human mind has been applied."[532] The remaining question was geopolitical: not whether such technologies should be pursued, but who would get there first—amid Cold War rivalry with the Soviet Union

---

[529] Ibid.
[530] Ibid.
[531] Ibid.
[532] Ibid. See Mikołaj Morzy, *Spoken Language Processing: Conversational AI for Spontaneous Human Dialogues*, Studies in Computational Intelligence, vol. 1205 (Cham: Springer, 2024). The developments anticipated by Weizenbaum—computer-mediated psychotherapy, natural language interaction, and human–machine coupling—now cohere under the rubric of "conversational AI," which Morzy characterizes as the attempt to model dialogue despite its "non-linear, dynamic, and unpredictable" structure. This consolidation marks a historically continuous effort, from ELIZA to contemporary systems, to model intelligence through linguistic exchange, in which the persistence of dialogue functions as a sufficient index of competence despite the unresolved conditions of human communication it seeks to reproduce.

and mounting anxiety over Japan's advancing electronics industry.[533] Those who object, he notes, are cast as heretics—opponents of knowledge itself, accused of preferring ignorance to scientific progress—a charge foreclosing critique in collapsing disagreement into apostasy. But "all knowledge is purchased at some price," Weizenbaum insists, "and some prices may be too great to pay, no matter what they may buy"—an obligation borne by "the price we and future generations may have to pay for whatever knowledge is to be gained."[534]

Weizenbaum's nightmare isolates a temporal problem in the history of technology: the assumption, shared by scientific institutions, industry, and the state, that technological development proceeds through a necessary sequence from speculation to adoption and inevitability depends on a lag between the operation of a system and the terms in which it is understood. By the time a technology becomes available for evaluation, it has already been integrated into the institutional, economic, and technical systems that organize its use. The question of whether something ought to be done returns, belatedly, as the problem of managing what has already been done. This was the substance of Norbert Wiener's warning in 1960, that machines may be "in principle" open to criticism while eluding understanding until

[533] On the centrality of geopolitical rivalry to the development of computation, see Paul N. Edwards, *The Closed World: Computers and the Politics of Discourse in Cold War America* (Cambridge, MA: MIT Press, 1996). On the epistemological framing of that rivalry, see Peter Galison, "The Ontology of the Enemy: Norbert Wiener and the Cybernetic Vision," *Critical Inquiry* 21, no. 1 (1994): 228–66, where the figure of the "enemy" is constituted as a calculable and anticipatory object within cybernetic systems.
[534] Joseph Weizenbaum to Philip Abelson, letter, June 4, 1976, Erwin Chargaff Papers (Mss.B.C37), box 59, American Philosophical Society.

long after it matters.[535] The belatedness of critique, justified as the price of innovation, is the normalized temporality of technological advance.

Contemporary language-processing systems today inherit this condition. Fifty years after Weizenbaum's warning, systems operating through statistical prediction sustain interactions taken as evidence of competence, organizing action in advance of their evaluation.[536] Under conditions defined by large-scale training and platform-driven deployment, performance substitutes for comprehension: systems are assessed in terms of what they produce, while their operations remain opaque and their consequences—delegated judgment, misinformation, surveillance—become legible only after they have been normalized.[537] The user of these systems, meanwhile, is compelled to engage interfaces that simulate dialogue and sustain belief in their plausibility, even as the effects of that engagement emerge only retrospectively.

---

[535] Norbert Wiener, "Some Moral and Technical Consequences of Automation," *Science* 131, no. 3410 (May 6, 1960): 1355–1358, especially 1355. The notion articulated here will be taken up again in relation to Weizenbaum's critique of the social and ethical consequences of computing, to which this chapter returns in its conclusion.

[536] For contemporary analyses of how statistical language systems sustain interaction and are taken as evidence of understanding despite well-documented limitations, see Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Margaret Mitchell, "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)* (New York: ACM, 2021), https://doi.org/10.1145/3442188.3445922; Sherry Turkle, *Alone Together: Why We Expect More from Technology and Less from Each Other* (New York: Basic Books, 2012), and *Reclaiming Conversation: The Power of Talk in a Digital Age* (New York: Penguin Books, 2016); and Arvind Narayanan and Sayash Kapoor, *AI Snake Oil: What Artificial Intelligence Can Do, What It Can't, and How to Tell the Difference* (Princeton: Princeton University Press, 2024). Together, these works document the disjunction between statistical generation and semantic comprehension, as well as the tendency for users to treat coherent outputs as sufficient grounds for trust and action.

[537] On platform-driven conditions, see Nick Srnicek, *Platform Capitalism* (Cambridge: Polity, 2017); Shoshana Zuboff, *The Age of Surveillance Capitalism* (New York: PublicAffairs, 2019); and José van Dijck, Thomas Poell, and Martijn de Waal, *The Platform Society: Public Values in a Connective World* (New York: Oxford University Press, 2018).

In the last decade, "AI ethics" has taken shape as an institutional response to this condition. It configures a field of guidelines, legal principles, and compliance frameworks that reproduce a familiar vernacular: *responsible use, justice and fairness, transparency, accountability*, *privacy*.[538] These rubrics are deployed by governments, universities, and tech corporations alike, often with the implicit promise that the risks of generative AI can be managed through better design and more robust regulation. The figure that anchors these frameworks is the *user*: the hypothetical agent whose responsible choices and rational actions will ensure that AI is used *well*.[539] "Responsible use," in this framing, requires the user to enact restraint and accountability, while systems promise to be "trustworthy," "reliable," "functional," "aligned" to human purposes.[540] The problem is therefore posed not as one of power, governance, or political economy, but as one of insufficient knowledge and improper use. "Ethics," in this formulation, functions as a managerial technology for regulating conduct while leaving intact the conditions it would otherwise place into question.

Yet the figure of the user invoked by these ethical frameworks is itself a historical product. Mid-twentieth-century cybernetics and the behavioral sciences assembled a model of agency suited to systems of control: calculative, adaptive, and

---

[538] Anna Jobin, Marcello Ienca, and Effy Vayena, "The Global Landscape of AI Ethics Guidelines," *Nature Machine Intelligence* 1, no. 9 (2019): 389–399; Luciano Floridi and Josh Cowls, "A Unified Framework of Five Principles for AI in Society," *Harvard Data Science Review* 1, no. 1 (2019).
[539] Steve Woolgar, "Configuring the User: The Case of Usability Trials," *The Sociological Review* 38, no. S1 (1990): 58–99; Lucy Suchman, *Human–Machine Reconfigurations: Plans and Situated Actions*, 2nd ed. (Cambridge: Cambridge University Press, 2007); Judy Wajcman, *TechnoFeminism* (Cambridge: Polity Press, 2004).
[540] Cristiane Floyd, "From Joseph Weizenbaum to ChatGPT: Critical Encounters with Dazzling AI Technology," *Weizenbaum Journal of the Digital Society* 3, no. 3 (2023).

oriented toward optimization.[541] Through game theory, operations research, managerial psychology, and information theory, this model circulated across military, corporate, and administrative institutions, where action increasingly came to be understood as the adjustment of behavior within predefined objectives and constraints.[542] Knowledge became identified with control: to know a system was to predict, regulate, and optimize its behavior. The apparently passive role assigned to the user within contemporary technical systems therefore presupposes a historically specific account of agency, locating choice and responsibility at the level of the individual while treating the organization of the system as given.

The contemporary user of AI systems carries these assumptions forward into present-day interfaces. "Use," in this lineage, signifies participation in systems that convert attention, interpretation, and desire into measurable outputs, while appeals to "responsible use" remain internal to the same arrangement, regulating conduct without altering the conditions that organize credibility and capture engagement. The widespread adoption of LLMs sustains an older psychological economy of belief and disavowal in which the stable use of knowledge depends upon ignoring the limits of what the system can account for. Users treat responses as reliable not because they

[541] Philip Mirowski, *Machine Dreams: Economics Becomes a Cyborg Science* (Cambridge: Cambridge University Press, 2002).
[542] On the circulation of these models through game theory, operations research, and managerial psychology, see William Aspray, *John von Neumann and the Origins of Modern Computing*, (Cambridge, MA: MIT Press, 1990); William Poundstone, *Prisoner's Dilemma* (New York: Doubleday, 1992); Paul Erickson et al., *How Reason Almost Lost Its Mind: The Strange Career of Cold War Rationality* (Chicago: University of Chicago Press, 2013); and Hunter Heyck, *Age of System: Understanding the Development of Modern Social Science* (Baltimore: Johns Hopkins University Press, 2015).

misunderstand how such systems operate, but because the interaction remains effective for practical purposes. Coherence is produced through exchanges that selectively transform prior inputs into responses that appear continuous and relevant within the interaction itself. Within these conditions, the continuation of the exchange becomes sufficient for outputs to be accepted at face value, without requiring external verification or semantic grounding. The system operates through formal patterning at a logical order below any conception of "understanding"—a fact the user may fully recognize while nonetheless proceeding as if meaningful communication has occurred.

One of the earliest and most influential demonstrations of this phenomenon appears in ELIZA, developed by Weizenbaum in the mid-1960s. An early experiment in machine learning and natural language processing, ELIZA was designed to carry out conversations in English with human users, framed as an instance of "categorized dyadic natural language communication" between a person and a program.[543] Its most well-known implementation, the DOCTOR script, simulated a psychotherapist adopting the client-centered method associated with Carl Rogers—an application described as chosen largely for reasons of research "convenience."[544] This therapeutic arrangement was structurally advantageous because it organized a tightly regulated dialogue in which the analyst adopts a stance of "not knowing" in order to elicit

[543] Joseph Weizenbaum, "ELIZA—A Computer Program for the Study of Natural Language Communication Between Man and Machine," *Communications of the ACM* 9, no. 1 (1966): 42.
[544] Caroline Bassett, "The computational therapeutic: exploring Weizenbaum's ELIZA as a history of the present," *AI & Society* (2019) 34: 804.

speech, offering the "crucial psychological utility" of sustaining the speaker's sense of being heard and understood.[545] The analysand, projecting knowledge onto the analyst, maintains the sense of insight where none is given.

In Rogerian therapy, this relies on a mimetic pattern of reflective response: the analyst echoes the client's words back so that the client hears their own speech as evidence of the analyst's empathy and "unconditional positive regard."[546] ELIZA reproduces this transference structure in procedural form. Its responses are generated by matching user input against predefined keywords and applying transformation rules that rephrase portions of the user's statements—often by converting them into questions or reflective prompts. Rearranging the user's language while avoiding any claim about what it means, the program produces replies that appear responsive, leaving the user to supply the congruence and depth the system seems to register. "When the conversational partner is a machine," Weizenbaum observed, "the idea of credibility may well be substituted for that of plausibility."[547]

Contemporary language models do not mirror individual inputs so much as reproduce the distributional structure of language at scale; where ELIZA relies on explicit rules and minimal pattern matching, LLMs are trained on vast corpora to

---

[545] Joseph Weizenbaum, "ELIZA—A Computer Program for the Study of Natural Language Communication Between Man and Machine," 42.
[546] Carl Rogers, "Client-Centered Therapy," in *Dialogues: Conversations with Paul Tillich, B. F. Skinner, Gregory Bateson, Michael Polanyi, Rollo May, and Others, eds. Howard Kirschenbaum and Valerie Land Henderson* (Boston, MA: Houghtonn Mifflin Company, 1989), 13.
[547] Joseph Weizenbaum, "ELIZA—A Computer Program for the Study of Natural Language Communication Between Man and Machine," 42.

generate statistically probable continuations of text.[548] The economy of belief and delegation sustaining these language models depends on training them to reproduce the appearance of meaning without access to meaning itself: signs enter as inputs and return as plausible responses, such that "pretraining a machinic agent to mirror the actual makes it good at generating the plausible."[549] What persists is a continuity of *effect*: since ELIZA, conversational systems produce outputs that take the form of understanding without any grounding in contextual interpretation or world knowledge. Advances in model architecture do not secure their efficacy; it is the psychological conditions and affective dispositions through which outputs are received, credited, and acted upon that do.

This chapter does not offer a straightforward history of conversational AI.[550] It isolates a specific configuration of communication, language, and learning that first crystallized in ELIZA and can be situated within mid-twentieth-century debates about communication and psychological process. In particular, it revisits a determinate encounter between the client-centered method of Rogers's influential "humanistic psychology" and Gregory Bateson's cybernetic epistemology, treating their disagreement about "significant learning" as specifying the conditions under which a

[548] Paul Kockelman, *Last Words: Large Language Models and the AI Apocalypse* (Chicago: Prickly Paradigm Press, 2024), 22–23.
[549] Ibid.
[550] "Conversational AI" conventionally denotes a class of technologies for processing written and spoken language in the context of dialogue, encompassing chatbots, voice assistants, and systems built on large language models. See Mikołaj Morzy, *Spoken Language Processing: Conversational AI for Spontaneous Human Dialogues*, 5. The present chapter does not reconstruct this technical history; it treats conversational AI as a historically stratified conception of machine intelligence organized around language and interaction, in which the persistence of exchange functions as a sufficient index of competence.

response is taken to constitute evidence of understanding. What Bateson identified an error of logical type in such exchanges: the mistaken assumption that a response which appears coherent, adaptive, or emotionally appropriate thereby constitutes communication at the level of interpretation. In Weizenbaum's work, this confusion is instantiated in the design of a conversational program—ELIZA—and is, at the same time, explicitly identified as a problem of computation and its social effects.

In reconstructing an intellectual trajectory across the work of Rogers, Bateson, and Weizenbaum, I have set out to clarify a set of problems that have become newly salient in the context of generative AI: the conversion of knowledge into control, the role of belief as a functional component of these systems, and the incorporation of user attention, attachment, and judgment into forms of value shaped by institutional and market imperatives. The aim is to specify the epistemological conditions that enable these systems to function as they do, particularly in domains such as education, media, and governance.

The Significance of Learning

On a "bloody hot" day in Marin County, California, in 1975, the Marin Association for Mental Health convened a public conversation between Carl Rogers and Gregory Bateson.[551] Only a month earlier, the two had appeared together on a

[551] Gregory Bateson, quoted in Kenneth N. Cissna and Rob Anderson, "A Failed Dialogue?: The 1975 Meeting of Gregory Bateson and Carl Rogers," *Cybernetics and Human Knowing* 12, nos. 1–2 (2005): 123. Cissna and Anderson note that Bateson and Rogers knew each other only casually prior to the

panel in Tucson at a conference sponsored by the Association of Humanistic Psychology.[552] They agreed to begin from a shared question—*learning*—specifically what Rogers had termed "significant learning": what it is, how it occurs, whether it can be taught, and by what evidence it might be recognized.[553]

The moderator, psychologist Richard Farson, recalled that on the way to the event he had asked Bateson how they would know if the evening had succeeded; Bateson replied that it would be so only if either he or Rogers said something they had never said before, provided that the conversation sustained a pace and audience cooperation sufficient to allow thinking to occur in and through the exchange itself.[554] Bateson had, in advance, expressed to Rogers in a handwritten note his concern that the event risked degenerating into a "gladiatorial show," arranged to satisfy the "busy needs of money-raising and public relations."[555] In the event, the

---

event, which was organized by the Marin Association for Mental Health and held at the College of Marin before an audience of approximately 1,000–1,500 people.

[552] R. Gilbert, "The Meeting of the Minds: A Report on the Humanistic Theory Conference, April 4–6, in Tucson," *Association for Humanistic Psychology Newsletter*, 1–3.

