Workshop

Agre After Techno-Utopianism

1-2. September 2022 (University of Siegen, Germany)

SFB1187 Media of Cooperation

2:00 pm	**Arrival**
2:15-2:45 pm	**Welcoming Remarks** by **Tatjana Seitz** and **Sam Hind** (both University of Siegen **Introduction** by **Geert Lovink** (Institute of Network Cultures/University of Amsterdam)
2:45-4:15 pm	**Workshop I: Agre as Internet Critic** with contributions by **Daniela van Geenen** (University of Siegen), **Jethro Masis** (Universidad de Costa Rica), **Neal Thomas** (Laurier University), and **Adam Hyland, Julie Vera and Brett Halperin** (University of Washington)
4:15-4:30 pm	Break
4:30-6:00 pm	**Workshop II: Capture and Capturism** with contributions **by Tobias Stadler** (University of Oldenburg), **Andreas Beinsteiner** (University of Vienna), and **Marina Gržinić** (Academy of Fine Arts in Vienna)
6:00-7:15 pm	Dinner
7:15-8:45 pm	**Early Agre** in conversation with **Lucy Suchman** (Professor Emerita, Lancaster University) and **David Chapman** (Technology Entrepreneur), moderation by **Tatjana Seitz** and **Sebastian Randerath** (University of Bonn)

11:30am-1:00 pm	**Workshop III: Reading Agre. Joint Discussion** P. Agre (1995) From High Tech to Human Tech P. Agre (1995) The Soul Gained and Lost P. Agre (1999) The Architecture of Identity: Embedding Privacy in Market Institutions P. Agre (2003) Your Face is Not a Bar Code
1:30-3:00 pm	Lunch
3:00-4:30 pm	**Mailing Lists and Netzkritik** in conversation with **Geert Lovink** (Institute of Network Cultures) and **Pit Schultz** (Internet Activist) with special contribution on the RRE mailing list by **Marc Tuters** (University of Amsterdam) and **David Gauthier** (Utrecht University), moderation by **Elena Pilipets** (University of Siegen)
4:30-4:45 pm	Break
4:45-6:15 pm	**Surveillance in Post Tech-Utopian Digital Cultures** in conversation with **David Murakami Wood** (Queen's University), moderation by **Armin Beverungen** (Leuphana University)
6:15-7:30 pm	Dinner
7:30-9:00 pm	**Rethinking Techno Futures** in conversation with **Marina Gržinić** (Academy of Fine Arts in Vienna), **Liliana Conlisk-Gallegos** (California State University), **Ricardo Dominguez** (University of California, San Diego), **Denis 'Jaromil' Roio** (Activist), moderation by **Marina Gržinić**

Contents

Workshop

It is hard to imagine digital culture without the work of Philip E. Agre. His description of the mutual dynamics of digital technology and ideology, so-called 'grammars of action' (Agre 1994), and the appeal for a critical technical practice (Agre 1997) have inspired scholars across media studies, HCI, and digital art and design for over 30 years. This workshop, 'Agre After Techno-Utopianism', seeks to evaluate his contribution to the study of technology, ideology, critique, and practice since the 'techno-utopia' of the early internet era ended, and more dystopic energies emerged.

The relevance of his work today is substantial. In *Surveillance and Capture* (Agre 1994), Agre saw the threats new workplace technologies posed that would mutate into examples of surveillance capitalism. In *Real-Time Politics* (Agre 2002), he wrote extensively on the downsides of digital cultures when the web was still considered a techno-utopia. In *Pengi* (Agre and Chapman 1987), Agre and David Chapman explored critiques of dominant AI conceptualizations. Together, these strands can be considered precursors to work, now commonplace, in software studies and integrated into computational methods for the study of digital culture. In *Toward a Critical Technical Practice* (Agre 1997), Agre famously offered a synthetic approach to studying technology, straddling the 'craft work of design' and the 'reflexive work of critique'. In *High Tech to Human Tech* (Agre 1995) the political economy of digital culture became an even greater interest, debunking the ideology of 'empowerment' in newly 'computerized' workplaces. Even lesser-known work on the *Networked University* (Agre 2000) offered a prescient insight into the 'promise and danger' of remote learning.

Agre's contribution to, as well as critique of, digital culture was just as significant. He ran the monthly mailing list *The Network Observer* (TNO) (1994-96) before starting the *Red Rock Eater News Service* (RRE) (1996-2001), offering regular insights into the 'social and political aspects of computing and networking'. Not only did Agre critique the emerging digital world, but contributed to the counterculture within it. Since this period, we have experienced the downsides of networks, social media, and platforms, with AI and sensory media accelerating capitalism further. Bringing the founders of the *nettime* (1995-present) mailing list into conversation with the history of *REE* we want to think about how list cultures equally manifest as cultures of resistance. In this, we want to re-discover ideas that resisted tech-utopian narratives, and practices that challenged these ideologies.

Collectively, we want to deepen our understanding of Agre's thinking and the significance of his work. From revisiting well-known texts to rediscovering less-popular work, and exploring the exciting interconnections between various disciplines and forms of 'net activism' that engaged with Agre's work from computational science to sociology, and from work in political economy to across the wider arts and humanities. Within the context of the contemporary platform condition we want to collectively reflect on the relevance, as well as limitations, of his work; continuing to debunk cyber utopias, whilst disassociating and rearticulating narratives of power.

Schedule

September 1st, 2022

Day 1

2pm	Arrival
2.15-2.45pm	Welcoming remarks, by Tatjana Seitz and Sam Hind Introduction, by Geert Lovink
2.45-4.15pm	Workshop I: 'Agre as Internet Critic', with contributions by Daniela van Geenen, Jethro Masís, Neal Thomas, and Adam Hyland, Julie Vera and Brett Halperin
4.15-4.30pm	*Break*
4.30-6pm	Workshop II: 'Capture and Capitalism', with contributions by Tobias Stadler, Andreas Beinsteiner, and Marina Gržinić
6-7.15pm	*Dinner*
7.15-8.45pm	'Early Agre' in conversation with Lucy Suchman and David Chapman, moderation by Tatjana Seitz and Sebastian Randerath

September 2nd, 2022

Day 2

11am-1.30pm	Workshop III: Reading Agre Agre (1995) *From High Tech to Human Tech* Agre (1995) *The Soul Gained and Lost* Agre (1999) *The Architecture of Identity: Embedding Privacy in Market Institutions* Agre (2003) *Your Face is not a Bar Code*
1.30-3pm	*Lunch*
3-4.30pm	'Mailing Lists and Netzkritik' in conversation with Geert Lovink and Pit Schultz with special contribution on the RRE mailing list, by Marc Tuters and David Gauthier, moderation by Elena Pilipets
4.30-4.45pm	*Break*
4.45-6.15pm	'Surveillance Cultures and Post Tech-Utopian Digital Cultures', in conversation with David Murakami Wood, moderated by Armin Beverungen
6.15-7.30pm	*Dinner*
7.30-9pm	'Rethinking Techno Futures', with Marina Gržinić, Ricardo Dominguez, Liliana Conlinsk-Gallegos, Ricardo Dominguez, Denis 'Jaromil' Roio; moderated by Marina Gržinić

End

Organizational information

Location

Room AH-A 217/18, Floor 2, Herrengarten (AH) 3, 57072

University of Siegen, Siegen, Germany

Zoom link: https://uni-siegen.zoom.us/j/63298485068?pwd=M2cxS0FJZVJZMEFaNlhRa21VQ2RSZz09

Meeting ID: 632 9848 5068 / Password: 4nyh+&G

Directions

From Siegen Hbf, 3-4 minute walk: https://goo.gl/maps/aJ1i7nkkCJGRcgSw9

Hotel

AMEDIA Hotel Siegen, Koblenzer Straße 135, 57072 [formerly Cityhotel Siegen]

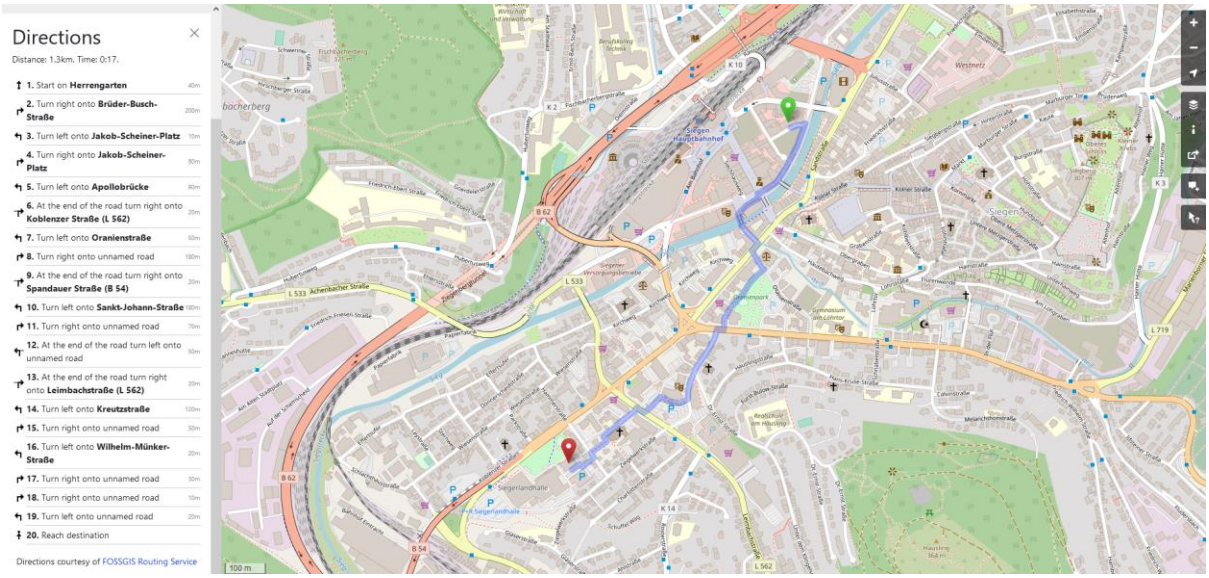
+49 271 400 3840 / siegen@amediahotels.com

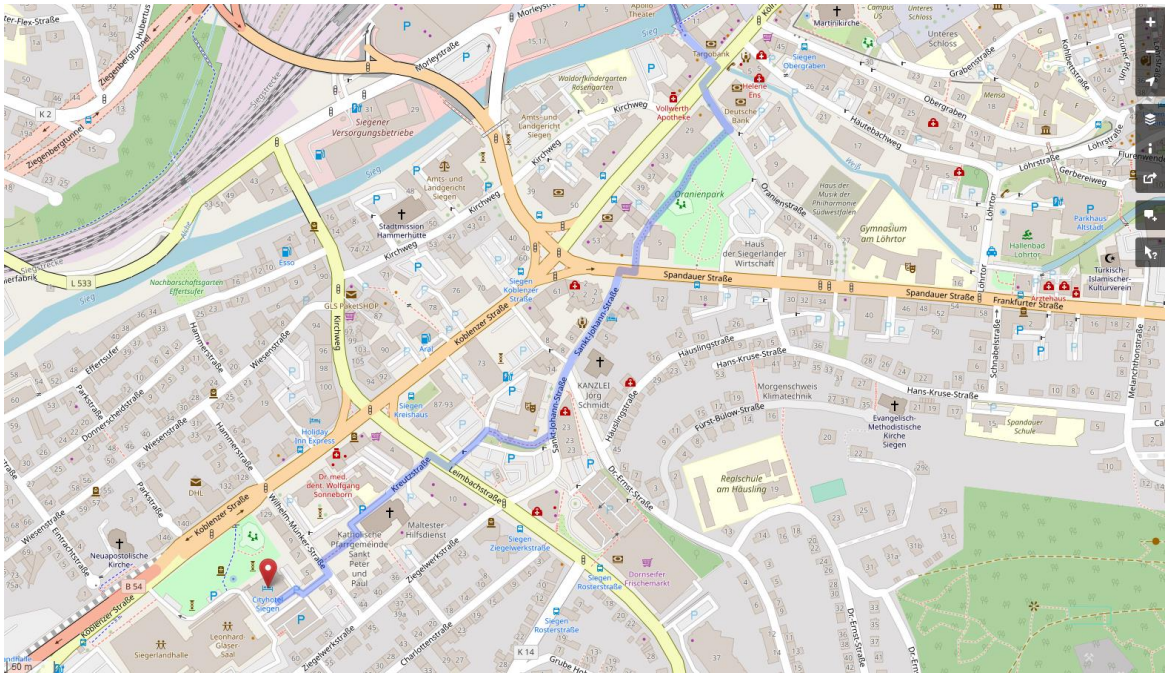
https://amediahotels.com/de/deutschland/siegen/

Directions

From Siegen Hbf, 14-15 minute walk: https://goo.gl/maps/5gpybBVRQoCEJuB87

From Herrengarten 3/workshop venue, 15-17 minute walk:
https://goo.gl/maps/PpJxY4UaJqcn6En99





Paper contributions

Critical Technical Practice: Philip Agre's Philosophy of Technology Reconsidered

Jethro Masís
University of Costa Rica
jethro.masis@ucr.ac.cr

Making AI Philosophical Again

Philip Agre received his doctorate in Electrical Engineering and Computer Science at MIT, so his interest in exploring the philosophical assumptions pervading technological practices was all the more surprising and unexpected for the people around him.[1] Agre deemed 'mistaken' to consider the Cartesian lineage of AI ideas as merely incidental and as having no purchase on the technical ideas that descend from it. On the contrary, argues Agre, "computer systems are thus, among other things, also philosophical systems—specifically, mathematized philosophical systems—and much can be learned by treating them in the same way" (1997, p. 41). Here a rather radical claim: "AI is philosophy underneath" (2005, p. 155). And this is why:

· AI ideas have their genealogical roots in philosophical ideas.

· AI research programs attempt to work out and develop the philosophical systems they inherit.

· AI research regularly encounters difficulties and impasses that derive from internal tensions in the underlying philosophical systems.

· These difficulties and impasses should be embraced as particularly informative clues about the nature and consequences of the philosophical tensions that generate them.

· Analysis of these clues must proceed outside the bounds of strictly technical research, but they can result in both new technical agendas and in revised understandings of technical research itself. (*idem*)

Influenced considerably by Dreyfus's pragmatization of Heidegger, Agre too understands *Sein und Zeit* as providing a phenomenology of ordinary routine activities, and believes Heidegger's *Analytic of Dasein* can provide useful guidance for the development of computational theories of interaction. Most importantly, a phenomenology of everyday life practice can also contribute to afford technology a historical conscience it overtly lacks, since "research like Heidegger's can have its most productive influence upon AI when AI itself recovers a sense of its own historical development" (1996, p. 25). This last critical and historical trait allows Agre, as a philosopher of computing, to hold that modern computational practices can be viewed as the resolute incarnation of a disembodied conception of

---

[1] In the informal essay, 'Critical Thinking for Technical People,' (originally an email message for the subscribers of the Red Rock Eater News Service), Agre tells the story "about how I became (relatively speaking, and in a small way) a better person through philosophy."

philosophy having Augustine, Descartes, and Turing as pivotal figures, with the opposition of body and soul at the core of their thinking:

> Each man's cultural milieu provided fresh meaning for this opposition: Augustine struggled to maintain his ideals of Christian asceticism, Descartes described the soldier's soul overcoming his body's fear as the Thirty Years' War raged, and Turing idealized disembodied thought as he suffered homophobic oppression in modern England (1997, p. 103).

For Agre, there is a historical tradition and discourse sustaining the practices of contemporary computational approaches, so by no means they can be said to sustain themselves exclusively on technical terms. The latter view is not only naïve but also dishonest. But unfortunately, Agre sees that computer science is utterly oblivious to "its intellectual contingency and recast itself as pure technique" (*idem*). Therefore, Agre castigates this forgetfulness of the assumptions running deep in AI, which more often than not are substituted by the formalist attempt to cleanse computational programs of the 'inexactness' of natural language, and to strip AI altogether of its historical and cultural underpinnings. It is by virtue of not paying attention to how their scientific practices are constituted that formalists attempt to liberate computational work precisely from the unruliness and imprecision of vernacular language, which appears foreign and annoying to their technical field. Moreover, "they believed that, by defining their vocabulary in rigorous mathematical terms, they could leave behind the network of assumptions and associations that might have attached to their words through the sedimentation of intellectual history" (2002, p. 131). This is why Agre believes such an attempt should be confronted by means of a 'critical technical practice:' the kind of critical stance that would guide itself by a continually unfolding awareness of its own workings as a historically specific practice (1997, p. 22). As such, "it would accept that this reflexive inquiry places all of its concepts and methods at risk. And it would regard this risk positively, not as a threat to rationality but as a promise of a better way of doing things" (1997, p. 23).

It is crucial for Agre to present his work as neither an internalist account *of* AI, nor as a philosophical study *about* AI, but as "actually a work *of* AI: an intervention within the field that contests many of its basic ideas while remaining fundamentally sympathetic to computational modeling as a way of knowing" (1997, p. xiv). Agre finds intents to intervene in the field to show practically how critical technical views might help develop better artificial systems. So both the critical intervention on the field and the fundamental sympathetic posture deserve, furthermore, a separate explanation.

With regard to the critical intervention, Agre notes that there is a certain mindset when it comes to what 'computer people' believe—and this is, of course, Agre's niche—regarding the aims and scope of their own work. This belief must be viewed in conjunction with the very nature of computation and computational research in general. According to Agre, computational research can be defined as an inquiry into physical realization as such. Moreover, "what truly founds computational work is the practitioner's evolving sense of what can be built and what cannot" (1997, p. 11). The motto of computational practitioners is simple: *if you cannot build it, you do not understand it. It must be built and we must accordingly understand the constituting mechanisms underlying its workings.* This is why, on Agre's account, computer scientists "mistrust anything unless they can nail down all four corners of it; they would, by and large, rather get it precise and wrong than vague and right" (1997, p. 13). However, Agre deems it too narrow to entertain just this sense of 'work.' Such conception of what counts as success is also ahistorical in that it can simply be defined as working because the program conforms to a pre-given formal-mathematical specification. But an AI system

can also be said to work in a wholly different sense: when its operational workings can be narrated in intentional terms by means of words whose meaning goes beyond the mathematical structures (which is, of course, a pervasive practice in cognitive scientific explanations of mechanism). For example, when a robot is said to 'understand' a series of tasks, or when it is proclaimed that AI systems will give us deeper insights about human thinking processes. This is indeed a much broader sense of 'work,' one which is not just mathematical in nature, but rather a clearly discursive construction. And it certainly bears reminding that such discursive construction is part of the most basic explanatory desires of cognitive science. So in the true sense of the words 'build' and 'work,' AI is not only there to build things that *merely* work. Let us quote Agre at length:

> The point, in any case, is that the practical reality with which AI people struggle in their work is not just 'the world,' considered as something objective and external to the research. It is much more complicated than this, a hybrid of physical reality and discursive construction. The trajectory of AI research can be shaped by the limitations of the physical world—the speed of light, the three dimensions of space, cosmic rays that disrupt memory chips—and it can also be shaped by the limitations of the discursive world—the available stock of vocabulary, metaphors, and narrative conventions. (1997, p. 15)

This also gives hints as to how exogenous discourses, like philosophy, are supposed to be incorporated into technological practices. Agre is of the opinion that the point is not to invoke Heideggerian philosophy, for example, as an exogenous authority thus supplanting technical methods: "the point, instead, is to expand technical practice in such a way that the relevance of philosophical critique becomes evident *as a technical matter*. The technical and critical modes of research should come together in this newly expanded form of critical technical consciousness" (1997, p. xiii). The critical technical practice Agre envisions is one "within which such reflection on language and history, ideas and institutions, is part and parcel of technical work itself" (2002, p. 131). More exactly, Agre's goal is "to do science, or at least something about human nature, and not to solve industrial problems" (1997, p. 17). And he adds: "but I would also like to benefit from the powerful modes of reasoning that go into an engineering design rationale" (*idem*). In such a way, Agre pretends to salvage the most encompassing claims of AI research—that it can teach us something about the world and about ourselves—by means of incorporating a self-correcting, history-laden approach combining both technical precision and philosophical rigor. By expanding the comprehension of the ways in which a system can work, "AI can perhaps become a means of listening to reality and learning from it" (Agre 2002, p. 141). But it is because of its not listening to reality that, for instance, Dreyfus (1992) launched his attacks against AI as an intellectual enterprise.

Agre contends that merely "lashing a bit of metaphor to a bit of mathematics and embodying them both in computational machinery" (1997, p. 30)—which is usually what computer scientists come up with—will not do the job of contributing to the understanding of humans and their world. So framed, the approach appears to Agre as too narrow, naïve, and a dire way of not listening to reality. So he has a more ambitious project: the very metaphors being lashed to a bit of mathematics that end up in machinery implementation must be investigated. Both physical reality and discursive construction must be taken into account. Although technical languages encode a cultural project of their own (the systematic redescription of human and natural phenomena within the limited repertoire of technical schemata that facilitate rational control), "it is precisely this phenomenon that makes it especially important to investigate the role of metaphors in technical practice" (1997, p. 34). At this point, Agre sounds strikingly similar to Blumenberg, whose metaphorological project "seeks to

burrow down to the substructure of thought, the underground, the nutrient solution of systematic crystallizations; but it also aims to show with what 'courage' the mind preempts itself in its images, and how its history is projected in the courage of its conjectures" (2010, p. 5). For Agre too metaphors play a role in organizing scientific inquiry or, to say it with Blumenbergian tones, metaphors are by no means 'leftover elements' (*Restbestände*) but indeed 'foundational elements' (*Grundbestände*) of scientific discourse.[2]

The sympathetic attitude towards computational modeling that Agre espouses takes as its point of departure the analysis of agent/environment interactions which accordingly should be extended to include the conventions and invariants maintained by agents throughout their activity. This notion of environment is referred to, with clear Husserlian overtones (see Husserl 1970), as *lifeworld,* and can be incorporated into computational modeling via "a set of formal tools for describing structures of lifeworlds and the ways in which they computationally simplify activity" (Agre & Horswill 1997, p. 111). From this follows that Agre's theoretical emphasis lies on the concept of *embedding*. This means that agents are not only to be conceived as embodied but more crucially as embedded in an environment. The distinction between embodiment and embedding can be explained as follows:

> 'Embodiment' pertains to an agent's life as a body: the finiteness of its resources, its limited perspective on the world, the indexicality of its perceptions, its physical locality, its motility, and so on. 'Embedding' pertains to the agent's structural relationship to its world: its habitual paths, its customary practices and how they fit in with the shapes and workings of things, its connections to other agents, its position in a set of roles or a hierarchy, and so forth. The concept of embedding, then, extends from more concrete kinds of locatedness in the world (places, things, actions) to more abstract kinds of location (within social systems, ecosystems, cultures, and so on). Embodiment and embedding are obviously interrelated, and they each have powerful consequences both for agents' direct dealings with other agents and for their solidarity activities in the physical world. (Agre & Horswill 1997, 111-112)

The importance for cognitive science of having a well-developed concept of the environment is not to be underestimated, since it seems that only prior a basic understanding of an agent's environment can a given pattern of adaptive behavior be figured out. Taking a stride towards defining the environment with at least a modicum of rigor amounts to developing "a *positive theory* of the environment, that is, some kind of principled characterization of those structures or dynamics or other attributes of the environment *in virtue of which* adaptive behavior is adaptive" (Agre & Horswill 1997, p. 113). Accordingly, Agre and Horswill lament that AI has downplayed the distinction between agent and environment by fatally reducing the latter to a discrete series of choices in the course of solving a problem, but "this is clearly a good way of modeling tasks such as logical theorem-proving and chess, in which the objects being manipulated are purely formal" (*idem*). AI can go on well without a well-developed concept of the environment but only at the price of focusing on mere toy-problems, microworlds, and toy-tasks within such artificial environments. It should then not come as a surprise that the situation changes dramatically for tasks involving physical activities, where "the world shows up, so to speak, phenomenologically: in terms of the differences that make a difference for *this* agent, given its particular representations, actions, and goals." (*idem*). Such environmental indexicality is

[2] It must be noted that, for the sake of the argument, reference is here made to early Blumenberg and his project in *Paradigmen zu einer Metaphorologie* (1960) of tracing absolute metaphors as those *Grundbestände* that cannot be conceptually reduced (*nicht in Begrifflichkeit aufgelöst werden können*) but rather function as constituting a catalytic sphere from which the universe of concepts continually renews itself.

often objected to by cognitivists as though agents performed tasks without any computation whatsoever, or as though agents inhabiting a lifeworld lived in an adamant reactive mode. But the point is rather that "the nontrivial cognition that people do perform takes place against a very considerable background of familiar and generally reliable dynamic structure" (Agre & Horswill 1997, p. 118). Now, precisely indexicality has been difficult to accommodate within AI research. With this in view, Agre has criticized the usual assumptions of the received view in technical practice as follows:

> · That perception is a kind of reverse optics building a mental model of the world by working backward from sense-impressions, inferring what in the world might have produced them.
>
> · That action is conducted through the execution of mental constructs called plans, understood as computer programs.
>
> · And finally, that knowledge consists in a model of the world, formalized in terms of the Platonic theory analysis of meaning in the tradition of Frege and Tarski. (2002, p. 132)

The dissociation of mind and body (the founding metaphor of cognitive science and modern philosophy) is here at work, precisely when traditional AI thinks of the mind roughly as a plan generator and the body as the executor of the plan. Moreover, AI is so framed in terms of a series of dissociations: mind versus world, mental activity versus perception, plans versus behavior, the mind versus the body, and abstract ideas versus concrete things (Agre 2002, p. 132). According to Agre, these dissociations are contingent and can be considered 'inscription errors' (Smith 1996): "inscribing one's discourse into an artifact and then turning around and 'discovering' it there" (Agre 2002, p. 130). That AI research has been framed along these contingent oppositions makes it clear that it is part of the history of Western thought. As such,

> it has inherited certain discourses from that history about matters such as mind and world, and it has inscribed those discourses in computing machinery. The whole point of this kind of technical model-building is conceptual clarification and empirical evaluation, and yet AI has failed either to clarify or to evaluate the concepts it has inherited. Quite the contrary, by attempting to transcend the historicity of its inherited language, it has blunted its own awareness of the internal tensions that this language contains. The tensions have gone underground, emerging through substantive assumptions, linguistic ambiguities, theoretical equivocations, technical impasses, and ontological confusions. (Agre 2002, p. 141)

Nevertheless, it is interesting to note that—for all his philosophical acumen—Agre himself has not been able to liberate himself from the persistence of a representational theory of cognition, even when his is certainly more concrete, more historically conscious, and more enactive than the one customarily held in the traditional view. As a result, the latter critical concepts grouped together conform the motivation for developing a concept of indexical-functional or deictic representation (Agre & Chapman 1987; Agre 1997), the main idea being that agents represent objects in generic ways through relationships to them (Agre & Horswill 1997, p. 118). On Agre's view, the concept of representation must be redefined to show what kind of representational activity is at work in interaction. Thus, the point is to criticize the underlying view of knowledge presupposed by the traditional theory of representation (that knowledge is picture, copy, reflection, linguistic translation, or physical simulacrum of the world), while suggesting that "the primordial forms of representation

are best understood as facets of particular time-extended patterns of interaction with the physical and social world" (Agre 1997, p. 222). Therefore, "the notion of representation must undergo painful surgery to be of continued use" (Agre 1997, p. 250). Given that this redefinition of representation by Agre has its own quirks, it must now be carefully explained.

The traditional theory of representation which has been put into work in AI is based on the notion of world model. This notion refers to some structure which is thought to be within the mind or machine that represents the outside world by standing in a systematic correspondence with it (Agre 1997, p. 223). As such, the assumption that there is a world model being represented by the mind is the epitome of mentalism (Agre 1997, p. 225). Mentalism was previously defined by Agre as the generative metaphor pervasive in cognitive science according to which every human being has an abstract inner space called a 'mind' which clusters around a dichotomy between outside and inside organizing a special understanding of human existence (Agre 1997, p. 49). On Agre's terms, giving preeminence to indexicality amounts to inverting this picture, since conceding that human beings are not minds that control bodies implies that interaction cannot be defined "in terms of the relationships among a mind, a body, and an outside world" (1997, p. 234), which is unfortunately so typical in cognitive scientific explanations. And here the key term is indeed interaction, understood not as the relation between the subjective and the objective, but rather as emerging from the actual practices people employ to achieve reference *in situ*. So indexicality "begins to emerge not merely as a passive phenomenon of context dependence but as an active phenomenon of context constitution" (Agre 1997, p. 233).

According to Agre, when indexicality is introduced as a constituting factor of interaction "human activities must be described in intentional terms, as being *about* things and *toward* things, and not as meaningless displacements of matter. Physical and intentional description are not incomparable, but they *are* incommensurable" (1997, p. 245). From this follows that the typical ascription of intentional states to nonembedded systems is absurd, precisely because embedding, and the interaction deriving thereof, is the condition of possibility of intentional comportment. When actual, concrete intentional activities are taken off the picture, representation is no longer connected with a lifeworld. Thus, the illusion can be then entertained that a semantic theory merely entails the categorization in some objective way of the ontology of a concrete situation, before the event of activity has taken place, or ignoring *tout court* the eventual character of activity. According to Agre, "when a speaker uses an indexical term such as 'I,' 'you,' 'here,' 'there,' 'now,' or 'then' to pick out a specific referent, this picking out is determined by relations between situations; it is not an *act* on the speaker" (1997, p. 233). So these interactions and how they shape situations must be clarified, since it can be said that "interaction is central, both to human life and to the life of any agent of any great complexity" (1997, p. 234). Embedded activities must be investigated in how they are structured, as well as the sort of representing which is most incumbent on them.

For Agre, the latter requires a proper theory of intentionality couched within the Heideggerian distinction between *Zuhandenheit* and *Vorhandenheit* (SZ § 15). Traditional AI research can be accused of having only paid attention to present-at-hand phenomena, thus attempting to model computationally what precisely appears salient objectively in perception. In contrast, Agre finds that this phenomenological distinction is not psychological nor mechanistic but a description of the structure of everyday experience which can be suitable for a new way of computational modeling of that experience. One could, *à la* Dreyfus (2002a; 2002b), identify respectively *Vorhandenheit* with representational intentionality and *Zuhandenheit* with a sort of nonrepresentational intentionality and so proclaim beforehand the failure of artificial systems propounding the accomplishment of high-level intelligence. For Agre, however, this is too radical and, above all, too pessimistic. What is needed

is a clarification of what kinds of representation exist and the role they play in real activities. Herein resides the importance of delving into experience and providing AI with a set of tools to enrich its vocabulary and metaphors. This is needed because "the philosophy that informs AI research has a distinctly impoverished phenomenological vocabulary, going no further than to distinguish between conscious and unconscious mental states" (1997, p. 239). Agre is onto something more important here, which is nothing less than making AI philosophical again: "technology at present is covert philosophy; the point is to make it openly philosophical" (1997, p. 240).

The traditional idea of representation understood it as a model in an agent's mind that corresponds to the outside world through a systematic mapping. Agre opines that AI research has been concerned only with a partly articulated view of representation. This is also the reason explaining why "indexicality has been almost entirely absent from AI research" (Agre 1997, p. 241). Moreover, "the model-theoretic understanding of representational semantics has made it unclear how we might understand the concrete relationships between a representation-owning agent and the environment in which it conducts its activities" (*idem*). On Agre's view, the reason why AI research has lagged behind a clear-cut understanding of representation and indexicality has not been its nondistinctiveness between mechanism and human phenomena. In fact, "AI research needs an account of intentionality that affords clear thinking about the ways in which artifacts can be involved in concrete activities in the world" (1997, p. 242).

Such account of intentionality was coined by Agre under the rubric of deictic representation as opposed to objective representation. First, two sorts of ontology are to be distinguished. According to an objective ontology, individuals can be defined without reference to activity or intentional states. A deictic ontology, by contrast, can be defined only in indexical and functional terms and in relation to an agent's location, social position, current goals and interests, and autochthonous perspective. Entities entering the space of whatever interaction with the agent, can only be understood correctly in terms of the roles they play in the agent's activities. In accordance with the aforementioned deictic notation, "some examples of deictic entities are *the-door-I-am-opening*, *the-stop-light-I-am-approaching*, *the-envelop-I-am-opening*, and *the-page-I-am-turning*. Each of these entities is indexical because it plays a specific role in some activity I am engaged in; they are not objective, because they refer to different doors, stop lights, envelopes, and pages on different occasions" (1997, 9. 242). Their nonobjective character, however, does not imply that, by contrast, indexical entities are to be considered as subjective and, for that matter, as phantasms or internal and intimate qualia. The idea behind this is precisely that a deictic ontology should not be confused with subjective, arbitrary musings of an encapsulated subject. In the first place, this is the ontology which can be most properly ascribed to routine activities. So it would be preposterous to suggest that they are private or ineffable. Routines activities are realized 'out there' in the world and, for that very reason, do not pertain to an internal mental game: they are, indeed, public. Accordingly, in routine activities the objective character of entities with which one copes, is not salient or important. Neither is their 'subjective feel,' nor the way they appear to me as individual. That their character is deictic means that what is most important is the role they play in the whole of activity. Therefore, hyphenated noun phrases like *the-car-I-am-passing* or *the-coffee-mug-I-am-drinking-with* are not mental symbols in the cognitivist sense. They designate "not a particular object in the world, but rather a role that an object might play in a certain time-extended pattern of interaction between an agent and its environment" (Agre 1997, p. 251).

Agre's alternative way of conceiving of activity and the express purpose of modeling it computationally is very attractive. As a matter of engineering, the leading principle is that of machinery parsimony: "choosing the simplest machinery that is consistent with known dynamics" (1997, p. 246). This view explicitly contrasts with the emphasis on expressive and explicit

representation typical of traditional AI, with all the inherent difficulties of programming beforehand —as scripts—all the situations an artificial agent might encounter when coping with the world. By clear contrast with traditional AI, "the principle of machinery parsimony suggests endowing agents with the minimum of knowledge required to account for the dynamics of its activity" (1997, p. 249). In such a way, Agre's approach also resonates with Brooksian tones (see Brooks 1999) of removing 'intelligence' and even 'reason' from the picture in order to render an account of interactive representation. Moreover, Agre sees deictic representation as changing the traditional view altogether since it presents us with the possibility, not of expressing explicitly and in every detail objective states of affairs, but of *participating* in them: "conventional AI ideas about representation presuppose that the purpose of representation is to express something, but this is not what a deictic representation does. Instead, a deictic representation underwrites a mode of relationship with things and only makes sense in connection with activity involving those things" (1997, p. 253). However, the objection may be raised that such a deictic approach violates the grand spirit of AI which seeks greater explicitness of representation and broader generality, since Agre's formula for design might simply contribute to model only special-purpose—and thusly limited—devices. But Agre responds that "the conventional conception of general-purpose functionality is misguided: the kind of generality projected by current AI practice (representation as expression, thought as search, planning as simulation) simply cannot be realized" (1997, pp. 249-250).

This is, of course, not just a series of theoretical postulates urged by Agre, since he distinguishes amongst levels of analysis (1997, pp. 27-28). The *reflexive* level, which has been already exhibited in the last pages of this exposition, provides ways for analyzing the discourses and practices of technical work. Given that technical language is unavoidably metaphorical, the reflexive level permits one to let those metaphors come to the surface and thus can they be taken into account when technical work encounters trouble in implementation. On the *substantive* level, the analysis is carried out with reference to a particular technical discipline, in this case AI. But Agre is primarily interested in proceeding, on top of the reflexive and substantive levels, on a *technical* level, in order to explore "particular technical models employing a reflexive awareness of one's substantive commitments to attend to particular reality as it becomes manifest in the evolving technical work" (1997, p. 28). On Agre's view, this partitioning of levels of analysis has not been conscientiously attended to by traditional AI practitioners. Particularly, the reflexive level that prescribes an awareness of the role of metaphors in technical work has been disdained, as though AI researchers could simply bootstrap their way to technical success without being aware of the underlying metaphors pervading their work. For Agre, this is specially problematic because "as long as an underlying metaphor system goes unrecognized, all manifestations of trouble in technical work will be interpreted as technical difficulties and not as symptoms of a deeper, substantive problem" (1997, p. 260).

As an exemplary case of technical work based on the aforementioned levels of analysis, Agre presents Pengi, a program designed by Chapman and Agre (1987) in the late 1980s under the rubric of being an implementation of a theory of activity. Pengi is a penguin portrayed in the commercial computer game Pengo, who finds itself in a maze made up of ice blocks that is surrounded by an electric fence. The maze is also inhabited by deadly bees that are to be avoided at all costs by Pengi and the task of the player is to maintain Pengi alive and defend it from such perils coming along the way. As defense, the bees can be killed by crushing them with a moving ice block or by kicking the fence while they are touching it. This momentarily stuns the bees and they can be crushed by simply walking over them. Agre agues that Pengo is an improvement on the blocks world, although it obviously fails to capture numerous elements of human activity. What is important is the combination of goal-directedness and improvisation involved in the game, from which Agre hopes to

learn some computational lessons. First of all, Agre and Chapman did not attempt to implement in advance everything they knew about the game, thus contradicting the mapping out beforehand which is typical in traditional AI systems. The point is to see Pengi as relating to the objects that appear in its world, not in terms of their resemblance to mental models which were beforehand programmed, but solely in terms of the roles they play in the ongoing activity. As such, what Agre and Chapman attempted to program was actually deictic representations: *the-ice-cube-which-the-penguin-I-am-controlling-is-kicking*, *the-bee-I-am-attacking*, *the-bee-on-the-other-side-of-this-ice-cube-next-to-me*, etc.

At any rate, Agre does not argue that this simple system can be regarded as intelligent: "Pengi does not understand what it is doing. No computer has ever understood to any significant degree what it was doing" (1997, p. 301). But the bottom line is straightforward enough to explain: the game constituting Pengi's world as agent is not made up of present-at-hand entities and processes, but more importantly of possibilities for action that require appropriate responses from the agent. This shows Agre's understanding of ready-to-hand entities as no entities at all, but as possibilities for action and subsequent responses to the demands of the situation at hand. Given that these possibilities for action are not objects at all and that usually this sort of open stance for responding skillfully to environmental challenges does not appear in propositional referring, it is understandable that they have been rather elusive for programers. After all, how can one program possibilities for action, since the focus is not on this particular object or the other but rather on the movement constituting the towards-which for-the-sake-of-which? The wellspring of this movement is all the more elided because, as Heidegger has it, precisely what is closest to us ontically is ontologically (and for that very reason) that which is farthest (SZ § 5, p. 15). This has been Agre's task, namely: to attempt to reveal the ontological dimension by means of technological implementation that does not obfuscate it but rather embracing it. By programming deictic representations instead of just objective ones, Agre argues, computational programs can learn this fundamental lesson: what was lacking in traditional AI systems was precisely a model to envision a specific relationship between machinery and dynamics based on the concept of interaction. This lesson, so the argument goes, can gradually dispel the need for mentalist approaches.

Conclusion

It should be noted that Agre was clearly influenced by Dreyfus's early critique of artificial reason (see Dreyfus 1992) but his path was individually constructed and his insights were also supported by different motivations. What is perhaps more interesting is his willingness to go on and program something, and this after researching with seriousness the history of the philosophy of mind and drawing even upon continental sources. Dreyfus credits him with expounding the philosophical debate and even understanding *Zuhandenheit* better than he himself did, since for Agre ready-to-hand is not a *what* but a *for-what* (2007, p. 252). Dreyfus has it that Agre was able to show how Heidegger wants to get at something more basic than simply a class of objects: equipment (*Zeug*). The entire point of the equipmental character of things in the world is not that they are entities with a function feature—this was Dreyfus's pre-Agrean interpretation of *Zeug* and *Zeugzusammenhang*—but rather that they open up possibilities for action, solicitations to act, and motivations for coping; an idea that Dreyfus takes admittedly from Agre's endeavors towards modeling *Zuhandenheit* on the basis of deictic intentionality. Nevertheless, Dreyfus is of the opinion that in attempting to program ready-to-hand, Agre succumbs to an abstract objectification of human practice, because affordances—inasmuch as they are not objects but the in-between interaction in which no subject nor object is involved—are not amenable to programming. That they are not is not something Agre seems to fully

understand, and this is why he thinks that somehow deictic representations must be involved in human understanding. According to Dreyfus, "Agre's Heideggerian AI did not try to program this experiential aspect of being drawn in by an affordance. Rather, with his deictic representations, Agre *objectified* both the functions and their situational relevance for the agent. In *Pengi*, when a virtual ice cube defined by its function is close to the virtual player, a rule dictates the response (e.g., kick it). No skill is involved and no learning takes place" (2007, p. 253). It must be admitted that a virtual world is not even slightly comparable with the complex dynamics of the real world. In a virtual world, the dynamics of relevance are determined beforehand, so a program like *Pengi* simply cannot account for the way human beings cope with new relevancies. Dreyfus concludes that Agre "finesses rather than solves the frame problem. Thus, sadly, his Heideggerian AI turned out to be a dead end. Happily, however, Agre never claimed he was making progress towards building a human being" (2007, p. 253).

Agre's contribution consists in his attempt to program *Zuhandenheit* instead of *Vorhandenheit*. That this can be made is, however, highly controversial. Certainly, what is deeply contentious is not that phenomenological insights can be brought to bear on cognitive science for a critical technical practice like the one Agre requires, but rather the assumption that the experiential dimension which phenomenology has revealed can be programmable. According to Heidegger, "*the essence of Dasein lies in its existence*" (SZ § 9, p. 42), which does not imply any "'properties' present-at-hand of some entity which 'looks' so and so and is itself present-at-hand" (*idem*). To exist as Dasein, then, implies that one's own existence has to be partly constructed, for existence is not already given and therefore is no program that can be run by any kind of hardware. So the Heideggerian *Sichöffnende und Offene* is perhaps not amenable to programing. To say it with Heideggerian overtones: programing can only be ontic but not ontological, since if the ontological were susceptible to programing, it would not be ontological. Agre has attempted to program routine activities and in doing so he has pragmaticized Dasein. But Heidegger himself warned specifically against this line of construing his philosophy, which reduces it to mere practical everyday activity:

> I attempted in *Being and Time* to provide a preliminary characterization of the *phenomenon of world* by interpreting the *way in which we at first and for the most part move about in our everyday world*. There I took my departure from what lies to hand in the everyday realm, from those things that we use and pursue, indeed in such a way that we do not really know of the peculiar character proper to such activity, and when we try to describe it we immediately misinterpret it by applying concepts and questions that have their source elsewhere. That which is so close and intelligible to us in our everyday dealings is actually and fundamentally remote and unintelligible to us. In and through this initial characterization of the phenomenon of world the task is to press on and point out the phenomenon of world as a problem. It never occurred to me, however, to try and claim or prove with this interpretation that the essence of man consists in the fact that he knows how to handle knives and forks or use the tram. (GA 29/30, pp. 262-263)

In conclusion, it must be noted that the reception of Heidegger's philosophy in cognitive science, in particular, and the use of phenomenological notions for enriching the cognitive landscape, in general, is not merely the putting into work of those insights but more often than not also a *translation* of those terms into cognitive ones. With this translation, something fundamentally phenomenological gets lost. Agre's work here discussed is a prominent example of how Heidegger's reception is rather 'analytic': 'analytic' not only in the sense that Heideggerian philosophy is appropriated by analytic-

trained Anglo-American philosophers or, in this case, by an MIT-trained engineer, but also in the decisive sense that the Heideggerian philosophy which is appropriated for the purposes of advancing the new paradigm of cognition, pays only attention to specific parts of Division I of *Sein und Zeit*; parts which, in the same vein, are also appropriated very selectively. The reception is 'analytic' in that it constitutes a very schematic version of Heidegger taking precisely his thought out of context.

However, Agre's work is full of important insights, the most important of which is his demand that technical practice be aware of the philosophical metaphors pervading research in technological implementation.

References

Agre, Philip (1994) "Surveillance and Capture. Two Models of Privacy." *Information Society*. Vol. 10, No. 2, pp. 101-127.

