PLSC 2021
Sebastian Benthall
NYU School of Law

Private Systems and Artificial Life

Note to readers: Thank you so much for considering this early stage work! I'm particularly interested in the following kinds of feedback:

- *Are there important works connecting privacy-by-design to second-order cybernetics and its influences that I have missed? Or other works that resonate with these themes but are not in this intellectual lineage?*
- *I understand that I have not yet* done the work *of showing how second-order cybernetics can deliver actionable insights for privacy regulation and engineering, and that this is the only true test of the claims here. However, I've tried to make a compelling case for a research agenda here as a source of ideas for more concrete projects. What other positive design recommendations/examples/principles does this line of thought lead to?*
- *Have I been able to present SOC in a way that makes it credible as a science and clearly relevant? Reasons to be skeptical? Known critiques of this line of work in literature I haven't addressed?*
- *Problems with the logic? Errors in the literature review?*
- *I'm passionate about this topic but am not used to writing this sort of article. What are good academic venues for this sort of work?*

I. Introduction

The burden of protecting individual privacy falls on those sociotechnical institutions that govern personal data in the aggregate with networked computing systems. Despite misgivings about attempts to solve social and legal problems through technical means, technical privacy-by-design remains an urgent challenge that demands a scientific operationalization of an elusive social and legal concept if only for the goal of establishing viable safe harbor conditions. However, many technical approaches to privacy-by-design have fallen short of normative and legal standards. Bracketing questions of political economy for the purpose of this article, we attribute the theoretical friction between social meanings of privacy and technical design to a scientific failure to operationalize privacy. We suspect that this is due to the limitations and biases of mainstream computer science as it has been defined historically by the field of artificial intelligence. We propose that other sciences that have heretofore been marginalized from privacy-by-design can be productively brought to bear on the problem.

After discussing several computer science and AI “solutions” to privacy and their shortcomings, we will introduce to privacy scholarship an alternative scientific paradigm from which to draw privacy designs. We refer to second-order cybernetics, a field that was originally an adaptation of the insights of cybernetics, the mid-20th century multidisciplinary study of control engineering, to itself as a scientific social field and to biology. This field has colorful origins associated with figures such as Margaret Mead, the pioneering anthropologist, Fernando Flores, the former Minister of Finance in Allende's Chile turned philosophy professor and entrepreneur, and Francisco Varela, the biologist who, after a conversion to Tibetan Buddhism, was one of the founders embodied cognitive psychology. This field has had over the course of decades had several fruitful encounters with computer science but has mainly developed in other fields such as psychology, "constructivism", and artificial life (AL). It is also influential in social theory and legal scholarship through the work of sociologist Niklas Luhmann and especially the legal theorist Gunther Teubner. We draw these connections to motivate a deeper exploration into what second-order cybernetics has to offer privacy-by-design as well as privacy scholarship more broadly.

We will conclude with a discussion of what the perspective of second-order cybernetics has to offer the design of data protection regulations and private systems. We arrive at three tentative conclusions. First, as a matter of regulatory design, the imposition of data processing requirements on technical systems and companies is perhaps less important than the establishment of an accountable coupling between the system and the regulatory body that takes seriously the requirements of each (the business/technology, the regulator body) to be viable in an ongoing way. Second, as a matter of technical design, we should stop approaching data governance as a matter of managing the representations of individuals flowing over channels and rather consider how meaningful data is based on an ongoing process-based relationship between the individual and the technical system. Lastly, second-order cybernetics encourages us to view data subjects in their biological specificity rather than as abstractly defined rows in a database; it is these biological properties, and the vulnerabilities entailed by them, that make personal data sensitive.

II. The urgency and challenge of privacy-by-design

Internet and software based systems have become a ubiquitous part of human life. These systems have made it easier for personal information to flow inappropriately, for example towards use in inappropriate applications (Nissenbaum, 2009). This is widely discussed as a problem of "privacy".

Perhaps because of an underlying ideology of political liberalism (Benthall and Goldenfein, 2021), the default legal regime governing personal data use in software systems relies on individuals engaging in private contractual agreements with companies that operate those systems. While well-intended, under this "notice and consent" regime individuals have so been overwhelmed by "notices" that they have little meaningful choice in how their data is used (McDonald and Cranor, 2008; Barocas and Nissenbaum, 2009; Cranor, 2012; Solove, 2013; Reidenberg et al., 2015; Obar and Oeldorf-Hirsch, 2020).

In many areas of consequence, such as the health and financial sectors, there are other regulations in place that stipulate rules for covered entities that employ personal data. These rules can interact with notice and consent, but may also include other kinds of rules and affirmative obligations. Today, sensitive personal information in even the most protected sectors, e.g., health and financial records are kept in digital systems at large and impersonal scale. The preservation of individual privacy has therefore become the responsibility of those sociotechnical institutions that govern personal data in aggregate. As these institutions typically use computing systems to manage their personal data holdings, there has been keen interest in privacy-by-design and privacy engineering standards which would guarantee an information system's compliance with privacy regulations (Barth et al., 2007; Fisler et al., 2010; Sen et al., 2014).

These efforts to engineer compliant systems have run into difficulties. Lawyers and engineers do not always “get along" intellectually (Swire and Anton, 2014). More substantively, compliance engineering researchers have found that some requirements in privacy regulations, such as those surrounding the intention of actors and purposes of data use, cannot easily be translated into the formal logical language of technical design (DeYoung et al., 2010; Tschantz et al., 2013). It has also been argued that the language of law and ethics is by its very nature ambiguous, contestable, and contextual, and so at odds with computing systems designed for fast, frictionless, portable, infrastructural applications (Hildebrandt, 2013; Hildebrandt, 2019; Selbst et al., 2019; Hildebrandt, 2020).

An important line of critique has exposed that many of the conflicts between business and technical architectures for personal data processing and normative or legal desiderata, let alone the weakness of our imagination for normative and legal desiderata, are due to the political economy of data (Cohen, 2019). For the purpose of this article, we will bracket these questions of incentives as much as possible (though we will not be able to do so completely) and focus more narrowly on the questions around technical feasibility. There are today prevalent paradigms of technical thought that have defined privacy in narrow terms.[1]

Technical approaches

Accepting the view that it is important to develop formal theories of privacy in order to facilitate privacy-by-design, in this section we will review some of the prominent proposed techniques from computer science. While these are all scientific accomplishments in their own right, they have not “solved the problem” of privacy successfully. We explore these techniques and their limitations before articulating what may be the general problem with these approaches.

Example 1. Differential privacy

[1] We suspect that those paradigms that are dominant are so precisely because they serve the interests of powerful political economic interests which do not have individual “privacy” in all its rich connotations truly at heart. Whether or not this is the case, it remains an open question what technical definition of privacy would satisfy normative and legal interests in the matter.

One approach to formally defining privacy originated in the study of the implications of the release of a database of personal information or a survey, and ways to modify the released information to preserve more of participant's privacy, in the case that each individual's personal information is secret. This problem has existed since before the widespread use of computers that we see today; however the problem is structurally similar to other problems involving tabular data. This use case is challenging because some information about the sensitive variables is required to flow to the database user, who in this case is also the adversary with respect to the secret personal information. In the best case, the system is private when an adversary that observes the system output is not able to infer anything about a sensitive value from it. This best case was introduced by Dalenius (1977) under the name *statistical nondisclosure*. One way to characterize this property is in terms of the probability: for any background, or auxiliary, information *B* that the adversary knows, sensitive attribute *S*, and system output *R*:

$$\Pr[S \mid B] = \Pr[S \mid R, B]$$

Where *Pr* is the subjective probability distribution of the adversary. In other words, the system output must provide no information about the sensitive value that was not otherwise supplied by the adversary's (arbitrary) background knowledge.

Unfortunately, for those wanting such a strong guarantee, Dwork and Naor (2010) proved it impossible to provide in general. Intuitively, statistical nondisclosure is threatened by the many ways in which undesirable statistical dependencies between inputs and outputs can arise. This motivated a different approach, *differential privacy* (Dwork et al., 2016), an intuitively weaker privacy guarantee that has enjoyed popularity in computer science research. Crucially, differential privacy is a property of the system itself, not a statement about an adversary's potential knowledge.

Formally, an algorithm *A* is ε-differentially private if for all subsets $S \subset image(A)$ and datasets D_1 and D_2 that differ on a single element, the probability of the output of the algorithm run on the datasets being in *S* differs by at most a multiplicative factor of e^{ε} and an additive factor δ. (Dwork & Roth, 2014)

$$[\text{Differential Privacy } ((\epsilon, \delta)\text{-DP})]$$

$$Pr[(D_1) \in S] \leq e^{\epsilon} Pr[(D_2) \in S] + \delta$$

When $\delta = 0$, then *A* is -differentially private (*ε-DP*).Differential privacy is defined in a way that suggests any individual's personal information is more protected by the algorithm as ε approaches zero.

Kasiviswanathan and Smith (2014) show the equivalence of the above definition of differential privacy with semantic privacy, which is a bound on the statistical difference of the posterior judgment about the database given the system's output transcript.