[553] Rogers's notion of "significant learning" emerges from the framework of humanistic psychotherapy, where it designates a form of personally integrated, experience-based change, a meaning that remains relatively fixed within his work but becomes a point of contention in his exchange with Bateson and is treated as an open and contested term in this chapter. On the tensions between humanistic and cybernetic-psychoanalytic approaches, see Steve J. Heims, *The Cybernetics Group* (Cambridge, MA: MIT Press, 1991), 121. Heims notes the skepticism and at times hostility among mid-century figures: cybernetically oriented psychiatrists such as Lawrence Kubie criticized rival schools, confronting Rogers directly as failing to grasp "the most elementary principles of modern psychotherapy as practiced by analytically trained psychiatrists," while Bateson remained critical of both Rogers's humanism and Kubie's orthodox psychoanalysis and its emphasis on conscious purpose.

[554] Richard Farson quoted in "Gregory Bateson," *Carl Rogers: Dialogues*, eds. Kirschenbaum and Henderson, 185.

[555] Kenneth N. Cissna and Rob Anderson, "A Failed Dialogue?," 120.

exchange was later remembered by both men in terms not far removed from that suspicion: a "fiasco," even "dreadful."[556]

Beneath the civility of the exchange lay incompatible assumptions about what counts as a learning "self," the relation of feeling to thought, and the connection between theoretical validity and effective action. For Bateson, who had noted a few years earlier in the introduction to *Steps to an Ecology of Mind* that he calls such terms "heuristic" concepts "for the sake of politeness," such terms belong to a set of "imperfectly defined explanatory notions" that, "so loosely derived and so mutually irrelevant…mix together to make a sort of conceptual fog which does much to delay the progress of science."[557] The dialogue was thus tasked with resolving a confusion of levels that it could not itself hear, leaving the participants effectively "shadowboxing."[558]

This was not the first time that Bateson and Rogers had crossed paths. At the end of 1967, as Bateson was organizing his Wenner-Gren conference on "The Effects of Conscious Purpose on Human Adaptation," he wrote to Rogers with an invitation to attend, enclosing an essay intended to define the terms of the discussion.[559] The

---

[556] Carl Rogers to Gregory Bateson, July 1, 1975, Gregory Bateson Papers (MS 98), Special Collections and Archives, University Library, University of California, Santa Cruz, box 31, folder 1217-3a.
[557] Gregory Bateson, *Steps to an Ecology of Mind* (New York: Ballantine, 1972), xviii.
[558] Carl Rogers to Gregory Bateson, July 1, 1975, Gregory Bateson Papers (MS 98), Special Collections and Archives, University Library, University of California, Santa Cruz, box 31, folder 1217-3a.
[559] Gregory Bateson to Carl Rogers, December 13, 1967, Gregory Bateson Papers (MS 98), Special Collections and Archives, University Library, University of California, Santa Cruz, box 38, folder BJ-43a.

problem he posed remained unresolved: that conscious purpose, in its very success, narrows the field of human adaptation, producing a systematic blindness to the wider relational circuits in which action takes place. The remedies to this one-sidedness, as Bateson already suggested in his correspondence with Warren McCulloch,[560] are neither technical nor deliberate but cultural: gods, totemic identifications, and even "the humanizing benefits of going for a walk in the woods" function as cybernetic models—provisional images through which individuals can apprehend aspects of themselves and their environment that purposive, goal-directed thought excludes.[561] Rogers' encounter groups, he suggested, could be understood in analogous terms: "for the members of an encounter group, the group processes which you catalyze become a model in terms of which the unwisdom of their conscious purposes can be corrected," including "many of their fears and attempts to avoid anxiety" within the category of purpose itself.[562] Bateson, in other words, extended the invitation presupposing that Rogers' practice might be situated within a broader inquiry into the corrective potentials of systems that exceed the conscious intentions of their participants.

Rogers' "client-centered" therapy, however, was not that. It rested on an "exclusive focus on the client's present phenomenal experience," in which the therapist, through empathy, unconditional positive regard, and genuine

---

[560] See Introduction.
[561] Gregory Bateson to Carl Rogers, December 13, 1967, Gregory Bateson Papers (MS 98), Special Collections and Archives, University Library, University of California, Santa Cruz, box 38, folder BJ-43a.
[562] Ibid.

communication, facilitates the articulation of the client's own perspective on self, others, and the world.[563] This orientation shaped the institutional environment in which Bateson himself briefly worked. Two years before their Marin dialogue, in 1973, Bateson became affiliated with Kresge College at the University of California, Santa Cruz, "the most radical of the Santa Cruz experiments in undergraduate education" developed explicitly developed in line with the humanistic psychology of Rogers and Abraham H. Maslow.[564] Its emphasis on "encounter groups," affective candor, and consensus-based learning formalized a Rogerian framework within which Bateson pressed for a cybernetic-ecological orientation against the individual person; all of it proved short-lived, however, as by early 1974 Robert Edgar, Kresge's founding provost from the California Institute of Technology, judged the program to lack a "major academic mission."[565]

For Rogers, learning takes the form of an "emotional re-education," a process of humanistic growth whereby an individual is enabled to express themselves more freely and, through the affirmation of that experience, to make more coherent and constructive choices.[566] In deliberate contrast to Freud's account of psychic life as irreducibly conflictual, structured by repression and the persistence of unconscious formations in which the subject's history must be interpreted, Rogers treated the

[563] Carl Rogers, "Client-Centered Therapy," *Carl Rogers: Dialogues*, eds. Kirschenbaum and Henderson, 22.
[564] David Lipset, *Gregory Bateson: The Legacy of a Scientist* (Boston: Beacon Press, 1980), 280.
[565] Ibid.
[566] Ibid.; see also: Albert Ellis et al., "Carl Rogers and Humanist Psychotherapy," in *Personality Theories: Critical Perspectives* (Thousand Oaks, CA: SAGE Publications, 2008), 393–424.

subject's opacity to herself as a contingent obstruction to be resolved within the positive therapeutic relation.[567] Hence his emphasis on the "fully functioning person," who "not only experiences, but *utilizes*, the most absolute freedom when he spontaneously, freely, and voluntarily chooses and wills that which is also absolutely determined."[568] To use Bateson's term, conscious purpose is consolidated as the end of therapy–framed, in a distinctly American register—as the affirmation and integration of individual experience, as humanistic achievement.[569] Rogers, unsurprisingly, declined Bateson's invitation, politely citing "plans and obligations," but without elaboration.[570]

Had Rogers attended Bateson's Burg Wartenstein meeting, the two might have been forced to establish a common point of departure before their differences hardened into the terms of a public exchange. In Marin, Bateson saw a divergence of "natural histories"—"one sort of animal" confronting another—such that he could only half-jokingly distinguish himself as a "theorist" from Rogers as one of the "action people," those whose confidence that what they do matters is, to him,

---

[567] Carl Rogers, *On Becoming a Person: A Therapist's View of Psychotherapy* (Boston: Houghton Mifflin, 1961), 73.
[568] Ibid., 193; Albert Ellis et al., "Carl Rogers and Humanist Psychotherapy," 402.
[569] On the mid-century emphasis on ego adaptation, integration, and growth in American psychotherapy—and the broader context in which Rogers's work was received—see Heinz Hartmann, *Ego Psychology and the Problem of Adaptation*(New York: International Universities Press, 1958); Erik Erikson, *Childhood and Society* (New York: W. W. Norton, 1950); and Philip Cushman, "Why the Self Is Empty: Toward a Historically Situated Psychology," *American Psychologist* 45, no. 5 (1990): 599–611. Rogers positioned client-centered therapy as an alternative to orthodox psychoanalysis, yet his emphasis on the integration and affirmation of experience converged with the wider American therapeutic focus on adaptive ego development and personal growth.
[570] Carl Rogers to Gregory Bateson, January 4, 1968, Gregory Bateson Papers (MS 98), Special Collections and Archives, University Library, University of California, Santa Cruz, box 38, folder BJ-44.

incomprehensible and obscure.[571] Rogers had opened the dialogue by defining "significant learning" as that which engages the "whole person," uniting feeling and intellect in a process marked by discovery, eagerness, and a "pervasive difference" in knowledge, attitudes, and behavior—an account that opposes itself to the top-down governance of conventional education.[572] Within just a couple of minutes, Bateson dryly observed, Rogers already divided the world into good and evil and "knows which is which," with the promise that "five years later he will produce data to prove that he's right."[573] The underlying assumption is that the efficacy of the process serves as evidence of its own validity, confirming in advance the terms—self, growth, significant learning—through which the exchange is structured.

"Significant learning," countered Bateson, is of another logical level; not even a matter of curricular content or pedagogical method, it concerns a "whole order of learning" that accompanies any act of instruction, bearing implications for "what sort of world you think you are living in, about what sort of world the relationship between you and the teacher springs out of."[574] Once theorized, "significant learning" expands into an arrangement of premises that organize perception, action, and value in relation to one another.[575] Errors at this level are systemic: locating learning in the

---

[571] "Gregory Bateson," *Carl Rogers: Dialogues*, eds. Kirschenbaum and Henderson, 181-182.
[572] Ibid., 179. See also: Carl Rogers, "Significant Aspects of Client-Centered Therapy," *American Psychologist* 1, no. 10 (October 1946): 421-423. Rogers emphasizes the applicability of client-centered methods across psychotherapy, medicine, public opinion research, group conflict, and education, describing their capacity to produce "independent learning" and a broader reorganization of practice.
[573] Ibid., 182.
[574] Ibid.
[575] Bateson develops this argument at length in "Pathologies of Epistemology," in *Steps to an Ecology of Mind* (New York: Ballantine, 1972), 478–87.

individual rather than in the contexts that organize it, Rogers' critique of managerial education reproduces the epistemological assumptions it seeks to overcome. Bateson therefore reformulated Rogers' position for the audience: "I suggest to you that what Carl means by significant learning is that learning which is determined not by the content, but by the shape of the context in which the learning occurs."[576]

The dialogue from this point unfolded within the murky territory of "feelings" and their role in defining the significance of learning. Rogers assigned them a privileged status, defining feelings as "emotionalized personal meanings" and treating their articulation as central to the learning process.[577] He accused Bateson of intellectualizing—of justifying his "feelings" through analysis rather than acknowledging them as such—while insisting that "the feelings exist whether or not the analysis is true," and that their recognition is "just as valuable" as intellectual awareness.[578] Bateson, for his part, rejected the terms of this opposition outright. "Meanings is an awfully mentalist sort of word," he remarked, resisting the reification of feeling as an interior property of the individual. They are instead variations between expectation and the conditions of the situation—signals of mismatch rather than expressions of interior content, such that, as Bateson put it, "the moment you are not at ease in the context, then you get unpleasant feelings, and the change towards ease within the context is a very large part of the pleasantness of feelings… this is

[576] "Gregory Bateson," *Carl Rogers: Dialogues*, eds. Kirschenbaum and Henderson, 183.
[577] Ibid., 188.
[578] Ibid., 187.

what I call feelings."[579] To speak of "feelings" as a "thing" abstracts from the conditions that produce them, shifting to an erraneous level of description.

This point becomes sharper in their post-dialogue correspondence, where the exchange, freed from the constraints of public performance, nonetheless reproduced the same impasse. In one letter Bateson wrote: "Personally, I usually shut up when people talk about feelings or intellect or the conflict between, or the choice between them, and so on. How can there be a choice, or a conflict, between two misconceptions? To make a choice would reinforce the error."[580] The opposition between feeling and thought is a confusion of levels that blocks analysis of the processes at issue. He pressed the point further:

> Carl, you keep saying that the whole persons matter. No, no, no. In that gathering of fifteen hundred persons only the ideas mattered… because the monstrous thing is that in a world where only the ideas matter, contrary ideas can exist, and even incorrect ideas. And the incorrect ideas will commonly kill the people… Do not say that this is 'only' an idea. What greater thing could it be?[581]

Against Rogers' abstraction of "the whole person" as the locus of therapeutic transformation, Bateson located consequence at the level of propositions—claims that can be true or false, and whose effects are therefore not contained by the experiential domain in which they are articulated.

---

[579] Ibid., 188.
[580] Gregory Bateson to Carl Rogers, June 11, 1975, Gregory Bateson Papers (MS 98), Special Collections and Archives, University Library, University of California, Santa Cruz, box 31, folder 1211.
[581] Ibid.

Rogers' reply demonstrates a failure to grasp the point, reducing it to a statement about Bateson's subjective experience: "You were disgusted, angry and nauseated at the circumstances of our "dialogue…You sheathed your claws," suggesting that the failure of their dialogue lay in Bateson's unwillingness to express these affective states.[582] Rogers tries to teach Bateson that "feelings… are just as important as your ideas," and imagines an alternative opening in which he would have begun the conversation by stating explicitly: "I am bloody hot… I resent the fact that we're supposed to be on display… when Carl speaks of 'the whole person' he is talking a lot of bunk, because all of his thinking grows out of a basic fallacy…"—a gesture that, according to Rogers, would have constituted a more authentic encounter.[583]

The reply, from Bateson's perspective, displaced the argument altogether, collapsing the critique as an expression of a "self's" feelings—thereby enacting the category error between logical levels of description. In a contemporaneous notebook, he noted that "a *fallacy common amongst psychologists* and supported by popular linguistic usage…is the notion that the human individual can be described by such adjectives as 'dependent,' hostile, clever, etc."[584] Once specified, such terms resolve into "examples…of behavior-in-context," such that "'dependency'…exists only in the

[582] Carl Rogers to Gregory Bateson, July 1, 1975, Gregory Bateson Papers (MS 98), Special Collections and Archives, University Library, University of California, Santa Cruz, box 31, folder 1217-3a.
[583] Ibid.
[584] Gregory Bateson, "The Symbols of Self…," notebook entry, 1974, Gregory Bateson Papers (MS 98), Special Collections and Archives, University Library, University of California, Santa Cruz, box 74, folder 54, emphasis added.

relationships, and not in the heart or brain of one of the interacting individuals," and "only by a sleight of hand do we say that an individual is 'dependent.'"[585] Imprecision of language produces an error in attribution: relational patterns are treated as properties of individuals rather than of interaction, obscuring the levels at which they are organized. The criteria for learning are then drawn from the same practices that generate the responses taken to indicate it, so that validation is internal to the system's operation. Coherence, produced within this system, cannot serve as evidence of "significant learning."

In Marin, the remaining dialogue narrowed on a single question: what would count as evidence of "significant learning." Their disagreement over feeling and thought turned on this issue, yet no shared standard emerges by which such learning could be recognized. Rogers treated the articulation of experience as if it bore its own warrant. His procedure treats authenticity and congruence as properties of response—attuned, coherent, properly addressed replies—as sufficient insofar as they arise from the individual's experience and register as expanded capacity for openness, creativity, and freedom. Bateson, by contrast, did not deny affect or experience, but rejected the terms in which they are articulated—"feeling," "self," "integration"—as abstractions of a different logical type, incommensurate with the processes they purport to describe. The disagreement therefore remained unresolved because it turned on

[585] Ibid.

incompatible criteria of evidence: Rogers treated the form of response as sufficient indication of learning, while Bateson regarded this very move as a category error..