_____. (1996) "Computational Research on Interaction and Agency." In: Phil Agre & Stanley Rosenscheim (eds.) *Computational Theories of Interaction and Agency*. Cambridge, MA · London: The MIT Press, pp. 1-52.

_____. (1997) *Computation and Human Experience*. Cambridge/New York: Cambridge University Press.

_____. (1998) "Beyond the Mirror World: Privacy and the Representational Practices of Computing." In: Philip Agre & Marc Rotenberg (eds.) *Technology and Privacy: The New Landscape*. Cambridge, MA/London: The MIT Press, pp. 29-61.

_____. (2002) "The Practical Logic of Computer Work." Cambridge, MA/London: The MIT Press, pp. 129-142.

_____. (2005) "The Soul Gained and Lost: Artificial Intelligence as a Philosophical Project." In: Stefano Franchi & Güven Güzeldere (eds.) *Mechanical Bodies, Computational Minds. Artificial Intelligence from Automata to Cyborgs*. Cambridge, MA/London: The MIT Press, pp.153-173.

Agre, Philip & David Chapman (1987) "Pengi: An Implementation of a Theory of Activity." *Proceedings of the Sixth Annual Meeting of the American Association of Artificial Intelligence*. Seattle: Morgan Kaufmann, pp. 268-272.

_____. (1988) "What are plans for?" MIT AI Laboratory Memo 1050a.

Agre, Philip & Ian Horswill (1997) "Lifeworld Analysis." *Journal of Artificial Intelligence Research*. Vol. 6, No. 1, pp. 111-145.

Blumenberg, Hans (2010) *Paradigms for a Metaphorology*. Trans. by R. Savage. New York: Cornell University Press.

Brooks, Rodney (1999) *Cambrian Intelligence. The Early History of the New AI*. Cambridge, MA: The MIT Press.

Dreyfus, Hubert (1992) *What Computers Still Can't Do*. Cambridge, MA: The MIT Press.

_____. (2002a) "Intelligence Without Representation: Merleau-Ponty's Critique of Mental Representation. The Relevance of Phenomenology to Scientific Explanation." *Phenomenology and the Cognitive Sciences*. Vol. 1, No. 4, pp. 367-383.

_____. (2002b) "Refocusing the Question: Can There Be Skillful Coping Without Propositional Representations or Brain Representations?" *Phenomenology and the Cognitive Sciences*. Vol. 1, No. 4, pp. 413-425.

_____. (2007) "Why Heideggerian AI Failed and how Fixing it would Require making it more Heideggerian." *Philosophical Psychology*. Vol. 20, No. 2, pp. 247-268.

Heidegger, Martin (SZ) *Sein und Zeit*. Tübingen: Max Niemeyer Verlag, 1979. [*Being and Time*. Trans. by J. Macquarrie & E. Robinson. Oxford · Cambridge, MA: Blackwell, 2001.]

_____. (GA 29/30) *Die Grundbegriffe der Metaphysik. Welt - Endlichkeit - Einsamkeit*. Gesamtausgabe Bd. 29/30. Ed. by F.-W. von Herrmann. Frankfurt am Main: Vittorio Klostermann, 1983. [*The Fundamental Problems of Metaphysics. World - Finitude - Solitude*. Trans. by W. McNeill & N. Walker. Bloomington · Indianapolis: Indiana University Press, 1995.]

_____. (Zoll) *Zollikoner Seminare. Protokolle-Zwiegespräche-Briefe*. Ed. by M. Boss. Frankfurt am Main: Vittorio Klostermann, 1994.

Husserl, Edmund (1970) *The Crisis of European Sciences and Transcendental Phenomenology. An Introduction to Phenomenological Philosophy.* Trans. by D. Carr. Evanston: Northwestern University Press.

Smith, Brian Cantwell (1996) *On the Origin of Objects*. Cambridge, MA · London: The MIT Press.

Action-grammars as sociogenesis

Neal Thomas – CAS Instructor
Department of Communication Studies
Wilfrid Laurier University
Waterloo, Ontario, Canada
ngthomas@wlu.ca

Recent scholarship on the algorithmitization of social life has focused on the risks of machine learning systems as *platforms*, through which institutional judgment enacts control in a variety of neoliberal contexts.[1] We have Philip Agre to thank in part for initializing this conversation; but equally we do for opening our eyes to the importance of the data relation itself, as a ground upon which these systems take flight. Agre's work is rightly lauded for having described the data relation as an event of 'capture,' through which social formations are translated, configured, and designed such that groups may come to reason and act in concert through practical systems of *formalization*.[2]

Unfortunately, despite traveling under the notionally appealing terms of connection and communication, the social costs Agre warned of have only intensified in the intervening decades of his writings. The data relation has proliferated via certain especially powerful 'action-grammars' in social media, search, and related automations of

[1] See eg. Dyer-Witheford, Nick, Atle Mikkola Kjosen, and James Steinhoff. *Inhuman Power: Artificial Intelligence and the Future of Capitalism*. Pluto Press, 2019, Noble, Safiya Umoja. *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press, 2018, and Mattu, Julia Angwin, Jeff Larson, Lauren Kirchner, Surya. "Machine Bias." ProPublica. Accessed April 1, 2022. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[2] Agre, Philip E. "Surveillance and Capture: Two Models of Privacy." The Information Society 10, no. 2 (April 1, 1994): 101–27. https://doi.org/10.1080/01972243.1994.9960162.

intellectual labour, to become a core site for value-extractive semantic classification—supporting modern life as we know it, but also disrupting it in deleterious ways.[3]

And yet, more than twenty-five years on from Agre's seminal paper on surveillance and capture, the underlying philosophical commitments of formalization and formatting continue to receive less scrutiny than perhaps they should.[4] As a redressive contribution along these lines, this paper returns to his conceptualization of grammars of action, to flesh out some of its ontological premises. It does so principally by setting the capture model alongside work on grammatization by the philosopher Bernard Stiegler.

Stiegler's approach shares important affinities with Agre's. These include a post-Heideggeran focus on the "orthographic"—roughly, how media recording technologies produce some principle of exactitude, and the historical consequences of this for human societies—and relatedly, how media technologies grammatize through their inscriptive and representational strategies. Echoing Agre, Stiegler writes for example that, "Writing has an essential performativity as formalization of grammar rules. But informatics must equally be understood as a formalization technique of the already-there, and thus of the production of the rules governing memory."[5]

---

[3] See eg. Washington Post. "He Predicted the Dark Side of the Internet 30 Years Ago. Why Did No One Listen?" Accessed May 3, 2022. https://www.washingtonpost.com/technology/2021/08/12/philip-agre-ai-disappeared/, and Couldry, Nick, and Ulises A. Mejias. *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. 1 edition. Stanford University Press, 2019.

[4] Notable exceptions here include Iliadis, Andrew. *Semantic Media.* Boston: Polity Press, forthcoming; Koopman, Colin. *How We Became Our Data: A Genealogy of the Informational Person*. First edition. Chicago: University of Chicago Press, 2019; and Jancovic, Marek, Axel Volmar, and Alexandra Schneider, eds. *Format Matters: Standards, Practices, and Politics in Media Cultures*. meson press, 2019. https://doi.org/10.14619/1556.

[5] Stiegler, Bernard. *Technics and Time, 2: Disorientation*. Translated by Stephen Barker. 1 edition. Stanford, Calif: Stanford University Press, 2008. P.110.

Along with their affinities, one should also note certain ways that Agre and Stiegler's ideas speak in different registers. Writing in an idiom of 1990s-era social computing and computer-supported cooperative work research, Agre tends to take a pragmatic, practitioner's view of the structures and operations of modern institutions, as these materialize grammars of action through their stages of analysis, articulation, imposition, instrumentation, and elaboration.[6] As already noted, he also spends time thinking through their political-economic consequences, their impact on productivity, and role in market relations more broadly.

In his more overtly philosophical, world-historical approach to grammatization, Stiegler focuses instead on the collectively somatic and epistemic commitments of *all* technology, and especially on how writing technologies offer representational strategies for collective memory that 'exteriorize' thinking, in a process that he called *epiphylogenesis.* Stiegler understood the latter as a relation between an organism and the environment, mediated by some tool for preserving memory that leads to a kind of artificial genetic evolution of tools and groups, and as he wrote elsewhere, as a fundamental form of time for humankind living in a given epoch. There are nevertheless important parallels here with Agre's work on the digital abstraction, as when he (Agre) writes that,

> "The tokens of mechanized logic are not exactly meaningless, since they are useless without the meanings that they are routinely treated as carrying. But the formal structures of logic have taken on a life of their own;

[6] Agre, Philip E. "Surveillance and Capture: Two Models of Privacy." The Information Society 10, no. 2 (April 1, 1994): 101–27. https://doi.org/10.1080/01972243.1994.9960162. P.110.

> they promise, in their self-sufficient simplicity, a generative principle that can be used to reconstruct and systematize the whole of reason."[7]

The hope here is that bringing together these thinkers, Agre's insights might be reinvested and extended along more radical lines laid down by Stiegler. Why might we want to do so? Arguably we have reached a certain point of saturation with digital media technologies, such that for the sake of planetary safety, democratic decision-making institutions under threat, and what Stiegler described as a horizon of 'generalized proletarianization,' there must be some kind of renewed, explicitly political claim made to re-conjugate economic productivity, knowledge production, and market-style action-grammars together. For Stiegler, this claim was centered in an ethical, global, and ultimately *sociogenetic*, organism-environment relation. Or again put in terms of Agre and Stiegler mirroring one another's ideas, consider how the former contends that people who engage in grammatized activity "are somehow induced to organize their actions so that they are readily 'parsable' in terms of the grammar,"[8] while the latter argues, on more ontological terms, that for all writing technologies, "The dynamic of the *who* itself redoubles that of the *what:* conditioned *by* the *what*, it is equally conditioned *for* it: within the transductive negotiation of terms, the issue is always one of co-individuation."[9]

---

[7] Agre, Philip. *Computation and Human Experience*. Cambridge, U.K.;New York: Cambridge University Press, 1997. P.96.
[8] Agre, p.110.
[9] Stiegler, *T&T2*, p.7.

Stiegler borrows this terminology from another philosopher of technology, Gilbert Simondon. Transduction for Simondon involves defining identity as *productive of form*, in an organism's attempts to resolve problems with its internal resonances. Individuation is a related term, describing a modulation between inside and outside of living beings as they negotiate their milieu, which according to Simondon is always simultaneously biological, social, and technical. In what follows I contend that the abiding power of Agre's work lies in its capacity to explain how information systems co-individuate us on sociogenetic terms. But we must take the important lessons learned from Agre's appeals to philosophy in the design of systems, especially Heideggerean phenomenology and certain strains of speech act theory that framed the debates of his era, and update them with a Stieglerian, post-phenomenological approach to technology and meaning.

Since first appealed to as a way of countering the cognitivist/rationalist paradigm that dominated late-1980s information systems design, especially in critical work done by Terry Winograd and Fernando Flores[10], one could argue that certain critical limits have been reached in design appeals to phenomenology, when it comes to addressing the pressing issues and future needs of societies undergoing such widespread datafication.[11] Ubiquitous platforms like Facebook and Google seem to enact their soft control on the very basis of a reflexively-constructed, embodied-experiential account of self and other

[10] See Winograd, Terry, and Fernando Flores. *Understanding Computers and Cognition : A New Foundation for Design*. Reading, Mass.: Addison-Wesley, 1987.
[11] For one type of critique of current approaches in HCI, see Keyes, Os, Josephine Hoy, and Margaret Drouhard. "Human-Computer Insurrection: Notes on an Anarchist HCI." In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 1–13. New York, NY, USA: Association for Computing Machinery, 2019. https://doi.org/10.1145/3290605.3300569.

sharing a lifeworld, and yet they evince many of the hallmarks of those rationalizing systems that a phenomenological approach to design sought to critique in the past. Said another way, if social life itself is now shot through with action-grammars that have become so supple, interconnected, and communicative—literally designed according to deep insights gleaned from phenomenology—then perhaps we should table individualistic concerns over privacy, important as they remain, to have a more frank philosophical conversation about alternative forms of grammatization that might be articulated in future information systems. This is perhaps the core problematic of the paper, so it makes sense to next characterize our current situation in a bit more detail.

However often and vehemently we may disagree and/or deceive one another online, a key feature of contemporary social technologies is that their grammatizing strategies are grounded in semantic consensus. Discrete factual entities with attributes, or at least their probabilistic inference in searching for 'exactly what we're looking for', are what enable the shareable flows of action upon which we now rely. As an optimal technical result, like-minded people come together on the terms of semantic consensus, through things like hashtags, library and information science principles of classification, and other collaborative platforms, to allow for serialized and threaded conversations, and document sharing, to take place with ease. Socially, this semantic focus helps reproduce the myths and realities of identity as naturally sited in the intentional sovereign individual.

The story goes that political activity and organization online, read socially in a liberal-pluralist way, is about achieving autonomy and enlightenment for individuals, who

exist in and in agonal relation to different groupings and collectivities. At the level of data technique however, the story is different. Today's data structures essentially function according to the highly formal commitments of a computer science subfield known as mereotopology, which as John Stell writes, combines "mereology (the study of parts in relation to wholes) with topological aspects of the description of space."[12] Mereotopological networks, more colloquially called knowledge and social graphs, have fast become the coordinating medium for everyday social life, married to a value-form that relies on these topologies to train and deploy machine learning on the dominant commercial platforms. The effect is for users to be constantly inhabiting a form of relation that organizes them as members of quasi-scientific communities that are constantly seeking to accumulate, connect, and use empirical facts about reality. Asking Siri for the nearest gas station, for instance, returns an accurate answer in part thanks to these interlocking networks, with our devices incentivizing us to communicate according to a particular philosophical narrative about the nature of qualitative experience, and its relationship to reasoning.

Though styled on more performative grounds, befriending someone on a social network also happens on this matter-of-fact basis. Free social platforms derive economic value from us as an audience-commodity insofar as they are able to make ever-more fine-grained inferences about our relationships to things, events, places, and one another.

[12] Stell, John G. "Mereotopology and Computational Representations of the Body : Computational Culture." Computational Culture, no. 6 (November 28, 2017). http://computationalculture.net/mereotopology-and-computational-representations-of-the-body/.

These are again based on our habitual willingness to register our beliefs, actions, and desires in terms of empirically-styled, part-whole relations. The overall effect is that what we say comes to matter foremost in the simple fact of our *having said it*, recorded into a clustered network of objective assertions.

Collective habits interlace with these systems to productive effect because, following a discrete-continuous logic, they differentiate what we say about specific entities in a generically epistemological way, and then re-integrate these entities into perpetually reclassified groups. We accept the resulting landscape of empirical consensus maintained by platform capitalism because it allows us to more efficiently manage and perform our lives relative to other people, locations, institutions, and even emotional valences, all read as data-objects.

Given these current circumstances, how might Stiegler respond from a grammatizing perspective? At the heart of his work is an extension of Derrida's concept of the *pharmakon* into the history of technology, especially writing technologies. Those who've wrestled with Derridean deconstruction at some point will appreciate that pharmakon is a Greek term having the opposed meanings of "cure" and "poison". Derrida deployed the concept to describe the play of binary oppositions that he felt was crucial to, but submerged in the Western logocentric tradition—paradoxical oppositions enacting a systematic play of difference and deferral in knowledge, between speech/writing, good/bad, interior/exterior, and so on. Stiegler's innovation was to extend the concept of the pharmakon into our relationship with writing technologies, by putting Derrida's work

into conversation with that of developmental psychoanalyst Donald Winnicott, in particular his theory of a child's transitional object.

For Winnicott, the latter was simply some material thing—a teddy bear, toy, or safety blanket—that comes to sit in-between the love of mother and child, as a means for the child to form healthy psychic relations to reality. These relations ideally carry on into adult life through sublimation, converting desires and impulses for objects into socially productive behavior. Stiegler writes that "Winnicott's great discovery" on the logic of the transitional object was that the knowledge of a mother for their child is "knowledge of that which, in the transitional object, consists, though it does not exist, and which gives to the child placed under this protection the feeling that 'life is worth living'."[13] Put another way, beyond its simple existence the transitional object plays an instituting role, by acting relationally and transcendentally as "both an external object on which the mother and child are dependent [...] and in relation to which they are heteronomous; and an object that, not existing but consisting, provides (through this very consistence) sovereignty to both mother and child..."[14]

From out of this scene of the "first pharmakon", Stiegler extracts the notion that we share a similar kind of consistence-without-existence relation, but as adults, through writing technologies. Over time these technologies have produced different, shared forms of relationality to entities in the world, each of which has variously fostered or hobbled the

---

[13] Stiegler, Bernard. *What Makes Life Worth Living: On Pharmacology*. Translated by Daniel Ross. 1st edition. Cambridge, UK: Polity, 2013. P.5
[14] Ibid, p.8.

interplay between heteronomy and autonomy in collective life. Although he approaches it from a more acutely pragmatic angle, Agre is likewise concerned with the philosophical constitution of relation; first through his grammars of action, and then in subsequent work on the formalization of deictic ontology. Moreover, he is concerned for many of the same socio-political reasons that troubled Stiegler. It's no secret that Stiegler despaired of our current situation under the disruptive conditions of today's computational capitalism, in which this interplay between heteronomy and autonomy has been colonized and perverted by social media systems and bureaucratic automation.[15] Through their grammars of action, these systems promote standardization and narcissism, producing what he lamented as a pharmacology of nihilism.

Bracketing this nihilism for the moment though, for Stiegler the pharmacological more generally is about establishing a kind of infinite consistency, which produces the conditions of possibility for both lived heteronomy and autonomy. Writing's pharmakon always involves making continuous reality discrete in some precise way, so as to produce a consistency for private thought; but as importantly, it sits at the core of the generativity of writing systems for acting as a kind of guarantor for ethical social relations, too. This bears repeating, differently: the important point for Stiegler, as demonstrated by his accounts of the alphabetic letter and the cinematic apparatus, was that processes of grammatization

[15] See Stiegler, Bernard. *The Age of Disruption: Technology and Madness in Computational Capitalism.* Translation edition. Medford, MA, USA: Polity, 2019.

constitute the conditions for thinking in historically different ways, by *breaking language into discrete elements that can then be coordinated together into a shared continuum*.

We can look to a couple of different moments in Agre's work where he expresses the same idea. The first is an aspect of his "capture model", where he writes that the driving aims of thinking from capture are "not political but philosophical, as activity is reconstructed through assimilation to a transcendent ("virtual") order of mathematical formalism."[16] Agre would later expand on this view in his book *Computation and Human Experience*, by detailing a conceptual distinction between objective and deictic representation. The former relies on quantified variables and first-order logic to define objects nonspecifically, so as to make them amenable to mathesis, while the latter "underwrites a mode of relationship with things and only makes sense in connection with activity involving those things."[17]

Taking both Agre and Stiegler into account as I've summarized them so far, I now want to return to the claim that we find ourselves living today according to a sociotechnically-embedded principle of consensus, rooted in a semantic view of language. Put in Stieglerian terms, a core principle of our collective grammatization is premised not so much on the discrete unit of the alphabetic letter, nor the base abstraction of binary zeroes and ones in computing, nor the performative unit operations embedded in contemporary social media (all of which Stiegler gestured to at one point or another in his

[16] Agre, Philip E. "Surveillance and Capture: Two Models of Privacy." The Information Society 10, no. 2 (April 1, 1994): 101–27. https://doi.org/10.1080/01972243.1994.9960162. P.107.
[17] Agre, Philip. Computation and Human Experience. Cambridge, U.K. ;New York: Cambridge University Press, 1997. P.253.

writing) so much as on the discrete unit of the *predicate assertion*. We can also bring in Agre's diagnosis of the objective-deictic distinction in grammars of action here. Through the central figure of knowledge graphs and social graphs, the predicate assertion emerges as a key, generic technical element driving our relationship to data ensembles. This has occurred because of the way that graphs seemingly straddle-and-constitute Agre's relational boundary between objective and deictic representation.

As already mentioned, the predicate assertion as grammatizing principle comes due in large part to the ascendency of relational databases over the past several decades, and more recently, to their adaptation into the inter-networked paradigm of graph databases. In both cases, systems design has tended to follow orthographic principles first laid down in canonical work on semiotics and logic by the early 20th century founder of American pragmatism, Charles Sanders Peirce. Their grammatizing effect is that when we make statements to one another that are recorded and subsequently analyzed and retrieved as data, we are primarily concerned to attribute properties to objects via general concepts—what Agre calls objective representation. Implicit in this move, and central to Peirce's ideas however, is that in so doing we are essentially delegating or automating the process of submitting our ideas to an *interpretive community*, who judges our assertions as contestable claims that can be true or false; this is an important social-experiential underpinning for intersubjective deixis.

This is all pharmacological in the sense that through Peirce's account of interpretation and representation, we come to inhabit a relation to entities that is at once

heteronomous, or governed by the will of all seeking to distinguish the false from the true, but also autonomous in the sense of achieving freedom and personal agency by submitting our thoughts to contestation over what is actually true. To put it another way, Peircean philosophy gives an account of meaning that yokes or bootstraps our most basic affects and habits up into the social conventions of knowledge production, and also makes them calculable, by giving a self-consistent formalizable explanation of how they can be related. Peirce's overarching goal was to establish a phenomenal basis for scientific reasoning, such that we might together share an expanding base of true facts, through a rigorous semiotic means of discerning between truth and falsity. Over time his methods of analysis and synthesis, and especially his work on existential graphs, have thereby come to form something of a bedrock for the relational technicity of our connected digital devices. Peirce's fallibilist approach to thinking had three main features:

1) for any proposition, it is possible to hold a mistaken belief;
2) for any mistaken belief, a society of inquirers can discern its own mistakes and progress toward discovering the true state of affairs, through self-correction; and
3) taken together, these two features amount to a kind of evolutionary individuation of human beings, defining continuous growth of scientific knowledge and the social constitution of the self through perfected inquiry.

It's clear that the mixture of hope and anxiety driving our collective relationship to the Internet participates in this fallibilist imaginary, so let us now circle back to Agre and Stiegler's shared concerns about our modern circumstances. Creeping political economy of surveillance, our loss of collective agency to automated bureaucratic systems, and/or a proletarianization of the general intellect all suggest that as a ground for *political*

subjectivation, Peirce's approach to meaning too quickly assumes that we all voluntarily participate in the technical art of producing "settled" signs, over the long term, as a universal community. Despite obvious and abiding merits for scientific practice, his semiotic system has less to say about truth and signification requiring the transformative unsettling of knowledge between groups, which may occur through an intrusive force of the new, to existentially disturb and disrupt through irreconcilable dissensus.

This is how one might begin to approach the problems of misinformation and polarization from a grammatizing perspective, for example. Via Stiegler's ideas married to Agre's, we are better able to see how the materialized philosophy of Charles Sanders Peirce functions as a collective principle of grammatization. But alongside the many strengths of his philosophy, we should also be probing for conceptual limitations that come baked into Peirce's particular philosophical brand of semiotic idealism, constantly projected as it now is into global electronic discourse. It's easy to get lost in the weeds of Peirce's ideas, but one can at least gesture to one or two critiques of his thinking, which analogize fairly well to the grammatizing-related issues outlined so far.

One place to look among people who've critiqued Peircean epistemology in useful ways is Jurgen Habermas, who devotes a chapter in his 1988 book *Postmetaphysical Thinking* to understanding the potentials and drawbacks of Peirce's system from the perspective of human communication. Compressing the chapter's argument greatly, Habermas's criticisms come down to two major issues. The first is that Peirce's decision to vest the interpretation of signs cosmologically into nature itself, rather than to delimit

certain conceptual differences between human beings as interpretants, and non-human entities like plants, animals, and natural phenomena as interpretants, comes at the price of our possessing collective agency with respect to knowledge only in the light of knowledge's slow evolution over the long term.

Second but related to the first, Habermas believes that because of this assumed scene of all humankind participating in a constant and universal "correction of the record" through better and better synthetic inference, there is a real deficiency in Peirce's account concerning what makes a person who they are. For Peirce, as Habermas puts it, everything that makes a person into an individual is defined purely negatively, in terms of its difference from what is general—in terms of the distance separating error from the truth and of that dividing the egoist from the community.

Consider this passage capturing Habermas's overall sense of Peirce's work and its limitations on these two fronts. Because Peirce's doctrine of synthetic inference through semiotic chains finds its foundation in the laws of natural evolution, and by extension, "If the learning processes of the human species merely continue those of nature in more reflexive form," then "...argumentation, and the power of the better argument to convince, both lose the weight and value that are proper to them." He continues that

> "Unforced agreement and individuals holding one another accountable (faced with opinions that differ from person to person), ought to issue from argumentation by virtue of the latter's specific character. But this specific achievement falls victim to the leveling force of a universalism propelling

> itself inferentially from within reality itself. Multivocality of intersubjectivity becomes an epiphenomenon." [18]

To put it another way, and paraphrasing James Williams on the limits of pragmatism, we want to be able to organize our truths not just so that they are subject to corrective revision, but also so that they can be perpetually called into question in a more fundamentally evolving way. In his work juxtaposing the technical semiotics of Peirce with other thinkers like Barthes and Deleuze, Williams approach shows conceptual affinities with Agre's objective/deictic distinction when he writes that

> "Two forms of truth can therefore be defined. There are secondary truths, which are determined by their quality of being about something. Such truths can be captured in truthful propositions about the world. There is also, though, primary truth, which is determined as the disturbance of secondary truths: their sundering and transformation. In this real truth we experience the dissolution of our secondary, propositional truths in the encounter with the sign."[19]

Finally, in articulating his position on grammatization Stiegler likewise turns to the philosophy of Gilles Deleuze, and his sometime writing partner Félix Guattari. While clearly respecting Peirce's ideas, even borrowing liberally from them (as Deleuze did in his cinema books, for example), Deleuze and Guattari heavily reconfigure Peirce's semio-logical system. If Peirce starts from a position framed in terms of the ideality of signs, from out of which representational difference can be relationally coordinated in reason, then Deleuze

---

[18] Habermas, Jürgen. *Postmetaphysical Thinking*. Translated by William Mark Hohengarten. Reissue edition. Cambridge, Mass.: MIT Press, 1994. P.109.
[19] Williams, James. "Barthes, Deleuze, and Peirce: Pragmatism in Pursuit of the Sign". In Bignall, Simone, Sean Bowden, and Paul Patton. *Deleuze and Pragmatism*. Routledge, 2014. P.49.

and Guattari focus instead on difference-in-itself, as it develops through the ontogenetic formation of *series,* which emerge according to the conditions of a given environment or situation. In many ways, they preserve the relational diagrammatic structure of Peirce's system, but importantly, sidestep what they see as its limiting focus on a dialectic of realism and idealism, maintained as it is in the linguistic exchange of representational signs between people.

As James Williams lays out, the move entails a shift away from focusing on truthful propositions about the world—for our purposes, perhaps best exemplified by the methodological commitment to 'aboutness' at the heart of library and information sciences, applied to all kinds of new contexts via the expansion of social computing—towards a focus on what he calls 'primary truth', or truth as the *disturbance* of the truths of aboutness.

A hastily concluding point to all of this discussion might go something like this: If Peirce's metaphysics ostensibly underwrite social computing's orthographic practices, then how can we be expecting knowledge organization, social media, and the inferential machine learning strategies to which they are increasingly connected, to foster the ends of people encountering one another in their radical alterity, when their underlying techniques so thoroughly commit us to the continuous refinement of past resemblance? Circling back to Stiegler's diagnosis of our disindividuating moment, the upshot is that through a return to Agre's work, updating his philosophical assumptions away from phenomenology and towards other thinkers like Stiegler and those upon whom he relies, we might begin to

better address political subjectivation over digital systems, and how we might want to memorialize the agonistics of politics through action-grammars, at the level of their formal representations in data and technique.

The Grammars of Alternative Networks

Tobias Stadler, University of Oldenburg

working abstract

Abstract

Alternative social networks often present themselves as contestants against the exploitative logics of digital capitalism. And while they definitely are laboratories of better digital futures, they also continue the ways *the social* was reformatted by commercial social networking platforms.

In this paper I am going to discuss this ambiguity by taking a closer look at the `activityPub` protocol and the *Fediverse* which grew around it, comprising of open source software such as *Mastodon*, *Peertube* or *Pixelfed*. To build the argument, I will first give a short overview over the *digital labour debate* and expand on the concepts of enclosure, real subsumption and subjectification, to arrive at what I will describe as *subroutines of ideology* - the ideas, institutions and apparatuses stabilizing, legitimizing and obfuscating the regime of digital capitalism.

From there I will give an introduction in the federated systems around the `activityPub`

protocol, taking a closer look at the defined vocabulary and it's network structure. There are some quite literal *grammars of action* to be found and discussed. Some of them seem hauntingly familiar, especially in the context of the all too familiar allegories used in the various interfaces of these alternative social networks, echoing their commercial counterparts, continuing their reformatting of the social.

From there I will circle back to subroutines of ideology to show how some of them are continued in the fediverse, and how some of them are fundamentally questioned. With this analysis I will try to overcome the "bifurcation of utopianism vs. anti-utopianism" (Agre 1994), and will deploy the image of the digital objects as "barricades" (Dean, Hands, and Jordan 2020), sites of struggle over around techno-social systems.

Digital Labour

Every action we take within the walled gardens of commercial social networking platforms (CSNPs) is organized in a way optimized for the production of data. These data are the basis for various business models from selling behavioral prediction and modification, to training neural networks (what is nowadays called an *AI*) and selling their abilities as services, or - most prevalent on social networking sites by far - personalizing ad space. The users' actions, activities and interactions become the basis for a far reaching production process. This means, users *work*, but they are never compensated. Their unpaid labour produces data commodities which are hardly ever directly sold, but build the economic basis for some of the biggest and

most valuable companies of our current phase of capitalism.[1] This form of exploitation of unpaid labour is the core interest of the concept of *digital labour*.

Christian Fuchs analyzes the user data as a commodity in the marxian sense, with both an exchange value and a use value - a *doubled* use value actually. Whereas it has a use for the platform to make predictions about categories of users (mostly to sell ad space in the case of CSNPs), it means access to *the social* and communities for the users themselves. As Fuchs put it: "on the one hand, users produce use-values for themselves and others; they create a social relation between users and public visibility. On the other hand, users produce use-values for capital, that is, targeted advertising space for the advertising industry" (Fuchs 2014, 257). This double use value also leads to a curiously inverted form of commodity fetish. The classic definition of the commodity fetish would be that the commodity hides the social relations it is embedded into behind the economic function it displays, and thereby also reflects the social relations of the people in its production, circulation, consumption and disassembly as relations between objects, not humans. But due to its informational and social nature, the "social media commodity" displays an *inverted* commodity fetish: It hides the economic relationship between users behind its social use value (Fuchs 2014, 261), the modelling, quantification and commodification of the economic relations disappear behind the warm glow of *the social*. The users' unpaid labour is appropriated by the CSNP and ideologically made invisible, while other users' labour is mirrored as the *service* of the

[1]This phase is what is variously called "platform capitalism" (Srnicek 2018), "surveillance capitalism" (Zuboff 2019) or "communicative capitalism" (Dean 2005). While most of these concepts highlight certain aspects of the digitization of our mode of production, they also tend to downplay or overlook other parts. The trope of a *historical break* is prevalent among them and leads to the obfuscation of continuing logics and structures of 'plain' capitalism. To mark both the continuity of capitalism and its reformatting through computational infrastructures, I will use the term *digital capitalism*.

platform, the incentive to be there.

New Enclosures

We were not forced by violence to change our social relations into this. Instead it was a slow process of *enclosure* and normalization. The enclosure happened with vast amounts of resources made available through venture capital made possible by financial politics after the recovery from the *dotcom* crash and the ongoing 2008 financial crisis. This made it possible for actors like MySpace, Google* and *Facebook* to replace the many small servers hosting the social interactions and collective knowledge production of the time: the many specialized discussion boards; the local event pages; the public chat sites. Instead they were centralized on various generic platforms and later on consolidated through technical and social network effects into just a handful of quasi-monopolies, who invited everyone and everything onto their platform, but made it incredibly hard and expensive to leave again. They did so by a process of "decentralizing platform features, while recentralizing platform ready data" (Helmond 2015). Decentralizing their features meant casting a wide net of capture technologies by embedding them into every other website, widget, mobile application or even IOT-device. This process can be analyzed as the first step of how capital acquires control over the labour process: By centralizing the production process and managing the location, time and way it takes place - as well as the way, the produced commodities will *realize* their value, how they enter the sphere of circulation. Marx called this step *formal subsumption*, and the social character of the activity becomes fundamentally changed: the *work* of social interaction is turned into the *labour* of data production, now functioning as both at the same time. Formal

subsumption hardly changes the production process - the social practices in our case - but organizes them in a way producing commodities, turning them into alienated labour (Marx and Engels 2014).

But capital's control over the production process becomes more an more intense. While under formal subsumption "capital subordinates labour on the basis of the technical conditions in which it historically finds it" (Marx and Engels 2014), the constant need to expand demands an intensification of the production process. The *grammars of action* imposed by these increasingly extensive capture systems get more and more granular. It is not enough to surveil the users of the platform, they need to be captured, their specific forms of social interactions changed to be more data intensive. This is what Marx called *real subsumption*, wherein the actual activities of the production process get changed, broken up and rearranged, optimized for the production of relative exchange value.[2] Agre noted, how such forms of interactions formalized for data production are enforced in ways "both social (explicit procedures backed up by certain relations of authority) and technical (whether through machinery or simply through physical barriers)" (Agre 1994). Whereas the technical side of this is done through the interfaces and underlying protocols of CSNPs, the social side is done by the users themselves - dropping out of the conversation because of some change in the design of a platform, in the legal terms of service or in the way certain interactions are made possible comes at a very high social cost. Changes seem to be normalized almost immediately.

Changing the places and practices of sociality, care and reproduction changes us as well.

[2]Absolute surplus value is the surplus value created by a single labourer in e.g one working day, which legally cannot be extended beyond a certain time, whereas relative surplus value is the surplus value produced in a fixed amount of time (e.g. one hour) and can be extended by intensifying the production process through machinery, or in our specific case via more extensive capture systems and more data intensive social practices.

The platform organizes not only our actions and relations, it also regularly calls upon as, designating as as a user. "What's on your mind, `$USER_NAME`?" is the question we first encounter, once we open our apps or web interfaces. Through this appellation, we internalize the CSNPs logics and grammars. We universalize the platforms interests as ours as well, demanding of other users to be active and responsive, socially pressuring each other into producing for capital. What Geert Lovink calls "becoming-user" (Lovink 2016) is part of our subjectification through computational infrastructures and digital social relations. The idea of the *netizen* was something very different from a *user* or a *follower*.

The enclosure of the social is a new form of *primitive accumulation*, wherein a collective resource gets appropriated, reorganized and privatized. In this notion primitive accumulation is not a historically completed process which happened only once. It is rather a recurring process, part of capitals response to crises and struggles. Erecting property regimes over resources, activities and places which were held in common before, is part of this strategy. This is part of a double movement which privatizes these resources on the one hand, and separates people using them from the production process on the other hand. *Becoming-user* is not just a question of subjectivity or belonging, but turns into a question of *class*, understood as an economic positioning in the production process - being on the wrong side of exploitation, power and self-determination. It is not land and subsistence we are being divorced and alienated from by capital, but the main places of social interaction, public discourse and collective production of knowledge and culture.

Ideology

Feminist social reproduction theory not only shows how people, bodies and communities are formed, stabilized and reproduced through sociality and care work. Next to the material and generational side of social reproduction, there is an ideological one as well - the two being thoroughly intertwined, as Julia Dück reminds us (Dück 2021). The approval of and subordination under the relations of production and power needs to be stabilized and reproduced constantly as well.

This is one of the things, ideology *does* in various ways. Ideology is a tricky concept though, wanting to be nothing and everything at the same time, contradictory and far reaching. To develop a non-monolithic notion of ideology, Terry Eagleton describes various "ideological strategies" to overcome the divide between *epistemological* ideas of ideology - "preoccupied with ideas of true and false cognition, with ideology as illusion, distortion and mystification" - and more *sociological* concepts of ideology - "concerned more with the function of ideas within social life (Eagleton 1991, 3). These ideological strategies include *unification*, *naturalization*, *universalization* or *action orientation*. Stuart Hall describes ideology in a very similar manner, speaking of different"effects" it has (Hall 2019). These effects contain some quite similar concepts: masking and displacing (through fragmentation and imaginary unity or coherence); generalization or universalization (of class interests) and structuring of consent (and consensus among the subjects of power).

In this sense I want to understand the various effects as *subroutines of ideology*.[3] A

[3]This is inspired by Alexander Galloway's description of the functionality of software and ideology, in the sense of a function `f()` as a block of code. Whereas Galloway's description especially points to the processing

subroutine is "a sequence of program instructions that performs a specific task, packaged as a unit. This unit can then be used in programs wherever that particular task should be performed."[4] Ideological circuits don't need to be consistent in itself, different subroutines are deployed at various moments and places, to ensure the ideological side of social reproduction.

I have already mentioned several subroutines of ideology in this analysis so far, specifically deployed to normalize, hide or legitimate the exploitation of digital labour in the computational networks of our social reproduction and the reformatting of our subjectivities and practices of care and sociality. These and others include the work of users being *made invisible* and framed as joyful consumption; the *alienation* from the activity of community building both by being distanced from the place this happens in and by the form our organizing takes; the obfuscation of the economic relations users of CSNPs get put into behind the social relations the production of social data[5] needs - the already discussed *inverse commodity fetish*; the *naturalization* of data - the products of this labour - as 'the new oil,' or any other natural resource just found, not produced; the *subjectification* as passive users, actively engaging with each other for production, but unable to take part in designing the infrastructures enabling this engagement.

These subroutines are not new, most of them were already used in prior historical phases of capitalism, some of them still are. A prominent example is the work of women in the household being made invisible by framing reproductive labour as part of a essentialistic

of inputs and outputs, the shift to subroutines stresses the modularity of these subroutines.

[4]See en.wikipedia.org

[5]*Social Data* is a term by Salome Viljoen (Viljoen 2021), which points to the fact that data produced in social networking contexts is never just about individual people, as their value is deeply *relational*. Their commercial use always impacts multiple groups of people at the same time.

notion of womanhood and it's caring love. But the way in which they are configured and interfaced with each other in digital social networks is not random. Through their techno-social embeddedness into our infrastructures of social reproduction, they have the function of normalizing the logics of digital capitalism, and universalizing the interests of those profiting of these historically specific regimes of labour and property. Wheres the accumulation of capital is the imperative of all forms of capitalism, Jathan Sadowski points to two specific imperatives of digital capitalism, namely "data collection" or production and "expanding control," which is about "creating systems that monitor, manage, and manipulate the world and people" (Sadowski 2020, 34). This is the essence of economic surveillance or capture, best built with monopolies through centralization on digital platforms.

Protocols not Platforms

The concept of the platform is based on centralization, enclosure and capture, so any alternative social network has to struggle against the powerful network effects - both technical and social - this organizational model brings with it. Robert W. Gehl also describes the "Killer Hype Cycle" (Gehl 2015) in which an alternative network or app gets hyped early, does not live up to the hype and is declared dead shortly after - soon to be forgotten. Some names may still ring a bell though, *diaspora** was once widely known, and still has a surprising number of active nodes.

Much like diaspora*, the "fediverse" is a *federated network*. In such a network, there doesn't need to be a central server-complex to which all other nodes are required to connect

to. Instead its a network of computers running different software which implements the same *protocol* - in our case `activityPub`. A protocol defines "a set of possible communication actions between computer systems" (Guy 2017), by declaring which part of the data package is the message, which is the sender's user name, which is the profile picture, and so on, as well as how incoming and outgoing data packages should be handled. The protocol sets the rules for interactions between two servers, even if they were previously unknown. A single user account may be located on one server implementing `activityPub`, but is able to interact with accounts on other such servers as well - you can follow, like, block or comment on all postings you come across, which can be from servers specialized for spreading events (like Mobilizon), sharing videos (like PeerTube), posting photos (like Pixelfed) or for microblogging (like Mastodon).

One of the most widely used digital communication infrastructures world wide is in fact such a federated network: There is no central email server which controls all traffic in the email network, and although it becomes increasingly difficult[6] it is possible to host one's own email server, many institutions - like most universities - still do. But the actual email ecosystem is heavily centralized by now. After a big number of relatively big email providers existed for quite some time (think of *yahoo*, *AOL*, *gmx* or your universities) big players like Microsoft or Google amassed a significant part of all email users through bundling mail with other services, sheer market dominance or building new features not compatible with users from other services but heavily pushed onto existing users. Control over such a big part of a network gives enormous power over how other actors in the network should behave,

[6]The email protocol shows its age and the many additions it had to endure to respond to things like security problems or spam.

and what rules they have to adhere, to be deemed a save node and not instantly be put on various privately controlled spam- or blocklists. But while those few giant corporations can make email part of their platform, they can't completely privatize and enclose the whole network. This is why many alternative social networks like the fediverse do not try to become another *platform*, but rather are built around open standards and *protocols* instead - like `activityPub`, which abstracts the basic functions of other social networks into an open standard for various kinds of interoperable server software.

activityPub

Every protocol is per definition a *grammar of action* - and while Agre describes those (and capture overall) as comprised of "linguistic metaphors for human activities, assimilating them to the constructs of a computer system's representation languages" (Agre 1994), their focus is not on symbols and language, but on activities and actions. The `activityPub` protocol is based on the set of activities (the *vocabulary*) defined in the earlier `activityStreams 2.0` format, and contains both a protocol for two servers to interact with each other, as well as a protocol for clients to interact with servers. It was published in 2018 as a recommendation by the standardizing institution *World Wide Web Consortium* (W3C), which also maintains other fundamental web standards and protocols such as `html`, `http` or `webRTC`.

While this gives the protocol and it's developers - the by now disbanded *W3C Social Web Working Group*[7] - some hefty credentials for defining the programmatic basics of our

[7] W3C Social Working Group Wiki

digital world, any protocol is just as good as its implementations and reach. There are hardly any commercial servers in the fediverse, for various reasons. The inability to lock users into separated networks, where the *user experience* can be further refined for the production and privatization of social data, makes the implementation of the protocol undesirable for most commercial projects looking to compete in the social media space. But most of the rhetoric surrounding various fediverse projects positions itself as *actively opposed* to current forms of commercial social networking. From the free/libre open source software projects to the various community run servers, most of them finance themselves via donations of money and free labour by volunteers, users and developers, supporting their own alternative infrastructures. While the fediverse had its smaller hype cycles[8] and several moments of big spikes in user influx, and thereby turned in the one of the biggest alternative social networks so far. Due to the decentralized and federated structure of the network it is difficult to produce reliable data on the size and activity of the fediverse, but projects like fediverse.party or the Fediverse Stats Bot list between 8000-9000 nodes and around five million accounts. This number is of course dwarfed by the many millions or even billions of users of CSNPs, but research projects by the European Union[9] and Twitter[10] into federated social networks and specifically `activityPub` show the public interest in such alternative ways to organize the digital infrastructures of our social lives.

[8]The most recent one as a billionaire announced his plans to buy the whole of twitter.

[9]The EU hosts two instances in the fediverse, one of them is for twitter-like microblogging (EU Voice using Mastodon) and one for social video sharing (EU Video using PeerTube): Press Release

[10]The Project Bluesky started as a research project looking into the possibility of Twitter implementing an existing federated social networking standard, but is by now a public company developing its own competing protocol.