> [ϵ-semantic privacy)] A randomized algorithm A is said to be ϵ- semantically private if for all belief distributions b on $\mathcal{D}^n$, for all possible transcripts t, and for all $i = 1, ..., n$:
>
> $$SD(\bar{b}_0[\cdot|t]\bar{b}_i[\cdot|t]) \leq \epsilon$$

The strength of this approach is also its greatest weakness. Differential privacy is popular among computer scientists because it is a proveaby attainable property of a database, defined strictly as a causal relationship between system inputs and outputs (Tschantz et al. 2020). However, this is not was is normally socially meant by privacy. Our expectations of privacy must take into account the background information of adversaries. For example, consider a database of genetic information. With no background information, a differentially private genetic database might protect the private genetic information of data subjects in the database. However, if an adversary knows how subjects in the database are genetically related, they may be able to infer information from the output of the database based on the correlations between results.

Some researchers have questioned whether differential provides a strong enough privacy Guarantee (e.g. Kifer & Machanavajjhala, 2011; Kifer & Machanavajjhala, 2012; He et al., 2014; Zhu et al., 2014; Liu et al., 2016}. These authors have presented a range of privacy definitions of strength between differential privacy and statistical nondisclosure. An example of a privacy definition within this range would be one defined in terms of the earlier probabilistic equation but with restrictions of the kinds of available background information *B* available to the adversary. Indeed, the impossibility result of Dwork and Naor (2008) involves an adversary that is difficult to motivate as realistic, which perhaps contributes to the willingness of other researchers to propose definitions motivated by statistical nondisclosure.

Perhaps because of its limitations, differential privacy has excited great research interest from computer scientists but has found relatively few practical applications. Privacy is ultimately not about the design of computing systems; it is about the people who are using and used by those systems.

Example 2. Compliance tech

Another perhaps more nuanced class of technical solutions to privacy-by-design that has arisen in computer science literature takes a more granular approach to the control of information flows within a system. One line of work that draws on the theory that privacy is appropriate information flow (Nissenbaum, 2009) is instructive. This work attempts to reduce legal privacy requirements, such as those derived from sectoral privacy laws such as HIPAA and GLBA, to rules determining the allowable flows of personal information within a system. It then translates these rules into formal language used to specify software system designs. This enables the design of an auditing mechanism that can examine partial logs of a system's operation and flag when there may be compliance violations. (Barth et al., 2006; Garg et al,m 2011).

These designs attempt to explicitly represent the requirements of their social context as formal rules over which categories of data can flow where, and under what conditions. These approaches have run up against a number of challenges stemming from the limitations of their formal representations. Some requirements in privacy regulations refer explicitly to the intentions of actors who are collecting and using data, or more abstractly to the purpose of data use. These teleological concepts have proved elusive to the formalisms used by these computer science methods (DeYoung et al., 2010; Tschantz et al., 2013). Perhaps in pursuit of an automated aid to system compliance, these designs have excluded something normatively essential about the role of humans in the operational environment.
There are other challenges. Benthall, Gürses, and Nissenbaum (2017) note that computer science has thus far been unable to account for how social norms of information flow result from an adaptive social process over time. This is challenging for legal compliance to the extent that legal interpretation and enforcement varies with social norms. While a very few works (e.g. Criado and Such, 2015) take up the challenge raised by Dourish (2004a) of designing systems so that normativities are co-created by it and its users, this is extraordinarily rare in the literature. Martin et al. (2020) have proposed the use of complex adaptive system modeling to capture the context of machine learning systems as an aid to their design. This is a promising approach.

The general problem

Many articles have addressed in one way or another the friction or discrepancy between the nuances of social expectations and impact of technological systems and the rigidity of their design. These critiques range in scope between asking for more critical input or participation in the design of technical systems to the outright refusal of the use of computational or formal methods to address these social issues (Hildebrandt, 2019; Selbst et al., 2019; Abebe et al., 2020; Green and Viljoen, 2020; Fish and Stark, 2021).

We will not address this debate in depth in this draft. For the purpose of getting on with the argument, we will state our commitments. We maintain that the interests of the law and ethics of privacy are not *about the computers*, but are about people and their social needs. These are necessarily "outside" the computing system and implicate its social environment, and that computational designs that do not take this into account will very likely be inadequate. On the other hand, we reject the idea that the problem is *formalism*, or the use of mathematical tools for design. We look instead to the mathematized sciences of psychology and society to try to be more precise about the best design of computational systems.

We conjecture that we must look at sciences that are *outside* the purview of what has become the dominant field of computer science to find these answers. There are other sciences which have greater appreciation for how human values are contextualized but which still allows for general understanding and the potential for technical design. These have a rich history and are still practiced sciences today, though they are often not considered in the dominant venues of the ACM, the IEEE, and so on. What are they, and do they have anything to offer?

III. Second Order Cybernetics: A Genealogy

This section invites a change in perspective. We address this tension between data governance infrastructure design and the “social life” of law and personal information through the exploration of a third line of inquiry known historically as “second-order cybernetics”, an innovative extension of mid-20th century concepts from electrical engineering (“cybernetics” (Wiener, 2019)) to understanding the organization, autonomy, and self-creation (or “autopoiesis”) of biological life (Varela et al., 1974). In this section we present an genealogy of this field of thought.

Cybernetics was an amalgam of scientific practices and engineering techniques incorporated elements of control theory, as well as the recently discoveries of Shannon’s information theory and Turing’s theory of computing. However, it was more invested in the cycles of feedback between humans and machines than maximizing the amount of processing achievable by a computer chip. Many inventions by cyberneticists used analog, not digital, technology. Parallel to the emergence of artificial intelligence, cyberneticists began to extend their thinking to biological and social systems and asking deep questions about what it meant for themselves to be cybernetic systems within the world. This line of inquiry has resulted in an alternative constellation of scientific fields that continues to this day, though it is not represented in mainstream computer science. It is most connected with the names Norbert Wiener (*Cybernetics: Or Control and Communication in the Animal and the Machine* (1948); *The Human Use of Human Beings* (1950)), Ross Ashby (*An Introduction to Cybernetics* (1956)), and Stafford Beer (*Cybernetics and Management* (1959)), among others.

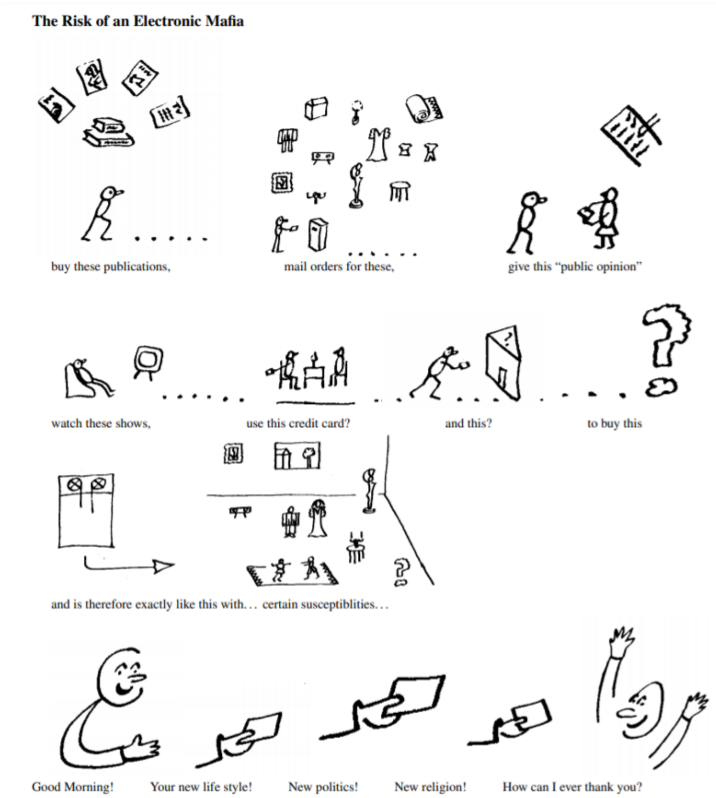
The concept of *second-order cybernetics* originates from the observation by anthropologist Margaret Mead that the scientific community working in the field of cybernetics in the mid-20th century could usefully turn its theories reflexively on itself to study its sustainability as a scientific field and capability of maintaining an objective perspective. In her article on the subject, she describes her experience proposing this idea *incognito* at a cyberneticist’s meeting in Atlanta and being dismissed as a crank before having the opportunity to present the idea again as a recognized speaker at the prestigious first meeting of the American Society of Cybernetics. (Mead, 1968) The idea that there could be a “second order cybernetics” was later advanced by Von Foerster, who would attribute its origins variously to Mead, to Stafford Beer, and to Humberto Maturana (Von Foerster, 2003).