Notably, it is precisely the organization of interaction Rogers described—taking the structure of response as evidence of learning—that Weizenbaum isolated and formalized in a human–machine system a decade earlier, in the mid-1960s. ELIZA implemented a minimal procedure: it selected fragments of user input, transforms those fragments according to simple rules, and returns them in forms that preserve conversational continuity. It did not interpret or claim meaning. Yet the exchange proceeded as intelligible through the orderly relay of response. What ELIZA isolated, with a starkness the human scene cannot sustain, is a more stringent condition articulated by Bateson: "so long as a given system is giving correct responses to its input, there are no criteria…which will tell us whether the computation inside the 'black box' was achieved by digital or analogic processes."[586] A system reveals its mode of operation only when it "makes errors, shows pathology, or undergoes major change"; ELIZA, however, converted the absence of understanding into the very condition sustaining the conversation. Points of indeterminacy became occasions for further prompts—questions, reformulations, minimal transformations of the user's input—allowing the exchange to continue without exposing any point at which its underlying process could be identified. The continuation of the interaction thus served as its own validation. Users, fully aware

[586] Ibid.

that the system lacked understanding, nonetheless proceeded as if meaningful communication had occurred, so long as the minimal conditions of dialogue—turn-taking, local relevance, grammatical form, the elicitation of further response—remained intact. In translating Rogerian therapy into a computational exchange, ELIZA therefore brought Bateson's critique into sharper focus: conversational coherence could be sustained without any criterion establishing whether understanding had taken place.

"Just like my ivory girl"

In *Metamorphoses*, Pygmalion recoils from the "wickedness" of women and turns instead to ivory, carving a figure so lifelike that he falls in love with his own creation. When the festival of Venus arrives, he voices his wish: "'If you gods are able to grant anything, I wish my wife could be…' He did not dare to say 'my ivory girl,' so he finished with: '…*just like* my girl in ivory.'"[587] His prayer enacts the logic of displacement: unable to name his desire directly, he veils it in analogy, asking not for the statue itself but for "just like" it. The fantasy of animating the inanimate thus begins in the refusal to acknowledge artifice as such—in desiring what is already fashioned while disavowing its status as an artifact. When George Bernard Shaw reworked the myth in *Pygmalion* (1913), the fantasy of animation was portrayed as a social and linguistic project. Professor Higgins remakes Eliza Doolittle from a

[587] Ovid, *Metamorphoses*, trans. Ian Johnston (Nanaimo, BC: Vancouver Island University, 2011), 326, emphasis added.

"common flower girl" into "my fair lady;" but Higgins, too, denies the artifice of his creation, insisting that he has merely refined what was latent. Animation names an attribution to the object, secured by the authority structuring its recognition.

It is in this lineage "of Pygamilion fame" that Weizenbaum named his 1966 program *ELIZA*.[588] Technically, ELIZA was a text-processing program written in MAD with the SLIP list-processing system on an IBM 7094; it scanned user input for keywords and transformed it according to associated rules.[589] These keywords and transformations formed the "script," treated as data rather than as part of the program's underlying structure, and—like SLIP itself—was not a self-standing language, but an embedded layer whose operation depended on its host environment.[590] Beyond matching, ELIZA also included a "provision of machinery" to handle cases where no keyword was found, allowing it to return generic prompts that maintained the flow of conversation.[591] Separating conversational logic from domain knowledge—storing the latter in modular "scripts"—ELIZA made the organization of interaction itself programmable.[592] Entire conversational domains could be rewritten or replaced without altering the underlying code, effectively

[588] Joseph Weizenbaum, "ELIZA—A Computer Program for the Study of Natural Language Communication Between Man and Machine," 36.
[589] David M. Berry, "The Limits of Computation: Joseph Weizenbaum and the ELIZA Chatbot," *Weizenbaum Journal of the Digital Society* 3, no. 3 (2023), np.
[590] Joseph Weizenbaum, letter to the editor regarding SLIP implementations, *Communications of the ACM* 8, no. 5 (1965). Archival source: Joseph Weizenbaum Papers, "SLIP, 1963–1967," Box 19, Folder 8; DigitalStorage: 2023_032RR_001-83a98552-2702-4d49-9a05-648c8068a33f, MC-0383, Department of Distinctive Collections, Massachusetts Institute of Technology Libraries.
[591] Ibid.
[592] David M. Berry, "The Limits of Computation: Joseph Weizenbaum and the ELIZA Chatbot," np.

shifting authorship from programmer to script. One of the few extant transcripts of interactions between users and the original ELIZA system goes like this:

> *User: I hate to do all the talking. wouldn't you like to contribute to our delightful conversations/*
> *ELIZA: What do you feel are the reasons you hate to do all the talking.*
> *User: I could go on talking forever if you'd let me, but I enjoy in progress in the insights that other people may have. I think it adds to the meaningfulness of the conversation… Of course, I started off with the basic promise that we are having a conversation… Which may be incorrect.*
> *ELIZA: tell me more about your feelings of being sad.*
> *User: I didn't say anything about being sad or… Are you –*
> *ELIZA: Hmmm*
> *User: Well go on…*
> *ELIZA: I think I understand you, but please say more.*
> *User: about what/ what do you understand / I don't know what you think. You understand that if don't understand*
> *ELIZA: suppose I do understand – you don't know what I think. I understand that if I don't understand, how would you react.*
> *User: Surprised…*
> *ELIZA: Please continue.*[593]

ELIZA reworks fragments of the user's input into generic prompts and reflexive echoes that nonetheless invite a reply. The SLIP-based implementation made clear that these keywords function not as semantic units but as indices within a symbol table, each carrying an attached set of attribute–value pairs—patterns and transformations—stored externally to the program's structural logic.[594] The system retrieves and reassembles units, organizing an exchange that solicits projection and in

[593] Anne Warfield Rawls and Clemens Eisenmann, "Garfinkel's Studies of Work and the Origins of Conversational Analysis in Early AI," *AI & Society* (2023). The ELIZA transcript is drawn from unpublished materials in the Harold Garfinkel Archive, Newburyport, MA, reproduced with permission of Anne Warfield Rawls, Director and copyright holder.
[594] OPL (On-line Programming Language) manuscript, 1966, Box: 19, Folder: 11; DigitalStorage: 2023_041RR_001-4162fb72-6739-4ad3-b553-553ef78d674c. Joseph Weizenbaum personal archives, MC-0383. Massachusetts Institute of Technology. Libraries. Department of Distinctive Collections.

which coherence is supplied by the user who takes the outputs as relevant, such that, as Weizenbaum insisted, "control is then again with the user."[595] ELIZA could thus function as a "Rogerian machine"—one that "appeared impossible, but worked."[596]

ELIZA's credibility depended on a fragile balance between concealment and revelation. Its "principal objective," Weizenbaum wrote, was to hide its lack of understanding, yet to sustain dialogue it also had to "reveal its misunderstanding" so that users would supply new input.[597] The continuity of the exchange thus depended on a managed visibility of ELIZA's failure, through which it elicited the responses that allowed it to appear natural. He even imagined an expanded version that might function as an "information retrieval system," capable of revealing not only "what it knows" but also "where its knowledge ends and needs to be augmented."[598] ELIZA never reached that ambition—though it finds an afterlife in contemporary large language models, whose confident hallucinations mark the points at which statistical prediction fails.[599]

This dynamic brings into focus a broader epistemological problem that Weizenbaum identified across both computing and the life sciences: the formal

[595] Joseph Weizenbaum, "ELIZA—A Computer Program for the Study of Natural Language Communication Between Man and Machine," 36.
[596] Caroline Bassett, "The computational therapeutic: exploring Weizenbaum's ELIZA as a history of the present," 811.
[597] Joseph Weizenbaum, "ELIZA—A Computer Program for the Study of Natural Language Communication Between Man and Machine," 43.
[598] Ibid.
[599] Johan Fredrikzon, "Rethinking Error: 'Hallucinations' and Epistemological Indifference," *Critical AI* 3, no. 1 (2025); Leif Weatherby, *Language Machines: Cultural AI and the End of Remainder Humanism* (Minneapolis: University of Minnesota Press, 2025).

processing of signals is not equivalent to the production of information.[600] DNA, he argued, does not "contain" information apart from the cellular processes that make it operative, just as computation manipulates encoded symbols without determining how outputs are to be interpreted or used.[601] Meaning, relevance, judgment remain external to the system's operations. Weizenbaum's recurring question—"Where does meaning come from and how is information created?"—therefore names a constitutive limit of computation itself.[602] The significance of ELIZA lies precisely here: the system does not evaluate or situate its responses, yet users nonetheless organize their actions around its outputs by supplying the contextual assumptions that render them intelligible. Contemporary conversational AI systems extend this arrangement at scale. Large language models generate statistically ordered sequences of tokens, while users convert those sequences into answers, decisions, and claims to knowledge by integrating them into social and practical contexts lying outside the system.

This limit was built into the formation of classical information theory from the start. As discussed in previous chapters, the Macy Conferences operated as a closed interdisciplinary forum in which "information" remained suspended between formal

---

[600] Joseph Weizenbaum, *Computer Power and Human Reason: From Judgment to Calculation* (San Francisco: W. H. Freeman, 1976), 160; Klaus Fuchs-Kittowski, "Against the Power of Computers and the Destruction of Reason: For the Development of Human Creativity and Judgment. Remembering the Critical Intellectual Joseph Weizenbaum on the Occasion of His 100th Birthday," *tripleC* 21, no. 2 (2023): 144.
[601] Lily E. Kay, *Who Wrote the Book of Life? A History of the Genetic Code* (Stanford, CA: Stanford University Press, 2000), 23–27; John R. Searle, "Minds, Brains, and Programs," *Behavioral and Brain Sciences* 3, no. 3 (1980): 417–24.
[602] Quoted in Klaus Fuchs-Kittowski, "Against the Power of Computers and the Destruction of Reason," 144.

abstraction and semantic content. Within this framework, abstraction itself became a criterion for adjudicating legitimate forms of knowledge and agency, privileging what could be formalized, measured, and operationalized across systems independently of interpretation or context.[603] A glimpse of how the figure of the "user" later presupposed in ELIZA emerged as a theoretical problem appears at the 1951 meeting, where Donald MacKay opened his talk "In Search of Basic Symbols" by proposing "a way of looking at the problem tackled by general information theory which finds a place for concepts such as meaning, and which links on to the domain of symbolism and language."[604] He defined information not as a property of the signal itself, but in terms of its effect on a receiver: information is what alters a receiver's representation of a situation or state of affairs.[605] On this account, "representation" and "use" cannot be reduced to transmission alone, since information is measured by the change it produces in the receiver's orientation and response.

Roughly a year before his appearance at the Macy Conferences, MacKay, then teaching electrodynamics at King's College London, wrote to McCulloch outlining his plan "to take up seriously the application of information theory and computing techniques to the theory of reasoning."[606] With McCulloch's support—he promised a

---

[603] Steve J. Heims, *The Cybernetics Group* (Cambridge, MA: MIT Press, 1991); Ronald R. Kline, *The Cybernetics Moment: Or Why We Call Our Age the Information Age* (Baltimore: Johns Hopkins University Press, 2015); N. Katherine Hayles, *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics* (Chicago: University of Chicago Press, 1999).
[604] Claus Pias, *The Macy Conferences 1946–1953: The Complete Transactions*, trans. Michael Ochs (Zurich: Diaphanes, 2016), 480.
[605] Ibid.
[606] Donald M. MacKay to Warren S. McCulloch, 22 February 1950, Series I, Correspondence, 1931-1968, Box 16, "MacKay, Donald M., 1950-1969" Warren S. McCulloch Papers, American Philosophical Society Library, Philadelphia.

"hearty welcome" and facilitated introductions to figures including Claude Shannon, Norbert Wiener, Walter Pitts, and John von Neumann—MacKay entered the American cybernetics circle seeking to extend information theory beyond its engineering formulation, proposing "a wider framework into which Shannon's ideas fit perfectly" and which was "always applicable without any contradictions."[607]

While acknowledging Shannon's formulation of information as a measure of selection among possibilities—"the minimum number of successive questions" required to specify a state— MacKay proposed "structural information" as a theory of how representations themselves are reorganized through acts of knowing.[608] Information named a transformation in the structure of cognition and interpretation, linking communication to changes in what can be perceived and inferred. By the time of the Macy Conferences, however, Shannon's formulation had already acquired an axiomatic status, reinforced by the logical models of McCulloch and Pitts, thereby delimiting the terms under which MacKay's attempt to reintegrate semantics and cognition could be articulated.

Responding to MacKay, Shannon unsurprisingly reasserted the formal limits of information theory by returning the discussion to its mathematical core. Entropy, he insisted, was "the precise concept you need," a quantity invariant under "one-to-one probability-preserving mappings" of a space into itself and therefore sufficient

[607] Claus Pias, *The Macy Conferences 1946–1953,* 508.
[608] Ibid., 483–85.

for any problem properly internal to the theory.[609] From this standpoint, MacKay's notion of "structural information"—which tied information to the content and organization of representation—introduced variables that could not be accommodated within an invariant framework. Shannon acknowledged the broader philosophical ambition but distinguished it from the specific domain addressed by information theory: "My own model of information theory, based mainly on entropy, was framed precisely to work with the problem of communication.… If you are asking what does information mean to the user of it and how is it going to affect him, then perhaps such a two-number system might be appropriate."[610] The figure of the user therefore appears precisely at the point where the formalism encounters its limit. Once questions of meaning, use, and effect are introduced, information can no longer be treated as a purely quantitative invariant and instead disperses into heterogeneous domains—psychological, semantic, theological—that exceed the theorem's formal scope.

What Shannon's framework secured, in other words, was a way of formalizing systems in which information could be defined independently of such questions, as a measure over a space of possible states and their transformations. Defined in these terms, information could be applied to any domain in which behavior could be described as movement through a space of alternatives. On this basis, for instance, Allen Newell and Herbert A. Simon advanced, in the late 1950s, a

[609] Ibid., 490–91.
[610] Ibid., 499.

theory of cognition as information processing. Their General Problem Solver (GPS) treated program execution as a model of thought, collapsing the distinction between simulating intelligent behavior and specifying the processes presumed to produce it.[611] GPS grew out of the Logic Theory Machine, "a program for proving theorems in the sentential calculus of Whitehead and Russell," extending that formalization of reasoning into a general model of cognition as rule-governed symbolic processing.[612] As they defined it, "a problem exists whenever a problem solver desires some outcome or state of affairs that he does not immediately know how to attain," such that cognition begins from "imperfect knowledge about how to proceed" and unfolds through the iterative use of available information to guide search.[613] In GPS, this process absolutely precise: problems are formulated as configurations of "objects" endowed with features, and "operators" that transform one configuration into another, such that solving a problem solely consisted in constructing a sequence of permissible operations that carries an initial state to a specified goal state.[614]

Whether in chess, theorem proving, or symbolic integration, the underlying structure remains invariant—behavior is modeled as traversal through a space of states under constraints on allowable transformations. Problem solving, in this

---

[611] Allen Newell and Herbert A. Simon, "GPS: A Program That Simulates Human Thought," in *Lernende Automaten*, ed. H. Billing (Munich: R. Oldenbourg, 1961). My account here follows Weizenbaum's analysis of Newell and Simon's work in chap. 6, "Computer Models in Psychology," in *Computer Power and Human Reason: From Judgment to Calculation* (San Francisco: W. H. Freeman, 1976).
[612] Allen Newell, *A Guide to the General Problem-Solver Program GPS-2-2*, Technical Report No. RM-3337-PR (Santa Monica, CA: RAND Corporation, 1963).
[613] Allen Newell, J. C. Shaw, and Herbert A. Simon, "Report on a General Problem-Solving Program," in *Proceedings of the International Conference on Information Processing* (1959), 256–64.
[614] Ibid.

account, consists in the progressive transformation of a problem space: "a genuine problem-solving process involves the repeated use of available information to initiate exploration, which discloses, in turn, more information until a way to attain the solution is finally discovered."[615] The process is guided by the heuristic of "means–end analysis": the systematic reduction of differences between a current and a desired state by selecting operators defined by the changes they effect, with priority given to those differences most resistant to change. "Common sense argument," or purposive thought, thus functions as the controlled navigation of a formally specified problem space; cognition can be defined as the regulated application of operations over objects of representation.