Critical Code Studies

This is one example of a publicly available status posting on the fediverse. I will not delve further into the specifics of this API response in this working paper, just point out the `content` section at line 20 and the `account.username` section at line 28, to get a grasp of whats going on here and how one `activity` shared between two servers can look like. The `jsonLD` syntax makes the shared data packages easily readable by demanding some predefined `key:value` pairs.[11] A big part of the standard is defining which of these pairs need to be present and which can be omitted, or which form the corresponding values should take. The web compatible version of the post can be viewed at the `url` at line 10, linked here.

```
0  {
1    "id": "107778817113430104",
2    "created_at": "2022-02-11T10:10:35.739Z",
3    "in_reply_to_id": null,
4    "in_reply_to_account_id": null,
5    "sensitive": false,
6    "spoiler_text": "",
7    "visibility": "public",
8    "language": "en",
9    "uri": "https://scholar.social/users/tobbsn/statuses/107778817113430104",
10   "url": "https://scholar.social/@tobbsn/107778817113430104",
11   "replies_count": 2,
12   "reblogs_count": 8,
13   "favourites_count": 23,
14   "edited_at": null,
15   "favourited": false,
16   "reblogged": false,
```

[11] The whole code block below is one *dictionary*, comprised of several key-value pairs. In line 7, the key is `"visibility"` and the value is `"public"`, which notes that everyone with the link is allowed to read the referenced object, the posting in this case. A value can be another whole dictionary, as is the case next to the key `"account"` in line 26, which describes the user who posted the status, which is me actually!

```
  "muted": false,
  "bookmarked": false,
  "pinned": false,
  "content": "<p>The thing I like most about online conferences and talks is the short moment when you can see the presenter&#39;s desktop or browser before they put their presentation into fullscreen. </p><p>That small glimpse of chaos is really reassuring, like: Okay, they lost control of their tabs as well, so thats kinda okay.</p>",
  "reblog": null,
  "application": {
    "name": "Sengi",
    "website": "https://nicolasconstant.github.io/sengi/"
  },
  "account": {
    "id": "189011",
    "username": "tobbsn",
    "acct": "tobbsn",
    "display_name": "tobbsn",
    "locked": false,
    "bot": false,
    "discoverable": null,
    "group": false,
    "created_at": "2020-08-26T00:00:00.000Z",
    "note": "<p><a href=\"https://scholar.social/tags/mediaStudies\" class=\"mention hashtag\" rel=\"tag\">#<span>mediaStudies</span></a> <a href=\"https://scholar.social/tags/digitalLabour\" class=\"mention hashtag\" rel=\"tag\">#<span>digitalLabour</span></a> <a href=\"https://scholar.social/tags/ideology\" class=\"mention hashtag\" rel=\"tag\">#<span>ideology</span></a> <a href=\"https://scholar.social/tags/infrastructure\" class=\"mention hashtag\" rel=\"tag\">#<span>infrastructure</span></a></p><p>having an account just to separate my academic stuff sometimes feels weird; main on cybre.space</p>",
    "url": "https://scholar.social/@tobbsn",
    "avatar": "https://cdn.masto.host/scholarsocial/accounts/avatars/000/189/011/original/e870f19a0a03e4d2.png",
    "avatar_static": "https://cdn.masto.host/scholarsocial/accounts/avatars/000/189/011/original/e870f19a0a03e4d2.png",
```

```
    "header": "https://cdn.masto.host/scholarsocial/accounts/headers/000/189/011/original/254a0934bd134e26.jpeg",
    "header_static": "https://cdn.masto.host/scholarsocial/accounts/headers/000/189/011/original/254a0934bd134e26.jpeg",
    "followers_count": 87,
    "following_count": 130,
    "statuses_count": 451,
    "last_status_at": "2022-08-04",
    "emojis": [],
    "fields": [
      {
        "name": "Prononuns",
        "value": "he/him",
        "verified_at": null
      },
      {
        "name": "Main",
        "value": "<span class=\"h-card\"><a href=\"https://cybre.space/@tobbsn\" class=\"u-url mention\">@<span>tobbsn@cybre.space</span></a></span>",
        "verified_at": null
      },
      {
        "name": "Writes in",
        "value": "English/German",
        "verified_at": null
      }
    ]
  },
  "media_attachments": [],
  "mentions": [],
  "tags": [],
  "emojis": [],
  "card": null,
  "poll": null
}
```

Into the Fediverse

I've already mentioned several different *implementations* of the `activityPub` protocol. An implementation is a piece of (server-)software, which is able to communicate its contents with other servers, by following the rules of a standard.[12] Due to the very abstract vocabulary of the protocol and its agnosticism regarding programming languages, there are wildly different ways to implement it, resulting in an wide system of microblogging, video sharing, discussion boards, long form blogs and event mobilization software.

These implementations cover a wide range of service which were centralized and privatized by corporate platforms. These alternative networks try to build infrastructures actively working *against* this centralized privatization. The official website of the `activityPub` protocol begins with the words: "Don't you miss the days when the web really was the world's greatest decentralized network? Before everything got locked down into a handful of walled gardens? So do we"[13]

This has several effects on the sociotechnical arrangements in these infrastructures. There are many small instances organized around certain topics or common interests like open source software (e.g. fosstodon.org) or crafts (like sunny.garden). Others are shared art projects like merveilles.town or are about geographical locations like berlin.social or wien.rocks. Some are political places like climatejustice.social. Different implementations have varying interfaces, but most of them make it possible to view all of the postings made on one server instead of

[12] While there are technical differences between a protocol and a standard (the former describing the latter), I use both terms mostly synonymously.

[13] see activitypub.rocks

only the accounts one follows, or all the visible posts from around the fediverse. Multiple timelines show people who you *follow* but also what other people on your server post. Due to the granular settings for post visibility, it is possible to communicate with only your server's users, forming tighter knit communities and closer social relations. This gives those communities the chance to form their own *governance*. Single users can more easily be part of the conversation on what kind of rules a server wants to implement and enforce. This reaches from *social rules*, like what kind of topics are not tolerated, how to handle posts of pictures without image descriptions, what kind of content is to be marked with a content warning - to *technical decisions*, like who gets to be a community moderator, which other instances are to be blocked on the server level, or who gets to join.

While there are are few instances with many thousands of users, most are in the few hundreds, many are even smaller, which makes it possible to know the people administrating and moderating one's social surroundings or have a relationship to them. The administrators' work becomes more visible, the repetitive activities of maintaining the server and moderating the community content are not hidden from the users sight anymore. This not only shows the relations of power between users and admins, but also how the work of administration and moderation for a community is a kind of *care work*. This aligns with Debbie Chachra description of infrastructures as more than just technological systems, but of "care at scale" (Chachra 2021). They put us into sustained relationships, help continuously fulfill the needs of our bodies but also materialize the ways, our needs and resources get structured by the interests of the designers, owners and maintainers of such infrastructures. The act of building alternative infrastructures turns into a political act, as it makes the exploitation inherent in

existing structures visible and tries to overcome them.

preferred writing

As already discussed above, protocols and their grammars are collections of allowed and forbidden activities and practices. This focus on activities and actions means normalizing certain forms of social interaction, specific activities and practices. But these practices were formally and really subsumed, meaning they were turned from simple social activity and unalienated work into exploitable labour optimized for the production of social data. These are the same forms of interactions, the `activityPub` protocol tries to abstract into an interoperable standard, which brings new contradictory moments into my analysis. Many of the same grammars of action are put into place and while the specifics and the scope may change, the core vocabulary remains the same, the basic organisation of "profile-centric liking and sharing" (Lovink 2022) is continued. Since most critiques of commercial social networking platforms stop at the liberal indignation over commercial surveillance violating the bourgeois private subject, they can't acknowledge the ideological rewiring of the social. This critique stops well before talking about production and exploitation.[14] which gives every interaction the form of a commodity. Even if it's exchange value may never be realized through actual exchange in the circulation sphere, the interaction is still in a commodified form. We still relate to each others as accounts - as objects, not as people.

[14]It is no coincidence that the star to liberal tech critique, Shoshana Zuboff, only analyzes the sphere of circulation in her book on "surveillance capitalism," as she does not really have a concept of capitalism, as Evgeny Morozov points out (Morozov 2019)

This comes with the industrialized production of *the social* - meaning the sum of our social interactions, our shared stories, our memes, our relations to ourselves and others. With Agre we can say that this production was reorganized to make the social *parsable*, machine-readable, ready to be used as a commodity. For a single user the exchange values of the data commodities they produce is almost negligible, since the use value of those *social data* only emerges in the relation of many such data sets - they are inherently relational. But the "cooperative peer-production" (Sevignani 2017) of the social still happens in alternative networks in mostly the same way, but in a somehow less alienated context.

The organizational basis of this production - and with the form of its infrastructure - is fundamentally changed by the federation part of the `activityPub` protocol. The social data and its production is distributed over a whole network, instead of a single private server complex. But while the technical infrastructure has changed in the fediverse, a big part of the protocol just reuses the *grammars of action* emerged from the real subsumption of the social. There are of course differences, like mitigations of some obviously undesirable effects or some shared cultural norms, but the form of the activity is mostly the same. The form of the social has become ideologically normalized, the interests behind seem universal: Sharing as much as possible with as many people as possible is a very dominant idea of social behaviour, numbers of followers and likes are indicators for social status. To paraphrase Wendy Chun: The protocol, the standard, the underlying grammars have become common sense, the "software is ideology" (Chun 2004). Ideological texts have to be read to function ideologically, and while Stuart Hall's concept of *decoding* showed that there is a possibility to alternative or counterhegemonic readings, he always insisted on what he called the "preferred

reading" of a text, one that transports the encoded ideological subroutines effectively. And even though most of the web is surprisingly passive consumption nowadays,[15] the focus on reading alone doesn't do the *read-write* web justice, and most social networks embody that ethos. This is why the analytical complex of real subsumption, grammars of action and software as ideology is so helpful in understanding how the ideological subroutines of digital capitalism are still effective in alternative social networks: They not only create preferred readings of media objects in the production of the the social, but also *preferred writings*.

Ambiguities

The ideological subroutines of digital capitalism get reproduced in alternative networks through the forms of *preferred writing* the established grammars of action impose on the users of social networking sites. While the production of social data does not have to be framed as oil, to hide the exploitation and dispossession necessary to privatize it, other forms of *naturalization* are working in the fediverse. Many implementations pride themselves with having no algorithms sorting the timelines of postings, arguing this would be more *organic* or *natural*. While the algorithmic sorting on CSNPs is one of the most insidious parts responsible for a lot of the problems associated with 'social media' nowadays, the idea of a more 'natural' way of using them is actively harmful too. There is *nothing* natural about the design of digital infrastructures and framing them as such is hiding the effects of the design choices made,

[15]The biggest users of global bandwith by far are streaming platforms like Netflix and YouTube. And while only a very small part of people on YouTube interact with the videos (liking, commenting, etc), the number of people actively producing and posting videos on the platform is even tinier. Netflix itself doesn't even give its users the possibility to interact with more than a thumbs up or down.

whether this is done actively or not. The same is true for the the commodified and therefore fetishized way of relating to one another, which is still dominated by a computer-formalized and heavily metrified idea of sociality.

The digital labour lens introduced in the beginning of this analysis and the ideology critique built on top of it helps us acknowledge bot the continuities of digital capitalism and the seeds of alternatives in alternative social networks. As laboratories of a better digital sociality, and tools for bringing about a postcapitalist world, they are sites of both struggles and enclosures. This is what Jodi Dean pointed towards with her image of digital objects as *barricades*: "As sources of enclosure as well as defenses for liberated space, barricades are erected where struggles are fierce and the stakes are high. They are necessarily partisan divides, different politicizations and deployments of a common surface." (Dean, Hands, and Jordan 2020). A materialist analysis helps in seeing these struggles and overcoming the "utopian and anti-utopian" genres, Agre pointed out with Kling and Dunlop as hindering the analysis of "computer technology in society" (Agre 1994).

References

Agre, Philip E. 1994. "Surveillance and Capture: Two Models of Privacy." *The Information Society* 10 (2): 101–27. https://doi.org/10.1080/01972243.1994.9960162.

Chachra, Debbie. 2021. "Care at Scale." *Comment Magazine*.

Chun, Wendy Hui Kyong. 2004. "On Software, or the Persistence of Visual Knowledge." *Grey Room*, no. 18: 26–51.

Dean, Jodi. 2005. "Communicative Capitalism: Circulation and the Foreclosure of Politics." *Cultural Politics* 1 (1): 51–74.

Dean, Jodi, Joss Hands, and Tim Jordan. 2020. "Digital Barricades: Interventions in Digital Culture and Politics - Series Preface." In *Furious - Technological Feminism and Digital Futures*, ix. London: Pluto Press.

Dück, Julia. 2021. "Mehr Als Erschöpfungen Im Hamsterrad - Soziale Reproduktion Und Ihre Krise(n)." In *Plattformkapitalismus Und Die Krise Der Sozialen Reproduktion*, edited by Moritz Altenried, Julia Dück, and Mira Wallis, 28–49. Münster: Westfälisches Dampfboot.

Eagleton, Terry. 1991. *Ideology: An Introduction*. London ; New York: Verso.

Fuchs, Christian. 2014. *Digital Labour and Karl Marx*. New York: Routledge.

Gehl, Robert W. 2015. "The Case for Alternative Social Media." *Social Media + Society* 1 (2): 205630511560433. https://doi.org/10.1177/2056305115604338.

Guy, Amy. 2017. "The Presentation of Self on a Decentralised Web." PhD thesis, University of Edinburgh.

Hall, Stuart. 2019. "Culture, the Media and the "Ideological Effect" (1977)." In *Essential Essays Vol. 1*, 298–35. Durham: Duke University Press.

Helmond, Anne. 2015. "The Platformization of the Web: Making Web Data Platform Ready." *Social Media + Society* 1 (2): 205630511560308. https://doi.org/10.1177/2056305115603080.

Lovink, Geert. 2016. "On the Social Media Ideology." *E-Flux*, no. 75.

———. 2022. *Stuck on the Platform: Reclaiming the Internet*. Amsterdam: Viliz.

Marx, Karl, and Friedrich Engels. 2014. *Das Kapital: Band 1 (1867)*. Dietz Verlag. Vol. 23. Marx-Engels Werke. Berlin.

Morozov, Evgeny. 2019. "Capitalism's New Clothes." *The Baffler*, February.

Sadowski, Jathan. 2020. *Too Smart: How Digital Capitalism Is Extracting Data, Controlling Our Lives, and Taking Over the World*. Cambridge, MA: MIT Press.

Sevignani, Sebastian. 2017. "Facetten Der Debatte über Das Digitale Arbeiten - Herausforderungen für Eine Kritische Theorie Des Informationellen Kapitalismus." *PROKLA* 47 (186): 43–62.

Srnicek, Nick. 2018. *Plattform-Kapitalismus*. Translated by Ursel Schäfer. 1. Auflage.

Hamburg: Hamburger Edition.

Viljoen, Salomé. 2021. "Data Relations." *Logic Magazine*, no. 13 (May): xx–yy.

Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*.

Imposition everywhere!
On the digital re-rendering of the lifeworld and the (lack of) resistance against it
Contribution to the Workshop "Agre After Techno-Utopianism" (Siegen, September 1-2)
by Andreas Beinsteiner

I want to start with an anecdote from personal experience: I am teaching media studies on several institutes. One of them emphasizes particularly the critical dimension of this discipline. On this institute, like everywhere, we were teaching remote during the pandemic. When I was back for an on-site-seminar the first time again, I wanted to get a drink during the break. I entered the room with the vending machines and was surprised to find out that, in the meantime, those machines had been modified so they would not accept cash anymore. There were still the same machines, one for coffee, one for soft drinks and one for snacks, yet to each of them was attached not only a panel for contactless paying, but also a mechanical barrier on the coin slot that would prevent people from inserting coins. So those modifications had been made in accordance with a conscious decision not only to offer new payment options, but also to restrict others. While their might be a little gain of efficiency for the company that operates those machines (their staff does not need to refill the coin stacks anymore, but nevertheless they have to come around frequently to re-fill the goods that are offered), this change has another obvious effect: on this institute with its critical stance towards various issues in contemporary society, it is not possible anymore to get a coffee from the machine without this act of consumption being recorded in terms of personal data.
With Philip Agre, this inconspicuous modification of the vending infrastructure, hardly recognized by most of the students and teachers who use it, can be read as an act of *imposition*. As a consequence of tiny modifications of infrastructure and of the affordances that the artifacts of everyday life offer, practices are modified so that they become feasible for the automatic capture of behavioral data. In this short paper, my assumption is that these tiny modifications, precisely because of their inconspicuousness, provide a central mechanism for gradually subjecting our lifes to total control: not only in the sense of comprehensive tracking of behavior, not only in terms of the exploitation of insights derived from psychometrics (hyper-nudging), but ultimately also by defining what options of action are granted to a given person in a given situation in the first place.[1] An important task for critical research on digital media is therefore to expose the oppressive implications of those modifications that, if noticed at all, get covered by rhetorics of smartness, usability and increased security. Agre's concept of imposition is important here because it makes these modifications readable as socio-technical moves in a power struggle.
The first section will indicate the sense in which imposition is a prerequisite for computational cognitions about human behavior. Section two poses the question why imposition, if recognized at all, seems to be met by so little resistance and tries to suggest some possible explanations. The concluding section argues for the need to complement resistance *within* digital technologies (which is the main focus of both existing critical technical practices and critical research) with resistance *against* digital technologies.

From cognition to imposition

Some years ago, N. Katherine Hayles published a paper with the title "Cognition everywhere: The rise of the cognitive unconscious and the costs of consciousness" (Hayles 2014) as part of her work on the nonconscious (which, unlike the psychoanalytic unconscious, is not related to human consciousness at all). Hayles argues that cognition does not only occur in the form of intentional thought, but is a

1 Beyond the scope of the present paper, this could be discussed as a personalization of what Agre calls "grammars of actions".

process that occurs pervasively in biological organisms and the more complex forms of interaction that emerge between them. In recent years, technology has become an important carrier of such nonconscious cognition. As technological devices are penetrating more and more regions of the world in general and the human lifeworld in particular, with smart homes, autonomous vehicles, smart cities, financial trading algorithms, airport safety procedures, etc., processes of cognition are established in all those domains. With pervasive computation, we are currently moving towards a situation of "cognition everywhere". Taking into consideration approaches on this proliferation of digital technology that are more sensitive to questions of economy and power than Hayles' - for example, Shoshana Zuboff's (2019) work on surveillance capitalism, it might appear that the term cognition is a little bit to unspecific and too innocent for many instances of this development. While diverse sensor systems might be involved in cognitive processes about environmental parameters, coordination between various autonomous devices etc, a big share of technological cognition has human practices as its object, striving to make behavior susceptible to analysis and manipulation. However, if we want to understand how such cognition is implemented technologically, Philip Agre's informed engineering perspective can contribute important aspects that remain neglected in Zuboffs account. Beyond generic accounts on technological cognition or a new economy centered around means of behavioral modification, Agre allows us to focus on specific forms of interplay between digital technology and human practices that grant the possibility of the production of behavioral data in the first place. And while it is plausible to claim that nowadays the proliferation of sensors and the increase in information processing between "smart" devices the capture model gives way to to the more radical paradigm of infrastructural surveillance (Geeker/Hind 2020), I would claim that to the extend that some important features of Agre's analysis still maintain their relevance – in particular to the extent that human practices are concerned.

In contrast to what popular discourses about big data are suggesting, human actions are far from being simply legible for computational systems. Rather, practices have to take on a specific form that enables the parsing of human activities into discrete and clear-cut units; a parsing for which Agre reserves the term "capture" to distinguish it from more traditional understandings associated with the notion of surveillance. Without going into the details of the interference between social factors (like certain paradigms in computer programming or specific economic contexts), we can assert that, as a condition of possibility for capture, practices have to be refashioned as sequences of human-computer-interaction. Where humans interacted with humans or other technologies hitherto, digital devices have to enter the scene for automatic capture processes to happen in the first place. This is a point that also Zuboff confirms: while she emphasizes the relevance of "the puppet master" and not of "the puppet" (2019, p. 14), that is, she locates the problem in the economic paradigm of surveillance capitalism and not on the computational technology it is based upon, she nevertheless admits that computational technology is the condition sine qua non of this paradigm: surveillance capitalism is "a market form that is unimaginable outside the digital milieu" (2019, p. 15). Wherever the capture of human actions shall take place, digital devices need to get involved. This is what happened in the in the case of the vending machines not accepting cash anymore, and it happens at innumerable other occasions. Catalyzed by pandemic measures, more and more social practices are refashioned in such a way that digital devices participate in them. I just mention three further examples in addition to the one described initially:

- In a growing number of cities, it is not possible anymore to buy public transport tickets directly on the bus. Things become cumbersome if one has to anticipate where one is going to go when, and to buy tickets beforehand. Many people thus adapt their bus riding habits to the online ticket-buying affordances of public transport apps on the smartphone.

- During 2021 there were phases when, at least in Austria, access to universities, restaurants, hairdressers, cultural events and even many people's workplace was only possible with a valid Covid19 test certificate. With test validity time spans of sometimes only 24 hours, people had to get tested on the same day before being allowed to enter one of the sites mentioned. This already was labourious if one had a smartphone with the possibility of mobile internet access so that it was possible to access the test result as soon as it was available. For people without smartphone, it was virtually impossible.
- More and more airports and airlines offer fast lanes for passengers who are ready to submit themselves to automatic biometrical recognition technologies. People who do not wish more of their biometrical data to be collected and processed than legislation and border transition requirements enforce anyway have to expect longer queue time and a higher risk of missing connection flights.

As those examples indicate, smartphones, cashless payment devices, and processes of biometrical recognition play a major role in transforming a large share of human practices in such a way that they allow for the automated capture of peoples actions. With Agre, the gradual and inconspicuous disappearance of possibilities to consume, travel or communicate without using electronic payment infrastructures or smartphone apps can be traced back to so many acts of imposition, enforcing people to transform their practices in such ways that they become digitally traceable. Thus, to the extent that it is concerned with human practices, "cognition everywhere" relies on "imposition everywhere" as its condition of possibility. Yet, in contrast to what Agre expected in the 1990ies, resistance against imposition hardly seems to be one of the essential spots of political struggle today.

Where is resistance against imposition?

Against the mythology, or even ideology, that claims practices and behavior would just be traced by digital technologies and remain the same they had been from the outset, Agre emphasizes the imposition of both a technological infrastructure and of a certain structure of behavior (so-called grammars of action) that remains parsable by this technology as prerequisite for any data-production from human actions: "it is crucial to appreciate the senses in which the imposition and instrumentation phases [for capturing some kind of activity] constitute a reorganization of the existing activity, as opposed to simply a representation of it" (Agre 1994, p. 108). Practices have to be modified to comply to demands of machine readability. While I believe that the aspect of the grammatization of human activities, subjugating them to a parsable grammar of action is still virulent today (though less openly perceivable, since the granularity of the grammars and the quantity of options offered to users is constantly increased, cf. Agre 1994, p. 114), in this short text I will restrain myself to the other, more basic, aspect of imposition: the fact that computational infrastructure needs to be introduced into a practice to make it feasible for capture. Agre mentions radio frequency beacons, barcodes, wireless networks, protocols that standardize procedures, software systems, and databases that all enable computers to keep track of state changes of an increasing range of entities; as it happens in accounting systems, scripted customer service, highway architecture, computer interfaces, communication protocols, restaurant software etc., with the tracking of employees being Agre's main example. Complying with imposition comprises acts of "data entry, […], swiping of cards through readers, aiming of sensors at bar codes, and so forth" (Agre 1994, p. 111) Our up-to-date examples above include the replacement of cash payment by electronic payment, of over the counter interactions by interactions with smartphone apps, and the delegation of identity checks from humans to biometric technology.

Agre expected that the processes of imposition that force people to adapt their practices as well as the resistance these processes provoke would become essential spots of political struggle. On an institutional level, interventions can range from "manipulation of the institutional procedures around

the system's use to lobbying for technical changes to overtly political campaigns to regulate uses of the technology" (Agre 1994, p. 112) – possibilities that seem to get enacted more by those striving to induce increasing imposition than by those who want to confine it. Below the institutional level, reactions of the ones who are exposed to exposition can follow different paths: On the one hand, users can resort to system-immanent tactics by twisting their practices in such a way that their performance is optimized according to certain parameters of their activity that get measured or tracked: "they will maintain an orientation to the 'image' they project to whoever is making use of the captured information, be it a boss, a colleague, an auditor, a regulatory agency, an insurance company, or whatever." (Agre 1994, 111) On the other hand, they can also choose to resist imposition more overtly: "They might attempt to minimize use of the computer […]. They might falsify certain information, they might delay entering it, or they might neglect entering it altogether." While we can observe a lot of tactics today that stay immanent to diverse systems of tracking and rating, there is little fundamental resistance to these systems as such.

The first option (optimizing ones behavior to increase performance as it is represented by the captured parameters) can hardly be understood as an act of resistance, since such optimization is exactly the outcome that system designs, in particular ones that are centered around rankings (of task delivery at work, of risk aversion in the case of driving behavior monitored by car insurance companies etc.), aim to induce. Research on the second set of tactics emphasizes particularly tactics of obfuscation and interruption as the ones presented by Brunton/Nissenbaum, rather than efforts not to be captured that much/in this way (to paraphrase Foucault); that is, not to have all that capture infrastructures imposed on ones life. Both in research and practice, tactics immanent to certain digital technologies seem to dominate over attempts of plainly refusing to participate in these dispositives. Why is this the case? There are (at least) three factors relevant for explaining this tendency:

- *the shift from workplace capture to leisure capture*

  Agre conceived capture as a management technology that he expected to get employed mainly in workplace contexts. Resistance to imposition in this context would be mainly a task for trade unions. But not only is solidarity among workers eroding in a post-industrial world eroding, as Gilles Deleuze famously noticed in his "postscript on the societies of control" (cf. Deleuze 1995); also has capture expanded to behavior far beyond the workplace, covering all kinds of everyday practices with a particular focus of leisure activities like consumption (of goods, services and information) and communication. While in workplace settings the repressive quality of acts of imposition remains quite evident, this is less the case with practices in which one engages on an seemingly voluntarily base.[2]

- *the emergence of new practices with new technological possibilities*

  In Agre, an important (though, as far as I can see, implicit) source for resistance against imposition seems to be the inertia of established ways of doing things. People want to perform their activities in the way the are used to; the need to modify practices is experienced as disturbing. (In accordance with this idea, it can be observed that political forces striving to defend the right of cash payment are frequently conservative or right-wing.) Yet many of the activities that are subjected to capture today are not long-established practices but rather new ones that emerge in interaction with new technological possibilities.

  This entails a potential for designing new practices that had been recognized in computer engineering already in the 1980ies. For example, the developers of the "communicator" software for intra-organizational communication state that "in our interpretation, technology is not the design of physical things. It is the design of practices and possibilities to be realized through artifacts."(Flores

2 Tellingly, efforts to legally secure a right to disconnect, as they are reported by Hesselberth 2018 (p. 1994f about Germany and p. 2005f about France) remain restrained to working contexts.

et al. 1988, S. 153) The use of all kinds of web platforms and smartphone apps is exemplary for designed and designable practices in this sense. While there may be several reasons why platforms and apps play such an important role in contemporary economy, on of them is certainly the fact that they allow it to design new social practices of unprecedented plasticity: “In the extreme case, design is able to prearrange in an anticipatory manner habits that do not exist yet. The mobile, so called ‘smart devices’ (‘smartphones’, ‘smartpads’) existing only since a few years, demonstrate impressively the emergent character of such transitional processes.” (Jörissen 2015, p. 221; my translation).

- *the stance of the humanities* (and the intellectual sphere more broadly)

Already in the 1950ies, philosopher of technology Günther Anders observed that “[t]here is nothing more precarious today, nothing that would render a person as immediately unbearable as the suspicion that he was a machine critic [Maschinenkritiker]. For as part of what class it, of what interest group, social system, in the context of what kind of political philosophy it may be – if one dares to present an argument about the ‘degrading effects’ of this or that device, one immediately obtains the reputation of a ridiculous luddite, and accordingly, one immediately is sentenced to intellectual, social and journalistic death. No suprise thus that most critics are silenced, and that a critique of technology has become a matter of civil courage these days. After all, the critic thinks, I cannot afford to be deemed by everyone to be the only one who stands in the way [in die Speichen falle] of world history, the only obsolete and reactionary one. And so he keeps his mouth shut – in order not to be regarded as reactionary.” (Anders 1962, p. 3, my translation) This assessment is appropriate no less today. There still is an intellectual climate that raises unease about what is called technological pessimism. It is no surprise under such conditions that among research discourses about digital resistance, the most advanced approaches are the ones who frame it as disruption or resistance *within* networks of digital technologies (in contrast to to empirical studies of technology non-use or critical accounts of resistance to media, if we follow the classification proposed by Hesselberth 2018): The intellectual common sense seems to be that resistance to technological progress is futile and that attempts to refuse/abolish certain technologies are reactionary romanticism.

There are, no doubt, good theoretical reasons for this common sense; i.a. various anthropological arguments stating that hominization is coextensive with technization (Arnold Gehlen, André Leroy-Gourhan, or, more recently, Bernard Stiegler). Yet to insist that there is no human life outside the sphere of technology does not straightforwardly imply the necessity to embrace all kinds of *digital* technology in all domains of life.

Another reason is that most critical research about technology and media emphasizes the difference between capitalist enactments of digital technology and alternative forms it can assume (cf. e.g. Lovink/Rasch 2013 and Stadler 2021 for alternative social networks), trying to keep faithful to the empowering potential that digital technologies once promised. For example, Zuboff argues against what she calls “inevitabilism” (2019, p.15) with regard to digital technologies, interpreting it as a discourse serving the interests of surveillance capitalism: “surveillance capitalists want us to think that their practices are inevitable expressions of the technologies they employ” (2019, p. 15). Presenting the way digital technology is effectively deployed currently as “inevitable when they are actually meticulously calculated and lavishly funded means to self-dealing commercial ends” (2019, p. 15) would thus block the perception and realization of different possibilities of arranging and using digital infrastructure.

Yet I would argue that Zuboffs argument might also work the other way round: Claiming that digital technology is something that does not pose problems in itself and that we should embrace, while almost all instances of this technology that are available for non-expert everyday users contribute to the extension and ossification of highly problematic asymmetries in knowledge and power, might

serve primarily the interests of surveillance capitalism. Only if refusing to use digital devices, applications and infrastructures becomes a realizable option, both in institutional/legal terms and on the micro-level of individual everyday practices, the spell of actually existing digital technology can be broken. By claiming that being digitally connected would be an irreducible prerequisite for social participation and thus should count as a human right, intellectual discourse unwittingly becomes an advocate of current power relations. On the contrary, the fight for a human right to be able to maintain a good life while having an unrestricted right to digital disconnection might be imperative today in order to achieve a true pluralization of (digital and non-digital) technologies, and thus, of ways of living with technology. My suggestion is that it might be useful to complement and extend the use of digital technology for resistance with resistance against the proliferation of digital technologies itself – at least in certain domains of life.

From resistance within to resistance against

A paradigmatic representative of the intellectual common sense outlined above is Alexander Galloway (2004), who, in his discussion of computational power as *protocological* (a concept rather similar to Agre's account of the role of grammars of action in computation) states that there can't be any resistance to protocol (just like there can't be resistance to gravity, cf. p. 147), so that "it is through protocol that one must guide one's efforts, not against it" (p. 17). In his account of the ideology of *computationalism*, David Golumbia (2009) questions this wholesale dismissal of resistance against the computational as such. Practices of resistance within the computational, like hacking, should be complemented by " resistance against what Galloway calls protocol, and what is more generally thought of as computerization" (p. 25). In the background here is the observation that "computationalism often serves the ends of entrenched power despite being framed in terms of distributed power and democratic participation" (p. 4). Golumbia refers in particular to the massive concentration of computation power (which has further increased since); leading to a severe power imbalance between individuals' and groups' subversive tactics and the accumulated power of companies like Alphabet or Meta, to name just the two most conspicuous. "[S]eeing public resistance too much in the context of computerization", he argues, "aligns too closely with the technological progressivism that conditions so much computational discourse" (p. 5)

I want to return to the simple (and maybe a little bit banal) observation made in the first section: the whole setup of surveillance capitalism would be impossible without computational technology as its condition sine qua non. Thus we might focus on the question in what contexts the introduction of digital technology is really desirable, and in what contexts it should rather be resisted. When are the benfits of computation really desireable, and where they are outweighed by the costs that come along with it (costs in terms of loss of privacy, susceptibility to behavior modification as Zuboff describes it, cognitive proletarianization in Stieglers terms, etc.)?

To what conclusions are we led if we address these questions in the spirit of Agre? At first sight, Agre's concept of a "critical technical practice" (Agre 1997), arguing for the inseperability of "the craft work of design" from "the reflexive work of critique" (p. 155) seems to clearly align him with the dominant tradition of resistance *within* digital technology. Furthermore, he indicates that the imposition of grammars of action, although deeply ingrained in software design, is a cultural heritage that is not without alternatives. Technologically, it should be possible to develop systems that capture less activity instead of more. However, if we look for alternative ways to design systems that do not impose grammars, we find "a trade-off that goes to the core of computing": the less the a system captures, the less functionality it can provide. Agre's respective examples point into the direction of systems that abstain from structuring human communication data in an detailed manner (e.g. plain email, voicemail or also shared drawing tools; cf. also Agre 1995). Promoting such computational technology that

captures less information instead of more is a rather radical orientation that strongly contradicts almost all contemporary computing practices with their ever more fine-grained capturing of activity and their massively centralized processing of the data captured. It could be argued that using less computational functionality instead of embracing everything that is presented as technically feasible would amount to practices of de-digitization and disconnection that strongly counter today's omnipresent equation of progress, digitization, and connectedness.

Hesselberth (2018) discusses various manifestation of what she calls the "paradox of dis/connectivity", which arises from the dissonance between the *gesture of disconnection* and the *norm of connectivity*. One of those manifestations, the one that characterizes the approach of resistance within digital technology, is constituted by the fact "to 'opt out' in the present context, or to invoke or secure ones 'right to disconnect,' one (first) has to connect more, that is, spend more time, energy, and effort engaging with these connective technologies, even if they are the very thing, or paradigm, one wishes to opt out from" (p. 2006). Following Agre's considerations as they have been presented above, this causality might be reversed: Precisely the intense engagement with the technical dimension of digital connectivity can cause the gesture of disconnection to appear as imperative.

A critical technological practice that resists both the imposition of formal grammatical structures on human activity and the data production enabled by this imposition would thus entail a radical refusal of almost all state-of-the-art proprietary centralized information technologies. It would include the readiness to restrain from utilizing a big share of technological functionality we use to take for granted. Furthermore, and to an increasing extent, it would need to encompass political and juridical struggles for a legally warranted right to disconnect (cf. Hesselberth 2018), which would have to reach out far beyond workplace settings and focus on protecting those from discrimination who prefer not to embrace all kinds of digital connectivity that are deemed to be state of the art today.

References

Anders, Günther (1961): Die Antiquiertheit des Menschen. Über die Seele im Zeitalter der zweiten industriellen Revolution. München: Beck.

Agre, Philip E. (1994): Surveillance and capture: Two models of privacy. In: The Information Society: An International Journal 10 (2): 101-127.

Agre, Philip E. (1995): From high tech to human tech: Empowerment, measurement, and social studies of computing. In: Computer Supported Cooperative Work (CSCW) 3 (1): 167-195.

Agre, Philip E. (1997): Toward a critical technical practice: Lessons learned in trying to reform AI. In: Bowker GC, Leigh Star S, Turner W and Gasser L (eds.): Bridging the Great Divide: Social Science, Technical Systems, and Cooperative Work. Mahwah, NJ: Lawrence Erlbaum Associates, pp. 131-158.

Brunton, Finn and Nissenbaum, Helen (2015): Obfuscation. A User's Guide for Privacy and Protest. Cambridge/London: MIT Press.

Deleuze, Gilles (1995): Negotiations 1972-1990. New York: Columbia University Press.

Flores, Fernando et al. (1988): Computer Systems and the Design of Organizational Interaction. In: ACM Transactions on Office Interaction Systems: 153-172.

Galloway, Alexander (2004): Protocol. How Control Persists after Decentralization. Cambridge/London: MIT Press.

Gekker, Alex and Hind, Sam (2020): Infrastructural surveillance. In: new media & society 22(8): 1414-1436.

Golumbia, David (2009): The Cultural Logic of Computation. Harvard: Harvard University Press.

Hesselberth, Pepita (2018): Discourses on disconnectivity and the right to disconnect. In: new media & society 20(5): 1994-2010.

Hayles, N. Katherine (2014): Cognition Everywhere: The Rise of the Cognitive Nonconscious and the Costs of Consciousness. In: New Literary History 45: 199-220.
Jörissen, Benjamin (2015): Bildung der Dinge: Design und Subjektivation. In: Jörissen, Benjamin & Mayer, Thorsten (Eds.): Subjekt Medium Bildung. Wiesbaden: Springer VS, pp. 215-224.
Lovink, Geert and Rasch, Miriam (Eds., 2013): Unlike Us Reader. Social Media Monopolies and their Alternatives. Amsterdam: Institute of Network Cultures.
Stadler, Tobias (2021): Ölstandsanzeiger - Über die Unsichtbarmachung und Naturalisierung der Produktion personenbezogener Daten. In: Bachor, Martina; Hug, Theo and Pallaver, Günther (Eds.): DataPolitics - Zum Umgang mit Daten im Digitalen Zeitalter. Innsbruck: innsbruck university press: pp. 163-176.
Zuboff, Shoshana (2019): The Age of Surveillance Capitalism. London: Profile Books.

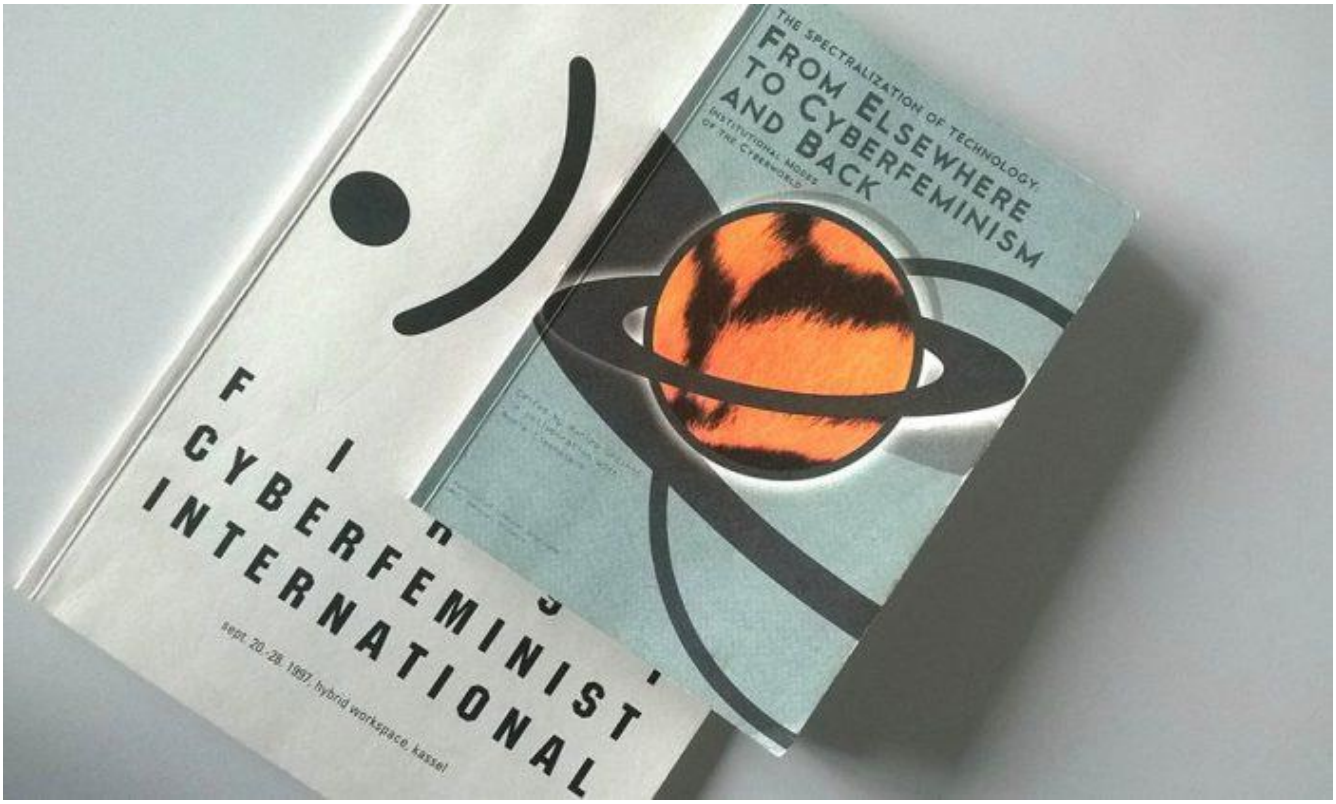
Technology, race, class, sexuality
CYBERMEDIA HISTORY 1990 on former Yugoslavia, former Eastern Europe and beyond, elsewhere

Marina Gržinić

When we think of public critical access to the internet and new media platforms, histories on this topic are mainly the stories of Western Europe and the first capitalist world; the other half of Europe, the now defunct former Eastern Europe, is rarely mentioned.[1] The last case is 1) the publication CYBERFEMINISM INDEX (see Seu, forthcoming [2022]; see also Cyberfeminism Index, n.d.), (2) the exhibition WETWARE: Cyberfeminism Index benefit sale, and (3) WETWARE opening event on Twitter Space (see @FeralFile, 2022). Mindy Seu though being informed, materials sent, omitted the second most important publication on the topic of the cyberfeminism conference that followed the pioneering one from documenta in Kassel in 1997 (see Sollfrank and Old Boys Network, 1998). In 1997 The Hybrid Workspace organized during the documenta X (the 10th edition of documenta) in Kassel was heavily based on the propositions and networking points developed by Nettime. Though documenta X in Kassel curated by Catherine David seriously reinvented the modernist setting of the documenta event with the hybrid workshops, it failed to invite important positions from former Eastern Europe. Nevertheless, documenta X hosted Makrolab by Marko Peljhan from Ljubljana (see Peljhan, 1997).

As mentioned above, in 1997 at documenta X in Kassel, Germany, the First Cyberfeminist International was held, organised by the Old Boys Network.[2] In 1999, in Maribor, Slovenia, as part of the Festival of Computer Arts, I co-edited with Adele Eisenstein the book *The Spectralization of Technology: From Elsewhere to Cyberfeminism and Back. Institutional Modes of the Cyberworld*, which involved cyberfeminists (Cornelia Sollfrank, Claudia Reiche, Eva Ursprung, Kathy Rae Huffman, Margarete Jahrmann, etc.). This was a commitment to feminism, technology, an entanglement of technology and gender, a revolt against digital patriarchy and capitalism discrimination not only in the real but as well in the virtual space. From that time capitalism, technology and gender have changed radically, the regime of Whiteness made many things exclusive and only for White women in the Occident; today, all of these notions and conditions are seized with persistently violent relations of power, management, etc. Both my artistic endeavours and my theoretical work are about life and death, about what it means to be a human being in the time of a hyper-digitalization, quantum computers, transhumanist discourses.

[1] Some of the following theses I present here were already developed in my article 'New media/internet platforms in former Yugoslavia, history and present' (see Gržinić, 2015).
[2] Old Boys Network (OBN) is considered the first international cyberfeminist alliance and was founded in Berlin in 1997. Since its beginnings, the network has constantly changed through changing members. OBN is a real and a virtual alliance of cyberfeminists. Under the term 'cyberfeminism', OBN contributes to the critical discourse on new media, especially with regard to its gender-specific aspects (OBN, n.d.).



Publication produced on the occasion of The First Cyberfeminist International (1997) and on the edited collection The Spectralization of Technology: From Elsewhere to Cyberfeminsm and Back. Institutional modes of the Cyberworld, Marina Gržinić (ed.), in collaboration with Adele Eisenstein, MKC Project, Festival of Computer Arts, Maribor, Slovenia 1999. Reference https://dimitragkitsa.net/projects/public-events/legacies-of-cyberfeminism-april-july-2016/

The reason for the omission of the former Eastern Europe and the former Yugoslavia could be the perception that stays glued on us and that is formulated as a state of backwardness, namely that the former Eastern Europe in the 1990s was in a transition from socialism (without democracy, freedom and etc.) to the 'happy' new world of capitalism. But let me note that in 1994/1995, when a fully commercial use of the internet began globally, the former Eastern Europeans or former Yugoslavs were shaped and changed by the way various subcultural groups and artists working with videos and experimental film could easily engage with the world through net art and activism, while leaving behind the outdated world of conservative modernist art. And more so, that despite the euphoria of Western multiculturalism, cyberfeminists, net activists, etc., were able to pinpoint the emergence of the harsh digital divide, processes of exclusion, control, and hegemony that will hit us all globally in the 21st century.