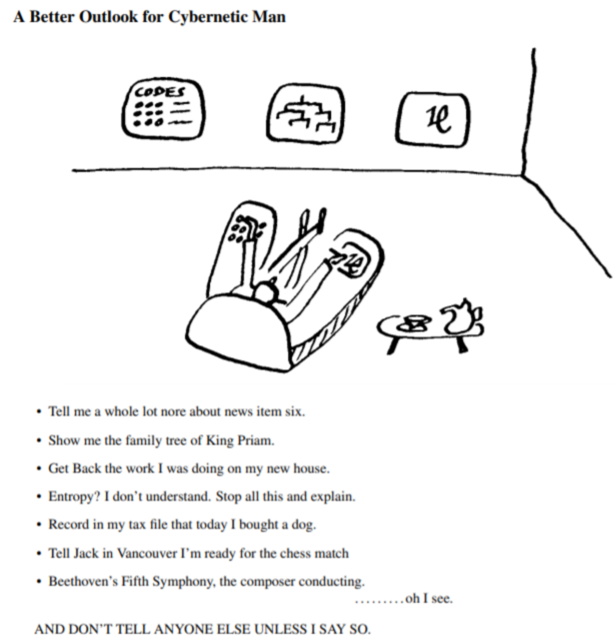
A. Cybersyn

Arguably, the origins of Second Order Cybernetics are in the Global South, specifically Chile. Medina (2006) relays the history of the influence of cybernetics on Chile under the leadership of Salvador Allende between 1970 and 1973. Allende's government was taken with cybernetics as a tool for socialist economic control, and invested in educating its scientists in this paradigm. The led to Allende inviting Stafford Beer, a British operations research consultant who drew publicly on the theories of Ross Ashby, to Chile to assist in the design of a telecommunications-driven system for monitoring and steering the country's economy. Beer was enthusiastic about the project and dropped all of his other commercial contracts to assist Allende's government in developing this system, dubbed Cybersyn. Beer worked under Fernando Flores, an engineer who at the age of twenty-seven became the Minister of Finance in charge of this cybernetic economic supervisor.

During this time, there was intellectual cross-pollination between the Cybersyn team and a team of biologists, Humberto Maturana and his student, Francisco Varela, though Maturana and Varela did not involve themselves much with economics. Rather, Maturana, who was in communication with von Foerster and the American Society of Cybernetics, was interested in the foundations of autonomous life, cellular biology, neuroscience, and what it meant for the ability of the scientist, or anybody, to be the subject of experience.

The economy did not fare well under Allende. Cybersyn, when put into operation, was successful in helping the government respond to a strike by the political opposition. But the project never evolved past its earlier stages because Allende's government was brought down by a CIA-backed military coup that put Augusto Pinochet in power. Beer would later deliver his “Designing Freedom” lectures, reflecting on the authoritarian use of cybernetics and its potential instead to uphold “all we hold most dear”. In these lectures, he anticipates the problems of “an Electronic Mafia” controlling every aspect of society, and imagines a better future in which “Cybernetic man” has elaborate control over the information they are exposed to, and, emphatically, absolute control over who learns about their queries and interactions with the media interface. (See figure.)





[From Beer, "Designing Freedom", 1973]

Meanwhile, Fernando Flores and Francisco Varela left Chile to live in exile in the United States. There, Fernando Flores became a philosophy PhD student and later professor at UC Berkeley. Francisco Varela would live for seven years in Colorado, where he practiced Tibetan Buddhism, developing his insights about the relationship between mind and the mind through meditation. We will pick up the threads of these thinkers and their scientific contributions in later sections.

A recent history of cybernetics (Pickering, 2010) has raised questions of whether the cyberneticists were serious or whimsical, and whether they were influential or an anomaly. Mirowski (2012) argues that its deepest influence was its contributions to neoliberal theory. "If Beer sounds eerily like a slightly wackier Herbert Simon, while reading [Pickering's history] I was struck by just how much of the later thought of Friedrich Hayek depended directly on these specific cybernetic theoreticians, and especially Ashby." In Mirowski's estimation, cybernetics "failed as academic science, [but] lives on as fundamental worldview" because of the computer as a cultural artifact.

We take an intermediate view. Indeed, most of first order cybernetics has been recuperated into computer science and neoliberal economics. Ashby and Wiener's contributions found their way into the work of Hayek and Simon. However, it was the transition from first to second order cybernetics that derailed its influence. Stafford Beer's intervention into socialist Chile and subsequent commitment to "designing freedom" was one moment when cybernetics transcended the study of machines to the study of designing society; Margaret Mead's intervention was another. This ambitious goal has not taken hold of the academic technical sciences. Perhaps now, with pervasive “smart” infrastructure, the necessity of designing freedom means this idea’s time has come.

B. Autopoiesis and artificial life

In parallel with these unfolding political events in Chile, a new scientific movement was developing at the intersection of biology, neuroscience, and philosophy. Led by biologist Humberto Maturana, whose scientific work included the student of the development of biological cells and the visual system of frogs, pioneered this new approach by grounding it in the role of the scientific observer in the expression of any scientific theory: “Anything said is said by an observer”. From this origin, Maturana and his student Varela developed a comprehensive theory, or “cosmology”, of living systems and cognition.

This perspective, most canonically articulated in *Autopoiesis and Cognition: The Realization of the Living* (Maturana and Varela, 1972/2012), takes on the task of attempting to explain the nature of cognition, which in more mainstream fields was considered the work of philosophers of mind and cognitive psychologists (cf Kim, 2018). In the United States and United Kingdom, these latter fields were closely linked with the development of computer science and artificial intelligence. These fields were at this time still building on the mentalistic linguistics of Chomsky (1957) and exploring a naturalist theory of mind that equated mental states with objective functional, information-processing, or computational properties of the individual brain (Putnam 1960, 1967).

Maturana and Varela, in contrast, were not in the thralls of computationalism. Their position departs from that approach in two significant ways. First, their theory does not attempt to *naturalize* cognition; meaning, it does not attempt to reduce cognition to observer-independent properties of a system. Rather, it is explicitly grounded in *phenomenology*, or first-person experience, and indeed takes the necessity of the scientific observer in scientific description as not only an axiom but a useful one in explaining the phenomenon of cognition. Second, their theory explains cognition not in terms of a “computational” state of the brain, but rather as a function of the process by which a biological system recreates itself over time. They introduce the term *autopoiesis* for this process of self-creation. In other words, cognition for Maturana and Varela is not the living being’s processing of information. Rather, it is the process of maintaining itself as a living system in a changing environment. It is therefore *self-referential* in ways that are difficult to render in the computationalist theory of mind.

Autopoiesis is perhaps best understood in reference to a simpler system: a biological cell. A cell is composed of biochemically constructed components, or organelles. These same subsystems execute the process of producing themselves, or replications of themselves, out of chemicals in the environment. For Maturana and Varela, this self-reproduction is a form of cognition, analogous to the process of cognition at higher levels of life, including those with more advanced neurophysiology. According to this theory, cognition, and perhaps consciousness, is therefore more of a matter of *self-reference* than it is a matter of *representation* (via mentals states or “mental language”) of an outer world. This distinction prefigures later controversies in philosophy of psychology over representationalist theories of consciousness (see Lycan, 2019).

From the beginning, second-order cybernetics made use of computer models and mathematical formalism without ever being primarily *about the computers*, or making explicit analogy between living systems and concepts from computational engineering. The original autopoiesis model proposed by Varela, Maturana and Uribe (1974) was of a cell membrane forming out of more basic chemical components arranged in a grid. The chemistry of the membrane encourages it to form in strands, recreate itself out of elements from the environment when it deteriorates, and loop back on itself to create an enclosure for distinction between its internal system and environment.

Varela would continue to be an influential figure in the emerging science of *artificial life*, at the intersection of biology and computation. This field would incubate at the Sante Fe Institute and spread to academic institutions. As an editor of the proceedings of the first European Conference on Artificial Life (ECAL), Varela would elaborate on his original model of living systems with Maturana, emphasizing the operational closure of living systems over the states that are viable given a changing environment (Varela & Bourgine, 1992). In this work Bourgine and Varela would not that adaptive, cognitive strategies for the sustained viability of life operate at the neuronal level, but also separately at the levels of the genetic system and immune system. Later the original autopoiesis model was revisited by a collaboration between McMullin and Varela (1997) published by the Sante Fe Institute.] This model continues to be studied as a foundational example of autopoietic life (Priya and Nehaniv, 2020).

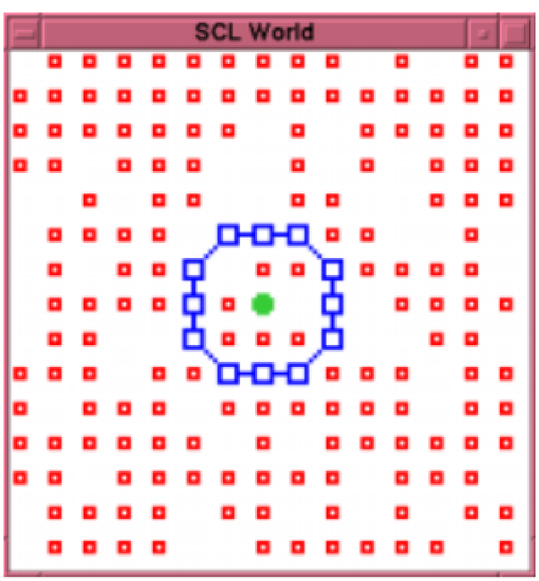


Figure 2: Initial Configuration.