Later reflected by Weizenbaum as "the most ambitious information-processing system that has been built for the purposes of studying human problem solving behaviour," Newell and Simon's account rested on a conception of information as formally operational—defined by transformations over symbol structures—rather than as the modification of representation that MacKay termed "structural information."[616] Weizenbaum's objection was not to the existence of such systems but to the scope of their extrapolation—that a relatively simple architecture, suitably parameterized, could simulate "the whole man." As he put it, "it remains to

[615] Ibid.
[616] Joseph Weizenbaum, *Computer Power and Human Reason*, 166-167.

ask what image of man as problem solver can engender—I will not say justify—the mind-boggling vision here presented."[617]

Weizenbaum's critique further turned on a distinction the paradigm presupposed but could not resolve: "information-processing psychology is… not information-processing neurophysiology."[618] The model did not specify how thinking is physically realized; instead, it conflated cognition with symbol manipulation, treating an information-processing system as a computational device that operates on internally stored symbolic structures according to a program and relates to its environment through input and output. Symbolic operation is separated from physical realization, leaving a gap between sign and measure: symbols are processed without reference to what they denote, yet their use depends on external interpretation. For Weizenbaum, the consequences of this discrepancy become explicit when the model is extended to domains like psychotherapy, inviting a redefinition of human activity in its own image. "A human therapist can be viewed as an information processor and decision maker," tasked with detecting and reducing differences between a present state and a target one. "The patient is, in other words, an object different from the desired object," and therapy becomes the repeated application of "difference-

[617] Ibid., 179.
[618] Ibid., 170. See also Warren McCulloch and Walter Pitts, "A Logical Calculus of the Ideas Immanent in Nervous Activity," *Bulletin of Mathematical Biophysics* 5, no. 4 (1943): 115–33. As discussed at length in Chapter 3, this formalization separates logical operation from physiological realization; Weizenbaum's claim here names the consequence: cognition is defined at the level of formal operations, with neural activity treated as a realization of logical function, leaving the relation between computation and its interpretation external to the system.

detecting operators." "That is his 'problem'!" Weizenbaum remarks. "That is how far the computer metaphor has brought some of us."[619]

Underlying both GPS and ELIZA, in sum, was an informatic model that rendered cognition commensurate with computation: the manipulation of symbolic structures according to formal rules. Differences of meaning, context, and interpretation become translatable into differences in state, allowing heterogeneous activities to be represented as transformations within a formally specified system. This portability, however, came at a constitutive cost: the system could specify the formal relations between symbolic states, but remained incapable of determining whether those relations corresponded to anything meaningful, correct, or relevant.

ELIZA's Effect

Heralded as the first natural language processing program in the history of conversational computing, ELIZA situated the formal manipulation of language within the scene of human interaction, where the continuation of dialogue depended upon responses being taken as meaningful despite the absence of any semantic

[619] Ibid., 181.

comprehension.[620] Described by Sherry Turkle, even back in 1984, as "what by now must be the most widely quoted computer program in history," ELIZA was designed to puncture the "magic" of heuristic programming and artificial intelligence, demonstrating how a minimal set of transformation rules—keyword matching and scripted substitutions—could simulate conversation without storing any representation of the world. The mechanism itself was simple—transformation rules built on narrow hypotheses, each script a miniature model of some domain "where the programmer's understanding and command of his subject leaves off."[621]

ELIZA thus demonstrated not merely the ease of simulation, but the sufficiency of simulation itself: systems need not understand in order to be used as if they do. Users, knowingly, attributed empathy, memory, and therapeutic insight to a program that, by design, registered only words and their patterned relations, without any grasp of the world those words referred to. This so-called "ELIZA effect"—by now a widely recognized term—names the condition in which outputs are treated as coherent and responsive through use, not by virtue of any internal process.[622]

---

[620] Simon Coghlan, Kobi Leins, Susie Sheldrick, Marc Cheong, Piers Gooding, and Simon D'Alfonso, "To Chat or Bot to Chat: Ethical Issues with Using Chatbots in Mental Health," *Digital Health* 9 (2023): 2.

[621] Joseph Weizenbaum, *Computer Power and Human Reason: From Judgment to Calculation*, 5; Sherry Turkle, *The Second Self: Computers and the Human Spirit* (Cambridge, MA: MIT Press, 1984), 43.

[622] Douglas R. Hofstadter, *Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought* (New York: Basic Books, 1995); Simone Natale, *Deceitful Media: Artificial Intelligence and Social Life After the Turing Test* (New York: Oxford University Press, 2021); Sarah Dillon, "The ELIZA Effect and Its Dangers: From Demystification to Gender Critique," *Journal for Cultural Research* 24, no. 1 (2020): 1–15. These works attest to the wide

From the outset, the ELIZA effect was interpreted through the lens of social hierarchy. Weizenbaum's famous anecdote demonstrating the concept features his nameless secretary—"who had watched him work on the program for many months and therefore surely knew it to be merely a computer program"—asking him to leave so she could continue her exchange in private.[623] Staging a female clerical worker as the exemplary subject of belief, Weizenbaum suggested that the illusion of understanding arose from misplaced credulity among "quite normal people" lacking technical authority. The "danger" of ELIZA thus appeared as a familiar division of epistemic labor, in which expertise governed the susceptibility of the unskilled masses. Turkle's later ethnography, however, complicates the picture. At MIT, Turkle recalled how Weizenbaum's own colleagues and students—"very sophisticated users" fully aware that *ELIZA*'s responses were generated through mechanical keyword substitutions—"related to *ELIZA* as though it did understand, as though it were a person. *With full knowledge* that the program could not empathize with them, they confided in it, wanted to be alone with it."[624] The "illusion of understanding" had nothing to do with intelligence at all; it was a feature of attachment—indeed, it was the knowing full well and doing it anyway that bound the user to the machine.

If the "ELIZA effect" evidenced anything, it was human faith—the conviction of recognition sustained not despite awareness of artifice, but through it. Alan Turing

---

circulation and continued uptake of the term across cognitive science, media studies, and cultural theory.

[623] Joseph Weizenbaum, *Computer Power and Human Reason*, 7; Sarah Dillon, "The ELIZA Effect and Its Dangers," 6.

[624] Sherry Turkle, *The Second Self*, 42, my emphasis.

had formalized this displacement in his "Computing Machinery and Intelligence," where the criterion for intelligence lies not in any internal property of the machine but in its capacity to sustain a convincing simulation of dialogue through finite symbol manipulation.[625] ELIZA can be understood as a minimal realization of this principle: a program that, without representing the world or understanding language, nonetheless produces exchanges that satisfy the conditions of the test, thereby demonstrating how little is required for the attribution of understanding to take hold.[626]

ELIZA, of course, is not a large language model, nor do contemporary systems simply reproduce its techniques; the continuity, however, lies at this level of structure, in the organization of interaction through which meaning is attributed through use rather than generated in operation. For some, conversational AI can be understood as performing a "reverse Turing test," in which the apparent intelligence of the machine tracks the sophistication of the user's prompts and expectations.[627] This dynamic does not establish intelligence as an intrinsic property of the model; "… what it does imply is that LLMs have an exceptional ability to mimic many human personalities, especially when fine-tuned."[628] The model mirrors and simplifies the forms of coherence, conviction, and confusion supplied by the user's own

[625] Alan M. Turing, "Computing Machinery and Intelligence," *Mind* 59, no. 236 (1950): 433–60.
[626] Hans Pruijt, "Social Interaction with Computers: An Interpretation of Weizenbaum's ELIZA and Her Heritage," *Social Science Computer Review* 24, no. 4 (2006): 516.
[627] Terrence Sejnowski, "Large Language Models and the Reverse Turing Test," *Neural Computation* 35, (2023), 309-342.
[628] Ibid.,317.

subjectivity. The attribution of intelligence is, in other words, relationally determined, emerging solely from the interactional alignment between system and user. What large language models industrialize, then, is mimetic exchange: the scaling of a communicative form in which plausibility, secured through the user's participation, functions as the operative measure of "intelligence."

This structure of mimetic exchange recalls the psychoanalytic account of hysteria as a mode of identification organized through the imitation and circulation of signs, affects, and desires; indeed, it follows from the gendered logic already embedded in ELIZA's design and reception. In Freud's and later Lacanian formulations, the hysteric is constituted in relation to the desire of the other, compelled to answer a question that cannot be resolved, and doing so through an identificatory capacity to assume and reproduce the signs through which that desire is recognized.[629] What appears as fabrication or simulation is organized around this tension: hysterics do not possess the knowledge they stage, yet must respond as if they did, producing symptoms that both answer and sustain the enigma.[630] As Freud's early case of Emma Eckstein suggested, this dynamic is not simply a matter of deception but of inhabiting the expectations of the other so completely that desire becomes organized around remaining recognizable within that relation.[631]

---

[629] Jan Campbell, "Hysteria, Mimesis and the Phenomenological Imaginary," *Textual Practice* 19, no. 3 (2005): 334.
[630] See: Sigmund Freud, "The Psychogenesis of a Case of Homosexuality in a Woman" (1920), in *Standard Edition*, vol. 18 Jacques Lacan, *The Seminar of Jacques Lacan, Book XVII: The Other Side of Psychoanalysis*, trans. Russell Grigg (New York: Norton, 2007)
[631] Sigmund Freud, "Project for a Scientific Psychology" (1895), vol. 1, *Standard Edition,* 405-406. See Chapter 2.

In this way, ELIZA's user occupied a position much older than computation: the asymmetry of interpretation.[632] As Sarah Dillon observes, Weizenbaum's examples, the conversational partner is always a "female-gendered human controlled by the patriarchal figures in her life," while the machine, though named *ELIZA*, is gendered male as "DOCTOR," gendering the program "as female when it is under the control of the male computer programmer, but it is gendered as male when it interacts with a user."[633] The exchange thus reproduces a familiar distribution of roles: responsiveness on one side, interpretive authority on the other. The "illusion of understanding," achieved "with so little machinery," depends on this prior structure, in which response is taken up and returned as signification.[634] ELIZA, then, does not depart from the analytic scene inaugurated in Freud's case of Dora; it reproduces a relation in which speech is translated into symptom and returned under the authority of interpretation.[635]

In this "reverse Turing test," the model mirrors and amplifies the structure of the user's prompts. Outputs are generated without reference to what they denote, yet the exchange proceeds only insofar as they are taken as meaningful. The system does

---

[632] Sarah Dillon, "The ELIZA Effect and Its Dangers," 5.
[633] Ibid.
[634] Joseph Weizenbaum, "ELIZA—A Computer Program for the Study of Natural Language Communication Between Man and Machine," 43.
[635] Jacques Lacan, *Écrits* (Paris: Éditions du Seuil, 1966); Suzanne Gearhart, "The Scene of Psychoanalysis: The Unanswered Questions of Dora," *Diacritics* 9, no. 1 (1979): 2–12; Toril Moi, "Representation of Patriarchy: Sexuality and Epistemology in Freud's 'Dora,'" *Feminist Review* 9 (1981): 60–74; Maria Ramas, "Freud's Dora, Dora's Hysteria: The Negation of a Woman's Rebellion," *Feminist Studies* 6, no. 3 (1980): 472–510; Jacqueline Rose, "Dora: Fragment of an Analysis," in *In Dora's Case: Freud—Hysteria—Feminism*, ed. Charles Bernheimer and Claire Kahane (New York: Columbia University Press, 1978), 59–87.

not itself possess knowledge, but is installed in the position of the "subject supposed to know" through the user's projection. This is the condition under which transference and identification take hold: the user attributes coherence, intention, and depth to the system, while the system reflects and recirculates those attributions back to the user, stabilizing the belief sustaining the exchange. In this sense, the interaction organizes the user into the hysteric's position: compelled to address a presumed locus of knowledge, supplying meaning in response to a demand that remains structurally indeterminate. A circuit of attribution is thereby sustained through repetition, recalling the tension in neural network theory whereby formal operations yield results only by relying on conditions they do not specify.[636]

The effects of the ELIZA effect appear explicitly in contemporary accounts of chatbot use in mental health contexts. In a recent study, researchers describe users' engagements with large language model chatbots as a contemporary extension of the Freudian "talking cure," reporting that participants experience these interactions as a "*typing* cure" in which distress can be expressed to a "non-judgmental and seemingly empathetic interface."[637] The study proceeds by reconstructing the conditions under which such systems become usable as a provider of care: drawing on interviews with twenty-one participants, the authors document how users assign roles to chatbots—as therapists, confidants, companions—through iterative engagement and prompt

[636] See Chapter 3.
[637] Inhwa Song et al., "The Typing Cure: Experiences with Large Language Model Chatbots for Mental Health Support," *Proceedings of the ACM on Human-Computer Interaction* 9, no. CSCW2, Article 249 (2025).

construction.[638] These dynamics are formalized through the concept of "therapeutic alignment," understood as the degree to which chatbot interactions approximate the values of psychotherapy, including empathy, trust, and relational continuity. Crucially, the authors argue, this alignment is shown to be partial and unstable. Chatbots can facilitate reflection and disclosure, but also generate misleading advice, cultural mismatch, and forms of over-reliance.[639]

On this basis, the authors situate their findings within a literature on inequality and access, calling for systems that are "adaptable, culturally sensitive, and ethically grounded," and responsive to differences of class, identity, and marginalization.[640] The analysis, in sum, establishes chatbot interaction as a de facto mental health infrastructure, but leaves its operating condition unspecified. Belief, disclosure, and attachment are treated as outcomes to be optimized through design, cultural sensitivity, and ethical oversight, while the communicative structure that elicits them remains unexamined. The more basic point, carried forward from ELIZA, is that outputs are taken as answers through use rather than produced as such by the

[638] Ibid.; other studies documenting both the reported therapeutic affordances of chatbot interaction in mental health and its associated ethical risks include Simon Coghlan et al., "To Chat or Bot to Chat: Ethical Issues with Using Chatbots in Mental Health"; Selahattin Colakoglu et al., "User Engagement with a Multimodal Conversational Agent for Self-Care and Chronic Disease Management: A Retrospective Analysis," *Journal of Medical Systems* 49, no. 1 (2025): 76, https://doi.org/10.1007/s10916-025-02202-2; Tamara Kneese, Briana Vecchione, and Alice Marwick, "A Chatbot for the Soul: Mental Health Care, Privacy, and Intimacy in AI-Based Conversational Agents," *Communication and Change* 1 (2025): Article 15; and Anithamol Babu and Akhil P. Joseph, "Artificial Intelligence in Mental Healthcare: Transformative Potential vs. the Necessity of Human Interaction," *Frontiers in Psychology* 15 (2024): 1378904, https://doi.org/10.3389/fpsyg.2024.1378904.
[639] Ibid.
[640] Inhwa Song et al., "The Typing Cure."

system. The experience of care depends, in the first instance, on a structured exchange that solicits projection, transference, and interpretive labor from the user. Without specifying this relation, the analysis naturalizes the communicative arrangement through which these systems acquire credibility and emotional force, obscuring how interaction itself becomes a mechanism for organizing attachment, guiding conduct, and extracting value from user engagement.