Much can be learned from the Balkan wars in the 1990s in the former Yugoslavia. Due to the war in the former Yugoslavia (which began in 1991), activists throughout the former Yugoslavia were forced to use new technology and its modes of networking to communicate, in particular the bulletin board system, or BBS, which is considered the forerunner of the World Wide Web[3]; BBS was clearly used

[3] BBS is a computer or application used for sharing or exchanging messages or files on a network. Originally, BBS was an electronic version of the noticeboard that hangs on the wall in many kitchens and workplaces, and was used to send simple

here for political reasons. In the context of critical internet platforms and digital activism, many groups and individuals came from subculture and alternative contexts in the former Yugoslavia, being affiliated with music, theory, critical discourses and video and independent publishing efforts. The event that decisively marked shaped the building of social networks and embracing the possibilities of internet in the former Yugoslavia while organizing a physical coming together of dozens of activists, artists, theorists, publishers and hackers from the former Yugoslavia and globally, was the conference with the title 'Beauty and the East' in Ljubljana in 1997.[4] We gathered in Ljubljana through an organizational effort of various positions, among which the Nettime group (a digital platform and email list; see Nettime, n.d.), the representatives of the Soros Foundation in Ljubljana and especially those involved in its media branch called Ljudmila were central.[5] Nettime was founded in 1995 at a meeting of artists, theorists and media activists at the Venice Biennale, and is considered a pioneering platform for building discourses of net critique, providing context for the emergence of net art and critical net culture in general. The central figure of Nettime at the time was (and still is) the media theorist Geert Lovink, and Pit Schultz and Diana McCarty, among others, were also involved. On the Ljubljana side, a key figure in the organization of the conference was Vuk Ćosić.

Mr. Soros, the 'uncle from America' (as we called him when we talked about his almost obscene position of power in Eastern Europe in the 1990s), financed the conference. In addition to the Soros Foundation in Ljubljana, Press Now and V2 (Lab for the Unstable Media, an interdisciplinary centre for art and media technology in Rotterdam) also financially supported the Nettime Conference, according to Diana M. Carty. To mark the occasion, 10,000 copies of the Nettime book of selected texts from Zentral Kommitee Proceedings were printed under title *ZKP4* (see Schultz et al., 1997). The evening programme, which was widely publicised, included pirate radio, talks by Peter Lamborn Wilson and the Critical Art Ensemble, and an all-night party with various DJs at K4, a club in the city centre. The conference presented positions that, when I wrote about this remarkable event, I placed in two broad lines of thought and action in global capitalism at the time and after.

The first line, which I called the 'Scum of Society Matrix', referred mainly to the positioning of the so-called Western participants, proposed autonomy and a progressive political option: a new autonomous internet economy and new structures from the appropriation and restructuring of the so-called old ones. Proposed was a return to writing only (email boxes) as a possible counterculture intercommunication strategy, rather than just improving the internet, i.e., deleting images and abolishing the pushy internet software industry. Under the guise of such utopian mind, it was possible to find genuine strategies of struggle and action, not just simple reproduction through technology.

The second line, mostly used by the former Eastern Europeans, I called the 'Matrix of the Monsters'. And why? As Peter Lamborn Wilson, also known as Hakim Bey, aptly stated in his well-received lecture in Ljubljana in 1997: 'The second world was erased (we are lucky, aren't we?) and what is left is the First and Third worlds'.[6]

Was this not the situation in 2015? Bey continued: 'Instead of the Second world, there is a big hole from which one jumps into the Third'. I called this hole and the second direction the 'Matrix of the Monsters'. I also directly referred to the general title of the Nettime conference in Ljubljana 'Beauty and the East' (a title that already paraphrases the well-known fairy tale 'Beauty and the Beast').

---

text messages between users. It was usually accessed via a dial-up modem. Before the advent of the World Wide Web, in the 1980s and early 1990s, BBS was the main type of online community (Zydyk, 2005).

[4] 'Beauty and the East', Nettime Conference, 22–23 May 1997, Ljudmila, Ljubljana, Slovenia (see Ljudmila, n.d.).

[5] Ljudmila (Ljubljana Digital Media Lab, part of Open Society Institute – Slovenia).

[6] Talk by Peter Lamborn Wilson, 'Beauty and the East', Nettime Conference, 22 May 1997, club K4.

We, the non-existent Second World, posed serious political and analytical questions related to the key issue: Who are the new and old actors who will build the 'Brave New World?'

'The Monsters', as I formulated our positions in the former Yugoslavia and the former Eastern Europe (also referring to Donna Haraway's 1992 text 'The Promises of Monsters: A Regenerative Politics for Inappropriate/d Others'), insisted on difference, a critical difference within rather than as a special method of classification marking the process of justifying difference, such as apartheid. The questions raised in the discussions concerned the role of the Soros Foundation and capital in the former Eastern Europe, the relationship between the state and NGOs and, last but not least, the dialectic of interpretation – who, when and how is allowed to write about the (art, cultural, political) history of the once known Eastern Europe.

In Budapest, the Hungarian Academy of Fine Arts opened a media arts department in 1991, and a few years later, in 1996, the Centre for Culture and Communication (C3) was founded in Budapest by the Open Society Institute – now known as Open Society Foundations (OSF) – to support media artists. In 1995, Ljudmila joined Soros' influential media initiatives in the region. At that time, however, Ljudmila was not yet conceived as an art project. An important reason for this, as Bosma pointed out, was the impossibility of obtaining funding for art projects, or the unavailability of funds from the patron of many former Eastern European media labs, George Soros. The E-Lab was founded already in 1996 in Riga, Latvia. A similar process took place in the Russian context. If we mention only two net art pioneers from Russia, Alexei Shulgin and Olia Lialina, it is significant that Lialina, a pioneer of internet art, was first one of the organisers and later director of Cine Fantom, an experimental cinema club in Moscow founded by Lialina in 1995 together with Gleb Aleinikov, Andrei Silvestrov, Boris Ukhananov, Inna Kolosova and others (Gržinić, 2015).

During the war- and nationalism-ridden 1990s in Croatia, the magazine *ARKzine* (a hybrid magazine where politics, culture, theory and media come together) was a single example dealing with Central and Eastern European NGOs and alternative media interests. The first two exclusively multimedia cultural spaces in Croatia were the Multimedia Institute (Mi2), opened in 1999, and Net Club Mama, opened in 2000. MaMa – Multimedia Institute is a net culture club in Zagreb founded in 1999 by Tomislav Medak, Marcell Mars, Teodor Celakoski, Petar Milat, Željko Blaće, Vedran Gulin, Vanja Nikolić and others. Since its opening in 2000, MaMa has been a meeting place for various local and international communities – from political activists, media artists, electronic music makers, theorists, hackers and free software developers to gay and lesbian support groups and the anime community.

Much earlier, in 1993, the association APSOLUTNO was founded in Novi Sad, Yugoslavia. The production of the association came about through the collaboration of its four members, Zoran Pantelić, Dragan Rakić, Bojana Petrić (and Dragan Miletić from 1995 to 2001), and the association stood in stark contrast to the claustrophobic context of Serbia. Out of the association and in connection with APSOLUTNO, an independent organisation called New Media Center_kuda.org was founded in Novi Sad in the late 1990s. It brought together artists, theorists, media activists, researchers and the general public in the field of information and communication technologies.

The war that began in 1991 in the former Yugoslavia was not only literally announced on television, but also demonstrated the essential importance of new media technology for activists. The Slovenian war, which was the first in a long series of wars in the Balkans, began as a 'fucking' disturbing lesson in Jean Baudrillard's theory about the function of television in capitalism and its use in practise. I was

breastfeeding my son in 1991 when the prime-time newscaster declared that the war had begun in Slovenia.

Vesna Janković (2013), founding member of the Antiwar Campaign Croatia (ARK) and founder and editor of *ARKzine* (1992–1998), writes that in the autumn of 1991 the telephone lines between Croatia and Serbia were cut.

> Indeed, new communication technologies, as a distinct constitutive element of new transnational movements, played an important role in the post-Yugoslav context as well. Maintaining communication between anti-war groups was one of the priorities of peace work, but this was increasingly difficult from the autumn 1991 because of the damaged post and telephone lines between Serbia and Croatia. The reasons behind this were more political, than technical. At the time, the official media in both countries were doing their best to demonize the "enemy" side, so it was important to block any information which could counter the official "truth". Eric Bachman, long term peace activist, involved in post-Yugoslav anti-war organizing from the very beginning as a nonviolent trainer and supporter, suggested a (globally) new technological solution — establishment of local BBSes (Bulletin Board System) which would be connected through the server in Germany. At the beginning of 1992, the first modem was attached to the computer in Antiwar Campaign office and BBS ZaMir (For Peace) Zagreb was formed. ZaMir Belgrade and ZaMir Ljubljana followed, and a transnational electronic network connecting post-Yugoslav peace, women's and human rights groups was born. By 1996 ZaMir had seven nodes (Sarajevo, Pristina, Tuzla and Pakrac, in addition to those already mentioned) and several thousand users. It became a firmly established transnational communication space, linking regional activists with each other and the world. (Janković, 2013: 3–4)

Janković asks in 2013, when elaborating on the 1990s: 'Can we talk about emergence of new political subject and new political arena? What is the impact of new electronic communications on organizational forms and structures of transnational organizing? What is the impact of transnational activism on national politics? […] And last, but not least what (semi-)periphery [that means the former Yugoslavia space] has to say about it?' (Janković, 2013: 5)

The relationship of new forms of subjectivity to capital is a process of establishing a set of strategies for dismantling mechanisms of exploitation, racialisation and dispossession. Donna Haraway's (1991) seminal text, 'A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late Twentieth Century', first published in 1985, invented the paradigm of the cyborg. Haraway's cyborg theory, further expanded in the 1990s with Trinh T. Minh-ha's (1986) paradigm of the 'inappropriate/d Others', rearticulates the problematics of the material-semiotic (sex, gender, technology, democracy and emancipation) in our present. The result is new directions within feminism and their connection with new media technology and theory, politics and subjectivity in the era of global capitalism. In short, in the present, we need to re-articulate the relations between the First, Second and Third Worlds and the relations of technology to the subjects, objects and things we conceive as such.

What is that we have today? The situation of the refugees and the statuses of their lives and bodies in the camps in Europe, at the borders of the European Union, or as corpses recovered from the Mediterranean Sea cannot be described solely as unwanted death, or destiny. These situations of massive suffering, death, and misery are also connected to certain historical situations that both differ from and are in continuity with what we have here and now. In other words, what we are seeing is a

process of persistent dehumanization at the centre of the more and more post-human, prosthetic, digitalized global capitalism. The basic relationship in these processes is the relationship between death and life. This is not only connected with immanently philosophical questions such as 'What is life?' and 'What is death?' but is also, increasingly, concerned with the ways of governmentality of life and death, of strategies and techniques through which life and death are managed, run, controlled by the State, by governments and their institutions.

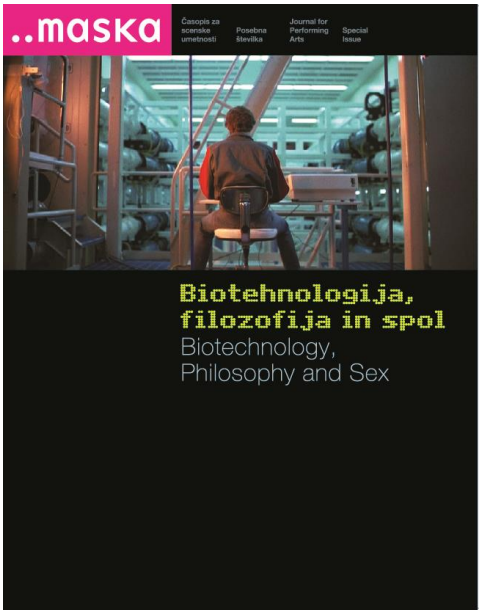
The all-female artist and activist collective VNS Matrix ('VeNuS' Matrix) in Adelaide, Australia, may have been the first to use the term with their 1991 (travelling) billboard manifesto, 'A Cyberfeminist Manifesto for the 21st Century' (see VNS Matrix, n.d.). In 1994, the 'Seduced and Abandoned: The Body in the Virtual World' conference in London saw cyberfeminism as a derivative of Haraway's cyborg feminism.[7] Ten years later, the ISEA2004 meeting in Helsinki celebrated the 10th anniversary of Cyberfeminist Theory.[8] Alla Mitrofanova, a critic and philosopher from St. Petersburg who works with media art, new ontologies and feminist philosophy, was also active in the cyberfeminist movement in the 1990s, as co-founder of the Cyberfemin Club in St. Petersburg (1994) and as a member of Cyberfeminist International. Other important names were Cornelia Sollfrank and Yvonne Volkart (1999).

In March 1999, the 'next cyberfeminist international' was held in Rotterdam, the Netherlands. It was the second major international meeting of cyberfeminists and came about thanks to the initiative of Corrine Petrus of TechWomen, Rotterdam. She had invited the Old Boys Network, the first international cyberfeminist organisation, to develop the concept for the conference, and provided local support (see Sollfrank and Old Boys Network, 1999). As mentioned above, in 1991, as part of the International Festival of Computer Arts (MFRU) in Maribor, Slovenia,[9] a conference was held and a book was published, co-edited with Adele Eisenstein: *The Spectralization of Technology*. At the conference in Maribor, we intensively questioned the politics of feminist, post-feminist and cyberfeminist positions from different geopolitical territories. The feminist movement in the 1970s in Belgrade and Zagreb at least marked a process of radical and avant-garde emancipation, but the cyberfeminist movements in Russia and Asia are waiting to be included in the official (Western and White) history of feminism and cyberfeminism, not to mention the important different perspectives of feminism and technology in African and African-American contexts. So, it is time to rewrite history and repoliticise technology and subjectivities.

[7] 'Seduced and Abandoned: The Body in the Virtual World', conference, Institute of Contemporary Arts (ICA), London, 12–13 March 1994.
[8] ISEA2004, 12th International Symposium of Electronic Arts, Helsinki, Tallinn and Baltic Sea, 14–22 August 2004.
[9] 5th MFRU, International Festival of Computer Arts 1999, Maribor, Slovenia, 9–15 May 1999.

Marina Gržinić and Adele Eisenstein (eds), *The Body Caught in the Intestines of the Computer and Beyond: Women's Strategies And/or Strategies by Women in Media, Art and Theory*, Maribor: MKC; Ljubljana: Maska, 2000

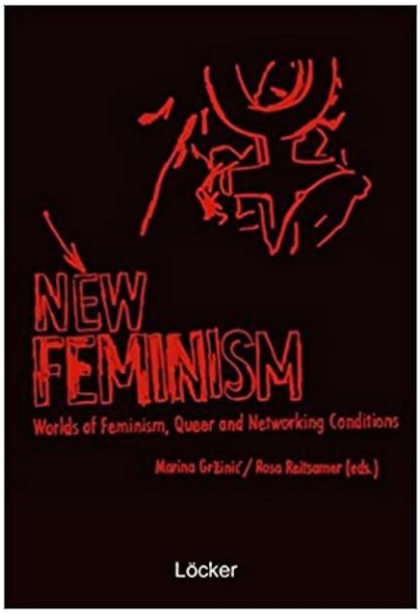
In 1999, I published an introduction entitled 'Kiborška politika in neprimerni / neprisvojeni drugi (Cyborg politics and inappropriate / unappropriated others) in Harraway's Slovenian translation of her book (see Gržinić, 1999) and in 2000 a co-edited publication of *The Body Caught in the Intestines of the Computer and Beyond: Women's Strategies And/or Strategies by Women in Media, Art and Theory* (see Gržinić and Eisenstein, 2000).



Marina Gržinić (ed), *Biotechnology, Philosophy and Sex*, special issue, *Maska*, no. 76–77, 2002

In 2002, a special issue of *Maska* entitled *Biotechnology, Philosophy and Sex* was published (see Gržinić, 2002), with its texts, which were also presented at an international conference,[10] enabling a renewed reflection on science and technology, bio-chips and organs, the male and the female (trans-sexuality) within a philosophical and interdisciplinary (artistic, cultural, political) context. Cloning and the different attitudes to globality have profound implications for the politics of representation, which focuses on the human body, developing systems and paradigms, structures and matrices of representation of bodies determined by history, social gender and class. Positions presented in the publication were: Marie-Luise Angerer (Germany), Caroline Bassett (Great Britain), Dominique Chateau (France), Sarah Franklin (Great Britain), Claudia Reiche (Germany), Marina Gržinić (Slovenia), Maria Klonaris/Katerina Thomadaki (France), Marie-José Mondzain (France), Karin Spaink (Netherlands), and Jackie Stacey (Great Britain).

In 2003 my book was published in Slovenian with the title *Estetika kibersveta in učinki derealizacije* (Esthetics of the Cyberworld and the Effects of De-realization) (see Gržinić, 2003); in 2005 it was translated into Croatian (see Gržinić, 2005a).



Marina Gržinić and Rosa Reitsamer (eds), *New Feminism: Worlds of Feminism, Queer and Networking Conditions*, Vienna: Löcker, 2007

In 2007, I co-authored with Rosa Reitsamer a book *New Feminism: Worlds of Feminism, Queer and Networking Conditions* (see Gržinić and Reitsamer, 2007), and authored *Re-Politicizing Art, Theory, Representation and New Media Technology* in 2008 (see Gržinić, 2008).

Important names and pioneer women are: Irina Aristarkhova, Nina Czgledy, Adele Eisenstein
Alla Mitrofanova, Boryana Rossa, Rasa Smite, Mare Tralla, Olia Lialin, etc.

[10] 'Biotechnology, Philosophy and Sex', international conference on trans-sexuality, new media technologies and gender, Slovenska kinoteka, Klub Cankarjev dom, Galerija Kapelica, Ljubljana, Slovenia, 10–12 October 2002.

In 2005 and 2006 under the editorial and production facilities of ESC im labor in Graz and its co-founder and producer Reni Hofmüller a [prologue] project started in Graz, Austria.

Two important numbers of the journal titled as well [prologue] were published in 2005 and 2006. I conceived and edited these two numbers under the support and overall editing of everybody included. Amazing pioneers of digital media and cyber worlds, artists, activists, theoreticians, producers. In the opening text, I wrote:

> Today every situated (that is, implicated, contaminated) art production and cultural initiative within the global world opens a space of collaboration, dedicated to life and resistance. The crucial point is to name the process and the new political subject. Furthermore, it is necessary to identify the ways in which this new subjectivity functions and what are the strategies implied in order to produce paradigms of resistance toward the Capital machine. Capital functions as the evacuation of spaces; it is a constant production of non-spaces. (Gržinić, 2005c, 1)

Other writers who engaged in the questions of [prologue] 1 project: 'about theory, activism, criticism, the visual and the political on new [Post] and [trans] feminist forces in the territory of [new] Europe' were Rosa Reitsamer and Jo Schmeiser, Suzana Milevska, Sol Haring, City of Women, Nataša Govedić, Vlatka Frketić, Seda Gurses, Eva Egermann and Emma Hedditch. As Hedditch (2005) writes, the journal gives us something to hold on to and refer to in these complicated discussions and experiences of feminist politics, performances and visual art works.

Prior to [prologue] 1, *in* 2004–2005, the project '[prologue] new feminism/ new Europe' emerged from the discussions held in 2004 at the Youth Cultural Centre Maribor (MKC Maribor), Slovenia, hosted by the director of MKC Maribor Dragica Marinič. The invited positions were: Reni Hofmueller (committed artist with a feminist background, Graz), Diana McCarty (member of Bootlab, Berlin), Kathy Rae Huffman (Director of Visual Arts, Cornerhouse, Manchester), Agnese Trocchi (member of Candida TV, Rome), Eva Ursprung (curator and artist for sound and free improvisation, Graz), Milena Deleva (InterSpace, Sofia) and Agnieszka Wolodzko (Laznia Centre for Contemporary Art, Gdansk), who were connected by e-mail, and me.

In 2005, [prologue], a symposium was held at ESC im labor in Graz (8 –10 May 2005), that featured an exhibition at ESC im labor entitled '[prologue] new feminism/new Europe' – which was featured also at the Cornerhouse, Manchester, UK (30 July –18 September 2005) – and a journal that served as an archive of current positions (see Gržinić, 2005b). In 2005 [prologue] exhibition (curated by Reni Hofmüller) introduced artists: Esther Straganza and Elke Auer, Senam Okudzeto, Doris Jauk-Hinz, Gertrude Moser-Wagner, Marina Gržinić and Aina Šmid, Nina Stuhldreher and Bonelli Bonelli.

[prologue] was a mutating platform for the formulation of a feminist aesthetic politics responding to the New Europe and how it affects women's lives – it was about reclaiming Europe from a gender perspective (Gržinić 2006, 1). 'Some of the women had already worked together, Diana McCarty and Kathy Rae Huffman with Vali Djordjevic started the women only mailing list FACES founded by in 1997. The online mailing list FACES was initiated in 1997 by women representing shifting political geographies: Kathy Rae Huffman (US citizen in Vienna that year), Diana McCarty (US citizen in Budapest at the time) and Valie Djordjevic (German citizen of Yugoslavian descendent, living in Berlin).[11] As Diana McCarty noted, this provided an important geographical link between East and

[11] The following part was published as a report of the symposium FACES: 20 Years, 13–15 October 2017 (see Gržinić, 2017).

West. The project was inaugurated at a Face Settings dinner in Vienna in 1997 (participants included Diana McCarty, Kathy Rae Huffman and Margarete Jahrmann).

The email list and platform bring together and connect media artists, activists, digital researchers and cyber entities who identify as women. In 2017, we celebrated FACES' 20th anniversary (13–15 October 2017) at Schaumbad – Freies Atelierhaus Graz. We (about 50 people) celebrated FACES with an exhibition, performances, screenings and a symposium: on art, gender and technology. The Graz meeting was a largely self-organised event, facilitated by Eva Ursprung, the director of Schaumbad and also one of the core members of FACES. Ursprung worked closely with another key figure and founder of FACES, Kathy Rae Huffman. Together they curated the exhibition and conceived the whole event with the support of three other key figures of FACES: Valie Djordjevic, Diana McCarty and Ushi Reiter.

FACES as a mailing list is not only a way of networking but also a digital method of documenting information, events, controversies; FACES calls for mobilisation. It enables for a certain conscious activation and empowerment or reconfiguration of access. Through FACES, important documents and facts about media art and cultural productions have been preserved, memorialised, and digitally catalogued. The mailing list consists of hundreds of online positions who declare themselves as women and are interested in being part of the network of women in media arts who connect in person and through the (non-public) FACES mailing list. The hosting of the online platform[12] and mailing list is done by servus.at[13] – a small, independent internet service provider from Linz, Austria, where Ushi Reiter has taken over the co-moderation of the list.

Inclusion in the list is thus determined by the declaration of a gender position that is not biologically conditioned, but is rather a moment of self-positioning. In a sense, it is determined by the definition of a certain political and also pragmatic condition and thus opens up a debate about what art and media practises by and for women are, what gender means as a constructed category and how we establish a relation to feminism as a political movement with a long history. It has become clear that the umbrella term 'woman' is not just a women's issue or a sexual or gender issue, and that gender is not binary, nor is it sex free of social and cultural construction. One of the questions that immediately arose was the discussion of feminism, its White lineage and Western history.

One possible response to the conflicting histories and agencies of feminism is the analysis of activist and theorist Luzenir Caixeta, who worked at and with maiz – Autonomous Centre by and for Migrant Women[14] – a self-organised association in Linz, Austria, founded in 1994, to help, educate and advise migrant women in Austria. Following another position, that of Beatriz Preciado, a transgender theorist who has since transitioned and is now Paul Preciado, Caixeta (2013) speaks of minoritised women, including immigrants, transgender people, sex workers, lesbians, etc., all bringing about a transformation in and of feminism.

In her 2011 essay, 'Minoritized Women Effect a Transformation in Feminism', reprinted in 2013, Luzenir Caixeta (2013) looks at dissident movements within feminism that transform its White, heterosexual, essentialised perceptions (based on characteristics seen as natural elements of a category called 'woman') into dissident feminisms (i.e., feminism in plural!). Or we can speak of the proletariat of feminisms to address today's terminal relations between labour and capital, governed by hyper-

[12] See FACES website at: https://www.faces-l.net/.
[13] See servus.at website at: https://core.servus.at/.
[14] See maiz website at: https://www.maiz.at/en.

precarity, and the constant exploitation of those considered non-citizens or second-class citizens under neoliberal global capitalism. The proletariat of feminisms is clearly related to the relationship between labour and capital under neoliberalism, which imposes the dictates of a constant underfunding and persistent marginalisation.

The Graz meeting showed that the members of FACES are by no means only the privileged White Western majority women who come exclusively from the centre of the imperial neoliberal capitalist world, but that we also come from the margins of the forgotten East of Europe as well as from the global South. Geopolitics has a strong impact on FACES.

This is to emphasise that the many histories of feminisms and media works by women are shaped along a geopolitical differentiation that runs on both sides of the so-called colonial/racial divide of neoliberal global capitalism and is underpinned by class differences. At the end of the event at the plenary session of FACES on 15 October 2017, we also discussed the old maladies of the European and global system of institutional, structural power, which again tried to marginalise the event greatly. Although FACES has grown through networking since its beginnings in 1997, the 20th anniversary celebration was underfunded and its historical significance was pressured to become less important (Gržinić, 2017).

In recent years, a number of authors have become known around the world who believe that the new feminism must go much further than the old demands of White, Western and heterosexual middle-class women for legal equality. Attention should be paid to women who have always been marginalised and the causes that led to differentiation based on class, ethnicity and gender should be fought (Caixeta, 2013: 146).[15]

Moreover, Caixeta straightforwardly subtitles a section in her above-mentioned text, 'Dissident Currents within Feminisms' and states that this section refers directly to the essay by the Spanish philosopher Paul Beatriz Preciado, 'Reportaje después del Feminismo. Mujeres en los márgenes' (Report after Feminism: Women on the Margins) (2007). Caixeta, refering to Preciado, argues that modern feminism, in opposition to an earlier feminism that developed its political discourse on the basis of a division 'between men (as dominators) and women (as victims) […] is developing new political concepts and strategies for action that call into question what has previously been regarded as generally true: namely that the political subject of feminism [was] women — meaning women in their predefined biological reality, but especially women according to a certain notion: white, heterosexual, submissive and from the middle class' (Caixeta, 2013: 146). The dissident demand in feminism thus calls for a process of radical differentiation. Preciado calls for 'feminisms for the excluded' (Preciado, 2007, quoted in Caixeta, 2013: 147). Or, as s/he argues in 'Pharmaco-Pornographic CapitalismPostporn Politics and the Decolonization of Sexual Representations', these new feminisms for the excluded are dissident projects for the 'collective transformation of the 21st century' (Preciado, 2013: 251). Dissident feminisms stand in opposition 'to a gray, normed and puritanical feminism, which sees in cultural, sexual or political distinctions a threat to its heterosexual and Eurocentric image of women' (Caixeta 2013: 146).

I emphasise that Preciado speaks of the 'proletariat of feminism' – a neologism s/he uses in reference to the writer and filmmaker Virginie Despentes – which includes all those 'monstrous bodies' excluded by puritanical Western feminism. These are subjects such as 'whores, lesbians, raped, butch, male and transgender women who are not white … in short, almost all of us' (Preciado, 2013: 251). Preciado

[15] I have developed some of the following thoughts in 'Dissident Feminisms, Anti-racist Politics and Artistic Interventionist Practices' (see Gržinić, 2014).

lists names and outlines a genealogy of positions that question the naturalness and universality of the feminine condition. I quote from Preciado's text 'Pharmaco-Pornographic CapitalismPostporn Politics and the Decolonization of Sexual Representations' in order to propose the genealogy of dissident feminisms:

> The first of these shifts is in the hands of theoretical gay and lesbian theorists like Guy Hocquenghem, Michel Foucault, Monique Wittig, Michael Warner and Adrienne Rich, who define heterosexuality as a political regime and a control device that makes the difference between men and women and transforms the resistance to gender standardization into pathology. Judith Butler and Judith Halberstam insist on the processes of cultural significance and stylization of the body through which the normalization of differences between the genders is effectuated, while Donna Haraway and Anne Fausto-Sterling bring into question the existence of two sexes as biological realities, regardless of the scientific-technical processes representation is constructed with. Moreover, along with the processes of emancipation of blacks in the United States and the decolonization of the so-called "Third World," the voices of criticism are also raised against the racist assumptions of colonial and white feminism. We have become empowered with projects and thoughts by Angela Davis, bell hooks, Gloria Anzaldúa and Gayatri Spivak, and black feminist, postcolonial, postChristian, postJewish, postMuslim projects, or those from the Diaspora that will require thinking gender in its constitutive relation to the geopolitical differences of race, class, sexuality, migration and human trafficking (Preciado, 2013: 251).

The first line of dissident feminist artistic interventions is characterised by those positions that escape 'from the academia to flourish in audiovisual production, literary, and performative action spaces' (Preciado, 2013: 252). As such, the goal of these queer-feminist projects is not so much the liberation of women or legal equality, 'but the dismantling of the political devices producing differences of class, race, gender and sexuality, thereby creating transfeminism art and action networks for decolonization politics' (Preciado, 2013: 252). The artists and their interventionist practises in the field of dissident feminisms seen in such a view are those associated with postporn strategies. As Preciado notes, we have a history going from the

> feminist kitsch porn movies of Annie Sprinkle, the docu-fiction by Monika Treut, literary works by Virginie Despentes, Dorothy Allison and Michelle Tea, Alison Bechdel's lesbian comics, photographs by Del LaGrace Volcano, dyke punk rock concerts by Tribe8, [to] the Gothic Revival sermons of Lydia Lunch, and the transgender porn science fiction of Shue-Lea Cheang, feminist postporn […] aesthetics made of traffic signs and cultural artifacts with the critical redefinition of code standards that traditional feminism considered improper for femininity. Some of the references of this aesthetic and political discourse are queer-horror-porn movies of Bruce La Bruce and Christopher Lee, the distortion of sex shops of Maria Llopis, PostOp or Orgy, Queer-Gothic literature, dildos as instruments of tectonic redefinition of the sexed body, trans-sexual vampires and monsters, cyborgs living queer punk, performance in public space as a useful political intervention … Sex was never so crude and gender has never been so cooked. (Preciado, 2013: 252)

The second line of dissident feminist artistic interventions that I would like to propose is even more closely linked to non-White feminists and activists working directly with anti-racism and post-colonial and decolonial positions. I refer here to names that Caixeta mentions as critical positions of migrant and Black women, such as Katharina Oguntoye, May Ayim, FeMigra, Lale Otyakmaz and Encarnación Gutiérrez Rodríguez. An important frame of reference for this development is the critique put forward by Women of Colour in the USA in the 1980s, including Combahee River Collective, Cherríe L. Moraga, Gloria Anzaldúa, bell hooks, and Angela Davis. And I add to these names María Lugones, Gayatri Spivak and Chandra Talpade Mohanty. Last but not least, artists in Austria who work in maiz, especially Marissa Lôbo, as well as The Research Group on Black Austrian History and Presence /

Pamoja from Vienna and its members Araba Evelyn Johnston-Arthur, Belinda Kazeem and Njideka Stephanie Iroh, to name but a few (see Gržinić, 2014).

The third line remains on the path of Liliana Conlisk Galegos and fantastic positions as Tjaša Kancler, Lucrecia Masson Córdoba, Esther (Mayoko) Ortega Arjonilla, YeLa An, Mirjam Wilhelm, Mika Maruyama, iki yos piña narváez funes, EsRap, Asma Aiad, Jennifer Ndidi Iroh, Jovita Pristovšek, etc.

These positions are important because they effectuate direct interventions in mainstream art production (co-opted and fully branded by the market) and in processes of dismantling interlinked capitalism and racism and Western occidental knowledge with White hegemonic social and institutional regimes such as universities, museums, etc.

A very good example of the importance of such work is the life and poetry of May Ayim (1960–1996), an Afro-German poet, educator, and activist. Margaret MacCarroll suggests in her 2005 MA thesis 'May Ayim: A Woman in the Margin of German Society' that 'Although there is a long history of dark-skinned people living in Germany, this study focuses primarily on the period after World War II and examines concepts of culture, race and ethnicity in order to determine what role these concepts play in the experiences of Afro-Germans like Ayim' (MacCarroll, 2005: 1). MacCarroll suggests that Ayim's life was marked by a sense of displacement and non-belonging as she desperately tried to find her place in German and African society. Thus, Ayim's tragic life and powerful art cannot be understood outside the genealogy of racism in Germany, which, as MacCarroll points out, works from '*Negerhuren* to *Mischlingskinder* to *Afro-Deutsche*' (MacCarroll, 2005: 3).

What is important for these and other dissident positions is that they cannot be confined to one artistic realm. In order to grasp their significance and the way they radically intervene in art, we must firstly dismantle a standard division of art disciplines and secondly always take into account a broad social, political and economic context of art.

References:

@FeralFile (2022), 'Starting now! WETWARE Opening Event @FeralFile for #cyberfeminismindex', *Twitter*, 15 Jun 2022. Available online: https://mobile.twitter.com/FeralFile/status/1537086890029072386 (accessed 22 August 2022).

Caixeta, L. (2013), 'Minoritized Women Effect a Transformation in Feminism', in Editorial Group for Writing Insurgent Genealogies (C. Agredo, S. Avraham, A. Cannito, M. Gerothanasis, M. Gržinić, M. Lôbo, and I. Marjanović) (eds), *Utopia of Alliances, Conditions of Impossibilities and the Vocabulary of Decoloniality*, 145–148, Vienna: Löcker.

Cyberfeminism Index (n.d.), *Cyberfeminism Index*. Available online: https://cyberfeminismindex.com/ (accessed 22 August 2022).

Gržinić, M. (1999), Kiborška politika in neprimerni / neprisvojeni drugi [Cyborg politics and inappropriate / unappropriated others], in D. J. Haraway, *Opice, kiborgi in ženske: reinvencija narave* [Apes, Cyborgs and Women: Reinventing Nature], 431–461, Ljubljana: ŠOU, Študentska založba.

Gržinić, M., ed (2002), *Biotechnology, Philosophy and Sex*, Special issue, *Maska*, no. 76–77.

Gržinić, M. (2003), *Estetika kibersveta in učinki derealizacije* [Esthetics of the Cyberworld and the Effects of De-realization], Ljubljana: Založba ZRC, ZRC SAZU.

Gržinić, M. (2005a), *Estetika kibersvijeta i učinci derealizacije* [Esthetics of the Cyberworld and the Effects of De-realization], Zagreb: Multimedijalni institut; Sarajevo: Košnica – centar za komunikaciju i kulturu.

Gržinić, M., ed (2005b), *Reclaiming Europe from a New Feminist Perspective: Theory, Activism, Criticism, [prologue]*, 1.

Gržinić, M. (2005c), 'Symposium, exhibition, screenings, journal', *[prologue]*, 1: 1.

Gržinić, M., ed (2006), *Performative and (Trans)Gender Politics & Migratory Bodies, [prologue]*, 2.

Gržinić, M. (2008), *Re-Politicizing Art, Theory, Representation and New Media Technology*, Publications of the Academy of Fine Arts Vienna, Vol. 6, Vienna: Schlebrügge.

Gržinić, M. (2014), 'Dissident Feminisms, Anti-racist Politics and Artistic Interventionist Practices', *p/art/icipate – Kultur aktiv gestalten*, 4. Available online: https://www.p-art-icipate.net/dissident-feminisms-anti-racist-politics-and-artistic-interventionist-practices/ (accessed 22 August 2022).

Gržinić, M. (2015), 'New media/internet platforms in former Yugoslavia, history and present', *Springerin*, 2. Available online: https://www.springerin.at/en/2015/2/monster-einst-und-jetzt/ (accessed 22 August 2022).

Gržinić, M. (2017), 'FACES: 20 Years of Interactions, Connections and Collaborations', Faces, 1 November 2017. Available online: https://www.faces-l.net/index.php/2017/11/01/faces-20-years-of-interactions-connections-and-collaborations/ (accessed 22 August 2022).

Gržinić, M., and Eisenstein, A., eds (1999), *The Spectralization of Technology: From Elsewhere to Cyberfeminism and Back. Institutional Modes of the Cyberworld*, Maribor: MKC.

Gržinić, M., and Eisenstein, A., eds (2000), *The Body Caught in the Intestines of the Computer and Beyond: Women's Strategies And/or Strategies by Women in Media, Art and Theory*, Maribor: MKC; Ljubljana: Maska.

Gržinić, M., and Reitsamer, R. (2007), *New Feminism: Worlds of Feminism, Queer and Networking Conditions*, Vienna: Löcker.

Haraway, D. (1991), *Simians, Cyborgs and Women: The Reinvention of Nature*, New York: Routledge.

Haraway, D. (1992), 'The Promises of Monsters: A Regenerative Politics for Inappropriate/d Others', in L. Grosberg, C. Nelson, and P. Treichler (eds), *Cultural Studies*, 295–337, New York: Routledge.

Hedditch, E. (2005), 'Locative Feminism'. *Mute*, 31 August 2005. Available online: https://www.metamute.org/editorial/articles/locative-feminism (accessed 22 August 2022).

Janković, V. (2013), ‘Post-Yugoslav anti-war movements as an early example of new transnational agency’, presentation at the International conference ‘EUROPE ON THE WAY: From solidarity to the European integration of the Balkans and Turkey’, Bergamo, 16 November 2013. Available online: https://www.academia.edu/5136376/Post_Yugoslav_anti_war_movements_as_an_early_example_of_new_transnational_agency (accessed 22 August 2022).

Ljudmila (n.d.), ‘nettime may meeting Beauty and the East’. Available at:
See: http://www.ljudmila.org/nettime/ (accessed 22 August 2022).

MacCarroll, M. (2005), ‘May Ayim: A Woman in the Margin of German Society’, MA diss., College of Arts and Sciences, Florida State University, Tallahassee.

Nettime (n.d.), ‘nettime mailing list’. Available online: https://nettime.org/ (accessed 22 August 2022).

OBN (n.d.), ‘FAQ___Frequently Asked Questions’. Available online:
https://obn.org/obn/inhalt_index.html (accessed 22 August 2022).

Peljhan, M. (1997), ‘INSULATION/ISOLATION PROCEEDINGS’, lecture for the 100 days-100 guests dX programme, 31 August 1997. Available online: http://ladomir.net/documenta-X-lecture-1997 (accessed 22 August 2022).

Preciado, P. B. (2007), ‘Reportaje después del Feminismo. Mujeres en los márgenes’ [Report after Feminism: Women on the Margins], *El País*, 13 January 2007. Available online:
https://elpais.com/diario/2007/01/13/babelia/1168648750_850215.html (accessed 22 August 2022).

Preciado, B. (2013), ‘Pharmaco-Pornographic CapitalismPostporn Politics and the Decolonization of Sexual Representations’, in Editorial Group for Writing Insurgent Genealogies (C. Agredo, S. Avraham, A. Cannito, M. Gerothanasis, M. Gržinić, M. Lôbo, and I. Marjanović) (eds), *Utopia of Alliances, Conditions of Impossibilities and the Vocabulary of Decoloniality*, 245–255,
Vienna: Löcker.

Schultz, P., McCarty, D., Lovink, G., and Ćosić, V., eds (1997), *ZKP4*, Ljubljana: Ljubljana Digital Media Lab.

Seu, M., ed (forthcoming [2022]), *Cyberfeminism Index*, Los Angeles, CA: Inventory Press.

Sollfrank, C. and Old Boys Network, eds (1998), *First Cyberfeminist International*, Hamburg: obn.

Sollfrank, C. and Old Boys Network, eds (1999), *next cyberfeminist international*, Hamburg: obn.

Sollfrank, C., and Volkart, Y. (1999), ‘Editorial’, in Sollfrank, C., and Old Boys Network (eds), *next cyberfeminist international*, 4–5, Hamburg: obn.

Trinh T. M-H. (1986), ‘She, the Inappropriate/d Others’, *Discourse*, 8: 86–87.

VNS Matrix (n.d.), ‘The Cyberfeminist Manifesto for the 21st Century’. Available online:
https://vnsmatrix.net/projects/the-cyberfeminist-manifesto-for-the-21st-century (accessed 22 August 2022).

Zydyk, M. (2005), ‘bulletin board system (BBS)’. Available online: https://www.techtarget.com/whatis/definition/bulletin-board-system-BBS (accessed 22 August 2022).

Reading Agre

Computer Supported Cooperative Work (CSCW) **3:** 167–195, 1995.

From High Tech to Human Tech: Empowerment, Measurement, and Social Studies of Computing

PHILIP E. AGRE
Department of Communication, University of California, San Diego La Jolla, California 92093–0503, U.S.A.
email: pagre@weber.ucsd.edu

(Received 27 July 1994; in final form 26 April 1995)

Abstract. "Empowerment" has become a pervasive term of art in business practice, particularly in the United States. The term traces its roots to the organizing models evolved by populist social movements, but within business discourse it refers to an emerging organizational philosophy that largely replaces conventional hierarchies with nominally autonomous teams. Proponents of empowerment frequently cite information technology as a crucial enabler of this shift without, however, spelling out fully the logic of the connection. A reconstruction of this logic provides evidence for the emergence of a novel vision of work-discipline, the empowerment and measurement regime. This regime is discussed in relation to market dynamics, Taylorism, and research on the social organization of information technology and its use.

Key words: Business discourse, capture, colonization, empowerment, grammars of action, taylorism, ideology, work measurement, work discipline

> "We are seeking high achievers. And high achievers love to be measured, because otherwise they can't prove to themselves that they are achieving. Measuring them says you care about them."
>
> – Robert Noyce of Intel, cited by Quinn (1992, p. 273).

1. Introduction

My title derives from a radio advertisement for a German luxury car. Its full slogan was: "The eighties were the decade of high tech; the nineties will be the decade of human tech". The phrase "human tech" obviously plays on a complicated set of symbols, beginning with the conventional opposition between human things and technical things. The oxymoronic tinge of "human tech" promises innovation at the meeting-point between two worlds. The radio ad was promoting what would traditionally have been called the ergonomics of the car: its ride, noise levels, control panels, and so forth. But the ad did not speak in traditionally ergonomic terms. The social project of ergonomics, particularly in its Cold War heyday, had been to insert human beings into a larger technical order (Edwards, 1995; Noble, 1989). As part of this project, people became objects of measurement in an extraordinary variety of ways. In earlier days, Lillian Gilbreth (1923) could operationalize the notion of human "individuality" by enumerating various human faculties, mea-

suring them for each person, and then matching that person to a job that requires all of those capacities and no more. Later on, the discourse grew in ambition and generality: everything was treated as a "system" and measurements defined people as components of that system (Lilienfeld, 1978). The automobile advertisement, by contrast, undertakes an inversion of this ideology. Where ergonomics sought ways to fit people into the needs of an overall system, this newer discourse speaks of fitting machinery to the needs of people (Norman, 1993). And where ergonomics disaggregates the human individual into a set of functional contributions to a larger system, now the individual acquires a curious opacity as the locus of a holistically understood "life". Technology is recoded as marketing and the instrumental individual becomes the "empowered" individual.