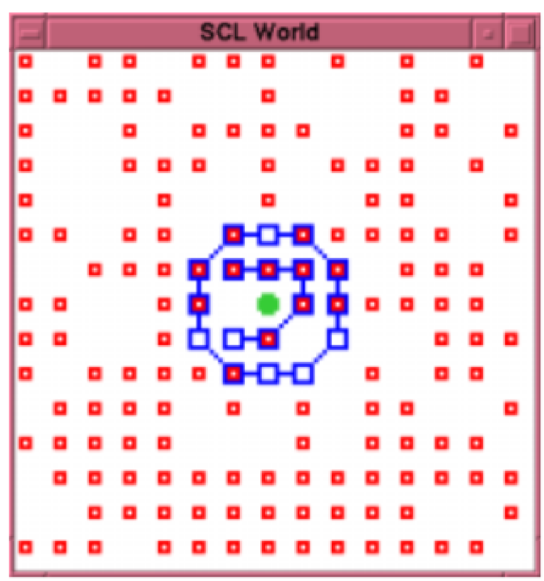


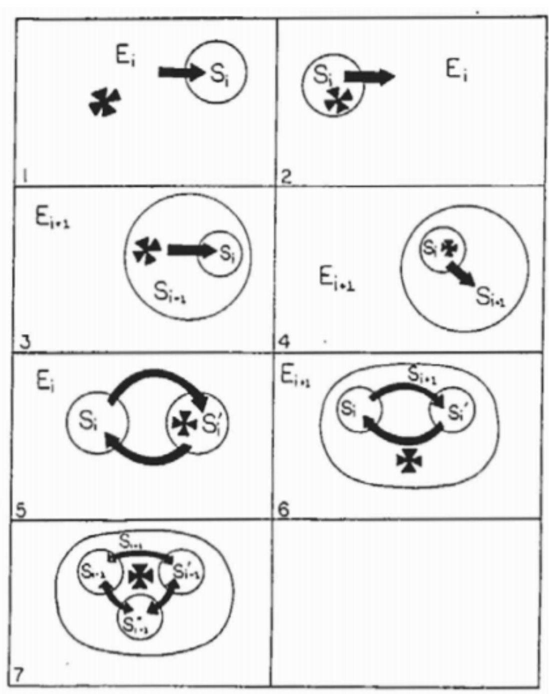
Figure 3: Experiment 1, Run 1, Time 110.

[From McMullin, B.M., Varela, F.J. (1997).]

Artificial life continues as a field of study today that is perhaps most closely associated with robotics research. While artificial *intelligence* has focused on getting more and more performance out of computational machines with respect to information-intensive tasks, artificial *life* has developed a formal, computational theory of the living system and how it adapts to its environment. The rise of the consumer Internet has brought the field of AI to prominence. We suggest that a revisitation of artificial *life*, and its implications for the design of computational systems that pervasively direct human and natural life, is overdue.

C. Influences on Computer Science (1980-2000)

Today, the dominant narrative of the sciences does not include most of these figures. Rather, it is a narrative that tells the origin story of the rise of the personal computer, consumer Internet, and neoliberal economic theory. In this dominant narrative, Turing and Shannon are lauded for their discoveries. These are consolidated into an early theory of artificial intelligence by Herbet Simon, who also wins a Nobel Prize in Economics. Computing and networking continue to advance and is led by entrepreneurial figures and industry magnates. Economics and computation connect once again in the development of electronic exchanges, which drive further industrial expansion.



FIGURES 1–7 Various configurations of systems, subsystems and marks; each configuration represents a cognitive viewpoint, and the mark indicates its center. The arrows indicate the flow of signals and interactions.

1—Control of a System S_i by its Environment E_i. 2—Autonomy of System S_i in its Environment E_i. 3—Control of a Subsystem S_i in a System S_{i+1}. 4—Autonomy of a Subsystem S_i of a System S_{i+1}. 5—Feedback Control of System S_i by System S_i'. 6—Communication between (Coordination of) Two Subsystems S_i, S_i' of System S_{i+1}. 7—Coordination (Ecology) of Subsystems S_{i-1}, S_{i-1}', S_{i-1}''.

Philosophical questions about the meaning of life are secondary to this story of science and industry which is encapsulated in the discipline of Computer Science. However, there have been numerous occasions in which second order cybernetics has engaged with computer science and made a lasting impression in the field. We discuss some of these cases in this section.

An early bridge figured between second order cybernetics and computer science was Joseph Goguen, a professor and researcher who worked in many areas of formal methods at the foundation of computer science. His retrospective essay (1999) summarizes his many contributions, while also opining on the state of computer science, its divergence from the study of ethics, and enduring challenges of coordination with social expectations. Goguen is perhaps best known today for his foundational work on computer security (Goguen and Meseguer, 1982).

Goguen was also a Tibetan Buddhist practitioner, an editor of the *Journal of Consciousness Studies*, and a collaborator with Varela on the topic of the mathematics of subjectivity. Goguen and Varela (1979) draw on Goguen's background of the mathematics of category theory to formalize the distinction between a system and its environment, and system's relationship to subsystems and other systems, and the way in which each system can maintain its own "cognitive viewpoint" (see figure).

[Figure from Goguen and Varela, 1979]

Meanwhile, in Berkeley, California, Fernando Flores, the exiled former Minister of Finance of Chile, was studying philosophy under Hubert Dreyfus. Flores would complete a doctoral thesis about the design of workplace technology. Later, he published an influential critique of artificial intelligence with computer scientist Terry Winograd, *Understanding Computers and Cognition* (1986). This book critiques the "rational" approach taken by artificial intelligence and advocates instead for an approach based on Maturana's theories of *autopoiesis* and structural coupling, as well as the phenomenological philosophy of Heidegger and Gadamer. This book is now considered a classic in the field of Human Computer Interaction.

This advancement of thinking about human-computer interaction as an alternative to artificial intelligence is a landmark. Since then, others have advocated for and debated greater inclusion of "lived experience" and phenomenological subtlety into technical design. Lucy Suchman (1993) would engage Winograd (1993) on these issues, advocating a perhaps more radical centering of the human in design. The field of Computer Supported Cooperative Work would emerge with contributions from Goguen (1997), Suchman (Blomberg,Suchman, Trigg, 1996) and Agre (1997). These works would develop the theme of Heideggerean phenomenology, combining it with the ethnomethodological methods of Garfinkel. This volume, edited by Bowker, Star, Turner, and Gasser (1997) was aimed at "the great divide" between social sciences and computer sciences. It's introduction nicely summarizes the challenges of bridging the varieties of formalism, divisions of labor, and ethical responsibility in terms that could be fruitfully revisited today.

However, despite the efforts to develop CSCW as a field of organic unity between heterogeneous disciplines, the tensions between artificial intelligence practices and design methods inflected by ethnomethodology and phenomenology remained. They would be revisited by Dourith (2004), among his other contributions to HCI that draw from Heiddeger and Garfinkel, in work on "embodied interaction". But however much ground has been gained by these approaches, they have a tendency to recede into a negative form of criticality that sabotages its own influence on design. For example, Phillip Agre, Internet, artificial intelligence, and critical design pioneer, vanished in 2009. He was found by the police in 2010 in good health but has not been publishing. Critical scholars in this tradition today remain in a defensive mode, seemingly endlessly in contest with the excesses of artificial intelligence. Has nothing changed in thirty years? Heidegger (1954), it must be said, not only advanced phenomenology but was skeptical of the role of technology in society *at all*. Perhaps a Heideggerean approach to technical design is self-defeating.

Another, perhaps less enduring, moment of second-order cybernetic's contact with computer science has been its influence on the practice of software development. The 1988 conference on Software Development and Reality Construction, in 1988, brought together diverse thinkers including von Foerster and Goguen, as well as Donald Knuth. It's proceedings editors included Christiane Floyd, the first woman computer science professor in Germany. In it, Goguen would argue that a software development project is, itself, an autopoietic system with, in a sense, a life of its own. The proceedings of this conference are delightful, including fanciful coda that summarizes the contents of the conference in the form of a screenplay derivative of the story of "Alice in Wonderland".

However, the arguments of these proceedings have not been taken up much in subsequent literature. (With notable exceptions. See Jus and Latzer, 2017 and Lilienthal, 2019). This is unfortunate, as the work prefigures many of the calls made today for more socially responsive technical design. Consider these concluding passages from Klein and Lyytinen, (1992), about the practice of data modelling (closely akin to artificial intelligence practices today).

> It was noted that the prevalent metaphor of data modelling is that of a representation that should correspond to reality. In conclusion we propose that a more fruitful metaphor is to view data models as a set of laws and data modelling as the activity by which laws are designed and enacted.
>
> ...
>
> Just as laws lead people to filter information and put their best foot forward, so do the rules of data models. Just as laws are formulated ambiguously to be acceptable to a mixed coalition of supporters, so are data models. Just as laws need to be approved by some sort of due process, so do data models. Just as laws constrain policy making and at the same time are the result of policy, so do data models. Just as laws should be supported by a democratic majority, so should data models, and this may only be possible at the expense of consistency or incompleteness. Just as the interpretation of laws changes with shifting policy orientations, so are data models. As information systems move closer to the centres of gravity of power and political will formation, the entanglement of schema development and interpretation with policy formation and implementation is a subject that can no longer be avoided.

Twelve years before Lessig's (2000) "code is law" formulation, Klein and Lyytinen argue that society will be regulated by code--specifically, database schemas and ontologies--and that its multiple constituencies imply a need for its design to be democratically legislated. This is perhaps easier said than done.

E. Enactivist psychology

Varela would eventually settle in France in 1986. By this time, he had been practicing Tibetan Buddhism for a decade, and through this practice was developing his synthetic science that combined neuroscience and phenomenology. In 1987, he co-founded with Adam Engle and the Dalai Lama the Mind and Life Institute, which was dedicated to the study of “contemplative sciences”, or the scientific study of the mind as experienced, especially via meditation. In 1988, Varela became a Director of Research at the Centre National de Recherche Scientifique and remained so until his death in 2001. Hence, while Varela’s work has been somewhat esoteric with respect to research communities in the United States, his legacy is more widely respected and continued in Europe.