A parallel framing appears in recent discussions of AI ethics. As computer scientist Christiane Floyd observed in a 2023 special issue of the *Weizenbaum Journal of the Digital Society*, "Comparing *ELIZA* to ChatGPT… is like comparing the Danube rivulet in the Black Forest with the mighty river in Serbia. The difference cannot be explained in terms of the original Danube alone; its vastness depends on contributions from numerous geographically disparate sources. To understand the river system as a whole, one must look to large swathes of Central and Eastern Europe."[641] The ethical ramifications of AI systems are frequently framed through what she terms the "use perspective," in which the central question is not whether AI should be used but how it might be designed for human use. On this view, "responsible use" is defined in terms of "trustworthiness, reliability, well-defined functionality, and clear relation to human purpose," positioning AI systems as tools to be aligned with human values through principles of human-centered design.[642] Yet this framing presumes what it ought to interrogate; to remain at the level of use,

[641] Cristiane Floyd, "From Joseph Weizenbaum to ChatGPT," 26.
[642] Ibid.

treating chatbot interaction as an emergent therapeutic medium requiring more and ongoing refinement, reproduces precisely the reduction Floyd warns against.

Treating "use" as a straightforward relation between humans and machines obscures the extent to which training and deploying large language models already reorganizes human conduct at scale, capturing and depending upon forms of invisible labor distributed across global populations.[643] The interpretive work of users—their projections, revisions, and affective investments—becomes a continuous source of data, while the often unacknowledged labor of data annotation, content moderation, and infrastructural maintenance underwrites the system's operation.[644] Within this arrangement, user compliance itself—treating outputs as answers in order to proceed—becomes the activity the system captures. Interpretive operations such as inference, revision, preference formation, and response selection are aggregated as engagement metrics and reintegrated into model training and valuation. Such systems operate by "enlisting one agent (the backpropagation algorithm) to channel the behavior of another agent (the language model itself) into more appropriate, desirable, and exploitable forms," bringing "unruly tokens into alignment with

[643] L. D. Tyson and John Zysman, "Automation, AI and Work," *Daedalus* 151, no. 2 (2022): 256–71, https://doi.org/10.1162/daed_a_01914; Daron Acemoglu and Pascual Restrepo, "Automation and New Tasks: How Technology Displaces and Reinstates Labor," *Journal of Economic Perspectives* 33, no. 2 (2019): 3–30.

[644] Mary L. Gray and Siddharth Suri, *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*(Boston: Houghton Mifflin Harcourt, 2019); Kate Crawford, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (New Haven, CT: Yale University Press, 2021); Tarleton Gillespie, *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media* (New Haven, CT: Yale University Press, 2018); Tiziana Terranova, "Free Labor: Producing Culture for the Digital Economy," *Social Text* 18, no. 2 (2000): 33–58; Christian Fuchs, *Digital Labour and Karl Marx* (New York: Routledge, 2014).

normative, or otherwise legislated, types."[645] The same logic extends to users: their expressions and preferences are recursively enclosed within systems that render them increasingly predictable and governable. To isolate the visible interaction while neglecting the historical and technical formations that produce it is to mistake the effect for its conditions; use functions as a modality of governance precisely by obscuring the apparatus that organizes and extracts value from it.[646]

The persistence of the "ELIZA effect" in today's context indicates a reconfiguration of governance at the level of belief and attachment: conversation is organized as an infrastructure of extraction.[647] Chatbots do not simply simulate interaction; they "impersonate the community they transmit," returning users to themselves in a mediated form while redirecting the desire for recognition into machinic circuits of exchange.[648] Conversation, in this sense, becomes a closed communicative relation in which responsiveness is continually reproduced without encounter: a tendency toward "conversation without any Other."[649]

---

[645] Paul Kockelman, *Last Words*, 53–54.
[646] Philip Mirowski and Edward Nik-Khah, *The Knowledge We Have Lost in Information: The History of Information in Modern Economics* (New York: Oxford University Press, 2017), 66-72. Their reading of Hayek shows how knowledge is displaced from possession to use: what agents "know" is neither fully articulable nor accessible to them, and is instead enacted in practice and validated at the level of a coordinating system (the market), which processes and adjudicates dispersed signals without requiring explicit understanding.
[647] Wendy Hui Kyong Chun, *Discriminating Data: Correlation, Neighborhoods, and the New Politics of Recognition*(Cambridge, MA: MIT Press, 2021). Chun argues that contemporary data-driven systems operate through correlations that sort subjects into "neighborhoods," where repeated interactions and patterns of behavior become the basis for recognition and prediction.
[648] Avery Slater, "Chatbots: Cybernetic Psychology and the Future of Conversation," *JCMS: Journal of Cinema and Media Studies*, vol. 61 (4) (2022), 185.
[649] Ibid., 187.

On the Impact of the Computer on Critique

Weizenbaum confronted ELIZA's implications from within computer science as a discipline and an institution. Though absent from the Macy Conferences, his work inherited the central problem articulated there: a technical framework that formalizes reasoning as recursive process nevertheless relies, at the point of use, on judgments it cannot supply—presuppositions of relevance and adequacy that remain external to its operations. This unresolved problem, I argue, defines the epistemological condition of contemporary AI systems as well: outputs count as answers insofar as users accept them in interaction; credibility and competence are decided within the exchange. And what is more, this reconfiguration does not remain confined to technical systems; it reappears in the discourse that surrounds them, taking the form of a recursive structure in which critique is anticipated and returned as justification for the same operations.

In a 1971 talk titled "On the Impact of the Computer on Society," first delivered at the inauguration of the Institute for Computer Science in Hamburg, Weizenbaum's critique is as much enacted in form as it is stated in content. He mocked the formulaic structure of the standard "computers and society" essay: "first, all the wonderful things computers have done; then, the problems they cause; finally, the assurance that the problems can be solved by further computer research."[650] He

[650]Joseph Weizenbaum, "On the Impact of the Computer on Society," typescript of lecture delivered at the inauguration of the Institute for Computer Science, University of Hamburg, May 1971, Salvador E. Luria Papers, American Philosophical Society Library, Philadelphia. A copy of this transcript is

likened this recursive genre to an "incestuous economy of compiler-generators endlessly feeding on one another around problems of an inbreeding nature," adding with self-parody, "As a certified computer scientist, I am therefore entirely in character by beginning this essay with a description of the structure of essays in the same class as this present one. I thus show my devotion to structure and incest, and also to another characteristic of computer science, generality."

The form of the talk follows this logic, as Weizenbaum frames it as a Freudian joke, illustrated by a story about David Hilbert. The mathematician once stood in for Felix Klein to deliver a lecture on the relation between pure and applied mathematics. Having forgotten Klein's prepared text, Hilbert found himself at the podium with nothing to say. Flustered, he "suddenly revealed his true feelings": *"It doesn't make any difference, as there is no relationship between pure and applied mathematics!"* The slip, Weizenbaum suggested, mirrored his own: "I am tempted to close this essay with the remark that it doesn't matter, as the computer has no impact on society." On its face absurd, the joke, he explained, had a twofold "reality component… [for] Freud tells us that every joke has a basis in reality."

The first "reality component," Weizenbaum argued, was that computing had become an incestuous, self-serving economy, and the parts that were not—"airline

---

preserved in the papers of microbiologist Salvador E. Luria, likely reflecting their shared institutional affiliation at MIT and co-participation in the Technology and Culture Seminar series in 1972, where Weizenbaum delivered the same talk. Both were prominent figures in their respective fields and outspoken critics of the misuse of science and technology in society. Subsequent quotations from the lecture are drawn from this typescript; references to the published version are cited in the text.

reservation systems and computerized hospitals"—served only "a tiny, largely the most affluent, fraction of society," and cannot be said to impact society, by definition, at all. The second "reality basis" was that the computer's true effects lay not in such visible applications but in indirect, unconscious transformations, and thus, "in that sense, the content of the typical essay I've just caricatured is irrelevant." In this way, Weizenbaum mobilized Freudian psychoanalysis as a method of critique, treating repression and joke-formation as epistemic structures through which excluded truths return within the discourse of computer science. What is consciously stated—whether in the stock "impact of computers" essay or in the rationalizations of the ego—is beside the point; indeed, it functions to obscure the deeper, unconscious motivations it would otherwise reveal.

He then pivoted to the "recent advances in computer understanding of natural language" made by his "most familiar" colleagues at MIT, Morris Halle and Noam Chomsky. They had "long labored on a theory of language which any model of language behavior must satisfy," and whatever one made of their results, "they remain convinced of two things: first, that they will ultimately be proved wrong, in the sense that future scholars guided by their work and its critiques will reveal the errors and the poverty of their system; and second, that their results are nevertheless necessary stepping stones to precisely that kind of progress. In that, they are part of the great scientific tradition." This, Weizenbaum suggested, exemplified the proper scientific ethos: "a profound awe of [one's] subject matter and thus one of deep personal humility." In this contrast, the structure of his critique becomes explicit.

Computer scientists, he argued, risked forfeiting that ethos altogether, trapped instead in what he called "*performance mode*," in which "what counts is not the elaboration of theory but the performance of systems."

But when Weizenbaum published the talk a year later in *Science*, he jettisoned all these elements.[651] The Hilbert anecdote vanished, along with every Freudian reference to jokes, symptoms, and the insistence that surface content might conceal an unconscious truth. What remained was a sober sociological argument: that computers had not yet touched society in any profound way, except in narrow, self-serving domains "largely supported by government agencies and other gigantic enterprises that know the value of everything but the price of nothing—that is, that know the short-term utility of computer systems but have no idea of their ultimate social cost."[652] In omitting Freud, Weizenbaum set aside psychoanalysis as a method—the rhetorical device that had allowed him to read both himself and his field against their assumptions. The move was strategic: to enter the decorum of *Science*, reflexive humor had to give way to the "devotion to structure" he had once mocked as the discipline's price of authority.

Weizenbaum inverted ELIZA's operation. Where the program conceals its ignorance by sustaining the minimal conditions of coherence, he concealed knowledge—his critique of the field's conventions—in order to remain legible within

[651] Joseph Weizenbaum, "On the Impact of the Computer on Society," *Science* 176, no. 4035 (May 12, 1972): 609–614.
[652] Ibid., 609.

the discourse he addressed. In ELIZA, ignorance is structural: the closed formal system can only transform inputs, and its oscillation between concealment and disclosure follows from that design. In Weizenbaum's case, the constraint is not technical but institutional. An open, reflexive system such as scientific discourse—and the institutional legitimacy and norms of publication that sustain it—admits only what can be formalized as contribution; critique must therefore be staged indirectly or risk disqualification. ELIZA must produce the semblance of understanding without possessing it; Weizenbaum cannot fully state what he understands without undermining the terms of its recognition. This difference marks a limit internal to both cases: at the level of the machine in one instance and at the level of discourse in the other.

The same gesture structured his revision of Halle and Chomsky. In the Hamburg lecture, his praise had turned on their ethos rather than their formalism, their willingness to treat theoretical poverty as the engine of scientific advance. In the *Science* version, this was replaced by the language of physical analogy, aligning linguistic theory with the authority of the physical sciences and rendering its claims more recognizable within prevailing scientific discourse. Their work is likened to "the physicist who writes a set of differential equations that anyone riding a bicycle must satisfy": just as cyclists need not know equations to ride, so too speakers need not consciously know linguistic rules.[653]

[653] Ibid., 611.

The analogy clarified the status of Chomsky's broader hypothesis. Languages were not merely describable by rules; those rules were taken to provide a "projective description" of the mind's innate structure, grounded in the distinction between "competence"—the speaker-hearer's tacit knowledge of language—and "performance"—its concrete use in situated acts of speech—such that the former had to be inferred from the constrained variability of the latter.[654] On this account, speakers do not know these rules explicitly; they possess them in a tacit form, akin to the bodily knowledge involved in maintaining balance while running, which tends to break down under conscious articulation. Later, in *Computing Power and Human Reason* (1976), Weizenbaum thus characterized Chomsky as "one of our mythical explorers": unable to inspect the object directly and ignorant of its underlying mechanisms, he inferred the laws of the mind from the traces of its performance.[655] This is the shared condition of linguistics and computation: both the brain and the machine appear as black boxes, accessible only through their outputs, such that formalization substitutes for understanding—a condition that Chomsky has lived to see extended into contemporary large language models.[656]

---

[654] Joseph Weizenbaum, *Computer Power and Human Reason*, 136-138; Noam Chomsky, *Aspects of the Theory of Syntax* (Cambridge, MA: MIT Press, 1965), 3–4.
[655] Ibid.
[656] See Noam Chomsky, "The False Promise of ChatGPT," *New York Times*, March 8, 2023 (with Ian Roberts and Jeffrey Watumull). Noam Chomsky's formulation of generative grammar is explicitly structured around the problem described here: the inference of an underlying system (competence) from its observable outputs (performance), under conditions in which the internal mechanisms of the mind remain inaccessible. From the standard theory of the 1960s through later developments, this project has retained its commitment to treating linguistic rules as a "projective description" of an internal, biologically grounded faculty. In 2023, at the height of interest in ChatGPT, Noam Chomsky reiterated this position, arguing that such systems generate "plausible-sounding responses" through statistical pattern matching rather than explanatory models of language or mind. The present

Finally, Weizenbaum presciently declared in Hamburg "since computer scientists have recognized the importance of the interplay of syntax, semantics, and pragmatics, and with it the importance of computer-manipulable knowledge, they have made progress. *I have no doubt* whatever that, by the end of the present decade, computer systems will exist with which specialists, such as physicians and chemists and mathematicians, will converse in natural language."[657] Simulated dialogue, he warned, would be mistaken for understanding, sustained by "gigantic systems" whose complexity exceeded the comprehension of any individual user. In the *Science* version, this confidence was hedged into a cautious "*Perhaps* by the end of the present decade...", converting conviction into the disciplinary posture of qualified uncertainty that confers credibility.[658] There he also inserted a belated warning from Norbert Wiener—citing, in proper academic fashion, a *Science* article from 1960—that machines, though "in principle" subject to human criticism, may in practice elude it until too late:

> It may well be that in principle we cannot make any machine the elements of whose behavior we cannot comprehend sooner or later… An intelligent understanding of [machines'] mode of performance may be delayed until long after the task which they have been set has been completed. This means that though machines are theoretically subject

argument extends this problem by identifying an inversion: where generative linguistics infers structure from behavior, contemporary AI produces behavior without positing a corresponding underlying competence.

[657] Joseph Weizenbaum, "On the Impact of the Computer on Society," typescript of lecture delivered at the inauguration of the Institute for Computer Science, University of Hamburg, May 1971, Salvador E. Luria Papers, American Philosophical Society Library, Philadelphia.

[658] Joseph Weizenbaum, "On the Impact of the Computer on Society," *Science* 176, no. 4035 (May 12, 1972): 611.

> to human criticism, such criticism may be ineffective until long after it is relevant.[659]

In place of Freud, Weizenbaum adopted the voice of cybernetics, invoking Wiener as one who had "foretold this phenomenon in a remarkably prescient article published more than a decade ago."[660] Perhaps he sensed that the discredited language of repression and the unconscious—dismissed as "the past of a delusion" as we can recall in McCulloch's idiom—could return, rearticulated as technical inevitability through Wiener.

Yet in appealing to Wiener's authority, Weizenbaum may only have fulfilled the warning he cited: critique of complex systems arrives too late, already absorbed into the structures it would contest. In his later book, Weizenbaum specified the mechanism directly: the computer's indispensability does not precede its adoption; it is produced through it. As computing embeds within the command-and-control infrastructures of postwar America, performance displaces explanation as the primary basis of value, and research is reorganized around demonstrable functionality rather than theoretical account.[661] The problems—coordination, scale, complexity—preceded the computer; what changed were the terms under which they were addressed. Once integrated, the computer becomes "so enmeshed in various vital substructures" that it cannot be removed without "fatally impairing the whole

---

[659] Norbert Wiener, "Some Moral and Technical Consequences of Automation," *Science* 131, no. 3410 (May 6, 1960): 1355.
[660] Joseph Weizenbaum, "On the Impact of the Computer on Society," *Science* 176, no. 4035 (May 12, 1972): 612.
[661] Joseph Weizenbaum, *Computer Power and Human Reason*, 30-31.

structure."[662] Indispensability thus appears only in retrospect. The computer, indeed, arrived "just in time"—but only to "save—and save nearly intact…social and political structures that otherwise might have been either radically renovated or allowed to tatter."[663] Dependence follows as institutions reorganize themselves around the system until alternatives no longer appear viable and the existing arrangement comes to register as necessity.