"Empowerment" is a pervasive term of art in American business at the moment. It indicates a complex set of practices through which workers are (it is said) liberated from conventional bureaucratic controls, thereby unleashed to do the job the (continually evolving) "right" way rather than the (once and for all) rationally prescribed way. While "empowerment", like all business keywords, is often employed in a superficial and even cynical way by managers who need to follow fashion, it also sometimes accompanies genuine changes in the material organization of work – as well as changes in the forms of subjectivity that companies attempt to instil in their workers. How this works out in practice is an important question. My own project here, though, is more modest. I propose to trace the outlines of a particular business vision of human relationships to technology. I refer to this vision as the "empowerment and measurement regime" because it couples the language of empowerment with a prescription for extensive measurement of work activities. The basic idea is that a decentralization of operational decision-making is coupled with an intensification of centralized control through measurement. On another level, I also wish to paint the economic and political context of the new business interest in the social dimensions of technology. Although my analysis may have broad application, I will place special emphasis on critically motivated studies of human-computer interaction. The body of research in this area is growing rapidly, and much of it is fueled by explicitly humanistic concerns. Yet the indisputably good intentions behind this work synchronize all too comfortably with some problematic reconceptions of human beings.

I will proceed as follows. Section 2 will sketch the business discourse around the emerging empowered individual, comparing and contrasting it with discourses of empowerment found in progressive populism and in conservative policy proposals. Section 3 presents some of the historical logic of manufacturing and marketing that provides the material background to the rise of empowerment in business discourse. Empowerment is part of a larger discursive construction of work-organization that I will refer to roughly as "empowered work". Empowered work turns out to stand in a complex relationship to the tradition of rationalized work based on time and motion studies, on ergonomics generally, and on the larger construction of people as technical components. Section 4 introduces the connection between

empowerment and measurement, outlining the largely unarticulated connection between them in contemporary management literature. Section 5 will introduce the connection between the empowerment and information technology by considering a range of opinions on the subject in the organizational literature. I will argue that basic assumptions in this literature are effectively being rethought in the emerging discourse. Section 6 will discuss the connection between empowered work and the process of "colonization" that large-scale distributed computing has made possible, particularly through the routine "capture" of work activities. Section 7 will conclude by drawing a lengthy comparison between the empowerment and measurement regime and certain critical literature on computing. The sudden popularity of these social studies of technology may reflect their capacity for legitimation as much as their capacity for critique. The point, though, is not to abandon them but to place the phenomena they describe in a fuller historical context.

Before setting off, I should make clear the limitations of this study. Since this is a study of business discourse and its material context, my argument will reflect an unreasonably clean discursive distinction between the constructs of rationalized and empowered work. The discourse of empowered work does not entail an end to either the discourse or the practice of rationalized work; rather in actual practice one will encounter, at best, complex transitions from one to the other. Furthermore, I will largely neglect the substantial variations in work organization and management culture among countries (Guillén, 1994), industries, and levels of professionalization, as well as the dynamics through which management innovations spread through specific organizational fields (Fligstein, 1990, p. 303).

2. The empowered individual

The term "empowerment" first arose among 1960's-era populist activists who sought a decentralized, participatory model of politics. Empowerment, for these people, is an experience rooted in locally organized political action. The term packs a remarkably complex system of meanings. To "empower" someone is to confer power on them, but in a specific sense of the term "power". One does not become empowered through the acquisition of money or status or skills, or even through the removal of external obstacles, but rather through a change in subjectivity. In realizing one's own innate capacity to change the world, one liberates a range of intrinsic but possibly dormant potentials. As such, the concept of empowerment presupposes a particular kind of equality among human beings: everybody has such potentials, simply for being human.

For the theorists of political populism, empowerment is not simply an individual psychological change, but is inseparable from political involvement on a community level that aims at gaining some control over what happens in one's life. One model (Riessman, 1986) traces a trajectory from the rise of mutual-aid groups, to the development of an advocacy focus as attempts at self-help raise awareness of local systems of domination, to an appreciation of the interconnection of issues

and the need for coalitions. Similar models have been applied in a wide variety of local political projects, including local economic development (Friedmann, 1992), social work (Mullender and Ward, 1991), workplace democracy (Bachrach and Botwinick, 1992), and community-based environmental movements (Szasz, 1994).

These movements, being rooted in local conditions and experiences rather than in theories, are often not very ideological. As such, as Riessman (1986) points out, the populist notion of empowerment has conservative potentials that can emerge as a generalized attack on government. Conservative theorists housed in policy institutes such as the Heritage Foundation have lately fashioned a distinct notion of "empowerment" which is wholly economic in character (Butler, 1990). Specifically, empowerment means participation in free markets; government programs, even voluntary programs ostensibly aimed at alleviating hardship, are viewed as exogenous constraints whose removal is *ipso facto* liberatory. It is proposed, for example, that government-sponsored housing projects be replaced with "housing vouchers". Theorists in this movement tend to minimize the social problems that have motivated government intervention, and they define the obstacles to social advancement purely in terms of individuals' relationships to the government.

The business use of the word "empowerment" does not coincide with either the populist or conservative uses. Instead, it refers to a process by which employees are freed of bureaucratic constraints and given control of their work in order to make decisions and reorganize their local work-processes in accord with their own judgement. The theme of empowerment is frequently bound up with the theme of employee "teams" (Wellins, Byham, and Wilson, 1991). Based on a tradition of organizational research on small groups (Lewin, 1951; Weisbord, 1991) and upon on observations of self-directed work groups by the founders of sociotechnical systems (STS) design (Trist, 1981), these teams are said to be empowered to organize work through consensus rather than through the imposition of hierarchical control. The idea with both empowerment and teams is that work activities cannot be wholly routinized, and the people who actually do the work are closest to it and in consequence are in the best position to judge how it should be done.

We can articulate precisely the ethical appeal of the word "empowerment". Inasmuch as "empower" is a transitive verb, to "empower" someone is to perform some action upon them. The liberal ideal of individual self-determination would normally object to this kind of operation. People are supposed to be able to define themselves and to choose their own identities and desires and intentions. And this ideal is normally violated when people show up as the objects of transformative verbs. But it is the special claim of empowerment to escape this objection. The person upon whom this action is performed, having been "empowered", is, by definition, in a position to take actions of his or her own. Indeed one might say that empowerment brings a proper, fully drawn human being into existence. Empowerment thus presents itself not as some kind of programming, but precisely as the removal of any susceptibility to programming.

It is striking that such a concept should find itself equally at home in both radical and business discourses. The common element is a claim about the relationship between human nature and particular forms of social order. In each case, a normative form of human organization is held to both presuppose and encourage a perfect clarity of individual consciousness through participation in large-scale collective action. The difference between the radical and business versions of empowerment, of course, lies in a contrast of social ontologies. Whereas the radical individual develops as part of a collective whole with converging interests and equally shared participation, the business individual aligns wholly personal interests with specific niches in an endlessly changing order. One is empowered to perform ("own") one's own job, albeit as part of a "team", and not to participate in choosing the larger purposes that the job serves. Clement (1994) expresses this point through his distinction between two conceptions of empowerment: functional empowerment, for which overall values and goals are established from the top down, and democratic empowerment, in which employees participate in all levels of decision-making.

Perhaps the most sophisticated theory of empowerment within the business literature is due to Block (1987). Directed principally at managers rather than line employees, this theory opposes empowerment to dependency. Block defines dependency as a subjective stance of placing all responsibility for organizational well-being on the managerial hierarchy and treating one's own situation as wholly determined by outside forces, thereby justifying a strategy of conformity, passivity, and manipulation. The empowered manager, by contrast, takes personal responsibility for the organization's success, articulates a vision of greatness for his or her own function, and maintains a rigorous policy of congruence between inner experience and outward action (Block, 1981). This deeply individualistic conception of personal accountability and integrity is similar to those found in a variety of personal growth organizations, from Alcoholics Anonymous to the Human Potential Movement, and it contrasts strikingly with the rational decision-making procedures that have characterized generations of prescriptive business thought.

The concept of empowerment, of course, has a genealogy within business discourse that is not wholly reducible to business appropriations of radical and countercultural language. Waring (1991), for example, traces a tradition of business "corporatism", according to which organizational conflicts can be overcome by treating employees decently and building a collective identity; representatives of this tradition include Mayo (1933), Maslow (1954), and McGregor (1960). More specifically, Rose (1991, pp. 80–108) traces the rise in 1950's Britain of notions of workplace democracy based on group psychological studies and attention to the social and psychological benefits of work. The next couple of decades saw these notions overtaken by a wave of personnel management techniques which conceived workers as essentially economic actors. Yet the emphasis shortly returned to work as a source of personal fulfillment with the rise of the Quality of Working Life movement. This emphasis persists under a variety of rubrics.

> No longer is there to be an antithesis between the motives of the individual and those of the organization. The citizen, at work as much as outside it, is engaged in a project to shape his or her life as an autonomous individual driven by motives of self-fulfillment. Individuals produce themselves in work; the organizational culture is to be reshaped in the name of a new psychological image of man (1991, p. 115).

The business notion of empowerment, then, is continuous with a larger history of discourses and "technologies of the self" (Foucault, 1988) that draws upon a variety of sources. Yet the concept of personal fulfillment through work that Rose describes is not yet identical with the concept of empowerment. Fulfillment is a psychological and perhaps spiritual goal; what remains is to specify the relationship between this goal and the actual material organization of work. The concept of empowerment elaborates the story by suggesting that workers or their teams will find fulfillment specifically in taking a significant degree of operational control over their work. The relatively specific sociotechnical systems theory of self-managed teams was now generalized and extended: empowerment is to be understood as a general trend toward the outward and downward delegation of decision-making authority in organizations.

3. The rise of empowerment

Although the reconstruction of this history is incomplete, we can be satisfied that the emergence of a business discourse of liberation has begun to make sense. But what has driven the development of this discourse? Ideologies of work have varied considerably (Bendix, 1956), but paid labor has not formerly been the object of such existential claims. Although the material background of this discursive innovation is presumably more complex than we know at present, nonetheless it is easy enough to identify a couple of its pieces, each of which expresses the continual intensification of market competition. One of them, on the level of production, is the emergence of processes for the continual, conscious evolution of work practices, first in manufacturing and now increasingly in services. The other, on the level of marketing, is the increasing attention and sophistication that large firms are bringing to their interactions with their customers.

These trends share an incompatibility with the conception of work implicit in Taylor's (1923) classical prescription for work rationalization based on time study. Central to Taylor's system was a separation between conception and execution. Whereas artisanal production was based on individual judgement in the deployment of inherited customary practices, rationalized production was based on engineers' once-and-for-all establishment of the "one best way" to perform a well-defined set of functions. For Taylor, this rational vision was an explicitly social project amidst a period of widespread industrial conflict in the United States: if the engineer could discover the optimal way to perform a task then surely management and workers would no longer have anything to fight about (Montgomery, 1987, pp. 252–254;

Nelson, 1980). This attempt at intellectual resolution of material conflict was bound not to work, but one must distinguish between two of its shortcomings. The first and more familiar is that the procedures of rationalization instantly became the locus of further conflict; to this day the stop-watch and (to a lesser extent) the video camera symbolize a crude type of social control through representation – in this case, the representation of work processes in technical terms (Montgomery, 1987, p. 221).

The second reason, often lost in the noise of the first, is that market production is inherently dynamic. New methods arise, old methods are improved, changing input prices modify the economic calculations that motivate specific choices of work methods, and the larger business cycle continually changes market demand and employees' bargaining power alike. So long as technology remained relatively primitive and the imperatives of management control required conception and execution to remain separated, this dynamism was slow enough for Taylorite engineers and shopfloor improvisation to accommodate (Hirschhorn, 1984, pp. 57–64; Montgomery, 1987, p. 249). Nonetheless, over the long haul, companies that found ways to internalize and promote the dynamism of production found a competitive advantage. The dynamism of production has also led to an increasing integration among the formerly compartmentalized stages of design, manufacturing, and distribution (Allen, 1994). As a result of these changes, the intensification of production has shifted from its abstract conception as a simple technical matter and revealed itself as a broader and more complex *socio*-technical matter. The key word in this transformation, remarkably enough, is "total". One practices, for example, "total quality management" (TQM), in which workers are required to participate in constant analysis and optimization of their own "performance" (Feigenbaum, 1983). These discourses may reject Taylorism, but just as often they propose to reclaim Taylorism by empowering work groups to apply Taylorite analysis to their own activities within a framework of group performance measurements and guarantees about job security (Adler, 1993; Halpern, Osofsky, and Peskin, 1989). While some of the representational devices might remain, their ideological framing changes considerably. A conventionally engineered system, however optimal its design, need only be understood as an inventory of components configured through rational procedures; an evolving system, though, must be understood holistically as a totality actively chosen through the continual reflection of its participants.

Just as the historical logic of manufacturing has brought human beings more fully into view as agents of production, parallel transformations in marketing have introduced more complex views of human beings into business discourse as well. Where Taylorite production understood people as components of optimally engineered systems, the imagination and practice of classical marketing understood consumers as mechanistic bundles of associations and needs. Although this view has certainly not disappeared, the last ten years or so have seen the rise within marketing discourse of a different kind of consuming subject. This new consumer, who

appears most frequently in discussions of customer service and public relations, is considerably more autonomous.

The key word for dealing with this consumer is, again, "total". One account of the process, for example, is entitled *Total Customer Service* (Davidow and Uttal, 1990). It emphasizes that customer service is based on a continually evolving understanding of the total relationship between the company and the consumer, and on a total understanding of the consumer's life insofar as it bears on this relationship. Some texts, rejecting the conventional distinction between manufacturing and service industries, dissolve the entire relationship between consumer and company into "service": the service rendered by an automobile, for example (transportation, comfort, entertainment, status, and so forth), is regarded as continuous with the services provided by sales, repair, and even advertising (Quinn, 1992, p. 7). No longer is the design of an automobile reducible to a set of ergonomic parameters. Instead, the object of design is the entire company – and, most particularly, the totality of the "face" that the company displays to the consumer, both through the physical product itself and through entire range of symbolic relationships in the life-cycle of buying and owning a car.

The impetus to construct and manage extended customer relationships on a mass scale, of course, derives from a variety of sources, including the growth of customer databases, the longer life-cycles of improved products, regulatory factors, and the evolving market for warranties. The point here is that the focus of this emerging logic of marketing is the autonomous consumer, understood as a complex, self-defining whole. The entire operation of the company, it is said, must revolve around this person. In some texts, the "service" relationship begins to seem downright sadomasochistic, as the company (through its human representatives) seeks endlessly to know and satisfy the desire of the customer. (For example, Davidow and Uttal (1990, p. 125)) note with approval the practice of hiring single women for these jobs, on the ground that they have a surplus of emotional energy available for investment in service relationships with customers.) This infinite task, like the infinite task of production, can no longer be submitted to once-and-for-all rational analysis. In both production and marketing, then, intensifying competition has brought a revised conception of work and of the agency of workers. This is precisely the "empowered" worker.

An extensive study could be made of the ideological means by which workers are encouraged to internalize the ideology of empowerment. Paid labor has always entailed strategies of legitimation, whether through the practical organization of work (Edwards, 1979) or the interactional regimes of management science (Hales, 1986) or otherwise. Empowerment, though, has been accompanied by a new generation of practices for the management of "corporate culture" through the programmatic creation of metaphors (Conrad, 1990), values (Quinn, 1992, pp. 156–157, 316–318), and stories (Armstrong, 1992). Collins and Porras (1995) are remarkably explicit in their analysis of "visionary" companies, whose unique combination of unity and flexibility they attribute to the cultivation of a "cultlike"

atmosphere in which all employees are "indoctrinated" into a "corporate ideology", with misfits being "expunged like an antibody". Whereas business theorists of an earlier era such as Barnard (1962 [1938]) and Simon (1947) investigated the conditions of employees' consent to organizational arrangements, the discourse of empowerment calls for active participation, secured not principally through coercion but through a technics of hegemony (Kunda, 1992). And within empowered teams, the team's creation of its own rules can endow those rules with a much more powerfully driven legitimacy than they would have possessed otherwise (Barker, 1993).

4. Empowerment and measurement

The rhetoric of liberation that surrounds the empowered worker has led to widespread hopes for democracy and other human values in work. It is certainly important to nurture such values. But it is also imperative to understand that, rhetoric notwithstanding, the empowered worker shares one thing with the rationalized worker: the ubiquity of measurement. When work processes are designed once-and-for-all, it is only necessary to record single measurements for each of those processes' atomic components. The literature describing the measurement of empowered work, though, focuses on outcomes as well as components, and empowered workers are expected to innovate in the service of these outcomes. Indeed, many business texts prescribe a new organization chart based on an "inverted hierarchy", with managers supposedly at the service of the productive workers at the "top" (Davidow and Uttal, 1990, pp. 90–91, 104–106; Quinn, 1992, pp. 129–134). Measurement of these workers, it is suggested, should no longer be administered from the outside by time-and-motion specialists; rather it should be continuous, built into the processes of work by their design. Measurement may take place at the level of individuals' actions or at the level of a team's joint actions; the variables being measured may focus upon inputs and outputs or upon actions between them, or both. Management employs these measurements in comparing work performance, benchmarking work procedures and arrangements, identifying "problems", evaluating innovations, and in an unbounded variety of other ways (Quinn, 1992, pp. 105, 117, 142–145, 166). The "output" of work in market terms may be measured directly through surveys of customers' perceptions and degrees of satisfaction (Davidow and Uttal, 1990, p. 197; Quinn, 1992, p. 117, 329–338). In contrast to earlier, frequently counterproductive computerized measurement schemes based on cost accounting (Grant, Higgins, and Irving, 1988), management discourse now emphasizes choosing measures with care, replacing them when necessary, and refining them as part of the "total" evolutionary process (Davidow and Uttal, 1990, pp. 185–205; Kaplan and Norton, 1993; Kaplan, 1990). Empowered work, then, is just as heavily monitored as rationalized work. Yet it does not afford the intellectual detachment that Lukács (1971 [1923], p. 89) could ascribe to the factory work of his day. Quite the contrary, it seeks the total involve-

ment of each individual's self. Empowerment and measurement, in short, together form a single disciplinary regime.

The empowerment and measurement regime is rarely identified as a single, coherent system in business writing. Usually the two themes are intertwined without the logic conjoining them being explicitly spelled out. This logic is most clearly in evidence in the literature on TQM. Texts for managers usually place great emphasis on measurement and its role in the continual analysis and improvement of work processes, with relatively short chapters encouraging that employees be empowered to make local decisions (Berk and Berk, 1993; Jablonski, 1992; Shores, 1990, pp. 90–91). Narrative accounts of TQM implementation, though, portray that process as a crusade of empowerment that shakes up hidebound bureaucratic structures and liberates their demoralized victims (Boyett et al., 1993; Frey 1993; Kearns and Nadler, 1992). Byham (1988) provides managers with a guide to interactional and rhetorical devices for stimulating the kind of initiative-taking that "quality" schemes require. (For a critical perspective see Parker (1985, pp. 9–22).) Regardless of the rhetorical approach, the argument for TQM is the same: hierarchical supervision can be replaced by team-building, non-coercive coaching, top-level strategy-setting, and thoroughgoing measurement.

The logic conjoining empowerment and measurement is negotiated in a variety of more complicated ways in the non-TQM business literature. In some publications, the connection is straightforward. Meyer (1994), for example, prescribes organizational procedures through which an empowered team, guided by strategic goals established by top management, should define its own measurements. Simons (1995) counts "diagnostic control systems" based on performance measurements and targets as one amongst four "control levers" in an organization of empowered employees. He takes for granted that diagnostic controls will be employed, but he emphasizes that they invite finagling when the other controls – beliefs systems, boundary systems (standing orders from management about what *not* to do), and interactive control systems (based on regular staff meetings) – are not in place.

In other cases, though, the connection is less clear. Lawler (1992, pp. 205–224), a proponent of employee involvement programs, argues that performance measures should be regularly communicated to employees. These measures, which are based largely on financial results and other firmwide outcomes, are much less detailed than those prescribed by TQM, and their purpose is described more in terms of morale than as concrete guidance for the conduct of work. (Whalley (1986) argues that such information can substitute to some extent for ideological control among technical workers, inasmuch as it makes tangible the higher discipline of the market.) Lawler also suggests that individual and group pay be related to skills and to various measures of performance (1992, pp. 192–200). He promotes widespread vertical and horizontal information flows in the organization, including individual performance measurements, without being specific about how this information should guide employees' work decisions.

Whereas Lawler is vague about the role of measurement in employee empowerment schemes, Johnson (1992) is elaborately conflicted. Having co-authored a damning history of management accounting (Johnson and Kaplan, 1987), Johnson seeks to free accounting from its profound connection with what he calls "remote-control management" (1992, p. 22) and the "pseudo-professionalism" of managers trained principally in economics (1992, p. 26). He emphasizes that measurement in itself does not guarantee that work processes will change for the better; quite the contrary, badly designed measurement schemes can reinforce existing patterns of thinking (1992, p. 105) and maintain a fragmented conception of the business (1992, p. 128). While he assumes that any business requires an elaborate management accounting system to track and measure work processes, he nonetheless rejects one accounting scheme after another, finding each too closely associated with the old static ways of remote control. In the end, he is left regretting that corporations are not democratic enough (1992, p. 102) without spelling out a substantive notion of democracy or identifying the underlying contradictions that work against the coherence of his discourse.

The relationship between empowerment and measurement is negotiated with special subtlety in the discourse of teams. Manuals for creating teams frequently employ the language of empowerment (Blanchard, Carew, and Parisi-Carew, 1990; Fisher, 1993) but rarely mention measurement. Certain elements of the team metaphor are explicit in the business literature; for example, a team has a clearly delineated membership, a common identity, ideals of smooth collaboration, and the subordination of individual personalities to the collective pursuit of a shared goal. Since this goal provides objective standards of performance, the manager's job loses much of its directly coercive quality and becomes more a "human" matter of coaching, teaching, counseling, and cheerleading. But the sporting metaphor has other elements as well: teams compete with one another through endless cycles of quantitative comparison. In this sense, measurement is a structural element of the discourse if not always overt in the texts.

5. Computers and empowerment

It is against this background that we can begin to understand something about the emerging role of computers in the business conception of empowered work. Historically, the broad existential claims now routinely made for paid labor coincide with the rise of decentralized computer technology. Much has been written, for example, about Apple Computer's ideology of empowerment and its opposition to the centralized world of IBM. Roszak (1986), for example, has derided the quasi-spiritual vocabulary that has long surrounded Apple computers (cf. Pfaffenberger, 1988). The Apple slogan, "The power to be your best", plays on the word "power" – referring either to computer power or personal empowerment – to make an existential promise: not simply *doing* but *being*.

But despite its notoriety, Apple and its slogans are nowhere near coextensive with the larger phenomenon of "empowerment" or the close association between empowerment and distributed computer technology. Its high-visibility advertising is aimed largely at individual professionals who have retained substantial degrees of autonomy and a relative freedom from the disciplines of measurement. Personal computers have often figured in institutional dramas pitting mainframe-oriented MIS managers against individuals trying to get work done at their desks. But irritating as they might be to some, MIS managers have rarely been the principal sources of constraint on anybody's work, and it is doubtful how great a contribution their putative overthrow has made to the actual reorganization of work activities. Changes in technology do not in themselves bring changes in organizational forms.

So it is worth taking a moment to consider the various explanations that have been offered of the connection between empowerment and computers. Within the management literature, the starting point for most of these explanations is Leavitt and Whisler's (1958) prognostications regarding "Management in the 1980's". In this article Leavitt and Whisler coined the term "information technology", which they identified as the combination of computer processing, mathematical programming and operations research, and the mechanized simulation of thought processes; and they issued a highly influential series of predictions about the evolution of organizations as a result of these technologies. Most importantly, they predicted a decimation of the middle ranks of management, with some jobs becoming more programmed and falling in status, while others jobs (especially programming and research jobs) rise in status. Moreover, managements that are inclined to recentralize decision-making authority, which had become decentralized under the weight of complexity, would be enabled to do so (cf George and King, 1991).

Although many have been struck by the accuracy of these predictions, the picture grows more complicated as we consider the reasoning on which they are based. Leavitt and Whisler view organizations as hierarchies of skill in which jobs that are "programmed" – that is, automated or Taylorized – are lower in the hierarchy and others are higher. Even if these assumptions are granted, Leavitt and Whisler might equally have argued for opposite conclusions. Consider, by contrast, the form of argument typified by Hirschhorn's (1984, pp. 57–58) analysis of automation. Hirschhorn notes that many manufacturing jobs have long been devoted to machine-tending of various sorts. As many of the simpler manufacturing operations have been automated, the resulting machinery increasingly employs advanced forms of feedback control, now generally computerized, that are more complex and difficult to understand and repair. As a result, many categories of machine-tending jobs have become inherently resistant to deskilling. Where Taylorite manufacturing depends on elaborate control regimes to channel machine-based work onto predetermined paths, the newer systems require workers to be empowered to anticipate problems and participate in continual improvement schemes. More recent authors

who predict a reduction in middle management through information technology arrive at this conclusion through different routes. Drucker (1988), for example, argues not that information technology automates middle-management reasoning but that it replaces many of the communicative functions of middle managers, thus allowing organizations to be "flattened" through an increase in the hierarchical span of control.

Implicit in the empowerment and measurement regime is a fresh approach to these questions. Consider, for example, the issue of centralization. Simon (1947, pp. 234–239) sought to impose conceptual order on this issue through his analysis of decision-making. He recognized, for example, that management can centralize decision-making authority either by limiting subordinates' decisions or simply by making the decisions themselves. And he discerned pressures in both directions: centralization of decision-making ensures that decisions are coordinated and performed with adequate expertise, and that responsibility for decisions is clearly defined; and decentralization of decision-making decreases costs by assigning decisions to lower-paid personnel and eliminating the overhead of communication up and down the hierarchy, as well as ensuring that decisions are made by individuals with immediate access to the "facts of the case". Faced with these competing considerations, the evident conclusion is that companies will seek the compromises that best suit their purposes. The empowerment and measurement regime, though, suggests an alternative approach: decentralize decision-making through empowerment while simultaneously centralizing control through measurement (Malone and Rockart 1993, p. 50). Such a scheme permits employees to use local knowledge in making decisions that directly affect products and customers while providing centralized experts with extensive data-streams to support overall strategy and coordination (Johnson and Kaplan, 1987, pp. 5–6).

Or consider the notion of organizational routines (cf Levitt and March, 1988). The general idea is that organizations' accumulated experience takes form in a set of routines that can outlast the particular organization members who might first develop them. The term is broader than the Taylorite notion of prescribed rote procedures; it also includes a broad range of conventions and policies, as well as the beliefs and cultures that support them. Routines are a conservative form of organizational memory that retains the "what worked" better than it remembers the history of "why". Contrasts between routine and non-routine activity are found in a wide variety of business literature (Cyert and March, 1963; Nelson and Winter, 1982; Sabel, 1982; Simon, 1960; Stinchcombe, 1990). Often, as in the case of Leavitt and Whisler (1958), routine or "programmed" activities are associated with lower strata of the organizational hierarchy. The empowerment and measurement regime, though, suggests supplying empowered workers with the steady stream of information and analytical skills they need to innovate and improvise in their local circumstances, using measurement to monitor the outcome rather than requiring conformance to routines. Of course, such a scheme is only preferable when the operational environment is complex and variable enough that local innovation

offers any advantages over evolved or scripted routines. It also entails the overhead of training and communication necessary to convey the "why" of organizational knowledge rather than simply reproducing the "what".

Finally, consider the distinction that numerous scholars have identified between two traditions of business thought; these traditions have been given many labels, including "direct control" versus "responsible autonomy" (Friedman, 1977), "theory X" versus "theory Y" (McGregor, 1960), and "bureaucracy" versus "corporatism" (Waring, 1991). It has been common to treat these opposed traditions as waves of fashion, each giving way to the other in a pendulum motion in which nothing really changes. But another approach is to trace the dialectical interaction through which the two traditions have become steadily integrated. The oscillation of fashions is real, of course, and each fashion predictably oversimplifies and overreaches and eventually discredits itself, but each such episode leaves a legacy of experience with attempted syntheses. The empowerment and measurement regime is precisely an advanced form of such synthesis, comprised of one theme drawn from the tradition of corporatism and autonomy – empowerment – and another theme drawn from the tradition of rationalization and control – measurement. The critical literature has commonly supposed that direct control through rationalization is a defining feature of capitalist production (Braverman, 1974), to which any experiment with workplace autonomy would necessarily revert once shopfloor resistance set in (Zimbalist, 1975). No doubt the empowerment and measurement regime encompasses a range of local varieties, no single one of which fully resolves the contradictions to which the two historical traditions of business thought have responded. But these contradictions must now be sought in different and more intricate spaces.

6. Accountability and colonization

The empowerment and measurement regime, then, can be viewed as an attempt to work through certain tensions in the management of large organizations. What remains in reconstructing this discourse is to specify the precise role envisioned for computing technology in implementing the emerging regime. What is at stake is not "computers" as a general category, but certain traditions of discourse and practice around the relationships between computing and human activity. As a provisional matter I wish to suggest that computers have been are distinctive in their role in a project of "colonization". I take this word from Habermas (1987, pp. 355–356), who speaks of the "colonization of the life-world" by the forces of technology. The colonization he has in mind is not simply the physical occupation of territory; it is, much more, the reorganization of communities' systems of meaning so that existing concepts are given technical definitions and thus subordinated to a technological order of knowledge and power. The communities I have in mind here are principally occupational communities: doctors, mechanics, accountants, secretaries, drivers, and so forth. The introduction of computers into a work community always brings

with it what Kling and Scacchi (1979) call "packages of social involvements" with other groups. But the introduction of computers into these trades has also brought new kinds of representation. To work with a computer has meant to continually re-represent certain aspects of one's activities for the computer's sake. This semantic or representational colonization has had a profound effect on many kinds of work and will continue to do so. The trend did not begin with computers and does not logically depend on them. The point, rather, is that particular deeply rooted computer system design practices have allowed the colonization process to intensify and spread. And this "quantitative" change has contributed to the "qualitative" change in the relations of work that I have been describing above.

One acute empirical study of empowerment is Suchman's discussion of "technologies of accountability" (1992). Her study concerns the operations room of an airline at a regional airport. The work in this room consists of coordinating "complexes" of airplanes which fly in to the "hub", exchange passengers and baggage while being serviced, and fly out again. This is a difficult and complex job. It is skilled work, the skill is hard to teach, and experience is at a premium. The workers often rotate among the various jobs. Constant choices are required, many of which interact with issues being dealt with simultaneously by other workers. The workers are thus, in particular senses, both autonomous and interdependent. In practice they spend most of their time communicating with one another and with workers at other sites (maintenance crews, pilots, operations people at other airports, etc) through several media (including radio, telephone, video links, and computers). The structures of authority are loose, with the workers all in earshot of one another and the supervisor standing largely in the background. And the work is intensively measured and monitored, with the numbers for the final outcomes – whether the planes depart within a certain margin of the scheduled times – continually available to the airline company's headquarters.

Although this is not a self-directed work group in the classical sociotechnical systems sense, it is nonetheless a reasonably good example of empowered work. Suchman's focus is on the phenomenon of "accountability". Her use of the word involves a studied ambiguity between two existing uses. The first use, in business discourse, refers to the reporting relationships through which authority is organized in a company. One is "accountable" (one is usually not "accountable to" anybody but simply "accountable") to the extent that one's activities are, as a practical matter, subject to the broader imperatives of the company as articulated by its management.

The second use of "accountability" derives from ethnomethodology (Garfinkel, 1984 [1967]). Here the word refers to a broad property of human activity, which roughly speaking is the responsibility to make one's actions intelligible to other people. One is "accountable" in the sense that one has the obligation of giving some account of one's actions. The management of mutual intelligibility and the moral order that surrounds this process are regarded by ethnomethodologists as the very "glue" of society.

As an ethnomethodological matter, accountability is a considerable burden for the airline operations workers in Suchman's study. Their work is "visible", through one medium of communication or another, to a wide variety of people, each of whom bears a different and equally complex relationship to it. The workers exhibit a thoroughgoing and remarkably sophisticated orientation to these burdens; they comport themselves in a way that all of their many audiences will find intelligible. In particular, the workers display an orientation to the mechanisms through which their work is measured, based on a complex formula which is intended to formalize the company's notion of "doing a good job of getting the planes in and out on time". The work group has a joint responsibility for satisfying this criterion, and their work activities are highly visible to one another. The workers are not orienting to any transcendent criterion of "a good job" and "on time", but rather to the ways in which particular versions of these notions are formulated and reported in practice. And the job is not done until suitable inputs have been provided to all of the relevant machinery of measurement.

This is where the two uses of "accountability" meet. For Suchman, the workers' accountability to management is something done *in* their work, indeed *as* their work. Their accountability, in other words, is not just a formal relationship or an outside force, but a practical process of exhibiting reality in such-and-such terms. The "technologies" of Suchman's title are designed specifically to facilitate the process of representing the work in just these terms. Although many technologies are involved, distributed computer technologies play a crucial role in creating, storing, accumulating, manipulating, and transmitting the representations through which the workers' activities are "accounted" to management.

I propose to extend Suchman's analysis through a more general consideration of the role of computers in making empowered work accountable. Like the Taylorite work it replaced, this newer style of work is heavily "staged". It does not just happen spontaneously, but is shaped through a great deal of preparation on the part of management. Much of this preparatory work is representational; managers and their staffs literally work up new representation schemes that are capable of expressing in formal terms the range of activities that occur within a given category of jobs. Then as now, this is understood as a process of breaking the work down into its smallest units. For Taylor these units are principally physical: hand movements, visual discriminations, reorientations of the body, and so forth. Generations of Taylor's followers extended and generalized this basic idea, evolving elaborate standardized symbolism and paperwork for representing the fine physical details of factory jobs (Gilbreth, 1912; Holmes, 1938; Shaw, 1952). In the design of empowered work, the units of work tend to be understood in informational and institutional terms: placing an order, acknowledging receipt, modifying a parameter, acknowledging a query, and so forth – the "minimum replicable units" from which work activities are composed (Quinn, 1992, pp. 103–109; Medina-Mora, Winograd, Flores, and Flores, 1992).

This contrast (physical versus informational units) should not be overdrawn; the emphasis in each case is on designing a "language" for representing activities. And in each case, the process begins with close observation and thorough analysis of the job as it is currently performed. At this point, though, Taylorite and "empowered" work design diverge. Where the Taylorite work-designer seeks the "one best way" to perform the work, the "empowered" work-designer attempts to construct a system of computers and communication devices to "capture" the work as it is performed. One readily observable example of this process can be found in just about any high-volume restaurant, where the waiters log each transaction, often by swiping a magnetized card through a computer workstation. These computers capture an extraordinary amount of information, both for measuring the productivity of the workers and for adjusting the restaurant's strategies by detecting and accommodating variations in customers' tastes.

The crucial difference between Taylorism and empowerment, then, lies with the ability of computers to "capture" work processes on a fine-grained level and then to actually make use of the resulting masses of data. Much has been written about the privacy concerns that result from the accumulation of this data (Clarke, 1989). My own topic here is not the contents of those databases but the processes through which the contents are kept up to date: the "capture" mechanisms at the meeting-points between the workers and the computers. Whereas Taylorite workers must simply execute a given series of movements, empowered workers must (in one way or another) continually "enter" aspects of their activities into a computer. To do so, they must understand their work within the terms of the machine's representations.

We might thus speak of the work-design process as establishing a "grammar of action" (Agre, 1994). This grammar will typically have been derived through the study of existing work-activities. But once the grammar is formalized and made the basis for interactions with a computer, workers' relationship to the grammar changes. Whereas once the grammar was something *posited* by researchers and *ascribed* to activities through some kind of coding procedure, now the grammar becomes a resource in the activity itself. In Suchman's terms it becomes the normative framework within which the activity becomes accountable through the mediation of particular technologies.

Grammars of action stand in particular relationships to the activities from which they are derived and upon which they are imposed. A grammar of action, *qua* formal system, is a kind of mathematical language. The vocabulary of this language is usually drawn, at least in large part, from the terms of art employed by the participants in those activities (Quinn, 1992, p. 104). A grammar of restaurant activities, for example, will speak of "orders", "changes", "items", "customers", "tabs", "tips", and so forth, drawing on the existing professional lingo of waiters, cooks, and managers (Quinn, 1992, p. 143). The representations of a given activity built within a given grammar of action (say, as printed out by a computer) will

often outwardly resemble the language in which the participants themselves would report the activity.

But the grammar of action is not the same as the workers' indigenous language. It is, instead, a mathematical formalization of it. Computers, after all, run on discrete mathematics; computer languages necessarily assimilate matters in the world to mathematically specified structures inside them. And mathematical languages are different from natural languages: they do not possess anything like the contextual flexibility that ethnomethodology (among other schools of empirical research on situated language use) has shown to be ubiquitous in workplace activities and other types of social interaction. This is because the formal system of mathematics strictly governs the inferences that are possible from a given statement without any attention to the broader interactional context within which the statement is made. This is not to say that mathematically formalized languages, when spoken by people in social interaction, are exceptions to the general observations of ethnomethodology. They are not. The point, rather, is that computerized "capture" of work activities requires workers to systematically gloss their actions in terms that "work" within the formal, "mechanical" order of the computer's inferences (Suchman, 1994; Winograd, 1994; Agre, in press).

These novel cognitive and linguistic constraints upon the conduct of work activities can cause immense confusion and turmoil, for all of the reasons that confound other attempts to mechanize human reasoning (Forsythe, 1993). But my own point here concerns not the failure, in certain terms, of such enterprises but their success, in other, equally important terms. These regimes of work reorganization *exist*; they function with considerable efficiency and are growing steadily more sophisticated as experience with them accumulates. The analysis above will certainly need to be extended and supplemented in many ways, but we are in a position now to identify the central *social* project that is at work in the midst of these schemes: it is the project I have referred to as colonization. Its basis is language, and specifically the redefinition in mathematical terms of the disciplinary languages of existing communities of practice. On one level this project is continuous with skill-extraction projects since the beginnings of Taylorism (Braverman, 1974; Thompson, 1989). What is new here is the attempt, so to speak, to re-inject the now-formalized conceptual system around these skills. Where Taylorized work attempted to circumvent the worker's cognition through regimented interactions with machinery, empowered work attempts to *regiment* the worker's cognition through *formalized* interactions with machinery.

Despite the deep connection between capture and computational formalization, it does not follow that computerization necessarily entails capture or colonization. When people exchange electronic mail, for example, they are using computers as a medium of communication that is indifferent to the content of their messages. Likewise, a shared drawing tool might capture certain aspects of a drawing, for example its basic geometrical units, without possessing any vocabulary with which to express higher-order aspects of the drawing. On the other hand, if a computer

cannot capture a given aspect of an activity then it cannot represent or reason about it either (Agre, 1994, pp. 113–114). For example, since an ordinary e-mail system does not recognize speech acts, it can neither provide the benefits nor impose the constraints of a system like The Coordinator (Flores et al., 1988) which does. Computing as currently understood thus possesses a profound duality and ambivalence (Feenberg, 1991, pp. 91–93 ff) which any critique or praxis must either negotiate or replace.

7. Empowerment and the critique of representation

In previous sections, a rational reconstruction of contemporary business discourse has brought to light a widespread and reasonably complicated story about work, technology, organizations, representation, and human agency. I would like to compare this story to the stories told about the same things in critical social studies of computer use. Central to these latter stories has been a critique of representation, and specifically a critique of the representations that are embodied in computers. This critique draws upon a number of intellectual sources, but at its core is an opposition to the view that representations transparently correspond to their objects. (This opposition is central to a variety of other projects as well, for example in Derrida (1978).) An extended analysis of this critique, and comparison with the business discourse of empowerment, will provide an occasion for rethinking both the critique itself and the theory of colonization that I have been developing so far.

To illustrate the critique of transparent representation, let us return momentarily to the foundations of ethnomethodology through Garfinkel's (1984 [1967]) initial formulations. Garfinkel's substantive portrayal of social interaction drew on the phenomenological tradition of Schutz (1967 [1932]), but in ethical terms his sociological project was a response to the positivist sociological tradition epitomized by the work of Talcott Parsons (1937). In developing his vast theoretical system, Parsons drew freely upon the resources of ordinary language to fashion terms ("action", "family", "crime", "power") that scientific sociology could define, operationalize, measure, and assemble into larger conceptual structures. Garfinkel insisted to the contrary that these terms should not be appropriated into such an artificially stabilized conceptual order, and furthermore that their use by the members of a given social setting should actually be the primary object of sociological research. Rather than seeking generalizations about families, for example, the sociologist should observe how particular groups of people displayed themselves to one another *as* a family as part of their ordinary activities. In this way, Garfinkel insists on the materiality of representation. His object of analysis is not language as such, or representation generally, but people using representations in particular settings. The actual force of particular words is routinely discovered to be indexical, ambiguous, and negotiated; their meanings are determinate only to the extent required by the practical purposes at hand.

Ethnomethodological research on computing, exemplified by Suchman's analysis of airport work discussed above, has translated Garfinkel's critique of sociological representation into a critique of computer system design (Suchman and Jordan, 1989; Button, 1993). In each case one encounters an organized practice of drawing upon indigenous terms to fashion theoretical categories (sociological concepts, computational procedures and datastructures), and in each case the underlying moral project is to vindicate the tacit expertise that the indigenous community employs in its seemingly mundane interactions. As my use of the term "indigenous" may suggest, this project is organized by an implicit narrative that my own analysis above has made explicit: technology as a form of semantic colonization. The prototype for this narrative is the appropriation of craft knowledge and the founding of rationalized work in Taylor's era. In its application to computer work, it reflects a particular kind of historical experience: the first arrival of computerization into a given site of practice. This narrative begins with a community, conceived as self-contained and autonomous, which has evolved a body of tacit skills and customary practices whose sophistication may not be recognized when set against abstract standards (Scribner, 1984). Into this community comes an invasion: new forms of power that represent the existing practices, employ the resulting representations in reorganizing those practices, and stabilize the new regime in part by taking control of the meanings of key indigenous terms. The community may resist its domination on other levels, but this resistance will eventually be subverted by the new cognitive order that is created by the new scheme of representation (Miller, 1991). This narrative form motivates a critical strategy with two parts: first, demonstrate how the colonization process fails by its own standards through its inability to fully represent the tacit skills of the natives and its failure to appreciate the complex materiality of representation in indigenous interactional practices; and second, exhibit the methods of "deleting the work" (Star, 1989a) through which computer practitioners maintain the illusion of representational transparency.

The colonization narrative has been deeply influential, but it has been joined by two others. Bowers (1992) has drawn upon a range of theoretical sources to paint an elaborate picture of the material use of representations in society. At a local level, once again, he categorizes the failures of transparency – indexicality, incompleteness, dialogicality, and so on. At a more macroscopic level, he invokes Latour's (1986) account of the "centers of calculation" into which representational materials ("immutable mobiles") are drawn through the channels of bureaucracy from far-flung locations throughout the social body. The narrative here is one of homeostasis and resembles the visions of social management in Foucault: a gaze cast asymmetrically upon the population, based everywhere in a local politics of representation, aggregated by an authority that is centralized but not personified, and serving to administer certain norms through the coordinated application of expertise. Unlike the narrative of colonization, this narrative of homeostasis posits no lost origins and risks no charges of romanticism. Bowers' critical object is not computing as such but formalism in general; he wishes to draw attention

to the extensive and laborious practical work of representation through which formalisms are stably connected to objects in the world despite their inherent indeterminacy. The historical experience to which he appeals is a situation already given, in which life is pervaded by the operations of knowledge and power through which formalisms are endowed with meaning and effectivity. The critical strategy recommended by this narrative entails, once again, exhibiting the deleted work of representation, but simultaneously renouncing any nostalgia for a precolonial era. Following Haraway (1991) he calls for a praxis whose starting point is the irreversible intermixing (sometimes called "postcolonial") between ourselves and our machines, and particularly the critical appropriation of formalism as a domain of positive action.