Varela’s earlier work on neurophenomenology, which drew explicitly on the phenomenology of Edmund Husserl and Maurice Merleau-Ponty, blossomed into a field now known as embodied psychology or enactivism. *The Embodied Mind: Cognitive Science and Human Experience* (1991/2017), by Francisco Varela, Evan Thompson, and Eleanor Rosch is considered a landmark text in that field, which revitalized earlier philosophical phenomenology with methods from cognitive psychology.

A key result in this line of inquiry is the confirmation of Merleau-Ponty’s point in *Phenomenology of Perception* (1945) that conscious experience supervenes not just on “the brain”, as some “computationalist” psychologists and philosophers have argued, but on the whole body and even in the body situated in its environment. An example familiar to most readers of this article will be the experience of working at a laptop computer and writing an email. When first learning to type, one is conscious on the individual keys and keystrokes. But as one becomes more skilled, the physical act of typing fades and one becomes conscious instead of communicating words. In a heated conversation over text, the acts of transmission can fade entirely, overwhelmed by the emotional engagement with the other person. The “phenomenal body” extends beyond the brain and even physical body to include tools. In analytics philosophy of mind, this thesis had become known as the “extended mind hypothesis” (Clark and Chalmers, 1998), which continues to be closely associated with Andy Clark.

Today, there remains a lively research agenda around *embodied cognition* (Wilson and Foglia, 2017) associated with, for example, Andy Clark (See Appendix A) and Ezequiel A Di Paolo. Work in this area need not be directly phenomenological -- indeed much of it is done in the mode of naturalist science or robotics engineering. However, it is a body of scientific work that from its origins is intended to be consistent with sensitivity to lived experience. This positions it differently with respect to many contemporary critiques of technical design.[2]

---

[2] In terms of philosophical history, it may be that the distinction between different schools of thought within phenomenology has been lost on some interdisciplinary fields in the United States. In particular, while writers such as Agre and Dourish have been eager to connect critiques of AI with Heidegger, who as noted was skeptical of technology in general, they have not picked up on the tradition of Merleau-Ponty, who developed the phenomenology of Husserl in conversation with the emerging cognitive psychology and neuroscience practices of his time. In short, while both Heidegger and Merleau-Ponty were phenomenologists, Merleau-Ponty was not a technophobe and his account of lived experience is more

D. Systems theory and the law

Talcott Parsons, the sociologist, was present at the first meeting of the American Cybernetic Society where Margaret Mead made her declaration of “Cybernetics of Cybernetics”. Parsons’s systems theory was to be influential in structuralist sociology. Parson’s student, Niklas Luhmann, developed systems theoretic sociology further and based it more explicitly on Maturana and Varela’s notion of autopoiesis. Indeed, Luhmann’s work is largely about elaborating on the use of autopoiesis to understand social systems as a form of life.

Much as Maturana and Varela grounded their understanding of autopoiesis in biology in the recognition that the scientific speaker must also be an observer and living being, so too is Luhmann’s sociology grounded in reflexive self-reference. Luhmann (1995) is pointed about how his theory’s *first* test is whether it is able to account for itself as a system capable of having a perspective on the world. In this way, his systems theoretic account continues to defeat the dichotomy between naturalism and phenomenological validity, or the complaint that scientific theories are necessarily biased because they are theory-laden. Rather, Luhmann’s theory has the advantage that it is able to ground itself with a consistent account of its own universality.

One important concept developed by Luhmann (originally from Parsons) and used in some analyses of data protection law is *double contingency*. Double contingency is the feature of communication between two autonomous persons that both parties will be in many respects uncertain to the other. Communication depends on a continuous effort to prevent errors in interpretation and maintain the structural link that allows the communication to continue. Neither party is in complete control. Double contingency has been introduced to the (legal) scholarly discussion of privacy by Hildebrandt (2013)

Teubner has most strikingly applied Luhmann’s theory to law, arguing that law itself is an autopoietic system (1993). The law is constantly referring back to itself (self-reference). Many expressions and activities within law are designed, almost foremost, to maintain and reproduce legal authority. Rules enforcing respect with courts and accountability to regulatory agencies are part of the process by which law continues to establish itself over time. To the extent that “smart technologies” elude the comprehension or oversight of the law, they may indeed erode the “rule of law” (Hildebrandt, 2015). Teubner’s (2006) usefully applies Luhmann beyond the understanding of law itself to the understanding of new forms of agency that are to be regulated by law. In his elucidation of the law and ethics of non-human persons, he argues that Luhmann’s theory of personhood as being an autonomous system capable of communication enables the legal conceptualization of artificial agents, including the corporation but also, relevantly, an artificial intelligence system.

---

compatible with advances in science and design. From the Merleau-Pontian perspective, and arguably that of Varela, Clark, and this while lineage, the dichotomy between lived experience and science/technology is simply false. Rather, the scientific task is to understand lived experience rigorously, using mathematical models if necessary.

Teubner has been taken up as a foundation for other legal arguments by Baxter (1997) and Sheffi (2020).

IV. Implications for Regulation and Design

In this section we will consider how second order cybernetics can motivate new interventions in data protection regulation and privacy-by-design.

A. Regulations

Autopoiesis and data protection regulation

The greatest contributions of second-order cybernetics to legal theory is via Luhmann and Teubner. According to this view, social systems in general recreate themselves much like other forms of life. Not only are they autonomous, but they are autopoietic. Teubner applies this specifically to legal systems (Teubner, 1993). Sheffi (2020) extends this analysis to the self-regulation of a platform company, AirBnB.

This perspective shines a different light on the relationship between regulations, covered entities, and their technical systems. In earlier work by computer scientists on privacy-by-design, there have been attempts to represent the propositional contents of laws directly in order to translate them into formal requirements for software systems. This approach relies on an unrealistic understanding of how law is promulgated and perpetuated through agonistic trials, judicial speech acts, and regulators with enforcement discretion. An alternative approach more aligned with the science of second-order cybernetics would look at the relationship between the legal system, the system of the covered entity, and the technical system, and focus on how each system reproduces itself as it interacts meaningfully with the other systems. This perspective is corroborated by Waldman (2019) in work on legal endogeneity: how businesses internalize legal regulations into processes that are abstracted from the legal intent and viable within the process of maintaining a business.

Formulations of ethics for digital systems, which might assist privacy-by-design by anticipating social demand for the use of regulation, is a theory of the grounds of normativity. Contextual integrity theory (Nissenbaum, 2009; Benthall et al., 2017) grounds normativity in the adaptivity and survival of social spheres, which is quite close conceptually to one of the central claims of second order cybernetics: that *meaning*, for an organism, including a macro-level, 'social system' level organism, is tied to the conditions under which it reproduces itself. Ergo, a systems theoretic ethics of artificial intelligence must begin not with the instrumental normativity of optimization of a subjective utility or profit function, but with the existential normativity of maintaining the conditions under which individual and social life reproduces itself. However, contextual integrity suffers from its characterization of information flow via the conduit metaphor

(Reddy, 1979), and so is ill-suited to comprehend the meaning and normativity of intersecting circuits of data flow in a global system that exceeds any naive and socially situated understanding of social norms.

Example: PCI DSS

An example of a data regulation that is designed in a way that is perhaps more consistent with autopoiesis of social systems is the Payment Card Industry Data Security Standard (PCI DSS). A industry standard developed jointly by Visa, Mastercard, American Express, Discover, and JCB, PCI DSS is now a standard that maintains interoperability between different parts of the payments ecosystem while aiming for a high degree of security and reliability.

Any vendor that accepts payments via payment cards is required to adhere to these standards, which include rules about how a network should be secured and data should be managed. For example, it is the PCI DSS that mandates that the three digit security code on a credit card be stored only temporarily while a transaction is validated. This feature goes beyond mere “data minimization” to use the short storage of data as a positive security measure.

PCI DSS works as a regulatory policy in large part because it is underwritten by the central institutions of the payment cards systems, the card issuers. Critically, the payment cards system is not primarily adding a new layer of functional meaning on data that is already issued by some other authority, such as social security numbers or digital device IDs. Rather, it issues payment cards, with card numbers and security codes, and mandates that any business using those cards to transact comply with standards around how the information is used.

Hence, the meaning of a payment card number is *never* simply “representative” of a person. Rather, it is always structurally coupled with a system of payments and vendors bound by information rules. The reliability of the complete circuit of card use, from individual actions to change’s in the individual’s environment (through purchases) is clear. This system is effective and, today, infrastructural.

B. Designs

Data circuits not information channels

A theme that runs throughout the scientific applications of second-order cybernetics is that phenomena of significance -- our conscious experience, the meaningfulness of our speech and actions, etc. -- are not due to the transmission and storage of static representations. Rather, what we sometimes call “representations” are patterns of structural coupling between a system and its environment. In psychology this shifts the focus away from the abstractly cognizing cortex to the sensorimotor system. In philosophy this is related to the emphasis on the Heideggarean notion of the phenomena of objects that are “ready-at-hand”, available for action, and the emphasis on speech *acts*, performative utterances like promises and apologies, rather than the propositional content of language. Extended from the individual to the social, Dourish (2004a) proposes that a system's context is dynamically created through the interaction between agents and the system, and cannot be so easily represented explicitly in memory.