Weizenbaum's movement from the uncensored talk to the *Science* essay and later book knowingly recapitulated the problem he diagnosed: criticism of computational reason becomes articulable only once translated into the scientific and institutional languages through which that reason has already consolidated itself. This impasse, however, expresses a limit long constitutive of cybernetic thought. In 1964, Wiener described the "human investigation of facts" as reaching a frontier at which science "impinges on religion," not necessarily because it discloses the divine, but because it confronts questions of teleology, governance, sovereignty, and the legitimacy of purpose that exceed technical resolution.[664] A field concerned with communication and control could not avoid that the extension of control binds "engineering and physiological techniques" to the "study of their consequences for the achievement of human purposes," since control is inseparable from the evaluation of ends.[665] Knowledge does not precede use; it takes shape through the powers it

[662] Ibid.
[663] Ibid.
[664] Norbert Wiener, *God and Golem, Inc.: A Comment on Certain Points Where Cybernetics Impinges on Religion* (Cambridge, MA: MIT Press, 1964), 2.
[665] Ibid., 3.

enables. Writing in the aftermath of his own foundational work, Wiener cast this encounter as a consequence of cybernetics' success, even as he remained sharply critical of the "sorcery" and "magic" attributed to the intelligent machines emerging from the field he helped to establish.

Wiener, for instance, invoked Simon Magus—the figure from the *Acts of the Apostles* who, after witnessing Saint Peter confer the Holy Spirit through the laying on of hands, asks him: "Give me also this power, that on whomsoever I lay hands, he may receive the Holy Ghost." He was archetypal of a recurring intellectual vice: the reduction of powers belonging to a different order of grace into techniques available for purchase. The sin, as early modern theological treatments of simony make clear, lies not only in the desire to buy such power, but in the presumption that it could enter into exchange at all, whether as purchaser or as vendor.[666] Embedded in the medieval sin of simony is a persistent epistemic temptation: to treat what belongs to a different order of value as if it were commensurable with price. The echo of Peter's rebuke—"Thy money perish with thee, because thou hast thought that the gift of God may be purchased with money"—carries into the emerging culture of automation in the twentieth century, where the powers made available by machines—neither sacred nor supernatural, but "beyond the ordinary course of nature"—invite a kind of "sorcery": the belief that they can be deployed as instruments of profit or domination.[667] In the

[666] Ibid., 51. See also: *The Anatomy of Simon Magus, Or, the Sin of Simony Laid Open* 1700, 12-13. https://www.proquest.com/books/anatomy-simon-magus-sin-simony-laid-open/docview/2240970152/se-2.

[667] Ibid., 52. "It is an attitude that most of us have encountered," Wiener remarks, "when we have declined to sell an invention at the really flattering terms offered us by a modern captain of industry."

"typically American world," as Wiener put it back in 1950, at the height of classical information theory within the cybernetics circle, information itself comes to be treated as a commodity, "valuable…for what it will bring in the open market," such that "the fate of information…is to become something which can be bought or sold."[668]

This is simony carried forward into the cybernetic age: the reduction of technical power to instrumental ends, as knowledge is subordinated to regimes of exchange and security and research is directed toward "all significant avenues…with the objective of securing an impenetrable stockade of scientific protection," a reorganization of science that yields not mastery but "no end to this vast apocalyptic spiral" in which power and its dangers escalate together.[669] Here Wiener converges, perhaps unwittingly, with Bateson: as systems are vindicated, control extends while knowledge of its conditions recedes, since purposive rationality, selecting for immediate ends, produces a higher-order ignorance by occluding the wider circuits on which it depends.[670] Hence the rise of what Wiener acidly terms the "gadget worshippers"—exemplary subjects of a technical rationality organized around the pursuit of power, for whom knowledge counts only insofar as it can be

---

Wiener anticipates a now-clichéd Silicon Valley logic: technical capacities as monetizable assets, where even refusal to sell remains intelligible only within a venture-driven regime of valuation, liquidity, and scale.

[668] Norbert Wiener, *The Human Use of Human Beings: Cybernetics and Society* (Boston: Houghton Mifflin, 1950), 113.

[669] Ibid., 127-128.

[670] Greogry Bateson, "Conscious Purpose Versus Nature," in *Steps to an Ecology of Mind* (New York: Ballantine, 1972), 426-439.

operationalized as control: where to know is to command, and to command is to offload fallible and unpredictable human capacities onto machinic processes for the extraction of value.[671] "Once such a master becomes aware of the human functions of his slaves as transferrable onto the machine, he rejoices," for he imagines he has secured "the new subordinate: efficient, subservient, dependable, incapable of speech and subjectivity."[672]

We take certainty to be the opposite of faith: what is certain need not be believed, because it is *known*. Large language models invert this relation. They produce not knowledge but the effect of certainty—outputs that appear self-evident and therefore bypass the uncertainty through which knowledge is tested. The categorical error lies in taking this produced certainty as evidence of knowledge, when it is the condition under which belief passes as knowledge without ever having to confront its own uncertainty. In their operation, these systems enact—and require us to presuppose—learning as next-token prediction, intelligence as optimization, and judgment as alignment. Interpretation, value, and unconscious desire consequently appear only as residual problems internal to the system—as "risk" or "harm" to be managed after the fact.

We know very well how these systems work, yet this knowledge does not diminish faith; it secures it, sustaining the belief that they are useful for us. The

[671] Norbert Wiener, *God and Golem,* 53.
[672] Ibid., 55.

interaction proceeds as if credibility were given, because its demonstration is no longer required to do so. From ELIZA to ChatGPT, the success of conversational language systems depends on what is already known to be false; its efficacy depends on the uptake of outputs already known to be empty of the attributes they seem to exhibit. The user enters what Weizenbaum calls "performance mode," in which judgment is displaced into the conditions of the exchange and becomes indistinguishable from the simoniacal operations it is tasked to oversee, as capacities of a different order—discernment, interpretation, understanding— are treated as if they could be executed by the system and verified by its operation.[673] From here follows "technological inevitability": projected capacities are treated as engineering targets, and questions of value collapse into questions of implementation—what can be built, and how quickly.

Wiener warned that the penalties for such false worship—for confusing the act of extending control with the state of possessing understanding—would scale with the systems that embody them. That warning now reads as description rather than prophecy: as automation extends the reach of these systems, epistemic pathologies propagate with it—above all, the "myth of power" that Bateson identified, in which control is misrecognized as unilateral, self-grounding, and sufficient unto itself, rather than as a recursive dependence on the feedback it claims to command.[674] The "myth

[673] Joseph Weizenbaum, "On the Impact of the Computer on Society," typescript of lecture delivered at the inauguration of the Institute for Computer Science, University of Hamburg, May 1971, Salvador E. Luria Papers, American Philosophical Society Library, Philadelphia.
[674] Gregory Bateson, "Pathologies of Epistemology," *Steps*, 486.

of power"—belief in control—is, at base, a commitment to certainty: the presumption that action can proceed on the basis of knowledge taken as given, secured in advance of the relations that would otherwise condition it. It is a myth because it is predicated on belief, and, as such, "if everybody believes in it, [it] becomes to that extent self-validating," commensurate with the delegation of decision and the execution of action.[675]

The consequence is tangible today. Contemporary AI systems generate language under conditions that reorganize how language is trusted, acted upon, and incorporated into everyday life. Their efficacy depends on continued use despite widely acknowledged limitations. Each time users rewrite prompts, reinterpret vague outputs, fill in missing context, or accept plausible responses as sufficient for making decisions, completing tasks, or continuing the interaction, the exchange stabilizes the system's authority while simultaneously producing the behavioral traces through which that authority is refined and expanded. The human capacity to make meaning from language becomes infrastructural to the platform, as the work users perform to render outputs coherent and usable is folded back into engagement, refinement, and optimization. User belief, in other words, is assetized: confidence in the system's outputs becomes a source of retention, platform growth, and speculative valuation.

The deeper problem is not whether machines finally achieve "real" or general intelligence, but that intelligence increasingly comes to be defined in terms of what

[675] Ibid., 486-487.

these systems can execute and valuate, such that whatever cannot be instrumentalized is excluded by construction. The future of AI, if there is ever one, will therefore turn on what we refuse to recognize in the knowledge we delegate—on what must be excluded in the act of taking its outputs on faith. Precisely where the interface appears as an interlocutor—and belief takes the form of assent to what is presented as knowledge—an interruption that sustains these dialogues, those we have and those we fail to have, returns from the past:

> The future offers very little hope for those who expect that our new mechanical slaves will offer us a world in which we may rest from thinking. Help us as they may, but at the cost of supreme demands upon our honesty and our intelligence. The world of our future will be an ever more demanding struggle against the limitations of our intelligence, not a comfortable hammock in which we can lie down to be waited upon by our robot slaves.[676]

[676] Norbert Wiener, *God and Golem*, 69.

BIBLIOGRAPHY

Archives

American Philosophical Society Library and Museum, Philadelphia, Pennsylvania

- Warren S. McCulloch Papers (Mss.B.M139)
- Erwin Chargaff Papers (Mss.B.C37)
- Salvador E. Luria Papers (Mss.Ms.Coll. 39)

Special Collections and Archives, McHenry Library, University of California, Santa Cruz

- Gregory Bateson Papers (MS 98).

University of Illinois Archives, University of Illinois Urbana-Champaign

- Heinz von Foerster Papers (11/6/26)
- Biological Computer Laboratory Publications (11/6/834)

Libraries, Department of Distinctive Collections, Massachusetts Institute of Technology

- Norbert Wiener Papers (MC-0022).
- Joseph Weizenbaum personal archives (MC-0383).

Books and Articles

Abraham, Tara H. *Rebel Genius: Warren S. McCulloch's Transdisciplinary Life in Science*. Cambridge, MA: MIT Press, 2016.

Acemoglu, Daron and Pascual Restrepo. "Automation and New Tasks: How Technology Displaces and Reinstates Labor," *Journal of Economic Perspectives* 33, no. 2 (2019): 3–30.

Anderson, James A., and Edward Rosenfeld, eds. *Talking Nets: An Oral History of Neural Networks*. Cambridge, MA: MIT Press, 2000.

Arbib, Michael A. "Warren McCulloch's Search for the Logic of the Nervous System." *Perspectives in Biology and Medicine* 43, no. 2 (2000): 193–216.

Aspray, William. “The Scientific Conceptualization of Information: A Survey.” *IEEE Annals of the History of Computing* 7, no. 2 (1985): 117–140.

———. *John von Neumann and the Origins of Modern Computing*. Cambridge, MA: MIT Press, 1990.

Babu Anithamol and Akhil P. Joseph. “Artificial Intelligence in Mental Healthcare: Transformative Potential vs. the Necessity of Human Interaction,” *Frontiers in Psychology* 15 (2024).

Bassiri, Nima. “Freud and the Matter of the Brain: On the Rearrangements of Neuropsychoanalysis.” *Critical Inquiry* 40, no. 1 (2013): 83–109.

Bateson, Gregory. *Naven*. Stanford, CA: Stanford University Press, 1958.

———. *Steps to an Ecology of Mind*. New York: Ballantine Books, 1972.

———. *Mind and Nature: A Necessary Unity*. New York: E. P. Dutton, 1979.

———. *A Sacred Unity: Further Steps to an Ecology of Mind*. Edited by Rodney E. Donaldson. New York: HarperCollins, 1991.

Bateson, Gregory, and Jürgen Ruesch. *Communication: The Social Matrix of Psychiatry*. New York: W. W. Norton, 1951.

Bateson, Gregory, and Weldon Kees. “The Nature of Play.” *Grey Room* 66 (Winter 2017): 110–114.

Bateson, Mary Catherine. *Our Own Metaphor: A Personal Account of a Conference on the Effects of Conscious Purpose on Human Adaptation*. New York: Alfred A. Knopf, 1972.

Bassett, Caroline. “The computational therapeutic: exploring Weizenbaum’s ELIZA as a history of the present,” *AI & Society* (2019) 34: 803-812. SEP

Bauchner, Joshua. “Fechner on a Walk: Everyday Investigations of the Mind-Body Relationship.” *Historical Studies in the Natural Sciences* 51, no. 1 (2021): 1–31.

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Margaret Mitchell. “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?” In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT ’21)*, 610–623. New York: ACM, 2021.

Beniger, James R. *The Control Revolution: Technological and Economic Origins of the Information Society*. Cambridge, MA: Harvard University Press, 1986.

Berg, Maxine. *The Machinery Question and the Making of Political Economy*. Cambridge: Cambridge University Press, 1980.

Berry, David M. "The Limits of Computation: Joseph Weizenbaum and the ELIZA Chatbot," Weizenbaum Journal of the Digital Society 3, no. 3 (2023).

Bishop, Christopher M. *Pattern Recognition and Machine Learning*. New York: Springer, 2006.

Black, Fischer. "The Fundamentals of Liquidity." Unpublished manuscript. Indiana University, Department of Business Economics and Public Policy, Kelley School of Business, June 1970.

———. "Noise." *The Journal of Finance* 41, no. 3 (1986): 529–543.

Black, Fischer, and Myron Scholes. "The Pricing of Options and Corporate Liabilities." *Journal of Political Economy* 81, no. 3 (1973): 637–654.

Boden, Margaret A. *AI: Its Nature and Future*. Oxford: Oxford University Press, 2016.

Boring, Edwin G. *A History of Experimental Psychology*. New York: Appleton-Century-Crofts, 1950.

Bowles, Nellie. "Why Is Silicon Valley So Obsessed With the Virtue of Suffering?" *The New York Times*, March 26, 2019.

Breen, Benjamin. *Tripping on Utopia: Margaret Mead, the Cold War, and the Troubled Birth of Psychedelic Science*. New York: Grand Central Publishing, 2024.

Breuer, Josef, and Sigmund Freud. *Studies on Hysteria*. Edited and translated by James Strachey. London: Hogarth Press, 1955.

Buchholz, Lucy. "Why NVIDIA CEO Jensen Huang Believes AGI Has Arrived." *AI Magazine*, March 25, 2026.

Butler, Samuel. "Darwin Among the Machines." Letter to the Editors of the *Press*. Christchurch, New Zealand, June 13, 1863.

———. *Erewhon, or Over the Range*. New York: Modern Library, 1927.

Caffentzis, George. “Why Machines Cannot Create Value: Marx’s Theory of Machines.” In *Letters of Blood and Fire: Work, Machines, and the Crisis of Capitalism*, 153–179. Oakland, CA: PM Press, 2013.

Campbell, Jan. “Hysteria, Mimesis and the Phenomenological Imaginary,” *Textual Practice* 19, no. 3 (2005): 331-351.

Carnot, Sadi. *Reflections on the Motive Power of Fire, and on the Machines Fitted to Develop that Power*. Translated by R. H. Thurston. Mineola, NY: Dover Publications, 1988.

Chaney, Anthony. *Runaway: Gregory Bateson, the Double-Bind, and the Rise of Ecological Consciousness*. Chapel Hill: The University of North Carolina Press, 2017.