The third narrative form in critical social studies of computing encourages a pluralism of interpretative communities. As articulated by Robinson and Bannon (1991) and Schmidt (1991), this narrative responds to experiences, especially common in recent years, of system development projects that attempt to integrate existing systems that have arisen one by one in particular settings. In such projects it often transpires that each local community has its own meanings for the terms that programmers have drawn upon in writing their programs. The accounting and manufacturing departments, for example, might have different ideas of what it means for an contract to be fulfilled, depending on the practical exigencies and professional ideologies that structure their work. The narrative begins with separate communities brought together (voluntarily or forcibly) and negotiating meanings (amicably or not) in particular situations where the dissonance of their respective usages becomes evident as an obstacle to their collaboration. System developers who convert the contested terms into representations and treat those representations as transparent reflections of organizational reality will risk calamity as each community carries on interpreting the term in its own way. Pluralistic critical strategies all begin by exhibiting these difficulties and vindicating the pluralism of meanings by legitimizing the "articulation work" (Schmidt and Bannon, 1992) at the boundaries between communities. But these strategies also exhibit a certain variety of emphases, depending on their theoretical orientation: for example, Star (1989b), working within symbolic interactionism, emphasizes the irremediability of semantic difference, while Schmidt (1991) exhibits a more functionalist orientation to the amelioration of these differences.

Each of these three critical practices, then, opposes itself to the ideology of transparent representation through a different narrative form and critical strategy. In consequence, each of them retains a powerful tie to the concept of transparency: transparency is the positive term to which each of them articulates a negative term, and each critical strategy is honed to revealing the things that technical practitioners obscure through through their pursuit of it. But precisely as a consequence, none of them affords an analysis of the historical dimension of the practices of representation. Ideologies of transparency are indeed found in many places and times, most particularly in the literature of computer science. Yet the business dis-

course of empowerment and measurement reflects a somewhat more sophisticated understanding of representation and its role in practice, one that evidences a history of practical struggle with the very phenomena that the critique of representation has revealed. And, strikingly, the recommendations of this discourse are aligned in many ways with those of the sociological critique.

Within the empowerment and measurement regime, formalism and representation are found most straightforwardly in the process of measurement. As the critique of transparency might suggest, widespread uncritical acceptance of the measurement schemes of cost accounting has long raised havoc with businesses' internal management controls. But a powerful critique of cost accounting has arisen within the business literature, leading to a variety of conceptual frameworks that acknowledge the contingent and instrumental nature of measures. Managers are now enjoined to select and evaluate particular measures based on their potential for finagling, the incentives they create, and the possibility that they do not actually measure the "right things" from the point of view of serving customers. Measurements are only held to be meaningful when the ideological controls exist to ensure adequate compliance with the practical arrangements that implement them.

The empowerment and measurement regime also exhibits considerable regard for the tacit skills of local work groups. The doctrine of empowerment aims at the mobilization and cultivation of these tacit skills, employing measurement to evaluate outcomes that affect customers, not for direct control. In this way, the empowerment and measurement regime joins the critical practices just summarized in its opposition to the traditions of work design that had taken for granted the transparency of their representations of work practices. Bannon's (1991) elegant call for a conceptual transition from "human factors to human actors" has its counterpart in numerous calls for businesses to empower employees by treating them like human beings and not like cogs in a machine.

A particularly remarkable point of convergence has lately emerged between business discourse and the critical narrative of interpretive pluralism. Despite his background in the relatively unsophisticated philosophy of business process redesign, Davenport's (1994) analysis of the information dynamics of organizations echoes the theories of Robinson, Bannon, and Schmidt in some detail. Davenport offers numerous examples of the diversity of meanings found across functions in an organization. By and large, though, he advocates negotiating common meanings once and for all rather than, legitimizing the diversity of meanings and providing technical support for their ongoing, local negotiation.

Clearly, then, the critique of transparency is limited in its capacity to make sense of the empowerment and measurement regime. The reason, I will suggest, is that it is ahistorical. It inherits this ahistorical character from the ideology of transparency itself, which proceeds by assimilating human reality to the ahistorical structures of mathematics and then effacing the difference between them. The critique of transparency recovers certain elements of the historicity of representation, but it does not in itself provide a positive historical analysis of the social relations

of representation. The narrative of colonialism, for example, posits a precolonial era that usually turns out to be much more complex and ambivalent than the before-and-after metaphor of invasion can express. The narrative of homeostasis, for its part, flattens the specific relations of representation in various places and times into a single network of measuring and gathering. And the narrative of interpretive pluralism points at difference among communities of interpretation without theorizing what these different communities *are*, where they came from, and what else might regulate their relationships beyond the random patterns of semantic discord and negotiation.

The point is not that these three critical procedures are misguided. The literature of empowerment and measurement is far from consistent in its critical attitude toward formalism, and assumptions of representational transparency can be found at many places throughout. One of the purposes of ideological control in this regime may be to encourage illusions of the transparency of measurement. Likewise, the grammars of action that support the measurements are rarely regarded as contingently useful representations of work activities, at least to the same extent that measurements are regarded as contingent. As an example, consider the application of conversation analysis "rules" to the design of natural language interfaces (Luff, Gilbert, and Frohlich, 1990), a proposal that has provoked controversy about the nature of these "rules", and specifically whether they can be regarded as a grammar of action in the sense required by a capture scheme (Button, 1990). In these ways, the critique of transparency defines a horizon within which further critical analysis might proceed.

Nonetheless, a historical analysis of the empowerment and measurement regime provides much that is missing from the various critiques of transparency. It places the social relations of workplace representation firmly in their historical context: within the professional traditions of engineering, the oscillations of fashion in management thought, the reinvention of organizational structure, and the complex history of workplace conflict and accommodation. Most particularly, in the terms of Gordon, Edwards, and Reich (1982), it suggests that we are observing the exploratory phases of a periodic reorganization in the dynamics relating industrial production to the structure of the labor force. Central to this process is what Feenberg (1991, pp. 108–109, 189–198), following Simondon (1958), calls "concretization", the development of synergies between technical systems and their natural and human environments. The empowerment and measurement regime represents a new level in the integration of productive technology with its ideological, interpersonal, organizational, and economic surroundings. Feenberg regards concretization as an immanent feature of technology, inasmuch as it responds to demands for technical efficiency and integration, *and* as a force for democracy, inasmuch as it partially reverses the "reified decontextualization" (1991, p. 189) of technical practice. This hopeful conclusion rests on a close identification between the reification of market commodities and the decontextualization of technical objects. The business literature on empowerment and measurement, though, clearly does not concur with

the identification of these two categories, since it recognizes a series of factors mediating between them. Most importantly, this literature, unlike the discourse of Taylorism whose historical facticity deeply influenced the rise of critical theory, does not treat human relations as a technical order commensurable with the workings of technical machinery. Rather, it defines distinct yet complementary orders of "technical" and "human" affairs bound together within a dynamic tension. Critical theory must reckon with this discourse and with the contradictory reality that it both mediates and refracts. Critical studies of human-computer interaction are remarkably respectable just now, and I think we should regard this state of affairs as a complicatedly mixed blessing. The ideology of the human worker in these studies is a dramatic improvement on that of Taylorism. Yet this same ideology risks appropriation within (and perhaps, to some extent, legitimation of) a new organization of work whose liberatory virtues are still far from proven. Taylorism has been superseded ideologically, if not always materially, because in some real sense it was mistaken. It was mistaken about the nature of markets, about the nature of work, and about the nature of human beings. Market competition, eventually and in its own complex way, compels large institutions to glimpse the truths underlying such mistakes, whereupon one ideology and material organization of activity gives way to another. During the moments of transition, many things are visible which are ordinarily obscured. Perhaps this is one of these moments, and perhaps it is possible to grasp the moment before it passes.

Acknowledgements

This is a substantially revised version of paper originally prepared for the Workshop on Social Science Research, Technical Systems and Cooperative Work, organized by Geof Bowker, Les Gasser, Leigh Star, and Bill Turner in Paris on 8–11 March 1993. During the meeting I benefitted from comments by Susanne Bodker, Eswaran Subrahmanian, and Claude Vogel. Subsequent revisions of the paper have benefitted from comments by Steve Bagley, John Batali, David Chapman, Carol Christopher, Diana Forsythe, Rebecca Henderson, Dave Kinnaman, and Lucy Suchman. I presented some of this material at a colloquium of the CORPS program at UC Irvine, and I would like to thank Jonathan Allen, Margaret Elliott, Rob Kling, Roberta Lamb, Ignace Snellen, and John Tillquist for the helpful commentaries they prepared for this occasion. Finally, Liam Bannon and three anonymous referees provided some exceptionally thorough and useful comments on the penultimate draft.

References

Adler, P. S. (1993): Time-and-motion regained. *Harvard Business Review*, vol. 71, no. 1, pp. 97–108.

Agre, P. E. (1994): Surveillance and capture: Two models of privacy. *The Information Society*, vol. 10, no. 2, pp. 101–127.

Agre, P. E. (1995): Accountability and discipline: A comment on Suchman and Winograd. *Computer Supported Cooperative Work (CSCW)*, vol. 3, no. 1, pp. 31–35.

Allen, J. (1994): Mutual control in the newly integrated work environments. *The Information Society*, vol. 10, no. 2, pp. 129–138.

Armstrong, D. M. (1992): *Managing by Storying Around*. New York: Doubleday.

Bachrach, P. and Botwinick, A. (1992): *Power and Empowerment: A Radical Theory of Participatory Democracy*. Philadelphia: Temple University Press.

Bannon, L. (1991): From human factors to human actors: The role of psychology and human-computer interaction studies in systems design. In *Design at Work: Cooperative Design of Computer Systems*, eds. Joan M. Greenbaum and Morten Kyng. Hillsdale, NJ: Erlbaum.

Barker, J. R. (1993): Tightening the iron cage: Concertive control in self-managing teams. *Administrative Science Quarterly*, vol. 38, no. 4, pp. 408–437.

Barnard, C. I. (1962): *The Functions of the Executive*. Cambridge: Harvard University Press. (Originally published in 1938).

Bendix, R. (1956): *Work and Authority in Industry: Ideologies of Management in the Course of Industrialization*. Berkeley: University of California Press.

Berk, J. and Berk, S. (1993): *Total Quality Management: Implementing Continuous Improvement*. New York: Sterling.

Blanchard, K., Carew, D. and Parisi-Carew, E., eds (1990): *The One Minute Manager Builds High-Performance Teams*. New York: Morrow.

Block, P. (1981): *Flawless Consulting: A Guide to Getting Your Expertise Used*. Austin: Learning Concepts.

Block, P. (1987): *The Empowered Manager: Positive Political Skills at Work*. San Francisco: Jossey-Bass.

Bowers, J. (1992): The politics of formalism. In *Contexts of Computer-Mediated Communication*, ed. Martin Lea. New York: Harvester Wheatsheaf, pp. 232–261.

Boyett, J. H., Schwartz, S., Osterwise, L. and Bauer, R. (1993): *The Quality Journey: How Winning the Baldrige Sparked the Remaking of IBM*. New York: Dutton.

Braverman, H. (1974): *Labor and Monopoly Capital: The Degradation of Work in the Twentieth Century*. New York: Monthly Review Press.

Butler, S. M. ed. (1990): *Agenda for Empowerment: Readings in American Government and the Policy Process*. Washington, DC: Heritage Foundation.

Button, G. (1990): Going up a blind alley: Conflating conversation analysis and computational modelling. In *Computers and Conversation*, eds. Paul Luff, Nigel Gilbert, and David Frohlich. London: Academic Press.

Button, G. (1993): *Technology in Working Order: Studies of Work, Interaction, and Technology*. London: Routledge.

Byham, W. C. (1988): *Zapp!: The Lightning of Empowerment*. New York: Harmony.

Clarke, R. A. (1989): Information technology and dataveillance. *Communications of the ACM*, vol. 31, no. 5, pp. 498–512.

Clement, A. (1994): Computing at work: Empowering action by "low-level users". *Communications of the ACM*, vol. 37, no. 1, pp. 53–63, 105.

Collins, J. C. and Porras, J. I. (1995): *Built to Last: Successful Habits of Visionary Companies*. New York: HarperCollins.

Conrad, C. (1990): *Strategic Organizational Communication: An Integrated Perspective*. Fort Worth: Holt, Rinehart, and Winston.

Cyert, R. M. and March, J. G. (1963): *A Behavioral Theory of the Firm*. Englewood Cliffs, NJ: Prentice-Hall.

Davidow, W. H. and Uttal, B. (1990): *Total Customer Service: The Ultimate Weapon*. New York: Harper.

Derrida, J. (1978): *Writing and Difference*, translated from the French by Alan Bass. Chicago: University of Chicago Press.

Drucker, P. F. (1988): The coming of the new organization. *Harvard Business Review*, vol. 66, no. 1, pp. 45–53.

Edwards, P. (1995): *The Closed World: Computers and the Politics of Discourse*. MIT Press.

Edwards, R. (1979): *Contested Terrain: The Transformation of the Workplace in the Twentieth Century*. New York: Basic Books.

Feigenbaum, A. V. (1983): *Total Quality Control*, third edition. New York: McGraw-Hill.

Feenberg, A. (1991): *Critical Theory of Technology*. New York: Oxford University Press.

Fisher, K. (1993): *Leading Self-Directed Work Teams: A Guide to Developing New Team Leadership Skills*. New York: McGraw-Hill.

Fligstein, N. (1990): *The Transformation of Corporate Control*. Cambridge: Harvard University Press.

Flores, F. Graves, M. Hartfield, B. and T. Winograd (1988): Computer systems and the design of organizational interaction. *ACM Transactions on Office Information Systems*, vol. 6, no. 2, pp. 153–172.

Forsythe, D. (1993): Engineering knowledge: The construction of knowledge in artificial intelligence. *Social Studies of Science*, vol. 23, no. 3, pp. 445–477.

Foucault, M. (1988): Technologies of the self. In *Technologies of the Self: A Seminar with Michel Foucault*, eds. Luther H. Martin, Huck Gutman, and Patrick H. Hutton. Amherst: University of Massachusetts Press.

Frey, R. (1993): Empowerment or else. *Harvard Business Review*, vol. 71, no. 5, pp. 80–82, 84, 86–89, 92, 94.

Friedman, A. L. (1977): *Industry and Labour: Class Struggle at Work and Monopoly Capitalism*. London: Macmillan.

Friedmann, J. (1992): *Empowerment: The Politics of Alternative Development*. Cambridge, MA: Blackwell.

Garfinkel, H. (1984): *Studies in Ethnomethodology*. Polity Press. (Originally published in 1967.)

Gasser, L. (1986): The integration of computing and routine work. *ACM Transactions on Office Information Systems*, vol. 4, no. 3, pp. 205–225.

George, J. F. and King, J. L. (1991): Examining the computing and decentralization debate. *Communications of the ACM*, vol. 34, no. 7, pp. 63–72.

Gilbreth, F. B. (1912): *Primer of Scientific Management*. New York: Van Nostrand.

Gilbreth, L. M. (1921): *The Psychology of Management: The Function of the Mind in Determining, Teaching and Installing Methods of Least Waste*. New York: Macmillan.

Gordon, D. M., Edwards, R. and M. Reich (1982): *Segmented Work, Divided Workers: The Historical Transformation of Labor in the United States*. Cambridge: Cambridge University Press.

Grant, R. A., Higgins, C. A. and R. H. Irving (1988): Computerized performance monitors: Are they costing you customers? *Sloan Management Review*, vol. 29, no. 3, pp. 39–45.

Guillén, M. F. (1994): *Models of Management: Work, Authority, and Organization in a Comparative Perspective*. Chicago: University of Chicago Press.

Habermas, J. (1987): *The Theory of Communicative Action, volume 2: Lifeworld and System: A Critique of Functionalist Reason*. Boston: Beacon Press.

Hales, M. (1986): Management science and the 'second industrial revolution'. In *Radical Science Essays*, ed. Les Levidow. London: Free Association Books, pp. 62–87.

Halpern, D. Osofsky, S. and M. I. Peskin (1989): Taylorism revisited for the 1990s. *Industrial Management*, vol. 31, no. 1, pp. 20–23.

Haraway, D. (1991): *A cyborg manifesto: Science, technology, and socialist-feminism in the late twentieth century*, in *Simians, Cyborgs, and Women: The Reinvention of Nature*. London: Free Association Books.

Henderson, A. (1991): A development perspective on interface, design, and theory. In *Designing Interaction: Psychology at the Human-Computer Interface*, ed. John M. Carroll. Cambridge: Cambridge University Press.

Hirschhorn, L. (1984): *Beyond Mechanization: Work and Technology in a Postindustrial Age*. Cambridge: MIT Press.

Holmes, W. G. (1938): *Applied Time and Motion Study*. New York: Ronald Press.
Jablonski, J. R. (1992): *Implementing TQM: Competing in the Nineties through Total Quality Management*. Second edition. Albuquerque: Technical Management Consortium.
Johnson, H. T. and R. S. Kaplan (1987): *Relevance Lost: The Rise and Fall of Management Accounting*. Boston: Harvard Business School Press.
Johnson, H. T. (1992): *Relevance Regained: From Top-Down Control to Bottom-Up Empowerment*. New York: Free Press.
Kaplan, R. S. ed. (1990): *Measures for Manufacturing Excellence*. Boston: Harvard Business School Press.
Kaplan, R. S. and D. P. Norton (1993): Putting the balanced scorecard to work. *Harvard Business Review*, vol. 71, no. 5, pp. 134–149.
Kearns, D. T, and D. A. Nadler (1992): *Prophets in the Dark: How Xerox Reinvented Itself and Beat Back the Japanese*. New York: Harper.
Kling, R. and W. Scacchi (1979): Recurrent dilemmas of routine computer use in complex organizations. In *Proceedings of the AFIPS National Computer Conference* 48. Arlington, VA: AFIPS Press, pp. 107–115.
Kunda, G. (1992): *Engineering Culture: Control and Commitment in a High-Tech Corporation*. Philadelphia: Temple University Press.
Lawler III, E. E. (1992): *The Ultimate Advantage: Creating the High-Involvement Organization*. San Francisco: Jossey-Bass.
Leavitt, H. J. and T. L. Whisler (1958): Management in the 1980's. *Harvard Management Review*, vol. 36, no. 6, pp. 41–48.
Levitt, B. and J. G. March (1988): Organizational learning. *Annual Review of Sociology*, vol. 14, pp. 319–340.
Lewin, K. (1951): *Field Theory in Social Science: Selected Theoretical Papers*, ed. Dorwin Cartwright. New York: Harper.
Lilienfeld, R. (1978): *The Rise of Systems Theory: An Ideological Analysis*. New York: Wiley.
Luff, P. Gilbert, N. and D. Frohlich eds. (1990): *Computers and Conversation*. London: Academic Press.
Lukács, G. (1971): *History and Class Consciousness: Studies in Marxist Dialectics*, translated by Rodney Livingstone. Cambridge: MIT Press. (Originally published in German in 1923).
Malone, T. W. and J. F. Rockart (1993): How will information technology reshape organizations?: Computers as coordination technology. In *Globalization, Technology, and Competition: The Fusion of Computers and Telecommunications in the 1990s*, eds. Stephen P. Bradley, Jerry A. Hausman, and Richard L. Nolan. Boston: Harvard Business School Press, pp. 37–56.
Maslow, A. H. (1954): *Motivation and Personality*. New York: Harper.
McGregor, D. (1960): *The Human Side of Enterprise*. New York: McGraw-Hill.
Mayo, E. (1933): *The Human Problems of an Industrial Civilization*. New York: Macmillan.
Medina-Mora, R. Winograd, T. Flores, R. and F. Flores (1992): The action workflow approach to workflow management technology. In *Proceedings of CSCW–92*, Toronto, 1992, pp. 281–288.
Meyer, C. (1994): How the right measures help teams excel. *Harvard Business Review*, vol. 72, no. 3, pp. 95–103.
Miller, A. G. (1991): Transformations of time and space: Oaxaca, Mexico, circa 1500–1700. In *Images of Memory: On Remembering and Representation*, eds. Susanne Kuchler and Walter Melion. Washington, DC: Smithsonian Institution Press.
Montgomery, D. (1987): *The Fall of the House of Labor: The Workplace, the State, and American Labor Activism, 1865–1925*. Cambridge: Cambridge University Press.
Mullender, A. and D. Ward (1991): *Self-Directed Groupwork: Users Take Action for Empowerment*. London: Whiting and Birch.
Nelson, D. (1980): *Frederick W. Taylor and the Rise of Scientific Management*. Madison: University of Wisconsin Press.
Nelson, R. R. and S. G. Winter (1982): *An Evolutionary Theory of Economic Change*. Cambridge: Harvard University Press.
Noble, D. D. (1989): Cockpit cognition: Education, the military, and cognitive engineering. *AI and Society*, vol. 3, no. 4, pp. 271–296.

Norman, D. A. (1993): *Things That Make Us Smart: Defending Human Attributes in the Age of the Machine, Reading*. MA: Addison-Wesley.

O'Reilly, C. (1989): Corporations, culture, and commitment: Motivation and social control in organizations. *California Management Review*, vol. 31, no. 4, pp. 9–25.

Parker, M. (1985): *Inside the Circle: A Union Guide to QWL*. Boston: South End Press.

Parsons, T. (1937): *The Structure of Social Action: A Study in Social Theory with Special Reference to a Group of Recent European Writers*. New York: McGraw-Hill.

Pfaffenberger, B. (1988): The social meaning of the personal computer: or, Why the personal computer revolution was no revolution. *Anthropological Quarterly*, vol. 61, no. 1, pp. 39–47.

Quinn, J. B. (1992): *Intelligent Enterprise: A Knowledge and Service Based Paradigm for Industry*. New York: Free Press.

Riessman, F. (1986): The new populism and the empowerment ethos. In *The New Populism: The Politics of Empowerment*, eds. Harry C. Boyte and Frank Riessman. Philadelphia: Temple University Press.

Robinson, M. and L. Bannon (1991): Questioning representations. In *ECSCW'91: Proceedings of the Second European Conference on Computer-Supported Cooperative Work*, eds. Liam Bannon, Mike Robinson, and Kjeld Schmidt. Dordrecht: Kluwer Academic Publishers.

Rose, N. (1990): *Governing the Soul: The Shaping of the Private Self*. London: Routledge.

Roszak, T. (1986): *From Satori to Silicon Valley: San Francisco and the American Counterculture*. San Francisco: Don't Call It Frisco Press.

Sabel, C. F. (1982): *Work and Politics: The Division of Labor in Industry*. Cambridge: Cambridge University Press.

Schmidt, K. (1991): Riding the tiger, or computer supported cooperative work. In *ECSCW'91: Proceedings of the Second European Conference on Computer-Supported Cooperative Work*, eds. Liam Bannon, Mike Robinson, and Kjeld Schmidt. Dordrecht: Kluwer Academic Publishers.

Schmidt, K. and L. Bannon (1992): Taking CSCW seriously: Supporting articulation work. *Computer Supported Cooperative Work*, vol. 1, nos. 1–2, pp. 7–40.

Schutz, A. (1967): *The Phenomenology of the Social World*, (translated by George Walsh and Frederick Lehnert). Evanston: Northwestern University Press. (Originally published in German in 1932.)

Scribner, S. (1984): Studying working intelligence. In *Everyday Cognition: Its Development in Social Context*, eds. Barbara Rogoff and Jean Lave. Cambridge: Harvard University Press.

Shaw, A. G. (1952): *The Purpose and Practice of Motion Study*. Manchester: Harlequin.

Shores, A. R. (1990): *A TQM Approach to Achieving Manufacturing Excellence*. Milwaukee: Quality Press.

Simon, H. A. (1947): *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organization*. New York: Macmillan.

Simon, H. A. (1960): The New Science of Management Decision, New York: Harper and Row.

Simondon, G. (1958): *La Mode d'Existence des Objets Techniques*. Paris: Aubier.

Simons, R. (1995): Control in an age of empowerment. *Harvard Business Review*, vol. 73, no. 2, pp. 80–88.

Star, S. Leigh (1989a): Layered space, formal representations and long-distance control: The politics of information. *Fundamenta Scientiae*, vol. 10, no. 2, pp. 125–155.

Star, S. Leigh (1989b): The structure of ill-structured solutions: boundary objects and heterogeneous distributed problem solving. In *Distributed Artificial Intelligence, Volume II*, eds. Les Gasser and Michael N. Huhns. San Mateo, CA: Morgan Kaufmann, pp. 37–54.

Stinchcombe, A. L. (1990): *Information and Organizations*. Berkeley: University of California Press.

Suchman, L. and B. Jordan (1989): Computerization and women's knowledge. In *Women, Work and Computerization: Forming New Alliances*, eds. Kea Tijdens, Mary Jennings, Ina Wagner, and Margaret Weggelaar. Amsterdam: North-Holland.

Suchman, L. (1992): Technologies of accountability: Of lizards and aeroplanes. In *Technology in Working Order: Studies of Work, Interaction, and Technology*, ed. Graham Button. London: Routledge.

Suchman, L. (1994): Do categories have politics?: The language/action perspective reconsidered. *Computer-Supported Cooperative Work (CSCW)*, vol. 2, no. 3, pp. 177–190.

Szasz, A. (1994): *Ecopopulism: Toxic Waste and the Movement for Environmental Justice.* Minneapolis: University of Minnesota Press.

Taylor, F. W. (1923): *The Principles of Scientific Management.* New York: Harper.

Thompson, P. (1989): *The Nature of Work: An Introduction to Debates on the Labour Process*, second edition. London: Macmillan.

Trist, E. (1981): *The Evolution of Socio-Technical Systems: A Conceptual Framework and an Action Research Program.* Toronto: Ontario Ministry of Labour.

Waring, S. P. (1991): *Taylorism Transformed: Scientific Management Theory Since 1945.* Chapel Hill: University of North Carolina Press.

Weisbord, M. R. (1991): *Productive Workplaces: Organizing and Managing for Dignity, Meaning, and Community.* San Francisco: Jossey-Bass.

Wellins, R. S. Byham, W. C. and J. M. Wilson (1991): *Empowered Teams: Creating Self-Directed Work Groups that Improve Quality, Productivity, and Participation.* San Francisco: Jossey-Bass.

Whalley, P. (1986): Markets, managers, and technical autonomy. *Theory and Society*, vol. 15, no. 2, pp. 223–247.

Winograd, T. (1994): Categories, descriptions, and social coordination, *Computer Supported Cooperative Work* vol. 2, no. 3, pp. 191–197.

Zimbalist, A. (1975): The limits of work humanization. *Review of Radical Political Economics*, vol. 7, no. 2, pp. 50–59.

The Soul Gained and Lost:
Artificial Intelligence as a Philosophical Project

Philip E. Agre
Department of Information Studies
University of California, Los Angeles
Los Angeles, California 90095-1520
USA

pagre@ucla.edu
http://polaris.gseis.ucla.edu/pagre/

This paper appeared in a slightly different form in a special issue of the Stanford Humanities Review, entitled "Constructions of the Mind: Artificial Intelligence and the Humanities", edited by Guven Guzeldere and Stefano Franchi. The official citation is *Stanford Humanities Review* 4(2), 1995, pages 1-19.

9400 words.

1 Introduction

When I was a graduate student in artificial intelligence, the humanities were not held in high regard. They were vague and woolly, they employed impenetrable jargons, and they engaged in "meta-level bickering that never decides anything". Although my teachers and fellow students were almost unanimous in their contempt for the social sciences, several of them (not all, but many) were moved to apoplexy by philosophy. Periodically they would convene impromptu Two-Minute Hate sessions to compare notes on the arrogance and futility of philosophy and its claims on the territory of AI research. "They've had two thousand years and look what they've accomplished. Now it's our turn." "Anything that you can't explain in five minutes probably isn't worth knowing." They distinguished between "just talking" and "doing", where "doing" meant proving mathematical theorems and writing computer programs. A new graduate student in our laboratory, hearing of my interest in philosophy, once sat me down and asked in all seriousness, "Is it true that you don't actually do anything, that you just say how things are?" It was not, in fact, true, but I felt with great force the threat of ostracism implicit in the notion that I was "not doing any real work".

These anecdotes may provide some sense of the obstacles facing any attempt at collaboration between AI and the humanities. In particular, they illustrate certain aspects of AI's conception of itself as a discipline. According to this self-conception, AI is a self-contained technical field. In particular, it is a practical field; to do AI is to prove theorems, write software, and build hardware whose purpose is to "solve" previously defined technical "problems". The whole test of these activities lies in "what works". The criterion of "what works" is straightforward, clear, and objective in the manner of engineering design; arguments and criticisms from outside the field can make no claim at all against it. The substance of the field consists in the "state of the art" and its history is a history of computer programs. The technical methods underlying these programs might have originated in other fields, but the real work consisted in formalizing, elaborating, implementing, and testing those ideas. Fields which do not engage in these painstaking activities, it is said, are sterile debating societies

which do not possess the intellectual tools -- most particularly mathematics -- to do more than gesture in the general direction of an idea, as opposed to really working it out.

I will be thought to exaggerate. Technicians will protest their respect for great literature and the attitudes I have reported will be put down to a minority of fundamentalists. Yet the historical record makes plain that interactions between AI and the humanities have been profoundly shaped by the disciplinary barriers that such attitudes both reflect and reproduce. Serious research in history and literature, for example, has had almost no influence on AI. This is not wholly due to ignorance on the part of technical people, many of whom have had genuine liberal educations. Rather AI, as a technical field, is constituted in such a way that its practitioners honestly cannot imagine what influence those fields *could* have.

Philosophy has had a little more influence. Research on AI's constitutive questions in the philosophy of mind is widely read and discussed among AI research people, and is sometimes included in the curriculum, but these discussions are rarely considered part of the work of AI, judging for example by journal citations, and any influence they might have had on the day-to-day work of AI has been subtle at best. Contemporary ideas from the philosophies of language and logic have been used as raw material for AI model-making, though, and philosophers and technical people have collaborated to some degree in specialized research on logic.

Perhaps the principal humanistic influence on AI has derived from a small number of philosophical critics of the field, most particularly Hubert Dreyfus (1972). For Dreyfus, the project of writing "intelligent" computer programs ran afoul of the critique of rules in Wittgenstein (1968 [1953]) and Heidegger's (1961 [1927]) analysis of the present-at-hand way of relating to beings (in this case, symbolic rules). The use of a rule in any practical activity, Dreyfus argues, requires a prior participation in the culturally specific form of life within which such activities take place. The attempt to fill in the missing "background knowledge" through additional rules would suffer the same problem and thus introduce a fatal regress [1]. Although most senior AI researchers of my acquaintance stoutly deny having been affected by Dreyfus' arguments, a reasonable amount of research has addressed the recurring difficulties with AI research that Dreyfus predicted. One of these is the "brittleness" of symbolic, rule-based AI systems that derives from their tendency to fail catastrophically in situations that depart even slightly from the whole background of operating assumptions that went into the system's design. For the most part, the response of AI researchers to these difficulties, and to Dreyfus' analyses generally, is to interpret them as additions to AI's agenda that require no fundamental rethinking of its premises [2].

Within the field itself, critical reflection is largely a prerogative of the field's most senior members, and even these papers are published separately from the narrowly technical reports, either in non-archival publications like the *AI Magazine* or in special issues of archival publications devoted to the founders' historical reflections. In 1990 I received a referee's report on a AAAI (American Association for Artificial Intelligence) conference paper that read in part,

In general, avoid writing these "grand old man" style papers until you've built a number of specific systems & have *become* a grand old man.

The boundaries of "real AI research", in short, have been policed with great determination.

Yet this is now changing. In part the current changes reflect sociological shifts in the field, in particular its decentralization away from a few heavily funded laboratories and the resulting, albeit modest, trend toward interdisciplinary pluralism. But the change in atmosphere has also been influenced by genuine dissatisfactions with the field's original technical ideas. AI's practices of formalizing and "working out" an idea constitute a powerful method of inquiry, but precisely for this reason they are also a powerful way to force an idea's internal tensions to the surface through prolonged technical frustrations: excessive complexity, intractable inefficiency, difficulties in "scaling up" to realistic problems, and so forth. These patterns of frustration have helped clear the ground for a new conception of technical work, one that recognizes the numerous, deep continuities between AI and the humanities. Although these continuities reach into the full range of humanistic inquiry, I will restrict myself here to the following five assertions about AI and its relationship to philosophy:

1. AI ideas have their genealogical roots in philosophical ideas.
2. AI research programs attempt to work out and develop the philosophical systems they inherit.
3. AI research regularly encounters difficulties and impasses that derive from internal tensions in the underlying philosophical systems.
4. These difficulties and impasses should be embraced as particularly informative clues about the nature and consequences of the philosophical tensions that generate them.
5. Analysis of these clues must proceed outside the bounds of strictly technical research, but they can result in both new technical agendas and in revised understandings of technical research itself.

In short, AI is philosophy underneath. These propositions are not entirely original, of course, and some version of them underlies Dreyfus' early critique of the field. My own purpose here is to illustrate how they might be fashioned into a positive method of inquiry that maintains a dialogue between the philosophical and technical dimensions of AI research. To this end, I will present a brief case study of one idea's historical travels from philosophy through neurophysiology and into AI, up to 1972. Although much of this particular story has been told many times, some significant conclusions from it appear to have escaped analysis. It is an inherently difficult story to tell, since it requires a level of technical detail that may intimidate the uninitiated without nearly satisfying the demands of initiates. It is a story worth telling, though, and I will try to maintain a firm sense of the overall point throughout. I will conclude by briefly discussing recent developments that have been motivated in part by critical reevaluations of this tradition, and by sketching the shape of the new, more self-critical AI that is emerging in the wake of this experience.

2 Rene Descartes: Criteria of intelligence

In a famous passage in his *Discourse on Method*, Descartes summarizes a portion of his suppressed treatise on *The World* as follows:

... the body is regarded as a machine which, having been made by the hands of God, is incomparably better arranged, and possesses in itself movements which are much more admirable, than any of those which can be invented by man. ... if there had been such machines, possessing the organs and outward form of a monkey or some other animal without reason, we should not have had any means of ascertaining that they were not of the same nature as those animals. On the other hand, if there were machines which bore a resemblance to our body and imitated our actions as far as it was morally possible to do so,

we should always have two very certain tests by which to recognize that, for all that, they were not real men. The first is, that they could never use speech or other signs as we do when placing our thoughts on record for the benefit of others. For we can easily understand a machine's being constituted so that it can utter words, and even emit some responses to action on it of a corporeal kind, which brings about a change in its organs; for instance, if it is touched in a particular part it may ask what we wish to say to it; if in another part it may exclaim that it is being hurt, and so on. But it never happens that it arranges its speech in various ways, in order to reply appropriately to everything that may be said in its presence, as even the lowest type of man can do. And the second difference is, that although machines can perform certain things as well as or perhaps better than any of us can do, they infallibly fail short in others, by the which means we may discover that they did not act from knowledge, but only from the disposition of their organs. For while reason is a universal instrument which can serve for all contingencies, these organs have need of some special adaptation for every particular action. From this it follows that it is morally impossible that there should be sufficient diversity in any machine to allow it to act in all the events of life in the same way as our reason causes us to act (page 116).

It is worth quoting Descartes' words at such length because they contained the seeds of a great deal of subsequent intellectual history. Distinctions between the body and the soul were, of course, of great antiquity, as was the idea that people could be distinguished from animals by their reasoned use of language. Descartes, though, extended these ideas with an extremely detailed physiology. His clearly drawn dualism held that automata, animals, and the human body could all be explained by the same mechanistic laws of physics, and he set about partitioning functions between body and mind [3] [4]. In establishing this partition, one of the tests was the conventional distinction between animal capabilities, which reside in the body, and specifically human capabilities, which required the exercise of the soul's faculties of reason and will. Thus, for example, automata or animals might utter isolated words or phrases in response to specific stimuli, but lacking the faculty of reason they could not combine these discrete units of language in an unbounded variety of situationally appropriate patterns. The soul itself has ideas, but it has no physical extent or structure. Thus, as Descartes explains in *The Passions of the Soul* (Articles 42 and 43), memory is a function of the brain; when the soul wishes to remember something, it causes animal spirits to propagate to the spot in the brain where the memory is stored, whereupon the original image is presented once again to the soul in the same manner as a visual perception.

The attraction of Descartes' proposals lay not in their particulars, many of which were dubious even to his contemporaries. Rather, Descartes provided a model for a kind of theory-making that contrasted with late scholastic philosophy in every way: it was specific and detailed, it was grounded in empirical physiology, and it was written in plain language.

3 Karl Lashley: Language as a model for action

The American cognitivists of the 1950's often modeled themselves after Descartes, and they intended their research to have much of the same appeal. Despite the intervening three centuries, the lines of descent are indeed clear. This is evident in the case of Chomsky, for example, who argued in explicitly Cartesian terms for a clear distinction between the physiology of speech, including the biological basis of linguistic competence, and the capacity for actually choosing what to say. While not a dualist, Chomsky nonetheless epitomized his conception of human nature in terms of "free creation within a system of rule" (1971: 50). Miller (1956) used Descartes' *Rules for the Direction of the Mind* to motivate his search for ways that people might more efficiently use their limited memories.

The first and most influential revival of research into mental mechanisms, Karl Lashley's 1951 paper "The problem of serial order in behavior", did not acknowledge any sources beyond the linguistics, psychology, and neurophysiology of the 1940's. Nonetheless, the underlying continuities are important for the computational ideas that followed. Despite his own complex relationship to behaviorism, Lashley's paper argued clearly that behaviorist psychology could not adequately explain the complexity of human behavior. Lashley focused on a particular category of behavior, namely speech. He pointed out that linguists could demonstrate patterns to the grammar and morphology of human languages that are hard to account for using the theory of "associative chains", whereby each action's effects in the world give rise to stimuli that then trigger the next action in turn. The formal structures exhibited by human language, then, were sufficient reason to restore some notion of mental processing to psychology.

Moreover, Lashley suggested that *all* action be understood on the model of language. He regarded both speech and physical movement as having a "syntax", and he sought the physiological basis of both the syntax of movement and the choice of specific movements from among the syntactically possible combinations. This suggestion was enormously consequential for the subsequent development of cognitivist psychology, and particularly for AI. Lashley summarized the idea in this way:

It is possible to designate, that is, to point to specific examples of, the phenomena of the syntax of movement that require explanation, although those phenomena cannot be clearly defined. A real definition would be a long step toward solution of the problem. There are at least three sets of events to be accounted for. First, the activation of the expressive elements (the individual words or adaptive acts) which do not contain the temporal relations. Second, the determining tendency, the set, or idea. This masquerades under many names in contemporary psychology, but is, in every case, an inference from the restriction of behavior within definite limits. Third, the syntax of the act, which can be described as an habitual order or mode of relating the expressive elements; a generalized pattern or schema of integration which may be imposed upon a wide range and a wide variety of specific acts. This is the essential problem of serial order; the existence of generalized schemata of action which determine the sequence of specific acts, acts which in themselves or in their associations seem to have no temporal valence (Lashley 1951: 122).

Two things are new to cognitive theorizing here, grammar as a principle of mental structure and the generalization of grammatical form to all action. But a great deal in Lashley's account is continuous with that of Descartes. To start with, it is an attempt at an architecture of cognition. Indeed, it is considerably less detailed than Descartes' architecture, although Descartes provided no account of the mechanics of speech. Both Lashley and Descartes assign the ability to speak individual words -- or in Lashley's case, to make individual discrete physical movements -- to individual bits of machinery, without being very specific about what these bits of machinery are like. And they both view the human capacity for putting these elements together as the signature of the mind. To be sure, Lashley's argument rests on the formal complexity of speech whereas Descartes points at the appropriateness of each utterance to the specific situation. In each case, though, what counts is the capacity of the mind to order the elements of language in an unbounded variety of ways.

The continuities go deeper. Lashley, as a neurophysiologist, shows no signs of believing in an ontological dualism such as Descartes'. Yet the conceptual *relations* among the various components of his theory are analogous to those of Descartes. In each case, the brain subserves a repertoire of bodily capacities, and on every occasion the mind orders these in

accord with its choices, which themselves are not explained. For Descartes the mind's choices simply *cannot* be explained in causal terms, though its operations can be described in the normative terms of reason, as for example in his *Rules for the Direction of the Mind*. Lashley does not express any overt skepticism about his "determining tendency", but neither does he have anything very definite to say about it; the concept stays nebulous throughout. This is not simply an incompleteness of Lashley's paper but is inherent in its design: the purpose of the determining tendency is not to *have* structure in itself but to *impose* structure upon moment-to-moment activities from the repertoire of action schemata made available to it by the brain.

In retrospect, then, Lashley's paper makes clear the shape of the challenge that the cognitivists had set themselves. They wished to rout their sterile behaviorist foes in the same way that Descartes had routed the schoolmen, by providing a scientific account of cognitive processes. The problem, of course, is that Descartes was not a thoroughgoing mechanist. So long as the cognitivists retained the relational system of ideas that they had inherited from Descartes, and from the much larger tradition of which Descartes is a part, each of their models would include a component corresponding to the soul. No matter how it might be squeezed or divided or ignored, there would always remain one black box that seemed fully as intelligent as the person as a whole, capable of making intelligent choices from a given range of options on a regular basis. As the field of AI developed, this recalcitrant box acquired several names. Dennett (1978: 80-81), for example, spoke of the need for "discharging the homunculus", something he imagined to be possible by dividing the intelligent homunculus into successively less intelligent pieces, homunculi within homunculi like the layers of an onion, until one reached a homunculus sufficiently dumb to be implemented in a bit of computer code. AI researchers' jargon spoke of subproblems as being "AI-complete" (an analogy: so-called NP-complete computational problems are thought to be unsolvable except through an enumeration of possible solutions -- an efficient algorithm for any one such problem would yield efficient algorithms for all of them). And several exceedingly skilled programmers devised computer systems that were capable of reasoning about their own operation -- including reasoning about their reasoning about their own operation, and so on *ad infinitum* (e.g., Smith 1985). In each case, the strategy was reducing the soul's infinite choices to finite mechanical means.

But beyond sketching the shape of a future problem, Lashley also sketched the principal strategy of a whole generation for solving it. The operation of the determining tendency might be a mystery, but the general form of its accomplishment was not. While the linguistic metaphor for action envisions an infinite variety of possible actions, it also imposes a great deal of structure on them. In mathematical terms the possible actions form a "space". The generative principle of this space lies in the "schemata of action", which are modeled on grammatical rules. A simple schema for English sentences might be

Sentence -> NounPhrase IntransitiveVerb .

That is, roughly speaking "one way to make a sentence is to utter a noun phrase followed by an intransitive verb". Other rules might spell out these various "categories" further; for example,
NounPhrase -> Article Noun
Article -> a
Article -> the
Noun -> cat
Noun -> dog

IntransitiveVerb -> slept
IntransitiveVerb -> died

These mean, roughly, "one way to make a noun phrase is to utter an article followed by a noun", "some possible articles are "a" and "the" ", "some possible nouns are "cat" and "dog" ", and "some possible intransitive verbs are "slept" and "died" ". And there might be other ways to make sentences; for example,
Sentence -> NounPhrase TransitiveVerb NounPhrase
TransitiveVerb -> saw
TransitiveVerb -> ate

This particular set of grammatical rules generates a finite space of English sentences; for example,
the cat saw a dog
a dog ate a dog

The process of "deriving" a sentence with these rules is simple and orderly. One begins with the "category" Sentence, and then at each step one makes two choices: (1) which category to "expand", and (2) which rule to apply in doing so, until no categories are left. For example, one might proceed as follows:
1. Sentence
2. NounPhrase TransitiveVerb NounPhrase
3. NounPhrase TransitiveVerb Article Noun
4. NounPhrase saw Article Noun
5. Article Noun saw Article Noun
6. the Noun saw Article Noun
7. the Noun saw Article dog
8. the cat saw Article dog
9. the cat saw a dog

The space of possible sentences, then, resembles a branching road with a definite set of choices at each point. The process of choosing a sentence is reduced to a series of much smaller choices among a small array of alternatives. The virtue of this reduction becomes clearer once the grammar generates an infinite array of sentences, as becomes the case when the following grammatical rules are added to the ones above:
Sentence -> NounPhrase CognitiveVerb that Sentence
CognitiveVerb -> thought
CognitiveVerb -> forgot

It now becomes possible to generate sentences such as
the cat thought that the dog forgot that a cat slept

Chomsky (1965: 8) in particular made a great deal of this point; following Humboldt, he spoke of language as making "infinite use of finite means". And although he believes that the mind ultimately has a biological (and thus mechanical) explanation (1979: 66, 97), he has focused his research on the level of grammatical competence rather than trying to uncover this explanation himself.