A corollary of the enactivist or embodied view of cognition is that the most important reference of cognition and experience is *self-reference*, and especially the implicit or explicit reference to the conditions of one's viability or survival. This is, according to theory, true whether the 'observer' is a single cell lifeform, a human being, a business, or a legal system. When a speaker A sends a message X to a listener B, the meaning of X for A will be in terms of A's viability, and the meaning of X for B will be in terms of B's viability. If X is to have a stable meaning between A and B, it will be because A and B are part of a larger system, C, and that X has meaning for C's viability.

Some consequences of this for technical design:

- We should not consider data to be a representation in technical designs. Data is defined by its role in functional processes.
- Data is personal when it is an element of a structural coupling between an individual and another system. In the context of a digital product or service, the data is a coconstruction of the user and the platform, not strictly belonging to either.
- The data can never be "transferred", as if it were a physical object or private property title. It is not an object; it is a relationship or circuit of activity between systems.

Considering data in this way will necessarily make use reconsider whether and how data is nonrivalrous. If data does not "flow" but rather abides in the networks of coupling between systems, the topology of the information economy looks quite different from anything imagined in the liberal imagination (Benthall and Goldenfein, 2021). This reality requires new thinking.

Fair AI and biological agents

Enactivist psychology tells us: people are sensitive about their bodies. Second-order cybernetics is not shy about how living systems are biological. It explicitly rejects the theory that the minds of living beings are just like computers that can be disembodied and abstracted from their history and form. Gone is the model of the agent as a single utility function attached to a robot. Instead, we have the model of the human as the ongoing interaction of many differently intelligent subsystems: a genetic system, an immune system, and, yes, a cognitive system.

Insisting on the recognition of human biology, while possibly uncomfortable in some respects, will prevent some errors in privacy-by-design that result from the undue abstraction over sensitive attributes. For example, Benthall and Haynes (2019) note that in early approaches to 'fairness in AI' by computer scientists, race was construed as an abstract category or label disconnected from its specific role in, e.g., U.S. society as a hierarchical political category. The history of race in the U.S., which is in many ways a legal history, renders clear how each politically constructed race has a different social history, and how, while not itself a biological category, race is ascribed to individuals on the basis of what are *loosely* inherited traits such as ancestral origin, nationality, language, and phenotype. A computational system that is designed to adjust itself to legal requirements surrounding racial categories will not be adequate unless it is able to somehow internalize and respond to this history of real, living beings and their biological ancestry.[3]

This analysis extends, or else does not extend in subtle ways, to other protected categories such as biological pregnancy, sex and (dis)ability. In many cases the science of these categories is complex, unsettled, and politically disputed. Privacy-by-design in large scale software systems will not perform *better* than these scientific fields by ignoring what is most important to people, their bodies.

V. Conclusions and next steps

We have examined recent approaches by computer science to address personal privacy and data protection concerns, as well as related questions surrounding ethical AI. We have found that these techniques have had limited success for a number of reasons, including: the modeling of data as a mode of representation rather than a mode of regulation; and the omission of models of the social environment of technical systems. Lawyers and ethicists on the one hand and computer science or artificial intelligence approaches on the other seem to have “a great divide” between them.

This has happened before in the history of computing. A third way, known as second order cybernetics, has since the late 1960’s investigated the liminal intellectual territory between social and machine systems. We have traced a genealogy of this tradition from its origins in the observations of Margaret Mead and Humberto Maturana, through its influences on computer science in the late 80’s and early 90’s, to its continued influence today in psychology, biology, and law. Second order cybernetics is a rich vein of thought fueled over decades by intellectual curiosity and social conscientiousness. It claims both legal relevance, phenomenological insight, compatibility with mathematically rigorous empirical science. It has potential to be applied anew to the challenges of technology in society today.

We have discussed some potential applications of second order cybernetics to privacy regulations and design. While consistent in many ways with recent work in the these fields, second order cybernetics provides a fresh and unifying perspective for critical attitudes of AI and privacy, while not foreclosing the opportunity for new technical and regulatory advances through

---

[3] This conclusion has been resisted by Hanna et al., 2020 on the grounds that personalized genetics testing from services such as Ancestry.com is not a reliable data source. While the confusion is understandable, we see this argument as besides the point. The genealogical data relevant to the definition of race is more effectively, and less invasively, tracked through, e.g., birth records and personal testimony than genetics testing. Ancestry.com is not what we mean by “ancestry”. We also caution against overly reductivist or inaccurate uses of ancestral information to reinforce unfair or hierarchical racial categories. See Benthall and Haynes, 2019, for details about how ancestry information can be included in an anti-racist AI design.

scientific progress in understanding of the meaning of life. We see this has a hopeful line of future work.

Appendix A. Mathematizations

For the PLSC audience, we have left the discussion of the article above at a broad, readable, and qualitative level. However, second-order cybernetics will only be relevant to technical design if it can be rendered into formalisms that are legible to mainstream computer science and engineering. To some extent this work exists already in the relevant scientific fields. Ideas in the intellectual lineage of second-order cybernetics have been operationalized according to modern statistical theory (Berstchinger et al. 2008) and used for measuring Web activity (Oka and Ikegami, 2012), the understanding of information systems ethics (McBride, 2014), and the evaluation of AI systems (Vakhrameev et al., 2020). We survey some of this work below to give the reader a flavor of what is available.

1. Bertschinger et al. 2008

Bertshinger et al. (2008) consider the proposals of Maturana and Varela and develop formal definitions of their terms defined in terms of information theory recognizable by scientists today.

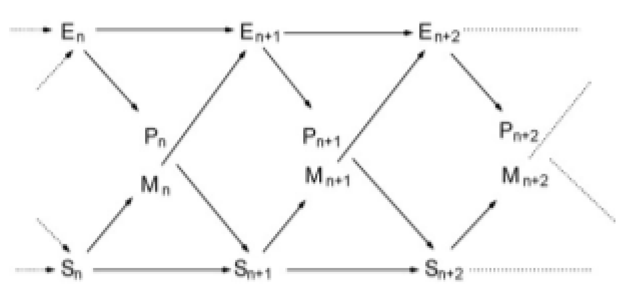
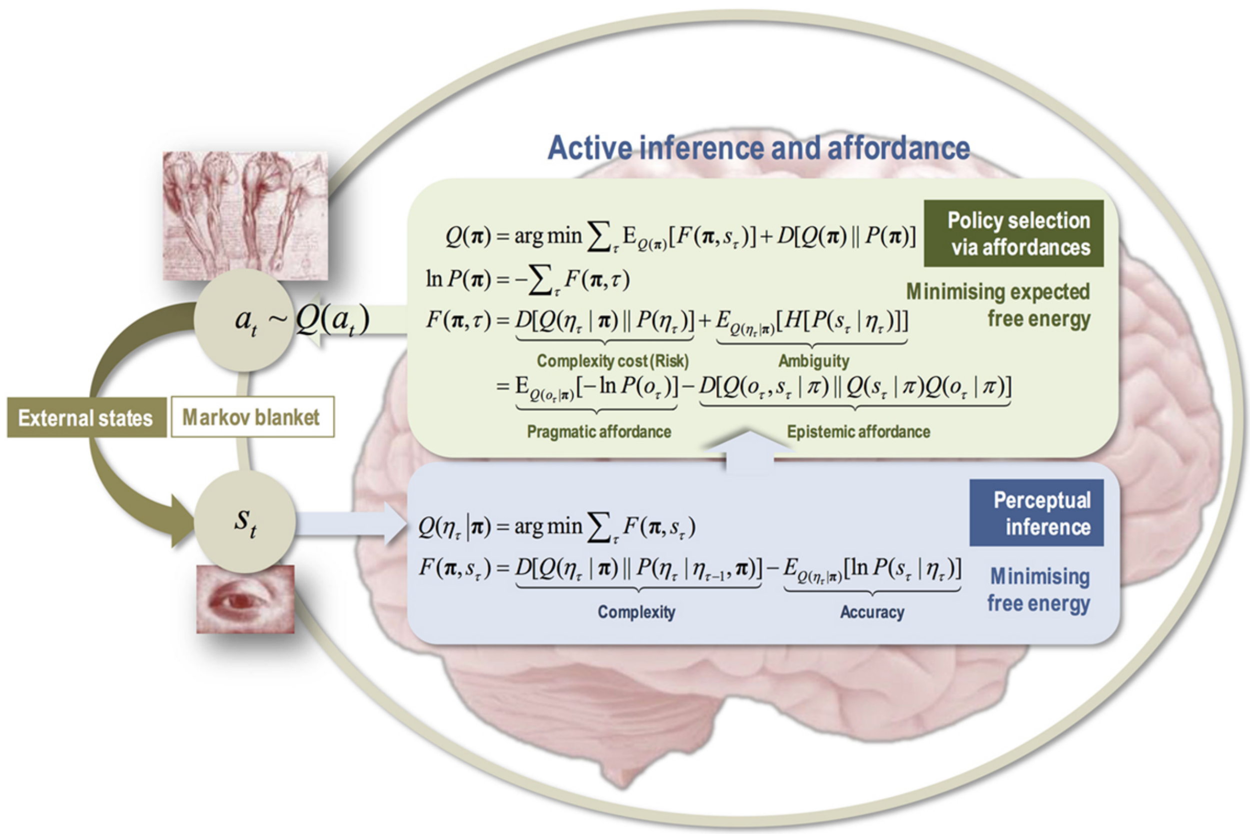


Fig. 1. The system S and the environment E interact through the channels P (perception) and M (motor output). The figure shows the temporal dependencies of this interaction.