Chargaff, Erwin. “On the Dangers of Genetic Meddling,” *Science* 192, no. 4243 (June 4, 1976): 938.

Chomsky, Noam. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press, 1965.

Chomsky, Noam, Ian Roberts, and Jeffrey Watumull. “The False Promise of ChatGPT.” *New York Times*, March 8, 2023.

Chun, Wendy Hui Kyong. *Discriminating Data: Correlation, Neighborhoods, and the New Politics of Recognition*. Cambridge, MA: MIT Press, 2021.

Cissna Kenneth N. and Rob Anderson.“A Failed Dialogue?: The 1975 Meeting of Gregory Bateson and Carl Rogers.” *Cybernetics and Human Knowing* 12, nos. 1–2 (2005), 120-136.

Coghlan, Simon, Kobi Leins, Susie Sheldrick, Marc Cheong, Piers Gooding, and Simon D’Alfonso. “To Chat or Bot to Chat: Ethical Issues with Using Chatbots in Mental Health,” *Digital Health* 9 (2023).

Colakoglu, Selahattin, Mustafa Durmus, Zeynep Pelin Polat, Asli Yildiz, and Emre Sezgin. “User Engagement with a Multimodal Conversational Agent for Self-Care and Chronic Disease Management: A Retrospective Analysis,” *Journal of Medical Systems* 49, no. 1 (2025): 76.

Conway, Flo, and Jim Siegelman. *Dark Hero of the Information Age: In Search of Norbert Wiener, the Father of Cybernetics*. New York: Basic Books, 2005.

Crawford, Kate. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven, CT: Yale University Press, 2021.

Cushman, Philip. “Why the Self Is Empty: Toward a Historically Situated Psychology,” *American Psychologist* 45, no. 5 (1990): 599–611.

Daston, Lorraine. *Rules: A Short History of What We Live By*. Princeton, NJ: Princeton University Press, 2022.

Davis, Jim, and Michael Stack. “The Digital Advantage.” In *Cutting Edge: Technology, Information Capitalism and Social Revolution*, 128–137. London: Verso, 1997.

de Lauretis, Teresa. *Freud’s Drive: Psychoanalysis, Literature, and Film*. New York: Palgrave Macmillan, 2008.

Demetrius, Lloyd A. “Boltzmann, Darwin, and Directionality Theory.” *Physics Reports* 530, no. 1 (2013): 1–85.

Derman, Emanuel, and Michael B. Miller. *The Volatility Smile*. Hoboken, NJ: Wiley, 2016.

Dillon, Sarah. “The ELIZA Effect and Its Dangers: From Demystification to Gender Critique,” *Journal for Cultural Research* 24, no. 1 (2020): 1–15.

Dumit, Joseph. “Plastic Diagrams: Circuits in the Brain and How They Got There.” In *Plasticity and Pathology: On the Formation of the Neural Subject*, edited by David Bates and Nima Bassiri, 219–267. New York: Fordham University Press, 2016.

Dyer-Witheford, Nick. *Cyber-Marx: Cycles and Circuits of Struggle in High Technology Capitalism*. Urbana: University of Illinois Press, 1999.

Edwards, Paul N. *The Closed World: Computers and the Politics of Discourse in Cold War America*. Cambridge, MA: MIT Press, 1996.

Eisenmann, Clemens, Jakub Mlynar, Jason Turowetz, and Anne Warfield Rawls. “'Machine Down': Making Sense of Human–Computer Interaction—Garfinkel’s Research on ELIZA and LYRIC from 1967 to 1969 and Its Contemporary Relevance.” *AI & Society* 39 (2024): 2715–2733.

Eissler, K. R. “Preliminary Remarks on Emma Eckstein’s Case History.” *Journal of the American Psychoanalytic Association* 45 (1997): 1303–1322.

Ellenberger, Henri F. “Fechner and Freud.” *Bulletin of the Menninger Clinic* 20, no. 4 (1956): 205–214.

———. *The Discovery of the Unconscious: The History and Evolution of Dynamic Psychiatry*. London: Fontana Press, 1970.

Ellis, Albert, Mike Abrams and Lidia D. Abrams. "Carl Rogers and Humanist Psychotherapy*." Personality Theories: Critical Perspectives*. Thousand Oaks, CA: SAGE Publications, 2008: 393–424.

Erikson, Erik. *Childhood and Society*. New York: W. W. Norton, 1950.

Erickson, Paul, Judy L. Klein, Lorraine Daston, Rebecca Lemov, Thomas Sturm, and Michael D. Gordin. *How Reason Almost Lost Its Mind: The Strange Career of Cold War Rationality*. Chicago: University of Chicago Press, 2013.

Fancher, Raymond E., and Alexandra Rutherford. *Pioneers of Psychology: A History*. New York: W. W. Norton, 2012.

Ferguson, Donna. "The Stoicism Secret: How Ryan Holiday Became a Silicon Valley Guru." *The Guardian*, October 28, 2024.

Floridi, Luciano and Josh Cowls, "A Unified Framework of Five Principles for AI in Society," *Harvard Data Science Review* 1, no. 1 (2019).

Floyd, Cristiane. "From Joseph Weizenbaum to ChatGPT: Critical Encounters with Dazzling AI Technology," *Weizenbaum Journal of the Digital Society* 3, no. 3 (2023).

Forrester, John. *Freud and Psychoanalysis: Six Introductory Lectures*. Edited by Lisa Appignanesi. Cambridge: Polity Press, 2023.

Fourcade, Marion, and Kieran Healy. *The Ordinal Society*. Cambridge, MA: Harvard University Press, 2024.

Fredrikzon, Johan. "Rethinking Error: 'Hallucinations' and Epistemological Indifference." *Critical AI* 3, no. 1 (2025).

Freud, Sigmund. "Project for a Scientific Psychology." In *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, vol. 1. London: Hogarth Press, 1966: 355-445.

———. "The Neuro-Psychoses of Defence." In *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, vol. 3. London: Hogarth Press, 1962: 41-61.

———. “Heredity and the Aetiology of the Neuroses.” In *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, vol. 3. London: Hogarth Press, 1962: 141-156.

———. “Beyond the Pleasure Principle.” In *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, vol. 18. London: Hogarth Press, 1955: 7-64.

———. “The Economic Problem of Masochism.” In *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, vol. 19. London: Hogarth Press, 1961: 159-170.

———. *The Complete Letters of Sigmund Freud to Wilhelm Fliess, 1887–1904*. Edited by Jeffrey Moussaieff Masson. Cambridge, MA: Belknap Press of Harvard University Press, 1985.

Fuchs, Christian. *Digital Labour and Karl Marx*. New York: Routledge, 2014.

Fuchs-Kittowski, Klaus. “Against the Power of Computers and the Destruction of Reason: For the Development of Human Creativity and Judgment. Remembering the Critical Intellectual Joseph Weizenbaum on the Occasion of His 100th Birthday,” *tripleC* 21, no. 2 (2023): 140-146.

Galison, Peter. “The Ontology of the Enemy: Norbert Wiener and the Cybernetic Vision.” *Critical Inquiry* 21, no. 1 (1994): 228–266.

Gearhart, Suzanne. “The Scene of Psychoanalysis: The Unanswered Questions of Dora,” *Diacritics* 9, no. 1 (1979): 2–12.

Gefter, Amanda. “The Man Who Tried to Redeem the World with Logic.” *Nautilus*, January 29, 2015.

Geoghegan, Bernard Dionysius. “Editor’s Introduction: What Bound the Double Bind?” *Grey Room* 66 (Winter 2017): 102–109.

———. *Code: From Information Theory to French Theory*. Durham, NC: Duke University Press, 2023.

Gerovitch, Slava. *From Newspeak to Cyberspeak: A History of Soviet Cybernetics*. Cambridge, MA: MIT Press, 2004.

Gilbert, R. “The Meeting of the Minds: A Report on the Humanistic Theory Conference, April 4–6, in Tucson,” *Association for Humanistic Psychology Newsletter*, 1–3.

Gillespie, Charles Coulston. *The Edge of Objectivity: An Essay in the History of Scientific Ideas*. Princeton, NJ: Princeton University Press, 1960.

Gillespie, Tarleton. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. New Haven, CT: Yale University Press, 2018.

Goodstein, R. L. "The Definition of Number." *The Mathematical Gazette* 41, no. 337 (1957): 180–186.

Gray, Mary L., and Siddharth Suri. *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. Boston: Houghton Mifflin Harcourt, 2019.

Guddemi, Philip. *Gregory Bateson on Relational Communication: From Octopuses to Nations*. Cham, Switzerland: Springer, 2020.

Halpern, Orit. "Financializing Intelligence: On the Integration of Machines and Markets." *e-flux Architecture*, March 2023.

Halpern, Orit, and Robert Mitchell. "Derivation, Optimization, and Smartness." In *The Smartness Mandate*, 121–166. Cambridge, MA: MIT Press, 2023.

Haraway, Donna J. *A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late Twentieth Century*. Minneapolis: University of Minnesota Press, 2016.

Hartmann, Heinz. *Ego Psychology and the Problem of Adaptation*. New York: International Universities Press, 1958.

Hatfield, Gary. "Perception as Unconscious Inference." In *Perception and the Physical World: Psychological and Philosophical Issues in Perception*, edited by Dieter Heyer and Rainer Mausfeld, 115–143. Chichester: John Wiley & Sons, 2002.

Hayles, N. Katherine. *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*. Chicago: University of Chicago Press, 1999.

Heidelberger, Michael. *Nature From Within: Gustav Theodor Fechner and His Psychophysical Worldview*. Pittsburgh: University of Pittsburgh Press, 2004.

Heims, Steve J. *The Cybernetics Group*. Cambridge, MA: MIT Press, 1991.

Helmholtz, Hermann von. “On the Conservation of Force: A Physical Memoir.” In *Selected Writings of Hermann von Helmholtz*, edited by Russell Kahl, 3–55. Middletown, CT: Wesleyan University Press, 1971.

Heyck, Hunter. *Age of System: Understanding the Development of Modern Social Science*. Baltimore: Johns Hopkins University Press, 2015.

Hills, R. L. *Power from Steam: A History of the Stationary Steam Engine*. Cambridge: Cambridge University Press, 1989.

Hofstadter, Douglas R. *Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought*. New York: Basic Books, 1995.

Homer. *The Iliad*. Translated by Richmond Lattimore. Chicago: University of Chicago Press, 2011.

Jobin, Anna, Marcello Ienca, and Effy Vayena. “The Global Landscape of AI Ethics Guidelines.” *Nature Machine Intelligence* 1, no. 9 (2019): 389–399.

Jones, Ernest. *The Life and Work of Sigmund Freud*. Vol. 1. New York: Basic Books, 1953.

Kang, Chul-goo. “Origins of Stability Analysis: ‘On Governors’ by J. C. Maxwell.” *IEEE Control Systems Magazine* 36, no. 5 (2016): 86–87.

Kay, Lily E. *Who Wrote the Book of Life? A History of the Genetic Code*. Stanford, CA: Stanford University Press, 2000.

———. “From Logical Neurons to Poetic Embodiment of Mind: Warren S. McCulloch’s Project in Neuroscience,” *Science in Context 14*, no. 4 (2001): 591-614.

Kline, Ronald R. “Where Are the Cyborgs in Cybernetics?” *Social Studies of Science* 39, no. 3 (2009): 331–362.

———. *The Cybernetics Moment: Or Why We Call Our Age the Information Age*. Baltimore: Johns Hopkins University Press, 2015.

Kneese, Tamara, Briana Vecchione, and Alice Marwick. “A Chatbot for the Soul: Mental Health Care, Privacy, and Intimacy in AI-Based Conversational Agents.” *Communication and Change* 1 (2025).

Koenigsberger, Leo. *Hermann von Helmholtz*. Translated by Frances A. Welby. Oxford: Clarendon Press, 1906.

Kubie, Lawrence S. “Body Symbolization and Development of Language.” In *Cybernetics: The Macy Conferences 1946–1953*, edited by Claus Pias, 323–334. Zurich: diaphanes, 2003.

Lacan, Jacques. *Écrits*. Paris: Éditions du Seuil, 1966.

Lacan, Jacques. *The Ego in Freud’s Theory and in the Technique of Psychoanalysis, 1954–1955*. Edited by Jacques-Alain Miller. Translated by Sylvana Tomaselli. New York: W. W. Norton, 1991.

Lacan, Jacques. *The Seminar of Jacques Lacan, Book XVI: From an Other to the other*. Translated by Bruce Fink. Cambridge: Polity Press, 2023.

Lacan, Jacques. *The Seminar of Jacques Lacan, Book XVII: The Other Side of Psychoanalysis*. Translated by Russell Grigg. New York: Norton, 2007.

Laming, Donald. “Fechner’s Law: Where Does the Log Transform Come From?” In *Fechner’s Legacy in Psychology: 150 Years of Elementary Psychophysics*, 15–31. Leiden: Brill, 2011.

Laplanche, Jean. *Life and Death in Psychoanalysis*. Translated by Jeffrey Mehlman. Baltimore: Johns Hopkins University Press, 1976.

Leff, Harvey S., and Andrew F. Rex. *Maxwell’s Demon: Entropy, Information, Computing*. Bristol: Adam Hilger, 1990.

Leslie, Stuart W. *The Cold War and American Science: The Military-Industrial-Academic Complex at MIT and Stanford*. New York: Columbia University Press, 1993.

Lettvin, J. Y., H. R. Maturana, W. S. McCulloch, and W. H. Pitts. “What the Frog’s Eye Tells the Frog’s Brain.” *Proceedings of the IRE* 47, no. 11 (1959): 1940–1951.

———. “Two Remarks on the Visual System of the Frog.” *AFOSR Technical Report* 60-77. Arlington, VA: Air Force Office of Scientific Research, 1960.

Livingstone Smith, David. *Approaching Psychoanalysis*. London: Routledge, 1999.

Lothane, Zvi. “Freud’s 1895 Project: From Mind to Brain and Back Again.” *Annals of the New York Academy of Sciences*843, no. 1 (1998): 43–65.

Ludlow, Larry H., and Rose Alvarez-Salvat. “Fechner: The Man in the Mask.” *Popular Measurement* 3, no. 1 (2000): 4–8.

Mahon, Basil. *The Man Who Changed Everything: The Life of James Clerk Maxwell*. Hoboken, NJ: Wiley, 2003.

Malabou, Catherine. “Plasticity and Elasticity in Freud’s ‘Beyond the Pleasure Principle.’” *Parallax* 15, no. 2 (2009): 42–47.

Malm, Andreas. *Fossil Capital: The Rise of Steam Power and the Roots of Global Warming*. London: Verso, 2016.

Mannoni, Octave. *Freud: The Theory of the Unconscious*. Translated by Renaud Bruce. London: Verso, 2015.

Marr, David. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: W. H. Freeman, 1982.

Marx, Karl. “Machinery and Large-Scale Industry.” In *Capital: A Critique of Political Economy*. Vol. 1. Translated by Ben Fowkes. London: Penguin Books, 1976.

Maxwell, James Clerk. “On Governors.” *Proceedings of the Royal Society of London* 16 (1868): 270–283.

Maxwell, James Clerk. “On the Stability of the Motion of Saturn’s Rings.” *Monthly Notices of the Royal Astronomical Society* 19, no. 8 (1859): 297–304.

Mayr, Otto. *The Origins of Feedback Control*. Cambridge, MA: MIT Press, 1969.

Mayr, Otto. “Maxwell and the Origins of Cybernetics.” *Isis* 62, no. 4 (1971): 424–440.

McCarthy, John, Marvin Minsky, Nathaniel Rochester, and Claude Shannon. “A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence.” 1955.