4 Allen Newell and Herbert Simon: The mechanization of the soul

Instead, the first steps in mechanizing this idea of a generative space were due to Newell and Simon (1963). Whereas Chomsky was concerned simply with the precise extent of the generative space of English grammar, Newell and Simon's computer program had to make actual choices within a generative space. And whereas Lashley posited the existence of a

"determining tendency" whose genealogical origins lay in a non-mechanical soul, Newell and Simon had to provide some mechanical specification of it. Here the generative structure of the space was crucial. Newell and Simon did not employ linguistic vocabulary. Nonetheless, just as grammatical rules and derivations provide a simple, clear means of generating any grammatical sentence, the application of "operators" provided Newell and Simon with a simple, easily mechanized means of generating any possible sequence of basic actions. Choosing *which* sequence of actions to adopt was a matter of "search". The mechanism that conducted the search did not have to make correct choices all the time; it simply had to make good enough choices eventually as it explored the space of possibilities.

Newell and Simon placed enormous significance on this idea (see, for example, Newell's comments in Agre (1993: 418)), and justifiably so. While maintaining the system of conceptual relations already found in Descartes, Lashley, and Chomsky, their program nonetheless embodied a serious proposal for the mechanization of the soul (cf Gallistel 1980: 6-7). Their strategy was ingenious: rather than endow the soul with an internal architecture -- something incomprehensible within the system of ideas they inherited -- they effectively proposed interpreting the soul as an epiphenomenon. Ironically, given Descartes' polemics against scholastic philosophy, the idea is approximately Aristotelian: the soul as the form of the person, not a discrete component. More specifically, rather than being identified with any particular device, the soul was *contained* by the generative structure of the search space and *manifested* through the operation of search mechanisms. These search mechanisms were "heuristic" in the sense that no single choice was ever guaranteed to be correct, yet the overall effect of sustained searching was the eventual discovery of a correct outcome. Despite the simplicity and limitations of their early programs, Newell and Simon were willing to refer to these programs' behavior as "intelligent" because they met this criterion. And they regarded their proposal as promising because so many human activities could readily be cast as search problems.

Up to this point, the story of the mechanization of the soul is a conventional chapter in the history of ideas: to tell this story, we trace the unfolding of an intellectual project within an invariant framework of continuities or analogies among idea-systems. With Newell and Simon's program, though, the story clearly changes its character. But how exactly? So far as the disciplinary culture of AI is concerned, the formalization and implementation of an idea bring a wholly new day -- a discontinuity between the prehistory of (mere) questions and ideas and the history, properly speaking, of problems and techniques. Once this proper history has begun, technical people can put their proposals to the test of implementation: either it works or it does not work.

Yet despite this conception, and indeed partly because of it, the development of technical methods can be seen to continue a along trajectory largely determined by the defining projects and internal tensions of the ancestral systems of ideas. In particular, these projects and tensions continue to manifest themselves in the goals and tribulations of AI's technical work. In the case of Newell and Simon's proposal, the central goals and tribulations clustered around the "problem" of *search control* -- that is, making heuristic search choices well enough -- not perfectly, just well enough -- to allow the search process to "terminate" with an acceptable answer within an acceptable amount of time. An enormous AI subliterature addresses this problem in a wide variety of ways. Within this literature, searches are said to "explode" because of the vastness of search spaces. It should be emphasized that mediocre search control ideas do not kill a mechanism; they only slow it down. Yet this research has long faced a troubling aporia: the more complicated the world is, the more choices become

possible at each point in the search, and the more ingenuity is required to keep the search process under control. The metaphors speak of a struggle of containment between explosion and control. Such a struggle, indeed, seems inherent in any theory for which action is said to result from formal reason conducted by a finite being (Cherniak 1986).

Newell and Simon's achievement thus proved tenuous. So long as AI's self-conception as a self-sufficient technical discipline has remained intact, however, these difficulties are readily parsed as technical problems seeking technical solutions. An endless variety of solutions to the search control problem has indeed been proposed, and each of them more or less "works" within the bounds of one or another set of "assumptions" about the world of practical activity.

5 Richard Fikes and Nils Nilsson: Mechanizing embodied action

To watch the dynamics of this process unfold, it will help to consider one final chapter: the STRIPS program (Fikes and Nilsson 1971). The purpose of STRIPS is to automatically derive "plans" for a robot to follow in transporting objects around in a maze of rooms. The program constructs these plans through a search process modeled on those of Newell and Simon [5]. The search space consists of partially specified plans, with each "operator" adding another step to the plan. Returning to the linguistic metaphor, the authors of STRIPS understand the robot's action within a grammar of possible plans. They refer to the units of action that Lashley called "expressive elements" as "primitive actions", and the "syntax of the act" strings these actions into sequences that can be "executed" in the same manner as a computer program. In Descartes' terms, the soul's faculty of reason specifies an appropriate sequence of bodily actions, each of which may well be complicated, its faculty of will decides to undertake them, and the body then physically performs them.

To those who have had experience getting complex symbolic programs to work, the STRIPS papers make intense reading. Because the authors were drawing together so many software techniques for the first time, the technically empathetic reader gets a vivid sense of struggle -- the unfolding logic of what the authors unexpectedly felt compelled to do, given what seemed to be required to get the program to work. A detailed consideration of the issues would take us much too far afield, but the bottom line is easy enough to explain. As might be expected, this bottom line concerns the technical practicalities of search control. A great deal is at stake: if the search can be controlled without making absurdly unrealistic assumptions about the robot's world, then the program can truly be labeled "intelligent" in some non-trivial sense.

Consider, though, what this search entails. The STRIPS program is searching for a correct plan -- that is, a plan which, if executed in the world as it currently stands, would achieve a given goal. This condition -- achieving a given goal -- is not simply a property of the plan; it is a property of the robot's interactions with its world. In order to determine whether a given plan is correct, then, the program must effectively conduct a simulation of the likely outcome of each action. For example, if a candidate plan contains the primitive action "step forward", it matters whether the robot is facing a wall, a door, a pile of rubbish, or an open stretch of floor. If "step forward" is the *first* step in the plan then the robot can predict its outcome simply by activating its video camera and looking ahead of itself. But if "step forward" is the seventh step in the plan, subsequent to several other movements, then complex reasoning will be required to determine its likely outcome.

This is a severely challenging problem, and Fikes and Nilsson approached it, reasonably enough given the state of computer technology in 1971, through brute force: they encoded the

robot's world in the form of a set of formulae in the predicate calculus, and they incorporated into STRIPS a general-purpose program for proving predicate-calculus theorems by means of a search through the space of possible formal proofs. This approach "works" in the same sense that any search method works: if the search ever terminates then the answer is correct, but how long this takes depends heavily on the perspicacity of the program's search control policies. And adequately perspicuous search control policies are notoriously elusive. As programmers like Fikes and Nilsson quickly learned, the trick is to design the world, and the robot's representations of the world, in such a way that long, involved chains of reasoning are not required to predict the outcomes of actions.

Yet predicting the outcomes of actions was, as programmers say, only the "inner loop" of the plan-construction process. Recall that the overall process of choosing possible actions is also structured as a search problem; extending Lashley's linguistic metaphor, it is as if the grammaticality of a spoken sentence depended on the listener's reaction to each successive word. Moreover, the space of possible plans is enormous: at any given time, the robot can take any of about a dozen primitive actions, depending on its immediate circumstances, and even a simple plan will have several steps. Once again, search control policies are crucial. At each point in the search process, the program must make two relatively constrained choices among a manageable list of options: it must choose a partially specified plan to further refine, and it must choose a means of further refining it -- roughly speaking, it must add another primitive action to the plan.

As with any search, making these choices correctly every time would require "intelligence" that no mechanism could probably possess. The point, instead, is to make the choices correctly *often enough* for the search to settle on a correct answer in a reasonable amount of time. This, once again, is the appeal of heuristic search: intelligent action emerges from a mass of readily mechanizable decisions. In other words, the problem for Fikes and Nilsson was that they had to write bits of code whose outcomes approximated two hopelessly uncomputable notions: "partially specified plan most likely to lead to a correct plan" and "best primitive action to add to this subplan". Their solution to these problems was unsurprising in technical retrospect, and the details do not matter here. Briefly, they chose whatever partially specified plan seemed to have gotten the furthest toward the goal with the smallest number of primitive actions, and they chose a new primitive action that allowed the theorem-proving program to make further progress toward proving that the goal had been achieved. Both of these criteria are virtually guaranteed to lead the plan-construction process down blind alleys (such as telling the robot to head for the door before getting the key). The important point is that these blind alleys did not hurt the robot; they only kept the robot waiting longer to be given a plan to execute.

How big a step was the STRIPS program toward mechanized intelligence? Reasonable people could disagree. It is certainly an impressive thing to watch such a program in operation -- provided you have long enough to wait. But the question of search control was daunting. To the AI research people of that era, search control in STRIPS-like plan-construction programs was a "problem" to be addressed through a wide variety of technical means. Yet this approach accepts as given the underlying structure of the situation: a steep trade-off between the complexity of the world and the practicality of the search control problem. If the robot can perform many possible actions, or if the results of these actions depend in complex ways on the circumstances, then the search space grows rapidly -- in mathematical terms, exponentially -- in size. And if it is impossible to predict the outcomes of actions -- say because the robot is not the only source of change in the world -- then the

search space will have to include all of the *possible* outcomes as well. In a prescient aside in the sequel to the original STRIPS paper, Fikes, Hart, and Nilsson (1972) pointed this out:

One of the novel elements introduced into artificial intelligence research by work on robots is the study of execution strategies and how they interact with planning activities. Since robot plans must ultimately be executed in the real world by a mechanical device, as opposed to being carried out in a mathematical space or by a simulator, consideration must be given by the executor to the possibility that operations in the plan may not accomplish what they were intended to, that data obtained from sensory devices may be inaccurate, and that mechanical tolerances may introduce errors as the plan is executed.

Many of these problems of plan execution would disappear if our system generated a whole new plan after each execution step. Obviously, such a strategy would be too costly, so we instead seek a plan execution scheme with the following properties:

(1) When new information obtained during plan execution implies that some remaining portion of the plan need not be executed, the executor should recognize such information and omit the unneeded plan steps.

(2) When execution of some portion of the plan fails to achieve the intended results, the executor should recognize the failure and either direct reexecution of some portion of the plan or, as a default, call for a replanning activity (1972: 268).

Thus, although they recognized the tension that was inherent in the system of concepts they had inherited, the technical imagination of that time provided Fikes, Hart, and Nilsson with no other way of structuring the basic question of intelligent action. It was fifteen years before the inherent dilemma of plan-construction was given definite mathematical form, first by Chapman (1987) and then more compactly by McAllester and Rosenblitt (1991). This kind of research does not decisively discredit the conceptual framework of planning-as-search; rather it clarifies the precise nature of the trade-offs generated by that framework. And indeed, productive research continues to this day into the formal structure of plan-construction search problems.

6 Beyond the Cartesian soul

The previous sections offer a critical reconstruction of a single strand of intellectual history, a single intellectual proposition worked out in increasingly greater technical detail so that its internal tensions become manifest. To diagnose the resulting impasse and move beyond it, it will be necessary to transcend AI's conception of itself as a technical, formalizing discipline, and instead to reconsider the larger intellectual path of which AI research has been a part. No matter how esoteric AI literature has become, and no matter how thoroughly the intellectual origins of AI's technical methods have been forgotten, the technical work of AI has nonetheless been engaged in an effort to domesticate the Cartesian soul into a technical order in which it does not belong. The problem is not that the individual operations of Cartesian reason cannot be mechanized (they can be) but that the role assigned to the soul in the larger architecture of cognition is untenable. This incompatibility has shown itself up in a pervasive and ever more clear pattern of technical frustrations. The difficulty can be shoved into one area or another through programmers' choices about architectures and representation schemes, but it cannot be made to go away.

This impasse, though, is not a failure. To the contrary, tracing the precise shape of the impasse allows us to delineate with particular confidence the internal tensions in the relational system of ideas around the Cartesian soul. According to this hypothesis, the fundamental embarrassment of Descartes' theory does not lie in the untenability of ontological dualism. Rather, it lies in the soul's causal distance from the world of practical action. As this world grows more complex (or, more precisely, as one's representational schemes reflect this world's complexity more fully), and as one becomes more fully immersed in that world, the soul's job becomes astronomically difficult. Yet Descartes performed his analysis of the soul in sedentary conditions: introspecting, visualizing, and isolating particular episodes of perception. When he did discuss complex activities, he focused not on the practicalities of their organization but on the struggles they engendered between the body and the soul (see, for example, *Passions of the Soul*, Article 47).

In order to impose intelligent order on its body's actions, the Cartesian soul faces a stern task. For example, to visualize a future course of events, the soul must stimulate the brain to assemble the necessary elements of memory. The reasoning which guides this visualization process must be based in turn upon certain knowledge of the world, obtained through the senses -- enough information to visualize fully the outcomes of the individual's planned sequence of actions. Our judgement that such a scheme places an excessive burden on the soul -- or, as technical people would say, makes the soul into a "bottleneck" -- is not a logical refutation; it is only an engineer's embodied judgement of the implausibility of a design. But within the logic of Descartes' project that is a lot.

The underlying difficulty takes perhaps its clearest form in Lashley. At the beginning of his lecture, he opposes his own view to the behaviorist and reflexological tale of stimuli and responses as follows:

My principal thesis today will be that the input is never into a quiescent or static system, but always into a system which is already actively excited and organized. In the intact organism, behavior is the result of interaction of this background of excitation with input from any designated stimulus. Only when we can state the general characteristics of this background of excitation, can we understand the effects of a given input (page 112).

In contradistinction to a scheme that focuses upon the effects of an isolated stimulus, Lashley proposes giving due weight both to a stimulus and to the ongoing flux of brain activity in which the stimulus intervenes. People, in other words, are always thinking as well as interacting with the world. Having said this, though, he immediately gives priority to the internal "background" of neural activity, and his paper never returns to any consideration of external stimuli and their effects. As with his silence about the nature of the determining tendency, this is not a simple omission but is intrinsic to his relational system of concepts. His analysis of action on the model of speech portrays speakers as laying out a complex series of sounds through internal processing and then producing them in a serially ordered sequence, without in any way interacting with the outside world [6]. As we have already seen in the case of STRIPS, this obscurity about the relationship between "planning" (of action sequences) and "interaction" (with the world while those actions are going on) structured cognitive theorizing about action, and AI research in particular, for many years afterward [7].

It is precisely this pattern of difficulty that has impelled an emerging interdisciplinary movement of computational modelers to seek a conception of intelligent behavior whose focal point is the fullness of embodied activity, not the reticence of thought. An organizing theme of this movement is the principled characterization of interactions between agents and

their environments, and the use of such characterizations to guide design and explanation. When the "agents" in question are robots, this theme opens out onto a systems view of robotic activity within the larger dynamics of the robot's world. When the "agents" are animals, it opens out onto biology, and specifically onto a conception of ethology in which creatures and their behavior appear thoroughly adapted to the dynamics of a larger ecosystem. When the "agents" in question are people, it opens out onto philosophical and anthropological conceptions of human beings as profoundly embedded in their social environments. In lieu of detailed references to these directions of research, allow me to me direct the reader to an issue of *Artificial Intelligence* on Computational Theories of Interaction and Agency that will appear in 1995.

7 AI and the humanities

I have argued that AI can become sterile unless it maintains a sense of its place in the history of ideas -- and in particular unless it maintains a respect for the power of inherited systems of ideas to shape our thinking and our research in the present day. At the same time, AI also provides a powerful means of forcing into the open the internal structure of a system of ideas and the internal tensions inherent in the project of getting those ideas to "work". Thus, AI properly understood ought to be able to participate in a constructive symbiosis with humanistic analyses of ideas.

Putting this mode of cooperative work into practice will not be easy. The obstacles are many and varied, but I believe that the most fundamental ones pertain to the use in AI of mathematics and mathematical formalization. This is not the place for a general treatment of these topics, but it is possible at least to outline some of the issues. The most obvious issue, perhaps, is the symbolic meaning attached to mathematics in the discursive construction of technical disciplines. Technical people frequently speak of mathematics as "clean" and "precise", as opposed to the "messy" and "vague" nature of the social world and humanistic disciplines. These metaphors clearly provide rich points of entry for critical research, but the important point here is that their practical uses go beyond the simple construction of hierarchies among disciplines. Most particularly, the notion of mathematics as the telos of reason structures AI researchers' awareness in profound ways.

To see this, let us briefly consider the role that mathematics plays in AI research. The business of AI is to build computer programs whose operation can be narrated in language that is normally used in describing human activities (Collins 1990). Since the function of computers is specified in terms of discrete mathematics, the daily work of AI includes the complex and subtle discursive practice of talking about human activities in ways that assimilate them to mathematical structures [8]. In the case of the computational models of action described above, this assimilation is achieved by means of a linguistic metaphor for action. This metaphor is not specific to AI; in fact it structures a great deal of the practice of applied computing (Agre 1994) [9]. And this fact in turn points to an inherent source of intellectual conservatism in AI: the field is not restricted *a priori* to speaking of human beings in particular terms, but it *is* restricted to speaking in terms that someone knows how to assimilate to mathematical structures that can be programmed on computers. In this way, the existing intellectual infrastructure of computing -- its stock of discursive forms and technical methods -- drags like an anchor behind any project that would reinvent AI using language drawn from alternative conceptions of human beings and their lives.

This observation goes far toward explaining the strange appearance that AI presents to fields such as literature and anthropology that routinely employ much more sophisticated and critically reflective conceptions of human life. The first priority for AI research is to get something working on a computer, and the field does not reward gnawing doubts about whether the conceptions of human life being formalized along the way are sufficiently subtle, accurate, or socially responsible -- thus the emphasis, mentioned at the outset, on "doing" as opposed to "just talking". Critical methods from the humanities are likely to appear pointless, inasmuch as they do not immediately deliver formalizations or otherwise explain what programs one might write. AI people see formalization as a trajectory with an endpoint, in which the vagueness and ambiguity of ordinary language are repaired through mathematical definition, and they are not greatly concerned with the semantic violence that might be done to that language in the process of formal definition. A word like "action" might present real challenges to a philosophical project that aims to respect ordinary usage (e.g., White 1968), but the assimilation of action to formal language theory reduces the word to a much simpler form: a repertoire of possible "actions" assembled from a discrete, finite vocabulary of "expressive elements" or "primitives". Having thus taken its place in the technical vocabulary of AI, the word's original semantic ramifications are lost as potential resources for AI work. The ideology surrounding formalization accords no intrinsic value to these left-over materials. As a result, formalization becomes a highly organized form of social forgetting -- and not only of the semantics of words but of their historicity as well. This is why the historical provenance and intellectual development of AI's underlying ideas claim so little interest among the field's practitioners.

What would a reformed AI look like? It would certainly not reject or replace mathematics. Rather, it would draw upon critical research to cultivate a reflexive awareness of how mathematical formalization is used as part of the engineer's embodied work of building things and seeing what they do. In particular, it would cultivate an awareness of the cycle of formalization, technical working-out, the emergence of technical impasses, the critical work of diagnosing the impasse as reflecting either a superficial or a profound difficulty with the underlying conception of action, and the initiation of new and more informed rounds of formal modeling. The privilege in this cycle does not lie with the formalization process, nor does it lie with the critical diagnosis of technical impasses. Rather, it lies with the cycle itself, in the researcher's "reflective conversation with the materials" of technical and critical work (Schon 1983).

Humanistic critical practice can take up numerous relationships, cooperative or not, to this cycle of research. My own analysis in this paper has employed a relatively old-fashioned set of humanistic methods from the history of ideas, tracing the continuity of certain themes across a series of authors and their intellectual projects. Since formalization is a fundamentally metaphorical process, discursively interrelating one set of things with another, mathematical set, it can be particularly fruitful to trace the historical travels of a given metaphor among various institutional sites in society, technical and otherwise (Martin 1987, McReynolds 1980, Mirowski 1989). The purpose in doing so is not simply to debunk any claims that technical institutions might make to an ahistorical authority, but to prevent the passage to formalism from forgetting the underlying commitments that a given way of speaking about human activities draws from its broader cultural embedding [10]. This contextual awareness will be crucial when the technical research reaches an impasse and needs to be diagnosed as a manifestation of internal tensions of the underlying system of ideas. Any given set of ideas will be more easily given up when they are seen as simply one path among many others not taken. Indeed this awareness of context will be crucial for

recognizing that an impasse may have occurred in the first place. Viewed in this way, technical impasses are a form of social remembering, moments when a particular discursive form deconstructs itself and makes its internal tensions intelligible to anyone who is critically equipped to hear them.

The cycle of reaching and interpreting technical impasses, moving back and forth between technical design and critical inquiry, can be practiced on a variety of scales, depending upon the acuity of one's critical methods. The example I traced in the body of this paper was extremely coarse: whole decades of research could be seen in hindsight to have been working through a single, clear-cut intellectual problem. The difficulty was not that AI practitioners were insulated from the philosophical critiques of Cartesian reason that might have provided a diagnosis of their difficulties and defined the contours of alternative territories of research. To the contrary, Hubert Dreyfus was articulating some of these critiques all along. The real difficulty was that the critical apparatus of the field did not provide its practitioners with a living, day-to-day appreciation for the contingent nature of their formalisms. Although they viewed formalization as conferring upon language a cleanliness and precision that it did not otherwise possess, the effect was precisely the reverse. Lacking a conscious awareness of the immense historicity of their language, they could not understand it as it called out to them the very things they had discovered. A reformed technical practice would employ the tools of critical inquiry to engage in a richer and more animated conversation with the world.

*** Footnotes**

[1] For a detailed analysis of this argument see Preston (1993).

[2] Dreyfus, in joint work with Stuart Dreyfus, has been cautiously supportive of one alternative AI research program, the "connectionist" attempt to build simulations of neural circuitry without necessarily formulating "knowledge" in terms of symbolic "rules" (Dreyfus and Dreyfus 1988). But as the Dreyfuses points out, this research program still faces a long, difficult learning curve and will not be discussed here.

[3] The terms "mechanism" and "mechanistic" require further analysis than space permits here. Suffice it to say that a mechanism is a physical object whose workings are wholly explicable in causal terms. To speak of something as a mechanism, furthermore, is to insert it into a rhetoric of engineering design, whether divine or human, and whether on the model of the clockmaker or the computer programmer. For the modern mathematical intepretations of the term, which are obviously relevant to the foundations of computing if not immediately to the genealogy being traced here, see Webb (1980).

[4] Note that the intellectual culture of Descartes' day did not distinguish between "mind" and "soul", and the two terms continue to be used interchangeably in Catholic philosophy to this day; see for example Holscher (1986). Even in the present day, these terms are usually not so much opposed as simply employed in different discourses with overlapping genealogies.

[5] Chapman (1987) presents a genealogy of the AI "planning" systems in this lineage.

[6] Actual human speakers frequently do interact with their addressees and others during the real-time production of their utterances (Goodwin 1981), but this fact is rarely taken into account in cognitive theories of grammar and speech.

[7] It is particularly clear in the opening chapter of Miller, Galanter, and Pribram's influential book *Plans and the Structure of Behavior* (1960).

[8] This is obviously an attribute that AI shares with a wide variety of other fields, for example mathematical economics, and much of the analysis here applies to these other fields as well. It should be noted that AI people themselves place great emphasis on a distinction between "neat" forms of AI, which openly avow their commitment to mathematical formalization and employ large amounts of mathematical notation in their papers, and "scruffy" forms, which do not (Forsythe 1993). My argument, though, applies equally to both forms of AI research. Regardless of whether its author was consciously thinking in terms of mathematics, a computer program is a notation whose operational semantics can be specified in mathematical terms. While the formalizations in "neat" research are frequently more consistent, systematic, and explicit than those of "scruffy" research, the design of any computer program necessarily entails a significant level of formalization.

[9] Different linguistic metaphors for human action are obviously possible, if perhaps equally problematic; see for example Ricoeur (1971).

[10] For an impressive cultural analysis of the origins of AI, see Edwards (1996).

*** Acknowledgements**

This paper has been improved by comments from Harry Collins, Guven Guzeldere, Scott Mainwaring, Beth Preston, and Joszef Toth.

*** References**

Philip E. Agre, Interview with Allen Newell, *Artificial Intelligence* 59(1-2), 1993, pages 415-449.

Philip E. Agre, Surveillance and capture: Two models of privacy, *The Information Society* 10(2), 1994, pages 101-127.

David Chapman, Planning for conjunctive goals, *Artificial Intelligence*, 32(3), 1987, pages 333-377.

Christopher Cherniak, *Minimal Rationality*, Cambridge: MIT Press, 1986.

Noam Chomsky, *Aspects of the Theory of Syntax*, Cambridge: MIT Press, 1965.

Noam Chomsky, *Problems of Knowledge and Freedom: The Russell Lectures*, New York: Pantheon, 1971.

Noam Chomsky, *Language and Responsibility*, translated from the French by John Viertel, New York: Pantheon, 1979.

Harry M. Collins, *Artificial Experts: Social Knowledge and Intelligent Machines*, Cambridge: MIT Press, 1990.

Daniel Dennett, Why the law of effect will not go away, in *Brainstorms: Philosophical Essays on Mind and Psychology*, Montgomery, VT: Bradford Books, 1978.

Rene Descartes, *The Philosophical Works of Rene Descartes*, translated by Elizabeth S. Haldane and G. R. T. Ross, volume 1, Cambridge: Cambridge University Press, 1972.

Hubert L. Dreyfus, *What Computers Can't Do: A Critique of Artificial Reason*, New York: Harper and Row, 1972.

Hubert L. Dreyfus and Stuart Dreyfus. Making a mind vs. modeling the brain: AI back at a branchpoint, *Daedalus* 117(1), 1988, pages 15-43.

Paul Edwards, *The Closed World: Computers and the Politics of Discourse in Cold War America*, Cambridge: MIT Press, 1996.

Richard E. Fikes and Nils J. Nilsson, STRIPS: A new approach to the application of theorem proving to problem solving, *Artificial Intelligence* 2(3), 1971, pages 189-208.

Richard E. Fikes, Peter E. Hart, and Nils J. Nilsson, Learning and executing generalized robot plans, *Artificial Intelligence* 3(4), 1972, pages 251-288.

Diane E. Forsythe, Engineering knowledge: The construction of knowledge in artificial intelligence, *Social Studies of Science* 23(3), 1993, pages 445-477.

C. R. Gallistel, *The Organization of Action: A New Synthesis*, Hillsdale, NJ: Erlbaum, 1980.

Charles Goodwin, *Conversational Organization: Interaction Between Speakers and Hearers*, New York: Academic Press, 1981.

Martin Heidegger, *Being and Time*, translated by John Macquarrie and Edward Robinson, Harper and Row, 1961. Originally published in German in 1927.

Ludger Holscher, *The Reality of the Mind: Augustine's Philosophical Arguments for the the Human Soul as a Spiritual Substance*, London: Routledge, 1986.

Karl S. Lashley, The problem of serial order in behavior, in Lloyd A. Jeffress, ed, *Cerebral Mechanisms in Behavior: The Hixon Symposium*, New York: Wiley, 1951.

Emily Martin, *The Woman in the Body: A Cultural Analysis of Reproduction*, Boston: Beacon Press, 1987.

David McAllester and David Rosenblitt, Systematic nonlinear planning, *Proceedings of the National Conference on Artificial Intelligence*, Los Altos, CA: Morgan Kaufmann Publishers, 1991, pages 634-639.

Paul McReynolds, The clock metaphor in the history of psychology, in Thomas Nickles, ed, *Scientific Discovery: Case Studies*, Dordrecht: Reidel, 1980.

George A. Miller, Information and memory, *Scientific American* 195(2), 1956, pages 42-46.

George A. Miller, Eugene Galanter, and Karl H. Pribram, *Plans and the Structure of Behavior*, New York: Henry Holt and Company, 1960.

Philip Mirowski, *More Heat Than Light: Economics as Social Physics, Physics as Nature's Economics*, Cambridge: Cambridge University Press, 1989.

Allen Newell and Herbert A. Simon, GPS: A program that simulates human thought, in Edward A. Feigenbaum and Julian Feldman, eds, *Computers and Thought*, New York: McGraw-Hill, 1963.

Elizabeth F. Preston, Heidegger and artificial intelligence, *Philosophy and Phenomenological Research* 53(1), 1993, pages 43-69.

Paul Ricoeur, The model of the text: Meaningful action considered as a text, *Social Research* 38, 1971, pages 529-562.

Donald A. Schon, *The Reflective Practitioner: How Professionals Think in Action*, New York: Basic Books, 1983.

Brian C. Smith, Prologue to *Reflection and Semantics in a Procedural Language*, in Ronald J. Brachman and Hector J. Levesque, eds, *Readings in Knowledge Representation*, Los Altos, CA: Morgan Kaufmann, 1985.

Judson Chambers Webb, *Mechanism, Mentalism, and Mathematics: An Essay on Finitism*, Dordrecht: Reidel, 1980.

Alan R. White, ed, *The Philosophy of Action*, London: Oxford University Press, 1968.

Ludwig Wittgenstein, *Philosophical Investigations*, third edition, translated by G. E. M. Anscombe, New York: Macmillan, 1968. Originally published in 1953.

The Architecture of Identity: Embedding Privacy in Market Institutions

Philip E. Agre
Department of Information Studies
University of California, Los Angeles
Los Angeles, California 90095-1520
USA

pagre@ucla.edu
http://polaris.gseis.ucla.edu/pagre/

***Information, Communication and Society* 2(1), 1999, pages 1-25.**

Please do not quote from this version, which may differ slightly from the version that appears in print.

8200 words.

Abstract. As the Internet becomes integrated into the institutional world around it, attention has increasingly been drawn to the diverse ways in which information technologies mediate human relationships. And as an increasingly commercial Internet has been employed to capture personally identifiable information, privacy concerns have intensified. To analyze these matters more systematically, this article considers the ideas about human identity that have been implicit in the development of economics and computer science. The two fields have evolved along parallel tracks, starting with an assumption of perfect transparency and moving toward a more sophisticated appreciation of individuals' private informational states. Progress in the analysis and resolution of privacy problems will require that this evolution be taken seriously and continued.

1 Introduction

A minor industry is trying to give intellectual shape to the widespread intuition that information technology is bringing about a revolution in human affairs. Even to call it a revolution, of course, is already to have a theory; the social changes that accompany information technology are likened to violent, discontinuous changes in a political order, or to epistemological discontinuities in science. One template for understanding the changes is the industrial revolution, with its extraordinary increases in productivity. The problem with this approach, notoriously, is that little evidence supports it: the aggregate productivity increases that the analogy would predict are nowhere to be seen (Landauer 1995, Madrick 1998; cf Brynjolfsson and Hitt 1998). Another approach sees the Internet writing its decentralized architecture onto the economic and political workings of the global society. This approach, too, conflicts with the evidence, as networked computing facilitates the most dramatic period of industrial concentration in human history.

Yet for all the cold water, revolution is still somehow in the air; we are still challenged to make sense of changes that, whatever their eventual magnitude may be, are strikingly pervasive. To understand them, we need to focus on another of the implications of this talk of revolution: the idea that human relationships, and perhaps human nature itself, will be transformed. Computers increasingly mediate human relationships, this argument goes, and so therefore the computers and the relationships, if they change, will necessarily change together. Revolutions, of course, have frequently been held to be impossible precisely because human nature does not change overnight. Perhaps human nature does not change at all past a certain point, and perhaps as a consequence the possibilities for computer-induced revolution are limited as well. I cannot evaluate these suggestions, but I can offer some account of what they mean.

No simple analysis of these matters will suffice. It is no good to infer changes in human life simply from the workings of machines; such an approach is inevitably reductionistic and can never do justice to the complexity of human life (Kling 1991). Machines cannot be understood as neutral tools, given the immense effort that organizations put into inscribing their own contingent selves into the machines' workings (George and King 1991). Nor can they be understood as mere epiphenomena of larger social processes, given the immense creativity that user communities put into appropriating them (Feenberg 1995).

To understand the place of information technology in society, we need to take an institutional approach. Institutions are the persistent structures of social life: social roles, legal systems, linguistic forms, technical standards, and all of the other components of the playing field upon which human relationships are conducted in a given society (Commons 1924, Goodin 1996, March and Olsen 1989, North 1990, Powell and DiMaggio 1991). Central to the institutional approach is an analysis of the relationship between institutions and people, and between institutions and information technology (King 1996, Kling and Iacono 1989). Institutions are not external to human affairs; they do not regulate our lives simply by reaching in to intervene. To the contrary, institutions go a long way toward defining us. To see this, consider Austin's (1962) analysis of speech acts. According to Austin, a given utterance does not count as a wedding vow or a jury verdict simply for arranging the correct words in the correct order. These speech acts are not felicitous unless certain institutional conditions obtain, and their result is to create new institutional facts in the world. The institutions, in other words, are constitutive of the acts themselves. The person who performs those acts does so in a particular institutional capacity, as the occupant and performer of a particular institutional role, in relationship to other individuals who are successfully occupying and performing complementary roles. Much of our lives are spent enacting institutional roles, and much of our own understandings of ourselves, and most of others' understandings of us, are bound up with those roles. That is not to say that we are puppets dangling on institutional strings. It is, however, to say that we have only a partial awareness of the depths to which our institutions define us. Nor is it to say that the institutions can live on without us. To the contrary, every institution is either reproduced or transformed on every moment through the actions of its participants. The relationship between individuals and institutions, in that sense, is reciprocal.

The institutional approach is particularly well-suited to the analysis of privacy issues. Privacy is patently an institutional matter; it pertains to the institutionally organized ability of individuals to negotiate their relationships with others. Philosophical and legal analysis has often identified privacy as a precondition for the development of a coherent self (Post 1989,

Schoeman 1984), and privacy issues have come increasingly to the fore as technologies for mediating human relationships have become more pervasive (Clarke 1994).

But the connection between institutions and privacy can also be found at a deeper level: both pertain to the construction of human identity. Information technology and privacy policy as we know them today both evolved under the assumption that computerized records fully and transparently identify the people whose lives they represent. Once those records are created, we have assumed, privacy is a matter of regulating the uses to which the records are put. Recent innovations in cryptography, however, have changed this picture considerably (e.g., Chaum 1985, 1992), and it has become clear that privacy can often be protected at in a more fundamental way by simply not creating individually identifiable information in the first place (Information and Privacy Commissioner and Registratiekamer 1995).

More generally, new cryptographic protocols have created a vast design space. Along one edge of this space lie the traditional technologies for creating personally identifiable records. Along the opposite edge lie technologies of anonymity, for example smart cards purchased with cash, for which it is nearly impossible to infer a user's identity. Between these two extremes lie numerous other possibilities. Digital cash (Chaum 1989), for example, makes it possible for a payer's identity to be traced, but only under specific conditions. Subsequent work (e.g., Maxemchuk 1994) has generalized these methods to the point where the identities of numerous participants to a transaction are protected, while still making it possible to reconstruct particular identities when certain complicated conditions obtain.

For these technologies to fulfill their promise, they must be integrated with the larger institutional world, including business models, regulatory systems, contractual language, and social customs. My purpose here is to prepare the conceptual ground for this integration. I will argue that privacy-enhancing technologies undermine some fundamental assumptions about both technology and economics. I will proceed in several steps, as follows. Section 2 will frame the contemporary policy debate by sketching some of the complexity hidden within the widespread notion that privacy can be regulated by means of markets. Section 3 will introduce the question of human identity more fully by drawing attention to certain features of the natural history of identity in ordinary settings and to the difficulties that surround current institutional practices of identification. Section 4 will recount the evolving place of identity in the history of computer science, focusing particularly on information technologists' long, slow transition from notions of representation based on an almost God-like objectivity to notions based on the perspectives of finite individuals. Section 5 will trace a parallel development in the history of economics, starting with the roots in physics of the neoclassical synthesis and moving toward the more complicated picture in contemporary game-theoretic models. Section 6 will draw these perspectives together by considering the construction of identity in economic institutions. Section 7 will open up the large question of the information-related dynamics by which economic institutions, and particularly their consequences for the construction of identity, are shaped. Section 8 will then conclude by sketching some new directions that these phenomena suggest for privacy policy.

2 The policy agenda

The conceptual basis of contemporary privacy policy originated in the late 1960's (Flaherty 1989). As the countries of northern Europe began building their welfare states, historical memory of the Nazi era led to concerns about centralized state databases. The "data protection" policies that arose in this environment, therefore, have been framed in political

terms: data subjects are first and foremost citizens who possess rights, such as the right to correct inaccurate information about themselves in bureaucratic records (Mayer-Schoenberger 1997). These policies also bear the marks of a technological environment characterized by small numbers of large, centralized, and poorly networked mainframe computers. As both the political and technological environments have evolved, the data protection model has retained its relevance because of its abstract nature. Rather than specify an architecture, the data protection model specifies technically neutral principles that can be -- indeed, that must be -- given content in each sector, public and private, and for each information processing technology.

Nonetheless, in recent years many commentators have suggested that the data protection model is out of date. The Clinton administration in particular has been vocal in urging industry to develop market-based self-regulatory mechanisms that achieve the goals of privacy protection without the purported downsides of a formal regulatory apparatus (Bernstein 1997). Indeed, when considered abstractly, the notion of rights to personal information as commodities that can be traded like any others in the marketplace has a compelling inner logic (Laudon 1996, Stigler 1980; cf Hine and Eve 1998, Miller 1969: 1223-1225). Different people value their personal information to different degrees, and many situations will arise in which the most stringent measures for privacy protection will require systems or procedures that are inherently more expensive, given the current state of technology, than those that offer less protection. Whether society should make an infinite investment in privacy protection, or zero investment, or some investment in between these extremes, it is argued, should be left to the same market mechanisms that determine society's investments in building cars and grilling burgers.

The principal challenges to this model are ethical and economic. The ethical arguments insist that the data protection model's political framing of privacy issues is the correct one, and that privacy is indeed fundamentally a right and not something that a person should have to buy at the going rate. Economic arguments point to the special properties of information. Talking of privacy as a commodity may make sense when we compare privacy to physical things like cars and burgers, but privacy rights are in fact special kinds of commodities because of the special economic properties of information (Arrow 1984, Baker 1997). Kang (1998), for example, observes that an individual will have a hard time evaluating whether to pay a company extra to keep her personal information to itself, given that many other companies that may possess similar information will most likely equivocate about their willingness to implement similar protections. Because information can be duplicated without being destroyed, even a single uncontrolled source of personal information can take up the slack for scores of other, more responsible organizations.

The problem here does not pertain to markets as such. Rather, scenarios like Kang's point to the necessity of institutional analysis. Simply speaking of "markets" is not to say much, given the diversity of actual and potential market institutions. Indeed, one way to interpret the data protection model is precisely as a set of ground rules for a market -- ground rules that aim to ensure that markets in privacy rights function correctly. If every firm is compelled to reveal its data-handling practices then obvious problems of asymmetric information in the privacy marketplace are alleviated and consumers can make their own choices.

Nor can the analysis end here. Quite the contrary, these elementary observations open up the complex issue of the role of personal information in the design of institutions generally and market institutions in particular. Market institutions provide contexts in which people come

together to transact business, and in doing so they create information whose embodiments and uses may concern the parties just as much as the goods and services that are bought and sold. In order to raise the question of markets in privacy, in other words, we should also investigate the question of privacy in markets.

The question of privacy in markets becomes particularly complicated in the context of privacy-enhancing technologies (Burkert 1997). One principle of data protection suggests that an organization conducting a transaction with an individual should only create and store the information needed for the purpose -- what Kang (1998), in the context of American privacy policy, calls "functionally necessary" information. Traditionally this concept has been construed to minimize the number of data fields that are created and stored within a particular transaction. But it can also be construed to minimize the degree to which the data subject is identified at all. Even if this principle is not legislated as an inherent property of market institutions, it should not be ruled out from the start. That is, market institutions should at least not make it technologically or administratively impossible a priori for buyers and sellers to contract for a minimal degree of identification. The great difficulty of conceptualizing this guideline in practice may provide us with the full measure of the difficulty of the topic.

3 The question of identity

Identity is central to social and institutional life, and yet the concept of identity is interpreted in quite different ways. A vast intellectual and popular discourse, for example, speaks of ethnic identities, of the role of identity in the construction of nation and community, and of the difference between the identities that are ascribed to one by others and the identities that one fashions for oneself. Identity in this sense is a public, symbolic phenomenon that is located in history, culture, and social structure.

This conception of identity contrasts strikingly with the conceptions that have historically been employed by information technologists. For information technologists, identity is epitomized by proper names. But identities in technical discourse are not the same as names: they are somehow purer, more mathematical. The identity of a human being, or for that matter a chair or a company, essentially consists in its being one single thing that is different from other, countable things. In the traditional ontology of computer science, in other words, the identity of a thing is strictly separate from, and prior to, its attributes.

To investigate the actual or potential role of information technology in real institutional life, therefore, our starting question should be something on the order of, "what does it mean to know who someone is?". The question answers itself automatically, of course, if we imagine that everybody has direct access to a realm of Platonic essences -- one for each thing in the world, and particularly each person. The evidence, however, does not favor that theory. Instead, questions of (what philosophers call) definite reference direct our attention to the social practices in which human beings are embedded, and particularly to the social networks and authority relationships through which human beings can make reference to people, places, and things that may be distant from their own experience (Donellan 1966, Kripke 1980). Simply to know a person's name is obviously not to know who that person is, even when the name in question is unique. Yet even a person whose name is very common can be identified reliably if the speaker and hearer share a reflexive orientation to the same institutional context and social network.

We can also know who someone is without knowing their name. A person can be a regular at a restaurant, eliciting greetings and a glass of beer, simply through the recognizability of her personal appearance. Indeed, in the big picture of the natural history of identity, face recognition occupies an important place in the aforementioned spectrum between complete identification and complete anonymity. Human beings can recognize one another's faces with remarkable skill, even if they cannot as readily attach names to them, and yet they find even familiar faces hard to describe. Walking into a restaurant, whether once or a hundred times, does not permit the employees or customers of the restaurant to create a representation of one's face that can readily be indexed into a database and connected to other sources of information. Police artists can make sketches of a suspect's face from witness reports, and members of the public can often recognize a face from sketches published in the newspaper, but making such sketches requires unusual skills and is expensive and fallible. Faces, in other words, have a one-way quality that is at least broadly reminiscent of the conditional traceability of identity in cryptographic protocols such as digital cash. And social customs depend on this quality. Social relationships unfold not through instantaneous access to one another's complete dossiers, but rather through an exchange of information that is incremental and negotiated. In face-to-face contexts, the first information that one reveals is very frequently the appearance of one's face, and the traditional customs for negotiating relationships would collapse if faces could be easily communicated in a uniquely precise way to others, much less connected by automatic means with computerized files.

In an institutional setting, to "know who somebody is" is roughly speaking the ability to get hold of them. Vendors may need to collect debts. Organizations that maintain customer accounts need to be assured that a person who shows up at a customer service point, or who calls on the telephone, is the same person who opened the account. The conventional strategy for establishing identity is, obviously, to create a unique representation of each individual. Each interaction between an individual and an organization then has (at least) two steps: first find out who the person is (that is, establish and authenticate an identifier) and then transact business with him or her (that is, work on the records that are indexed using that identifier). And yet, in actual institutional practice, this strategy has been implemented in a remarkably slapdash manner. Simsion (1994), for example, observes that real systems have tended to employ identifiers such as name-plus-birthdate that are not necessarily unique, and that can even change as an individual changes social status or when errors are discovered (cf Agre 1997c). Many organizations employ identifiers such as the Social Security Number that were never intended for such purposes and are poorly suited for them. What is more, organizations that issue identification mechanisms such as driver's licenses are frequently plagued by corruption (Ellis 1998). As a result, many institutional systems of identification are riddled with opportunities for error and fraud, and organizations often attempt to strengthen their own identification mechanisms by requiring individuals to display identification from other organizations.