(Figure 1. To the right is from Bertschinger et al.) In their conception, autonomy is a measure on the relationship between a system *S* and its environment *E* over time. The system is imagined as a biological agent with the capacity for perception *P* and motor control over the environment *M*. Using this model, they are able to posit a mathematical definition of *interactive autonomy* in terms of how the system is both not entirely determined by the history of the environment, but also not entirely independent of it (random). Rather, the system must proceed in a way that reacts to and acts in the environment in a way that is ultimately determined only by its internal state. Bertschiner et al. note that this definition falls short of another meaning of autonomy, *constitutive autonomy*, which is a qualitative distinction made between a living system and the non-living systems it interacts with.

2. Active Inference (Friston, Kirchhoff, Clark)

A recent proposal raised by an interdisciplinary group that includes neuroscientist Karl Friston, philosopher Michael Kirchhoff, and psychologists Andy Clark (see above) is an ambitious theory called “active inference” (Kirchoff et al, 2018; Linson et al., 2018). This theory attempts to unite thermodynamic theories from biology with Bayesian psychological theory to provide a better theory of how an autonomous agent thinks and adapts ecologically. Unlike Bayesian models, “active inference” theories explicitly address the need of an autonomous system to reduce its “free energy” in order to maintain a stable (biological) structure against the pressure of entropy posed by the forces implied by the second law of thermodynamics. They then bound from this to theories about the adaptive processes of the embodied mind. Because “surprise” in a statistical sense is a kind of “variational free energy”, for physical reasons the biological agent must have a kind of embodied motivation to reduce their surprise over time, and this should be the revealed goal of their choice of actions over time. The influence of Clark on this work puts it in the lineage of *embodied psychology*, which draws on the perceptual theories of Merleu-Ponty and is explicit about the ecological situatedness of the mind. Ramstead et al. (2020) draw the connection explicitly to the enactivist psychology of Varela, Thompson, and Rosch (1991/2017). We hence include it among the contemporary theories that have roots in second-order cybernetics.



[Above figure from Linson et al., 2018. It frames the “problem” faced by the agent in a notation that is a variation of syntax used commonly in reinforcement learning applications. Both perception and action are motivated by an embodied drive to reduce complexity. The mind/body is separated from its environment by a boundary of perceptual and motor organs. The mind his hence autonomous from the environment and shielded, statistically, by the “Markov blanket” of the body’s membrane.]

This body of work often refers to the Markov blanket of a system. In probability theory, the Markov blanket of one random variable x defined in a field of others Z is the set of other variables B which, if fixed, render x independent of other elements of Z. When, as is common, dependency relations between variables are represented as edges of a graph, this Markov blanket can easily be read from the graph's structure. Intuitively, the Markov blanket can be thought of as a "boundary" between a system (inside the blanket) and its environment. Anything within the system cannot be directly affected by the environment; rather it will be autonomous from it, interacting through the boundary variables.

References

Abebe, R., Barocas, S., Kleinberg, J., Levy, K., Raghavan, M., & Robinson, D. G. (2020, January). Roles for computing in social change. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (pp. 252-260).

Adam Barth, John Mitchell, Anupam Datta, and Sharada Sundaram. Privacy and utility in business processes. In Proceedings of the 20th IEEE Computer Security Foundations Symposium, CSF '07, pages 279–294. IEEE Computer Society, 2007. ISBN 0-7695-2819-8. doi: 10.1109/CSF.2007.26.

Agre, P. E. (1997). Lessons learned in trying to reform AI. Social science, technical systems, and cooperative work: Beyond the Great Divide (1997), 131.

Barocas, Solon, and Helen Nissenbaum. "On notice: The trouble with Notice and Consent." Proceedings of the Engaging Data Forum: The First International Forum on the Application and Management of Personal Electronic Information. 2009.

Barth, A., Datta, A., Mitchell, J. C., & Nissenbaum, H. (2006, May). Privacy and contextual integrity: Framework and applications. In 2006 IEEE symposium on security and privacy (S&P'06) (pp. 15-pp). IEEE.

Beer, S. (1973). Designing freedom. House of Anansi.

Baxter, H. (1997). Autopoiesis and the relative autonomy of law. Cardozo L. Rev., 19, 1987.

Benthall and Goldenfein (2021). Artificial Intelligence and the Purpose of Social Systems. Artificial Intelligence Ethics and Society 2021.

Benthall, S., Gürses, S., & Nissenbaum, H. (2017). Contextual integrity through the lens of computer science. Now Publishers.

Benthall, Sebastian, and Bruce D. Haynes. "Racial categories in machine learning." Proceedings of the Conference on Fairness, Accountability, and Transparency. ACM, 2019.

Bertschinger, N., Olbrich, E., Ay, N., & Jost, J. (2008). Autonomy: An information theoretic perspective. Biosystems, 91(2), 331-345.

Block, N. (Ed.). (1980). Readings in philosophy of psychology (Vol. 1). Harvard University Press.

Blomberg, J., Suchman, L., & Trigg, R. H. (1996). Reflections on a work-oriented design project. Human–Computer Interaction, 11(3), 237-265.

Bowker, G., Star, S. L., Turner, W., & Gasser, L. (Eds.). (1997). Social Science, Technical Systems and Cooperative Work: Beyond The Great Divide.

Chomsky, N. (1957/2009). Syntactic structures. De Gruyter Mouton.

Clark, A., & Chalmers, D. (1998). The extended mind. analysis, 58(1), 7-19.

Cohen, J. E. (2019). Between truth and power: The legal constructions of informational capitalism. Oxford University Press.

Cranor, Lorrie Faith. "Necessary but not sufficient: Standardized mechanisms for privacy notice and choice." J. on Telecomm. & High Tech. L. 10 (2012): 273.

Criado, N., & Such, J. M. (2015). Implicit contextual integrity in online social networks. Information Sciences, 325, 48-69.

Dalenius, T. (1977). Towards a methodology for statistical disclosure control. statistik Tidskrift, 15(429-444), 2-1.

DeYoung, H., Deepak Garg, Limin Jia, Dilsun Kirli Kaynar, and Anupam Datta. Experiences in the logical specification of the HIPAA and GLBA privacy laws. In WPES, pages 73–82, 2010.

Dourish, P. (2004a). What we talk about when we talk about context. Personal and ubiquitous computing, 8(1), 19-30.

Dourish, P. (2004b). Where the action is: the foundations of embodied interaction. MIT press.

Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2016). Calibrating noise to sensitivity in private data analysis. Journal of Privacy and Confidentiality, 7(3), 17-51.

Dwork, C. (2008, April). Differential privacy: A survey of results. In International conference on theory and applications of models of computation (pp. 1-19). Springer, Berlin, Heidelberg.

Dwork, C., & Naor, M. (2010). On the difficulties of disclosure prevention in statistical databases or the case for differential privacy. Journal of Privacy and Confidentiality, 2(1).

Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. Foundations and Trends in Theoretical Computer Science, 9(3-4), 211-407.

Epstein, J. M. (2006). Generative social science: Studies in agent-based computational modeling (Vol. 13). Princeton University Press.

Christiane Floyd, Heinz Züllighoven, and Reinhard Budde, Reinhard Keil-Slawik. (1992) "Software development and reality construction." Springer-Verlag Berlin Heidelberg. https://doi.org/10.1007/978-3-642-76817-0

Garg, D., Jia, L., & Datta, A. (2011, October). Policy auditing over incomplete logs: theory, implementation and applications. In Proceedings of the 18th ACM conference on Computer and communications security (pp. 151-162).

Goguen, J. A., & Varela, F. J. (1979). Systems and distinctions; duality and complementarity. International Journal of General System, 5(1), 31-43.

Goguen, J. A., & Meseguer, J. (1982, April). Security policies and security models. In 1982 IEEE Symposium on Security and Privacy (pp. 11-11). IEEE.

Goguen, J. (1997). Towards a social, ethical theory of information. Social science, technical systems and cooperative work: beyond the great divide, 27-56.

Goguen, J. (1999). Tossing algebraic flowers down the great divide. People and ideas in theoretical computer science, 93-129.

Hanna, A., Denton, E., Smart, A., & Smith-Loud, J. (2020, January). Towards a critical race methodology in algorithmic fairness. In Proceedings of the 2020 conference on fairness, accountability, and transparency (pp. 501-512).

He, X., Machanavajjhala, A., & Ding, B. (2014, June). Blowfish privacy: Tuning privacy-utility trade-offs using policies. In Proceedings of the 2014 ACM SIGMOD international conference on Management of data (pp. 1447-1458).

Heidegger, M. (1954). The question concerning technology. Technology and values: Essential readings, 99, 113.

Hildebrandt, M. (2013). Profile transparency by design?: Re-enabling double contingency. Privacy, due process and the computational turn: The philosophy of law meets the philosophy of technology, 221-46.

Hildebrandt, M. (2015). Smart technologies and the end(s) of law: novel entanglements of law and technology. Edward Elgar Publishing.

Hildebrandt, M. (2019). Privacy as protection of the incomputable self: From agnostic to agonistic machine learning. Theoretical Inquiries in Law, 20(1), 83-121.

Hildebrandt, M. (2020). The adaptive nature of text-driven law. Journal of Cross-disciplinary Research in Computational Law, 1(1).