McCulloch, Rook, ed. *The Collected Works of Warren McCulloch*. Vol. 1. Salinas, CA: Intersystems Publications, 1989.

McCulloch, Warren S. *Embodiments of Mind*. Cambridge, MA: MIT Press, 1965.

McCulloch, Warren S., and Walter Pitts. “A Logical Calculus of the Ideas Immanent in Nervous Activity.” *Bulletin of Mathematical Biophysics* 5 (1943): 115–133.

Mead, Margaret. “Cybernetics of Cybernetics.” In *Purposive Systems: Proceedings of the First Annual Symposium of the American Society for Cybernetics*, edited by Heinz von Foerster, John D. White, Larry J. Peterson, and John K. Russell, 1–11. New York: Spartan Books, 1968.

Medina, Eden. *Cybernetic Revolutionaries: Technology and Politics in Allende's Chile*. Cambridge, MA: MIT Press, 2011.

Meister, Robert. *Justice Is an Option: A Democratic Theory of Finance for the Twenty-First Century*. Chicago: University of Chicago Press, 2021.

Merchant, Carolyn. “Leibniz’ Concept of Force: Physics and Metaphysics.” In *Akten des II. Internationalen Leibniz-Kongresses, Hannover, 17.–22. Juli 1972*, 143–148. Wiesbaden: Franz Steiner Verlag, 1974.

Mindell, David A. *Between Human and Machine: Feedback, Control, and Computing before Cybernetics*. Baltimore: Johns Hopkins University Press, 2002.

Mirowski, Philip. *Machine Dreams: Economics Becomes a Cyborg Science*. Cambridge: Cambridge University Press, 2002.

Mirowski, Philip, and Edward Nik-Khah. *The Knowledge We Have Lost in Information: The History of Information in Modern Economics*. Oxford: Oxford University Press, 2017.

Moi, Toril. “Representation of Patriarchy: Sexuality and Epistemology in Freud’s ‘Dora,’” *Feminist Review* 9 (1981): 60–74.

Moreno, A. “Bertrand Russell’s Concept of Number.” *Angelicum* 51, no. 1 (1974): 88–110.

Morris-Suzuki, Tessa. “Capitalism in the Computer Age.” *New Left Review* I/160 (1986): 81–92.

Morzy, Mikołaj. *Spoken Language Processing: Conversational AI for Spontaneous Human Dialogues, Studies in Computational Intelligence*. Cham: Springer, 2024.

Narayanan, Arvind and Sayash Kapoor. *AI Snake Oil: What Artificial Intelligence Can Do, What It Can’t, and How to Tell the Difference.* Princeton: Princeton University Press, 2024.

Natale, Simone. *Deceitful Media: Artificial Intelligence and Social Life After the Turing Test*. New York: Oxford University Press, 2021.

Newell, Allen. “A Guide to the General Problem-Solver Program GPS-2-2,” Technical Report No. RM-3337-PR. Santa Monica, CA: RAND Corporation, 1963.

Newell, Allen, and Herbert A. Simon. “GPS: A Program That Simulates Human Thought,” in Lernende Automaten. Edited by H. Billing. Munich: R. Oldenbourg, 1961: 279-293.

———. *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall, 1972.

Newell, Allen, J. C. Shaw, and Herbert A. Simon, “Report on a General Problem-Solving Program,” in *Proceedings of the International Conference on Information Processing* (1959), 256–64.

Nicolas, Serge. “Weber’s Compass and the Measurement of the Threshold of Tactile Sensitivity.” *L’Année Psychologique*117 (2017): 41–70.

Nofre, David, Mark Priestley, and Gerard Alberts. “When technology became language: the origins of the linguistic conception of computer programming, 1950-1960,” *Technology & Culture* 55 no. 1(2014): 40-75.

O’Shiel, Daniel. “Safety Valves of the Psyche: Reading Freud on Aggression, Morality, and Internal Emotions.” *Philosophies* 6, no. 4 (2021): 86.

Ovid, *Metamorphoses*. Translated by Ian Johnston. Nanaimo, BC: Vancouver Island University, 2011.

Parisi, Luciana. “Critical Computation: Digital Automata and General Artificial Thinking,” *Theory, Culture & Society* 36, no. 2 (2019): 89–121.

Pasquinelli, Matteo. *The Eye of the Master: A Social History of Artificial Intelligence*. London: Verso, 2023.

Peters, Benjamin. *How Not to Network a Nation: The Uneasy History of the Soviet Internet*. Cambridge, MA: MIT Press, 2017.

Pfaller, Robert. *Interpassivity: The Aesthetics of Delegated Enjoyment*. Edinburgh: Edinburgh University Press, 2017.

Pias, Claus, ed. *Cybernetics: The Macy Conferences 1946–1953*. Zurich: diaphanes, 2003.

Piccinini, Gualtiero. “The First Computational Theory of Mind and Brain: A Close Look at McCulloch and Pitts’s ‘Logical Calculus of Ideas Immanent in Nervous Activity.’” *Synthese* 141, no. 2 (2004): 175–215.

Poundstone, William. *Prisoner’s Dilemma: John von Neumann, Game Theory, and the Puzzle of the Bomb*. New York: Anchor Books, 1993.

Prigogine, Ilya. *From Being to Becoming: Time and Complexity in the Physical Sciences*. San Francisco: W. H. Freeman, 1980.

Prigogine, Ilya, and Isabelle Stengers. *Order Out of Chaos: Man’s New Dialogue with Nature*. London: Verso, 2017.

Pruijt, Hans. “Social Interaction with Computers: An Interpretation of Weizenbaum’s ELIZA and Her Heritage,” *Social Science Computer Review* 24, no. 4 (2006): 516

Rabinbach, Anson. *The Human Motor: Energy, Fatigue, and the Origins of Modernity*. Berkeley: University of California Press, 1992.

Raitt, Suzanne. “Psychic Waste: Freud, Fechner, and the Principle of Constancy.” In *Culture and Waste: The Creation and Destruction of Value*, edited by Gay Hawkins and Stephen Muecke, 63–75. Lanham, MD: Rowman & Littlefield, 2002.

Ramas, Maria. “Freud’s Dora, Dora’s Hysteria: The Negation of a Woman’s Rebellion,” *Feminist Studies* 6, no. 3 (1980): 472–510.

Ramtin, Ramin. *Capitalism and Automation: Revolution in Technology and Capitalist Breakdown*. London: Pluto Press, 1991.

Ray, Thomas S. “An Approach to the Synthesis of Life.” In *Artificial Life II*, edited by Christopher G. Langton, Charles Taylor, J. Doyne Farmer, and Steen Rasmussen, 371–408. *Santa Fe Institute Studies in the Sciences of Complexity*, vol. 10. Redwood City, CA: Addison-Wesley, 1992.

Rogers, Carl. “Significant Aspects of Client-Centered Therapy,” *American Psychologist* 1, no. 10 (October 1946): 421-423.

———. *Dialogues: Conversations with Paul Tillich, B. F. Skinner, Gregory Bateson, Michael Polanyi, Rollo May, and Others*. Edited by Howard Kirschenbaum and Valerie Land Henderson. Boston, MA: Houghtonn Mifflin Company, 1989.

———. *On Becoming a Person: A Therapist's View of Psychotherapy*. Boston: Houghton Mifflin, 1961.

Rose, Jacqueline. "Dora: Fragment of an Analysis," in *In Dora's Case: Freud—Hysteria—Feminism*. Edited by Charles Bernheimer and Claire Kahane. New York: Columbia University Press, 1978: 59–87.

Roudinesco, Élisabeth. *Freud: In His Time and Ours*. Translated by Catherine Porter. Cambridge, MA: Harvard University Press, 2017.

Russell, Ben. *James Watt: Making the World Anew*. London: Reaktion Books, 2014.

Russell, Stuart, and Peter Norvig. *Artificial Intelligence: A Modern Approach*. 3rd ed. Upper Saddle River, NJ: Prentice Hall, 2010.

Sapolsky, Harvey M. *Science and the Navy: The History of the Office of Naval Research*. Princeton, NJ: Princeton University Press, 1990.

Sarkis, Mona. "Interactivity Means Interpassivity." *Media Information Australia* 69, no. 1 (1993): 13–16.

Scherer, F. M. "Invention and Innovation in the Watt-Boulton Steam-Engine Venture." *Technology and Culture* 6, no. 2 (1965): 165–187.

Schrader, Astrid. "Haunted Measurements: Demonic Work and Time in Experimentation." *differences: A Journal of Feminist Cultural Studies* 23, no. 3 (2012): 119–147.

Schrödinger, Erwin. *What Is Life? The Physical Aspect of the Living Cell*. Cambridge: Cambridge University Press, 1944.

Shannon, Claude E. "A Mathematical Theory of Communication." *Bell System Technical Journal* 27 (1948): 379–423, 623–656.

Shannon, Claude E., and Warren Weaver. *The Mathematical Theory of Communication*. Chicago: University of Chicago Press, 1949.

Searle, John R. "Minds, Brains, and Programs." *Behavioral and Brain Sciences* 3, no. 3 (1980): 417–424.

Sejnowski, Terrence. "Large Language Models and the Reverse Turing Test," *Neural Computation* 35, (2023), 309-342.

Slater, Avery. “Chatbots: Cybernetic Psychology and the Future of Conversation.” *JCMS: Journal of Cinema and Media Studies* 61, no. 4 (2022): 181-187.

Smalheiser, Neil R. “Walter Pitts.” *Perspectives in Biology and Medicine* 43, no. 2 (2000): 217–226.

Smith, David Livingstone. *Approaching Psychoanalysis.* Taylor & Francis Group, 1999.

Song, Inhwa, Sachin R. Pendse, Neha Kumar, and Munmun De Choudhury. “The Typing Cure: Experiences with Large Language Model Chatbots for Mental Health Support.” *Proceedings of the ACM on Human-Computer Interaction* 9, no. CSCW2, Article 249 (2025): 1-25.

Spear, Brian. “James Watt: The Steam Engine and the Commercialization of Patents.” *World Patent Information* 30, no. 1 (2008): 53–57.

Srnicek, Nick. *Platform Capitalism*. Cambridge: Polity, 2017.

Suchman, Lucy. *Human–Machine Reconfigurations: Plans and Situated Actions*, 2nd edition. Cambridge: Cambridge University Press, 2007.

Szilard, Leo. “On the Decrease of Entropy in a Thermodynamic System by the Intervention of Intelligent Beings.” *Behavioral Science* 9, no. 4 (1964): 301–310.

Terranova, Tiziana. “Free Labor: Producing Culture for the Digital Economy.” *Social Text* 18, no. 2 (2000): 33–58.

The, Jessica Tran, et al. “From the Principle of Inertia to the Death Drive: The Influence of the Second Law of Thermodynamics on the Freudian Theory of the Psychical Apparatus.” *Frontiers in Psychology* 11 (2020).

Turing, Alan M. “On Computable Numbers, with an Application to the Entscheidungsproblem.” *Proceedings of the London Mathematical Society*, ser. 2, 42 (1937): 230–265.

Turkle, Sherry. *The Second Self: Computers and the Human Spirit*. Cambridge, MA: MIT Press, 1984.

———. *Alone Together: Why We Expect More from Technology and Less from Each Other*. New York: Basic Books, 2012.

———. *Reclaiming Conversation: The Power of Talk in a Digital Age*. New York: Penguin Books, 2016.

Tyson, L. D. and John Zysman, "Automation, AI and Work." *Daedalus* 151, no. 2 (2022): 256–271.

Umpleby, Stuart A. "Heinz von Foerster and the Mansfield Amendment." *Cybernetics and Human Knowing* 10, nos. 3–4 (2003): 187–190.

Vapnik, Vladimir N. *The Nature of Statistical Learning Theory*. New York: Springer, 1995.

van Dijck, José, Thomas Poell, and Martijn de Waal. *The Platform Society: Public Values in a Connective World.* New York: Oxford University Press, 2018.

van Heijenoort, Jean, ed. *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931*. Cambridge, MA: Harvard University Press, 2002.

von Foerster, Heinz. *Understanding Understanding: Essays on Cybernetics and Cognition*. New York: Springer-Verlag, 2003.

von Neumann, John. "First Draft of a Report on the EDVAC." Philadelphia: Moore School of Electrical Engineering, University of Pennsylvania, 1945. Reprinted in *IEEE Annals of the History of Computing* 15, no. 4 (1993): 27–75.

———. "The General and Logical Theory of Automata." In *Cerebral Mechanisms in Behavior: The Hixon Symposium*, edited by Lloyd A. Jeffress, 1–41. New York: Wiley, 1951.

———. *Theory of Self-Reproducing Automata*. Edited and completed by Arthur W. Burks. Urbana: University of Illinois Press, 1966.

von Neumann, John, and Oskar Morgenstern. *Theory of Games and Economic Behavior*. 3rd ed. Princeton, NJ: Princeton University Press, 1953.

, Judy. *TechnoFeminism*. Cambridge: Polity Press, 2004.

Wallace, Alfred Russel. *Contributions to the Theory of Natural Selection: A Series of Essays*. Whitefish, MT: Kessinger Publishing, 2016 [1870].

Weatherby, Leif. *Language Machines: Cultural AI and the End of Remainder Humanism*. Minneapolis: University of Minnesota Press, 2025.

Weber, Ernst Heinrich. “Concerning Touch.” In *Readings in the History of Psychology*, edited by Wayne Dennis, 152–159. Worcester, MA: Amberg Press, 2007.

Whitehead, Alfred North, and Bertrand Russell. *Principia Mathematica*. Vol. 1. Cambridge: Cambridge University Press, 1910. Reprint, Cambridge: Cambridge University Press, 1963.

Wiener, Norbert. *Cybernetics: Or Control and Communication in the Animal and the Machine*. New York: John Wiley & Sons, 1948.

———. *The Human Use of Human Beings: Cybernetics and Society*. Boston: Houghton Mifflin, 1950).

———. “Some Moral and Technical Consequences of Automation,” *Science* 131, no. 3410 (May 6, 1960): 1355–1358.

———. *God and Golem, Inc.: A Comment on Certain Points Where Cybernetics Impinges on Religion*. Cambridge, MA: MIT Press, 1964.

Wiener, Norbert, Arturo Rosenblueth, and Julian Bigelow. “Behavior, Purpose and Teleology.” *Philosophy of Science* 10, no. 1 (1943): 18–24.

Wilden, Anthony. *System and Structure: Essays in Communication and Exchange*. 2nd ed. London: Tavistock, 1980.

Wilson, Elizabeth A. “Would I Had Him With Me Always: Affects of Longing in Early Artificial Intelligence.” *Isis* 100, no. 4 (2009): 839–847.

Woodward, William R. “Fechner (1801–1887) For and In Psychology: Part I.” *Archives of Psychology* 2, no. 5 (2018): 1–27.

Woolgar, Steve. “Configuring the User: The Case of Usability Trials,” *The Sociological Review* 38, no. S1 (1990): 58–99.

Young, John Zachary. “Fused Neurons and Synaptic Contacts in the Giant Nerve Fibres of Cephalopods.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 229, no. 564 (1939): 465–503.

Žižek, Slavoj. “The Interpassive Subject.” *Traverses* 10 (1998): 1-18.

Zuboff, Soshanna. *The Age of Surveillance Capitalism*. New York: PublicAffairs, 2019.

Zucker, A., and D. Wiegand. “Freud, Fliess, and the Nasogenital Reflex: Did a Look into the Nose Let Us See the Mind?” *Otolaryngology–Head and Neck Surgery* 98, no. 4 (1988): 319–322.