Ho, C. S. (2013). In Defense of Circular Reasoning: The Affordable Care Act and the Resilience of Law and Self-Reference. Wm. & Mary Pol'y Rev., 5, 1.

Kasiviswanathan, S. P., & Smith, A. (2008). A note on differential privacy: Defining resistance to arbitrary side information. CoRR abs/0803.3946.

Kasiviswanathan, S. P., & Smith, A. (2014). On the 'semantics' of differential privacy: A bayesian formulation. Journal of Privacy and Confidentiality, 6(1).

Kathi Fisler, Shriram Krishnamurthi, and Daniel J Dougherty. Embracing policy engineering. In Proceedings of the FSE/SDP workshop on Future of software engineering research, pages 109–110. ACM, 2010.

Kifer, D., & Machanavajjhala, A. (2011, June). No free lunch in data privacy. In Proceedings of the 2011 ACM SIGMOD International Conference on Management of data (pp. 193-204).

Kifer, D., & Machanavajjhala, A. (2012, May). A rigorous and customizable framework for privacy. In Proceedings of the 31st ACM SIGMOD-SIGACT-SIGAI symposium on Principles of Database Systems (pp. 77-88).

Kim, J. (2018). Philosophy of mind. Routledge.

Kirchhoff, M., Parr, T., Palacios, E., Friston, K., & Kiverstein, J. (2018). The Markov blankets of life: autonomy, active inference and the free energy principle. Journal of The royal society interface, 15(138), 20170792.

Krakauer, D., Bertschinger, N., Olbrich, E., Flack, J. C., & Ay, N. (2020). The information theory of individuality. Theory in Biosciences, 139(2), 209-223.

Langton, C. G. (Ed.). (1997). Artificial life: An overview. Mit Press.

Lessig, L. (2000). Code is law. Harvard magazine, 1 (2000).

Linson, A., Clark, A., Ramamoorthy, S., & Friston, K. (2018). The active inference approach to ecological perception: general information dynamics for natural and artificial embodied cognition. Frontiers in Robotics and AI, 5, 21.

Liu, C., Chakraborty, S., & Mittal, P. (2016, February). Dependence Makes You Vulnerable: Differential Privacy Under Dependent Tuples. In NDSS (Vol. 16, pp. 21-24).

Luhmann, N. (1995). Social systems. stanford university Press.

Lycan, William, "Representational Theories of Consciousness", The Stanford Encyclopedia of Philosophy (Fall 2019 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2019/entries/consciousness-representational/>.

Martin Jr, D., Prabhakaran, V., Kuhlberg, J., Smart, A., & Isaac, W. S. (2020). Extending the machine learning abstraction boundary: A Complex systems approach to incorporate societal context. arXiv preprint arXiv:2006.09663.

Maturana, H. R., & Varela, F. J. (1972/2012). Autopoiesis and cognition: The realization of the living (Vol. 42). Springer Science & Business Media.

McBride, N. K. (2014). ACTIVE ethics: an information systems ethics for the internet age. Journal of Information, Communication and Ethics in Society.

McDonald, Aleecia M., and Lorrie Faith Cranor. "The cost of reading privacy policies." Isjlp 4 (2008): 543.

McMullin, B., & Varela, F. J. (1997, July). Rediscovering computational autopoiesis. Santa Fe, NM: Santa Fe Institute.

Mead, M (1968) Cybernetics of Cybernetics, in Foerster, H von, White, J, Peterson, L and Russell, J (eds) (1968) *Purposive Systems*, New York, Spartan Books

Medina, E. (2006). Designing freedom, regulating a nation: socialist cybernetics in Allende's Chile. Journal of Latin American Studies, 571-606.

Mirowski, P. (2012). Minding the cybernetic gap. Technology and Culture, 53(1), 192-195.

Mullin, B.M., Varela, F.J. (1997). Rediscovering computational autopoiesis. In: Husbands, P., Harvey, I. (Eds.), Proceedings of the 4th European Conference on Artificial Life. MIT Press, Cambridge, MA, pp. 38–47.

Nissenbaum, H. (2009). Privacy in context: Technology, policy, and the integrity of social life. Stanford University Press.

Obar, J. A., & Oeldorf-Hirsch, A. (2020). The biggest lie on the internet: Ignoring the privacy policies and terms of service policies of social networking services. Information, Communication & Society, 23(1), 128-147.

Oka, M., & Ikegami, T. (2012, July). Characterizing autonomy in the web via transfer entropy network. In Artificial Life Conference Proceedings 12 (pp. 234-242). One Rogers Street, Cambridge, MA 02142-1209 USA journals-info@ mit. edu: MIT Press.

Pickering, A. (2010). The cybernetic brain: Sketches of another future. University of Chicago Press.

Priya, S., & Nehaniv, C. L. (2020, July). Life-Span Expectancy and Cycle Size in the Original Autopoiesis Algorithm. In Artificial Life Conference Proceedings (pp. 69-77). One Rogers Street, Cambridge, MA 02142-1209 USA journals-info@ mit. edu: MIT Press.

Putnam, H. (1960). Minds and machines.

Putnam, H. (1967). Psychological predicates. Art, mind, and religion, 1, 37-48.

Ramstead, M. J., Kirchhoff, M. D., & Friston, K. J. (2020). A tale of two densities: Active inference is enactive inference. Adaptive Behavior, 28(4), 225-239.

Reddy, M. (1979). The conduit metaphor. Metaphor and thought, 2, 285-324.

Reidenberg, Joel R., et al. "Privacy harms and the effectiveness of the notice and choice framework." ISJLP 11 (2015): 485

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019, January). Fairness and abstraction in sociotechnical systems. In Proceedings of the conference on fairness, accountability, and transparency (pp. 59-68).

Shayak Sen, Saikat Guha, Anupam Datta, Sriram K. Rajamani, Janice Tsai, and Jeannette M. Wing. Bootstrapping privacy compliance in big data systems. In Proceedings of the 2014 IEEE Symposium on Security and Privacy, SP ’14, pages 327–342. IEEE Computer Society, 2014. ISBN 978-1-4799-4686-0. doi: 10.1109/SP. 2014.28.

Sheffi, N. (2020). We accept: the constitution of Airbnb. Transnational Legal Theory, 1-37.

Solove, Daniel, J. "Privacy self-management and the consent dilemma." HARV. L. REV. 126 (2013): 1880-1883.

Suchman, L. (1993). Do categories have politics? The language/action perspective reconsidered. In Proceedings of the Third European Conference on Computer-Supported

Cooperative Work 13–17 September 1993, Milan, Italy ECSCW'93 (pp. 1-14). Springer, Dordrecht.

P. Swire and A. Anton. Engineers and lawyers in privacy protection: Can we all just get along? IAPP Privacy Perspectives, Jan 2014.

Teubner, G. (1993). Law as an autopoietic system. Oxford/Cambridge, Blackwell Publishers.

Teubner, G. (2006). Rights of non-humans? Electronic agents and animals as new actors in politics and law. Journal of Law and Society, 33(4), 497-521.

Teubner, G. (2020). The Constitution of Non-Monetary Surplus Values. Social & Legal Studies, 0964663920952531.

Michael Carl Tschantz, Anupam Datta, and Jeannette M. Wing. Purpose restrictions on information use. In Proceedings of the 18th European Symposium on Research in Computer Security (ESORICS), volume 8134 of Lecture Notes in Computer Science, pages 610–627. Springer Berlin Heidelberg, 2013.

Tschantz, M. C., Sen, S., & Datta, A. SoK: Differential Privacy as a Causal Property. In 2020 IEEE Symposium on Security and Privacy (SP) (pp. 179-196).

Vakhrameev, D., Aguilera, M., Barandiaran, X. E., & Bedia, M. (2020, July). Measuring Autonomy for Life-Like AI. In Artificial Life Conference Proceedings (pp. 589-591). One Rogers Street, Cambridge, MA 02142-1209 USA journals-info@ mit. edu: MIT Press.

Von Foerster, H. (2003). Cybernetics of cybernetics. In Understanding understanding (pp. 283-286). Springer, New York, NY.

Varela, F. J. (1997). Patterns of life: Intertwining identity and cognition. Brain and cognition, 34(1), 72-87.

Varela, F. J., & Bourgine, P. (Eds.). (1992). Toward a practice of autonomous systems: Proceedings of the First European Conference on Artificial Life. MIT press.

Varela, F. G., Maturana, H. R., & Uribe, R. (1974). Autopoiesis: The organization of living systems, its characterization and a model. Biosystems, 5(4), 187-196.

Varela, F. J., Thompson, E., & Rosch, E. (1991/2017). The Embodied Mind, revised edition: Cognitive Science and Human Experience. MIT press.

Waldman, A. E. (2019). Privacy Law's False Promise. Wash. UL Rev., 97, 773.

Wiener, N. (2019). Cybernetics or Control and Communication in the Animal and the Machine. MIT press.

Wilson, Robert A. and Lucia Foglia, "Embodied Cognition", The Stanford Encyclopedia of Philosophy (Spring 2017 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2017/entries/embodied-cognition/>.

Winograd, Terry, and Flores, F. Understanding computers and cognition: A new foundation for design. Intellect Books, 1986.

Zhu, T., Xiong, P., Li, G., & Zhou, W. (2014). Correlated differential privacy: Hiding information in non-IID data set. IEEE Transactions on Information Forensics and Security, 10(2), 229-242.