In short, institutional mechanisms of identification, particularly in the United States, are undergoing a slow-motion catastrophe. And with the spread of information technologies that can mediate institutional relationships across great distances, the problem is likely to become severe. Approaches to alleviating the problem can be sorted, at least for heuristic purposes, into two broad categories: centralized and decentralized. A centralized system would be a complete, functioning, and unified version of the scheme just described: one single unique identifier for each person, used consistently for every purpose. A decentralized system, by contrast, would resemble much more closely the "pseudonym" scheme devised by Chaum (1990), in which each individual is assigned a distinct identifier by every organization with

whom she does business, such that these identifiers are incapable of being linked to one another without the individual's permission.

Although the notion of a "pseudonym" would normally conjure notions of advanced cryptography or other special trickery, in fact something like the decentralized picture follows naturally so long as the various systems are developed independently of one another. Because of this, the institutional mechanisms of identification that we actually have in 1998 fall somewhere between the centralized and decentralized extremes, and their lack of robustness can be understood precisely in terms of the haphazard and accidental nature of this compromise. Organizations are well known to suffer from "islands of automation" within themselves, and in many cases the computers in the manufacturing divisions of different firms are more compatible than are the computers of the manufacturing and finance divisions of the same firm. A minor industry has grown up to patch over these separate spheres of representation (e.g., Brackett 1994), and this industry must contend with the extraordinary inertia induced by legacy systems and the sprawling networks of skills and practices that have usually grown up around them.

The project of privacy protection, in other words, does not begin from a clean slate, or from any other ideal-typical case. To the contrary, the current situation is a historical muddle of some complexity. To move forward, we need to comprehend the current situation on several levels. To keep the remainder of my discussion within manageable bounds, I will concentrate on only a few of these levels. Specifically, I want to consider some of the intellectual inheritances that shape our understandings of these problems. I will argue that those inheritances are themselves internally complex, and that only a proper historical analysis of them will make them useful in guiding the future evolution of both institutions and technology.

4 Technical evolution

Conventional histories of computing often focus their attention on the machinery itself and on the mathematical basis of logical design and programming. Important as this level of analysis is, we cannot comprehend the institutional history of computing without also attending to the specifically institutional ideas of its inventors. Schaffer (1994), for example, recounts the role of religious ideas and political economy in Babbage's understanding of the computer. Babbage followed in a long line of engineers by understanding his own work on the analogy of God's creation of the earth (Noble 1997). The factory was a microcosm, and the engineer's job was to impose upon it a perfect rational order, arranging the machinery and activities within it without any regard for the subjectivity of the workers.

The computer epitomized for Babbage this God-like ordering principle, and this conception of the computer set in motion a certain conception -- tacit to be sure, but powerful and pervasive nonetheless -- of the nature of computational representation. The word "representation", let us note, can take on several meanings. In the law, a representation is a social action that carries certain responsibilities, some of which can be enforced legally (Shapo 1974). But a representation can also be a thing: an artifact that has been designed to refer in some systematic way to circumstances in the world. And that is the sense in which computer scientists have understood the representations that they have embodied in their machines. Indeed, although the word "representation" is employed routinely by practitioners of artificial intelligence (Agre 1997a), practicing mainstream computer professionals tend to

use words such as "file" and "record" that name the representational artifacts of traditional institutional practices.

For Babbage and his descendents, however, the purpose of computerized representations was not simply to replicate the practices of bureaucracies but transparently to mirror the world (Agre 1997c). Even though the contemporary discourse of computing has secularized much of Babbage's theology, there remains the assumption, or ideal, that computers take up a God's-eye view of the world (cf. Edwards 1996, Haraway 1991, Hayek 1963). This approach to representation made a certain sense in the military and industrial settings in which computing was first developed, where power relations did subject all activities to omniscient monitoring. Privacy problems arise precisely when this model is transferred into settings where such relations of representation are taken less for granted.

This, however, is not the end of the story. Computer technology has obviously evolved a great deal since Babbage's day, and even since the days when the practices of organizational computing first took shape at IBM. Among the many lines of evolution, the ones that are relevant here are the ones that begin to recognize the perspective of individual human beings. The conception of individual human beings that is implicit in the conventional practices of computer systems design, in other words, is changing, and yet much of that change has escaped notice. I have already mentioned one area of innovation that has not escaped notice -- the emerging generation of cryptographic protocols that permit designers to underwrite a remarkably wide range of informational relationships among persons. More recently these developments have been formalized into architectural frameworks that promote the routinized design of such schemes (Blaze, Feigenbaum, Resnick, and Strauss 1997; Roscheisen and Winograd 1996; Stefik 1997).

Important as these developments are, they also depend upon a more fundamental shift that has been little recognized. Mainframe computers presupposed (and still do) that they are embedded in an institutional framework in which somebody authenticates the identities of users before issuing accounts to them. The machinery itself, of course, does not obligate the organization that owns it to establish the absolute, Platonic, once-and-for-all identity of every user, whatever that would mean. Nonetheless, the designers of timesharing operating systems have assumed that they are dealing with multiple users, and that one of their central tasks is to distinguish clearly between these users and to build barriers that prevent the users from interfering with one another (Agre 1998b). A mainframe may not exactly know who its users are, but it presupposes that somebody has done enough work to tell the users apart.

Personal computers are designed on no such assumption. The notion of distinct users is not central to the ontology upon which personal computer operating systems are designed, and those personal computers that do distinguish among users do so superficially, perhaps through a password protocol in a screen saver. Your personal computer truly does not know who you are. The Internet, for its part, was first pioneered during the mainframe era. The early Internet derived its security largely through social mechanisms -- by peer pressure within the small world of Internet users and by the institutions that selected people to be users in the first place -- and it derived its technical means of security largely from the timesharing operating systems through which its users gained access to it. Yet that all changed with the introduction of personal computers. Individuals on personal computers could gain access to the Internet without logging in to anything, and the concept of logging in to the Internet itself did not exist. The most widely used Internet software packages arose in this setting, and it is this historical circumstance that explains, in sharp contrast to the picture of traditional

timesharing systems, the remarkably poor facilities that the Internet provides to enable people to create boundaries for themselves. Whereas the cultural norms and cognitive practicalities of face-to-face interaction make it possible to negotiate incremental access to oneself, Internet users find themselves inadequately equipped to defend themselves against forgery, spam, and other aggressively antisocial practices.

The fundamental point is this: whereas mainframe operating systems represent their users, personal computers do not. Lest personal computers seem too strange as a result of this difference, however, we should recognize that the mainframes are the exception. Personal computers need not identify their users because the continuity of a user's relationship to a personal computer is provided for by the brute fact of her or her physical access to the machine. A personal computer does not understand its user as "John" or "Mary", any more than a car or an electric razor does, but rather as something more like "the person who is using me". The designer may well, of course, have an elaborate story in mind about the attributes and relationships of this person (Sharrock and Anderson 1994), and this story may well have been inscribed in the device itself (Akrich 1992), but the user himself or herself is still characterized in indexical terms through a certain definite description ("the person who ..."). The concept of indexicality is derived from linguistics, where it refers to any aspect of grammar whose referent depends on the circumstances in which it is used (Hanks 1990). The crucial point, in Smith's (1996) provocative formulation, is that the laws of physics are themselves indexical: they depend not on particular places and times but on "here" and "there" and "now" and "then".

Nor is this trend toward indexicality in representation confined to the tacit workings of personal computers. Research in artificial intelligence has long presupposed that human beings and other "intelligent agents" employ representations that resemble the "view from nowhere" (cf Porter 1994) that AI people call a "world model". And yet human beings, like cats and robots, are finite; they have bodies and are located physically and epistemically in space. For this reason and others, attempts to build intelligent machinery have been compelled over time toward a more explicitly indexical understanding of representation (Agre 1997a, Lesperance and Levesque 1995). The point is not exactly that anonymity is the natural order of things. We can, after all, recognize one another's faces. What is unnatural, so to speak, is precisely the attempt to establish a God-like representational perspective that gives all things their true names.

5 Economic evolution

Although many of the details of this evolution are specific to the institutions and practices of computing, nonetheless economics has traveled a remarkably analogous road. As Mirowski (1989) has observed, the neoclassical synthesis in economics that got its start with Cournot and that emerged full-blown at mid-century (Samuelson 1947) was, at its origin, explicitly modeled in both its methodology and its mathematics upon the notion of equilibrium in the classical theory of mechanics in physics. Thus, even though neoclassical economics is commonly described as an individualistic theory of human beings, it is equally important to understand the sense in which it is nothing of the sort. Central to Cournot's methods was the notion that everybody's utilities can be maximized at the same time as so many simultaneous equations. The famous idealizations of neoclassical economics -- perfect information, zero transaction costs, and so on -- all originate as the conditions under which this generalized and universal notion of equilibrium can be made to work out. Human beings in this scheme have no privacy, indeed no internal states at all, and they have no room to engage in strategic

behavior. Their market-relevant attributes are perfectly and instantaneously known to everybody, or these attributes are at least taken fully into account in everybody's judgements, and all of the attributes enter into a perpetual, holistic, instant-by-instant reshuffling of the great allocative machinery of the marketplace. Thus Epstein (1993), in his economic analysis of law, can analyze the perfectly functioning market in terms of the allocative choices that would be made by a hypothetical "single owner" of all the world's resources. Posner (1981) and Stigler (1980), likewise, have employed the neoclassical economic framework in characterizing privacy as harmful to the efficient functioning of the market.

It is only with subsequent developments that mainstream economics begins to construct a nontrivial picture of the individual human person. Whereas economists of a collectivist orientation have labored to bring the supposedly atomized individual of neoclassical theory into some kind of social relationship with others, the actual development of neoclassical theory has been close to the opposite of this: a gradual paring away, like a sculptor chipping at stone, of the many idealizations that merge the individual into one vast collective consciousness. These developments take place largely in the context of the economics of information. It has long been understood, for example, that information has a paradoxical place in neoclassical theory: markets clear in a general way only if market-relevant information is perfect and therefore free, but the necessary information will only be produced and distributed in sufficient quantity and quality if it is a commodity like any other. Stigler (1961), for example, pioneered the analysis of search costs within a neoclassical framework, and Grossman and Stiglitz (1980) argued on formal grounds that the economy can at best alternate between information that is free and information that is adequate. Boyle (1996) has developed the point philosophically.

These difficulties, of course, do not invalidate the neoclassical model. The neoclassical idealizations provide a formally attractive starting point from which a multitude of divergences can be explored -- each with an attendant increase, often substantial, in the complexity of the model. (In this, too, the neoclassical model is analogous to artificial intelligence: both fields attempt to formalize rational behavior, and AI as well has its canonical model, the formally attractive but frequently impractical "Good Old-Fashioned AI" architecture of world models and planning from which numerous divergences have been explored.) Institutional changes are also viewed as approximating the world ever more closely to the neoclassical ideal (North 1990, Williamson 1985). In other work (Agre 1997b), I have developed the consequences of these ideas, and particularly the information-centered model of Casson (1994), for the analysis of privacy problems. Briefly put, the optimistic prediction of the model is that markets with ever-lower information costs can become ever more efficient through the ever-greater pervasiveness of contractual monitoring, but only to the extent that individual consumers are willing to depart cheaply with their personal information. Whether that optimistic prediction holds true to the reality is a crucial topic for research.

Rather than analyze these models any further, however, let us briefly consider the even clearer emergence of a distinct human person in subsequent developments in mainstream economics. Although first invented fifty years ago, game theory (Von Neumann and Morgenstern 1947) has lately been developed into a general tool for the analysis of problems of asymmetric information among the participants in an institutionalized relationship (Hillier 1997, Rasmusen 1994). Phlips (1988: 8-9), for example, makes the useful distinction between imperfect information, which arises within an established framework of rules when participants are uncertain about one another's past and present behavior, and incomplete

information, which arises when the participants do not know some of the elements (payoffs, available strategies, number of other players) that define the game itself. Game theory still employs a holistic form of analysis derived (by Nash) from Cournot, and, as Nermuth (1982) among others has remarked, the mathematics of strategic interaction can become extraordinarily complex. Nonetheless, with this kind of analysis of asymmetric information and the varieties of strategic behavior that asymmetric information permits, one finally begins to get a sense of the private realm of the individual (Phlips 1988: 2-3). This sense of the private individual is also found in the study of incentives that govern the revelation of information and the economic institutions (prototypically the Vickrey auction) that encourage individuals to make accurate representations of their preferences (see Campbell 1995).

6 Institutions and identity

In sketching this history, I hope to have indicated the remarkable analogies between the intellectual development of both computer science and economics. Each has started from a transparent, God's-eye conception of human beings that subsumes them into a wholly rational universal system, and moved toward an increasing appreciation that human beings are in fact finite, located, and embodied. This progression is, one likes to think, science in action, the truth coming out. If so, it is a truth that challenges some assumptions of both technical design and public policy, both of whose public articulations remain very much rooted in the older, obsolescent conceptions of their respective fields.

In order to become practically useful, however, this emerging understanding must be embedded in an institutional context. This is an intellectual challenge, given that mainstream computer science and economics have also been similar in their foreshortened understandings of the institutional organization of technical and economic practice in the real world. Economics, for example, was long content to understand markets in terms of Walras' fictional auctioneer who could be assumed to bring buyers and sellers together at agreed prices. A "market" simply consisted of some buyers, some sellers, and some goods, and economic theory could proceed with astoundingly little curiosity about the actual material workings of marketplaces, job interviews, auto dealers, and the like. Of course, as Hodgson (1988: 182) points out, an auction is itself an institution; the institutional organization of markets was always tacitly presupposed in the theory. Now, however, it is emerging much more explicitly as a topic of research. Much of the recent economic concern with institutions, as I have already noted, has been driven by a teleology whereby technological advance progressively reduces friction in the marketplace (the term, now celebrated due to its use by Gates (1995), appears to originate with Williamson (1985)) and thereby causes the economy to approximate the neoclassical ideal ever more closely. Institutions are, on this view, a passing phase, albeit an important one.

Another version of this teleology is the widespread notion of disintermediation, which holds that the principal effect of information technology on institutions of all types is the cutting-out of intermediaries (cf Halper 1998). Although wildly oversimplified, not least because the same mechanisms that make it easier to cut out intermediaries can make the intermediaries more efficient as well, this notion has nonetheless led -- at last, though better late than never -- to principled and general analyses of the role of intermediaries in actual markets (Casson 1997; Spulber 1996, 1998).

The public prominence of information technology, then, has made it considerably harder to neglect the institutional organization of markets. Picot, Ripperger, and Wolff (1996), for

example, following in the tradition of Coase's (1937) analysis of the effects of the telegraph and telephone on the boundaries of the firm, have argued that information and communication technologies cause the boundaries of the firm to fade altogether, as firms develop more complex linkages among themselves and begin to adopt market-like mechanisms within themselves. Much more work is needed, however, not least conceptual work that recovers a full awareness of the many aspects of market institutions to which mainstream economics has been oblivious (Granovetter 1992).

Once the need for an analysis of institutions has been realized, it becomes possible to develop a sophisticated analysis of the nature and role of identity. Identity becomes an issue in markets, for example, when personal reputations serve as proxies for information about the quality of goods, or when contracts call for particular parties to perform their part despite having already obtained the consideration they wanted. In economic models of markets, however, the identities of the participants are almost invariably formalized in a primitive way. Buyers and sellers, for example, might be numbered from 1 to n, and all parties will be assumed to know the numbers and possess the ongoing ability to assign the right numbers to the right players. This formalization is often wrong and invariably incomplete. So long as the formalism is restricted to a single market -- meaning, roughly, the market in a single class of goods -- many privacy issues -- especially those pertaining to the secondary use of information -- do not arise. Even within attempts to model strategic behavior in a single market, individual players' actions can reveal information that provides others with an advantage in subsequent exchanges, thereby leading to a space of trade-offs between short-term optimality and long-term strategy whose Nash equilibrium can be exquisitely complex to model. Already, however, it is assumed -- unnecessarily from a technical perspective, at least in many settings -- that players can identify one another from round to round.

Market institutions can become progressively more complex as they take into account the forms of identity that market participants can construct with the aid of privacy-enhancing technologies. Consider, for example, the potential roles for cryptographic protocols that produce "credentials" (Chaum 1990). Recall that the traditional method for establishing the various properties of a customer to a transaction has been to identify the customer and then look up the records associated with that identity. Although conceptually convenient, this two-step process collects more information than is required to achieve its purpose. A more direct mechanism would employ indexical representations -- not "this is John and John is 21 so John can enter the bar", but rather "this person is 21 so he can enter the bar". And this is what credentials provide -- assuming, of course, that the necessary institutional framework is in operation.

More sophisticated institutional analyses of identity are possible. Poster (1990), for example, outlines a theory of databases as what Foucault calls "discourses". For Foucault, a discourse is not simply a linguistic phenomenon but includes the full ensemble of practical arrangements by which individuals are positioned in institutional roles -- teacher and student, judge and jury, doctor and patient, and so on. These arrangements increasingly involve databases, of course, and Poster argues that we miscomprehend the nature of databases if we simply look at them as patterns of bits in a computer. Instead, he points out, databases derive their usefulness from their position in larger circuits of relationships and interactions among people. This is Poster's version of the proposition, introduced at the outset, that human beings and information technologies are coevolving through the mediation of institutions and the social roles that they define. Information technologists, on this view, are insufficiently aware of the institutional arrangements that actively reconfigure the world to make the necessary

capture of data possible in practice (Agre 1994). And neoclassical economists themselves take an unreasonably narrow view of institutions, even market institutions, in their failure to analyze the elaborate connections between the allocative mechanisms of the market and the identity-sustaining mechanisms of other spheres of life. It is here that the two understandings of identity -- the cultural and the technical -- begin to converge into a single complex phenomenon.

Poster's analysis also draws attention to the roles of the body in constructing identity. One such role obviously pertains to the state: among the traditional reasons to identify the parties to a contract is to make it possible to hale them into court if they violate the contract. Identity, in other words, is employed as a means of access to a person's body, and many contracts require an individual to commit to the practical arrangements that may one day make this access possible. With privacy-enhancing technologies, however, it is entirely feasible to gain access to an individual's identity, and thereby to their person, only upon a breach of the contract, and not before.

Once again, we should recognize that this kind of wrangling over the establishment of identity is not the natural way of things. An enormous variety of practical arrangements make it possible for people to conduct complex business anonymously by taking advantage of the properties of bodies. A supermarket checkout aisle, for example, provides physically for queueing and effectively traps each successive customer in a confined space while the transaction is being conducted. The point, of course, is not to imprison anybody in an extreme sense, but rather to provide for the continuity of the indexical relationship between this-customer and this-clerk while the various steps in the transaction are being executed, thereby providing, using a term from computer science, for the atomicity of the transaction. Slips of paper serve much the same purpose: sufficiently unforgeable for a given degree of risk, they provide continuity through the several steps of a transaction such as attending a movie. Even though words of computer memory were originally modeled on the entries in a paper form, many of the security problems with computers derive precisely from their unpaperlike ability to be modified without leaving any eraser marks. Here again cryptographic protocols have begun to recover some like the protections that were already a natural part of the pre-electronic world (Hettinga 1998).

7 Shaping institutions

These remarks may provide some idea of the stakes in the coevolution between market institutions and the construction of identity. Even if we can speak reasonably about privacy as a commodity to be allocated in the market, we must first speak about privacy as an attribute of market institutions. And this, in turn, requires us to speak about the market in marketplaces. Where do market institutions come from, why do they embody the relationships and rules that they do, and how do they evolve from there? The issues become particularly acute as market institutions are increasingly mediated by information technology. Despite information technology's revolutionary reputation, many of its institutional dynamics are in fact obstinately conservative in character. Arrow (1984: 142) observes, for example, that "information creates economies of scale throughout the economy, and therefore, according to well-known principles, causes a departure from the competitive economy" -- a trend, in other words, toward increasing market power on the part of particular players.

As Mansell (1996) points out, market power can manifest itself not simply in excessive prices but also in the shaping of an unlevel playing field in electronic commerce. Market

mechanisms are network technologies (Katz and Shapiro 1994) in that the benefit from adopting them depends heavily on the number of others who have also adopted them, and this dynamic may lead to the locking-in of mechanisms that institutionalize bias. Among the important biases of an electronic commerce technology are those that pertain to information extraction. A technology, for example the SET payment protocol (Phillips 1997), can become established through the coalition strategies of interested players, even though the technology inherently extracts and retains far more information about purchasers and their transactions than necessary. Callon (1991) understands these phenomena more broadly as the "irreversibilization" of a heterogeneous set of elements that become interlocked with one another as a new social practice settles into a pattern.

Even before electronic commerce technologies come into the picture, market institutions are already subject to biases from the unequal position of the players, and it would seem important to have a general theory of the process by which these biases are institutionalized. Such biases cannot exist on the neoclassical model, and so the phenomena in question must be sought in reality's departures from that model. One starting-place might be the distinction, observable in almost all real markets, between specialists -- those who spend much of their lives playing their particular role in a given type of transaction -- and generalists -- those who play their own role in that type of transaction only rarely. Someone who sells used cars for a living is a specialist in that topic, whereas almost everyone else is a generalist who purchases many different types of goods including the very occasional used car.

Specialists and generalists, obviously, occupy asymmetric locations in an institution, and this asymmetry entails informational differences. By playing the same role in a long series of substantially analogous interactions, a specialist can learn by degrees how to "work the angles" of the interaction. The specialist may choose the wording in a form contract, design those aspects of the physical environment that determine what a customer can see and when, choose the order in which various items of information are revealed during negotiation, identify phrases and other selling techniques that various types of customers find most persuasive, and so on. Consultants can make a living by moving from one specialist to the next, transferring expertise through a kind of information arbitrage (Bowker 1994). Specialists also have a much greater interest than generalists in associating and organizing among themselves, in building social networks, and in sharing experiences and generally engaging in collective cognition (Agre 1998c). It follows that, even in the context of ordinary contract negotiation, individualistic theories of unequal bargaining (e.g., Cartwright 1991, Trebilcock 1993) are woefully inadequate unless they comprehend these systematic sources of inequality.

Political scientists have understood these phenomena in terms of the collective action problems posed by organized interest groups in a democracy (Olson 1965), but that is only one manifestation of a deeper problem. Something of the shape of the problem is conveyed by Commons' (1924) use of collective bargaining as a model for the analysis of all social institutions. The point of Commons' analysis becomes visible in technological contexts through the dynamics of technical standards-setting. Conflicts over open standards, for example, can be understood as an attempt by information technology customers to engage in collective bargaining with vendors. The ideology of open standards, unclear and shifting as its meanings have often been (Abbate 1995, Cargill 1994), has largely been an attempt by information technology customers to fashion a loose coalition and present a unified front to vendors. Lacking formal organization and established forums for bargaining, the two sides in this negotiation have engaged in complicated strategic behavior that cannot be properly

modeled without a full sense of each player's awareness of network effects and their consequences (Agre 1998a, Grindley 1995).

8 Conclusions

The conditions that affect privacy are integral to the construction of market institutions, and should therefore not be regarded as mere products of natural market forces. The principles of data protection, in particular, may provide guidance for institutional design. To the extent that market institutions are embodied in electronic commerce technologies, any intervention that hopes to influence the direction of those technologies needs to be timed correctly; once a standard is entrenched in market institutions, it may be for all practical purposes irreversible. Finally, the promotion of privacy-enhancing technologies should aim to incorporate them into the design of institutions generally and market intermediaries in particular.

* Acknowledgements

This paper was prepared for the Telecommunications Policy Research Conference, Alexandria, Virginia, October 1998. I appreciate the comments of the anonymous referees.

* References

Janet Abbate, "Open systems" and the Internet, Society for Social Studies of Science / Society for History of Technology Conference, Charlottesville, October 1995.

Philip E. Agre, Surveillance and capture: Two models of privacy, *The Information Society* 10(2), 1994, pages 101-127.

Philip E. Agre, *Computation and Human Experience*, Cambridge: Cambridge University Press, 1997a.

Philip E. Agre, Introduction, in Philip E. Agre and Marc Rotenberg, eds, *Technology and Privacy: The New Landscape*, Cambridge: MIT Press, 1997b.

Philip E. Agre, Beyond the mirror world: Privacy and the representational practices of computing, in Philip E. Agre and Marc Rotenberg, eds, *Technology and Privacy: The New Landscape*, Cambridge: MIT Press, 1997c.

Philip E. Agre, The Internet and public discourse, *First Monday* 3(3), 1998a.

Philip E. Agre, Yesterday's tomorrow, *Times Literary Supplement*, 3 July 1998b, pages 3-4.

Philip E. Agre, Designing genres for new media, in Steven G. Jones, *CyberSociety 2.0: Revisiting CMC and Community*, Newbury Park, CA: Sage, 1998c.

Madeleine Akrich, The de-scription of technical objects, in Wiebe E. Bijker and John Law, eds, *Shaping Technology/Building Society: Studies in Sociotechnical Change*, Cambridge: MIT Press, 1992.

Kenneth J. Arrow, *Collected Papers*, Volume 4: *The Economics of Information*, Cambridge: Harvard University Press, 1984.

J.L. Austin, *How to Do Things with Words*, Cambridge: Harvard University Press, 1962.

C. Edwin Baker, Giving the audience what it wants, *Ohio State Law Review* 58(2), 1997, pages 311-417.

Nina Bernstein, Goals clash in shielding privacy, *New York Times*, 20 October 1997, page A16.

Matt Blaze, Joan Feigenbaum, Paul Resnick, and Martin Strauss, Managing trust in an information-labeling system, *European Transactions on Telecommunications* 8, 1997, pages 491-501.

Geoffrey C. Bowker, *Science on the Run: Information Management and Industrial Geophysics at Schlumberger, 1920-1940*, Cambridge: MIT Press, 1994.

James Boyle, *Shamans, Software, and Spleens: Law and the Construction of the Information Society*, Cambridge: Harvard University Press, 1996.

Michael H. Brackett, *Data Sharing: Using a Common Data Architecture*, Wiley, 1994.

Erik Brynjolfsson and Lorin M. Hitt, Beyond the productivity paradox: Computers are the catalyst for bigger changes, *Communications of the ACM* 41(8), 1998, pages 49-55.

Herbert Burkert, Privacy-enhancing technologies: Typology, critique, vision, in Philip E. Agre and Marc Rotenberg, eds, *Technology and Privacy: The New Landscape*, Cambridge: MIT Press, 1997.

Michel Callon, Techno-economic networks and irreversibility, in John Law, ed, *A Sociology of Monsters: Essays on Power, Technology and Domination*, London: Routledge, 1991.

Donald E. Campbell, *Incentives: Motivation and the Economics of Information*, Cambridge: Cambridge University Press, 1995.

Carl Cargill, Evolution and revolution in open systems, *StandardView* 2(1), 1994, pages 3-13.

John Cartwright, *Unequal Bargaining: A Study of Vitiating Factors in the Formation of Contracts*, Oxford: Oxford University Press, 1991.

Mark Casson, Economic perspectives on business information, in Lisa Bud-Frierman, ed, *Information Acumen: The Understanding and Use of Knowledge in Modern Business*, Routledge, 1994.

Mark Casson, *Information and Organization: A New Perspective on the Theory of the Firm*, Oxford: Oxford University Press, 1997.

David Chaum, Security without identification: Transaction systems to make Big Brother obsolete, *Communications of the ACM* 28(10), 1985, pages 1030-1044.

David Chaum, Privacy protected payments: Unconditional payer and/or payee untraceability, in D. Chaum and I. Schaumuller-Bichl, eds, *Proceedings of Smart Card 2000*, North Holland, 1989, pages 69-93.

David Chaum, Showing credentials without identification: Transferring signatures between unconditionally unlinkable pseudonyms, in J. Seberry and J. Pieprzyk, eds, *Advances in Cryptology: Auscrypt '90*, Berlin: Springer-Verlag, 1990, pages 246-264.

David Chaum, Achieving electronic privacy, *Scientific American* 267(2), 1992, pages 96-101.

Roger Clarke, The digital persona and its application to data surveillance, *Information Society* 10(2), 1994, pages 77-92.

Ronald H. Coase, The nature of the firm, *Economica* NS 4, 1937, pages 385-405.

John R. Commons, *Legal Foundations of Capitalism*, New York: Macmillan, 1924.

Keith S. Donnellan, Reference and definite descriptions, *Philosophical Review* 75(3), 1966, pages 281-304.

Paul N. Edwards, *The Closed World: Computers and the Politics of Discourse in Cold War America*, MIT Press, 1996.

Virginia Ellis, Thriving trade in fake drivers' licenses poses tough problem for DMV, *Los Angeles Times*, 5 April 1998, pages A1, A26.

Richard A. Epstein, Holdouts, externalities, and the single owner: One more salute to Ronald Coase, *Journal of Law and Economics* 36, 1993, pages 553-586.

Andrew Feenberg, *Alternative Modernity: The Technical Turn in Philosophy and Social Theory*, Berkeley: University of California Press, 1995.

David H. Flaherty, *Protecting Privacy in Surveillance Societies: The Federal Republic of Germany, Sweden, France, Canada, and the United States*, University of North Carolina Press, 1989.

Peter Galison, The ontology of the enemy: Norbert Wiener and the cybernetic vision, *Critical Inquiry* 21(1), 1994, pages 228-266.

Bill Gates, *The Road Ahead*, New York: Viking, 1995.

Joey F. George and John L. King, Examining the computing and decentralization debate, *Communications of the ACM* 34(7), 1991, pages 63-72.

Robert E. Goodin, ed, *The Theory of Institutional Design*, Cambridge: Cambridge University Press, 1996.

Mark Granovetter, Economic action and social structure: The problem of embeddedness, in Mark Granovetter and Richard Swedberg, eds, *The Sociology of Economic Life*, Boulder: Westview, 1992.

Peter Grindley, *Standards Strategy and Policy: Cases and Stories*, Oxford: Oxford University Press, 1995.

Sanford J. Grossman and Joseph E. Stiglitz, On the impossibility of informationally efficient markets, *American Economic Review* 70(3), 1980, pages 393-408.

Mark Halper, Middlemania, *Business 2.0* 3(7), 1998, pages 45-60.

William F. Hanks, *Referential Practice: Language and Lived Space Among the Maya*, Chicago: University of Chicago Press, 1990.

Donna J. Haraway, Situated knowledges: The science question in feminism and the privilege of partial perspective, in *Simians, Cyborgs, and Women: The Reinvention of Nature*, Routledge: New York, 1991.

Friedrich A. Hayek, *Individualism and Economic Order*, Chicago: University of Chicago Press, 1963.

Robert Hettinga, Call for founders: An email list (and subsequent conferences) on digital bearer settlement, electronic message to the DBS mailing list, March 1998. On file with the author.

Brian Hillier, *The Economics of Asymmetric Information*, St. Martin's Press, 1997.

Christine Hine and Juliet Eve, Privacy in the marketplace, *The Information Society* 14(4), 1998, pages 253-262.

Geoffrey M. Hodgson, *Economics and Institutions: A Manifesto for a Modern Institutional Economics*, Cambridge, UK: Polity Press, 1988.

Information and Privacy Commissioner (Ontario) and Registratiekamer (The Netherlands), *Privacy-Enhancing Technologies: The Path to Anonymity* (two volumes), Information and Privacy Commissioner (Toronto) and Registratiekamer (Rijswijk), 1995.

Jerry Kang, Information privacy in cyberspace transactions, *Stanford Law Review* 50(4), 1998, pages 1193-1294.

Michael L. Katz and Carl Shapiro, Systems competition and network effects, *Journal of Economic Perspectives* 8(2), 1994, pages 93-115.

John L. King, Where is the payoff from computing?, in Rob Kling, ed, *Computerization and Controversy: Value Conflicts and Social Choices*, second edition, Academic Press, 1996.

Rob Kling and Suzanne Iacono, The institutional character of computerized information systems, *Office: Technology and People* 5(1), 1989, pages 7-28.

Rob Kling, Computerization and social transformation, *Science, Technology, and Human Values* 16(3), 1991, pages 342-367.

Jack Knight, *Institutions and Social Conflict*, Cambridge: Cambridge University Press, 1992.

Saul Kripke, *Naming and Necessity*, Cambridge, MA: Harvard University Press, 1980.

Thomas K. Landauer, *The Trouble with Computers: Usefulness, Usability, and Productivity*, Cambridge: MIT Press, 1995.

Kenneth C. Laudon, Markets and privacy, *Communications of the ACM* 39(9), 1996, pages 92-104.

Yves Lesperance and Hector J. Levesque, Indexical knowledge and robot action: A logical account, *Artificial Intelligence* 73(1-2), 1995, pages 69-115.

Jeff Madrick, Computers: Waiting for the revolution, *New York Review of Books*, 26 March 1998, pages 29-33.

Robin Mansell, Designing electronic commerce, in Robin Mansell and Roger Silverstone, eds, *Communication by Design: The Politics of Information and Communication Technologies*, Oxford: Oxford University Press, 1996.

James G. March and Johan P. Olsen, *Rediscovering Institutions: The Organizational Basis of Politics*, New York: Free Press, 1989.

Nicholas F. Maxemchuk, Electronic document distribution, *ATT Technical Journal* 73(5), 1994, pages 73-80.

Viktor Mayer-Schoenberger, Generational development of Data Protection in Europe, in Philip E. Agre and Marc Rotenberg, eds, *Technology and Privacy: The New Landscape*, Cambridge: MIT Press, 1997.

Arthur R. Miller, Personal privacy in the computer age: The challenge of new technology in an information-centered society, *Michigan Law Review* 67(6), 1969, pages 1089-1246.

Philip Mirowski, *More Heat than Light: Economics as Social Physics, Physics as Nature's Economics*, Cambridge: Cambridge University Press, 1989.

Manfred Nermuth, *Information Structures in Economics: Studies in the Theory of Markets with Imperfect Information*, Berlin: Springer-Verlag, 1982.

David Noble, *The Religion of Technology*, New York: Knopf, 1997.

Douglass C. North, *Institutions, Institutional Change, and Economic Performance*, Cambridge: Cambridge University Press, 1990.

Mancur Olson, Jr., *The Logic of Collective Action: Public Goods and the Theory of Groups*, Cambridge: Harvard University Press, 1965.

David J. Phillips, Cryptography, secrets, and the structuring of trust, in Philip E. Agre and Marc Rotenberg, eds, *Technology and Privacy: The New Landscape*, Cambridge: MIT Press, 1997.

Louis Phlips, *The Economics of Imperfect Information*, Cambridge: Cambridge University Press, 1988.

Arnold Picot, Tanja Ripperger, and Birgitta Wolff, The fading boundaries of the firm: The role of information and communication technology, *Journal of Institutional and Theoretical Economics* 152(1), 1996, pages 65-79.

Theodore M. Porter, Information, power and the view from nowhere, in Lisa Bud-Frierman, ed, *Information Acumen: The Understanding and Use of Knowledge in Modern Business*, London: Routledge, 1994.

Richard A. Posner, *The Economics of Justice*, Cambridge: Harvard University Press, 1981.

Robert C. Post, The social foundations of privacy: Community and self in the common law tort, *California Law Review* 77(5), 1989, pages 957-1010.

Mark Poster, Databases as discourse, or, Electronic interpellations, in *The Mode of Information: Poststructuralism and Social Context*, Cambridge: Polity Press, 1990.

Walter W. Powell and Paul J. DiMaggio, eds, *The New Institutionalism in Organizational Analysis*, Chicago: University of Chicago Press, 1991.

Eric Rasmusen, *Games and Information: An Introduction to Game Theory*, second edition, Cambridge, MA: Blackwell, 1994.

Martin Roscheisen and Terry Winograd, A communication agreement framework for access/action control, *Proceedings of the IEEE Symposium on Research in Security and Privacy*, Oakland, 1996.

Paul A. Samuelson, *Foundations of Economic Analysis*, Cambridge: Harvard University Press, 1947.

Simon Schaffer, Babbage's intelligence: Calculating engines and the factory system, *Critical Inquiry* 21(1), 1994, pages 201-228.

Ferdinand Schoeman, ed, *Philosophical Dimensions of Privacy: An Anthology*, Cambridge: Cambridge University Press, 1984.

Marshall S. Shapo, A representational theory of consumer protection: Doctrine, function and legal liability for product disappointment, *Virginia Law Review* 60(7), 1974, pages 1109-1386.

Wes Sharrock and Bob Anderson, The user as a scenic feature of the design space, *Design Studies* 15(1), 1994, pages 5-18.

Graeme C. Simsion, *Data Modeling Essentials: Analysis, Design, and Innovation*, New York: Van Nostrand Reinhold, 1994.

Brian C. Smith, *On the Origin of Objects*, Cambridge, MA: MIT Press, 1996.

Daniel F. Spulber, Market microstructure and intermediation, *Journal of Economic Perspectives* 10(3), 1996, pages 135-152.

Daniel F. Spulber, *The Market Makers: How Leading Companies Create and Win Markets*, New York: McGraw-Hill, 1998.

Mark Stefik, Trusted systems, *Scientific American* 276(3), 1997, pages 78-81.

George J. Stigler, The economics of information, *Journal of Political Economy* 69(3), 1961, pages 213-225.

George J. Stigler, An introduction to privacy in economics and politics, *Journal of Legal Studies* 9(4), 1980, pages 623-644.

Michael J. Trebilcock, *The Limits of Freedom of Contract*, Cambridge: Harvard University Press, 1993.

John Von Neumann, and Oskar Morgenstern, *Theory of Games and Economic Behavior*, Princeton: Princeton University Press, 1947.

Oliver E. Williamson, *The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting*, New York: Free Press, 1985.

Your Face Is Not a Bar Code:
Arguments Against Automatic Face Recognition in Public Places

Philip E. Agre
Department of Information Studies
University of California, Los Angeles
Los Angeles, California 90095-1520
USA

pagre@ucla.edu
http://polaris.gseis.ucla.edu/pagre/

Version of 10 September 2003.
5700 words.

An abridged version of this article appeared in *Whole Earth* 106, Winter 2001, pages 74-77.

Given a digital image of a person's face, face recognition software matches it against a database of other images. If any of the stored images matches closely enough, the system reports the sighting to its owner. Research on automatic face recognition has been around for decades, but accelerated in the 1990s. Now it is becoming practical, and face recognition systems are being deployed on a large scale.

Some applications of automatic face recognition systems are relatively unobjectionable. Many facilities have good reasons to authenticate everyone who walks in the door, for example to regulate access to weapons, money, criminal evidence, nuclear materials, or biohazards. When a citizen has been arrested for probable cause, it is reasonable for the police to use automatic face recognition to match a mug shot of the individual against a database of mug shots of people who have been arrested previously. These uses of the technology should be publicly justified, and audits should ensure that the technology is being used only for proper purposes.

Face recognition systems in public places, however, are a matter for serious concern. The issue recently came to broad public attention when it emerged that fans attending the Super Bowl had unknowingly been matched against a database of alleged criminals, and when the city of Tampa deployed a face-recognition system in the nightlife district of Ybor City. But current and proposed uses of face recognition are much more widespread, as the resources at the end of this article demonstrate in detail. The time to consider the acceptability of face recognition in public places is now, before the practice becomes entrenched and people start getting hurt.

Nor is the problem limited to the scattered cases that have been reported thus far. As the underlying information and communication technologies (digital cameras, image databases, processing power, and data communications) become radically cheaper over the next two decades, face recognition will become dramatically cheaper as well, even without assuming

major advances in technologies such as image processing that are specific to recognizing faces. Legal constraints on the practice in the United States are minimal. (In Europe the data protection laws will apply, providing at least some basic rights of notice and correction.) Databases of identified facial images already exist in large numbers (driver's license and employee ID records, for example), and new facial-image databases will not be hard to construct, with or without the knowledge or consent of the people whose faces are captured. (The images need to be captured under controlled conditions, but most citizens enter controlled, video-monitored spaces such as shops and offices on a regular basis.) It is nearly certain, therefore, that automatic face recognition will grow explosively and become pervasive unless action is taken now.

I believe that automatic face recognition in public places, including commercial spaces such as shopping malls that are open to the public, should be outlawed. The dangers outweigh the benefits. The necessary laws will not be passed, however, without overwhelming pressure of public opinion and organizing. To that end, this article presents the arguments against automatic face recognition in public places, followed by responses to the most common arguments in favor.

Arguments against automatic face recognition in public places

* The potential for abuse is astronomical. Pervasive automatic face recognition could be used to track individuals wherever they go. Systems operated by different organizations could easily be networked to cooperate in tracking an individual from place to place, whether they know the person's identity or not, and they can share whatever identities they do know. This tracking information could be used for many purposes. At one end of the spectrum, the information could be leaked to criminals who want to understand a prospective victim's travel patterns. Information routinely leaks from databases of all sorts, and there is no reason to believe that tracking databases will be any different. But even more insidiously, tracking information can be used to exert social control. Individuals will be less likely to contemplate public activities that offend powerful interests if they know that their identity will be captured and relayed to anyone that wants to know.

* The information from face recognition systems is easily combined with information from other technologies. Among the many "biometric" identification technologies, face recognition requires the least cooperation from the individual. Automatic fingerprint reading, by contrast, requires an individual to press a finger against a machine. (It will eventually be possible to identify people by the DNA-bearing cells that they leave behind, but that technology is a long way from becoming ubiquitous.) Organizations that have good reasons to identify individuals should employ whatever technology has the least inherent potential for abuse, yet very few identification technologies have more potential for abuse than face recognition. Information from face recognition systems is also easily combined with so-called location technologies such as E-911 location tracking in cell phones, thus further adding to the danger of abuse.

* The technology is hardly foolproof. Among the potential downsides are false positives, for example that so-and-so was "seen" on a street frequented by drug dealers. Such a report will create "facts" that the individual must explain away. Yet the conditions for image capture and recognition in most public places are far from ideal. Shadows, occlusions, reflections, and multiple uncontrolled light sources all increase the risk of false positives. As the database of

facial images grows bigger, the chances of a false match to one of those images grows proportionally larger.

* Face recognition is nearly useless for the application that has been most widely discussed since the September 11th attacks on New York and Washington: identifying terrorists in a crowd. As Bruce Schneier points out, the reasons why are statistical. Let us assume, with extreme generosity, that a face recognition system is 99.99 percent accurate. In other words, if a high-quality photograph of your face is not in the "terrorist watch list" database, then it is 99.99 percent likely that the software will not produce a match when it scans your face in real life. Then let us say that one airline passenger in ten million has their face in the database. Now, 99.99 percent probably sounds good. It means one failure in 10,000. In scanning ten million passengers, however, one failure in 10,000 means 1000 failures -- and only one correct match of a real terrorist. In other words, 999 matches out of 1000 will be false, and each of those false matches will cost time and effort that could have been spent protecting security in other ways. Perhaps one would argue that 1000 false alarms are worth the benefits of one hijacking prevented. Once the initial shock of the recent attacks wears off, however, the enormous percentage of false matches will condition security workers to assume that all positive matches are mistaken. The great cost of implementing and maintaining the face recognition systems will have gone to waste. The fact is, spotting terrorists in a crowd is a needle-in-a-haystack problem, and automatic face recognition is not a needle-in-a-haystack-quality technology. Hijackings can be prevented in many ways, and resources should be invested in the measures that are likely to work.

* Many social institutions depend on the difficulty of putting names to faces without human intervention. If people could be identified just from looking in a shop window or eating in a restaurant, it would be a tremendous change in our society's conception of the human person. People would find strangers addressing them by name. Prospective customers walking into a shop could find that their credit reports and other relevant information had already been pulled up and displayed for the sales staff before they even inquire about the goods. Even aside from the privacy invasion that this represents, premature disclosure of this sort of information could affect the customer's bargaining position.

* The public is poorly informed about the capabilities of the cameras that are already ubiquitous in many countries. They usually do not realize, for example, what can be done with the infrared component of the captured images. Even the phrase "face recognition" does not convey how easily the system can extract facial expressions. It is not just "identity" that can be captured, then, but data that reaches into the person's psyche. Even if the public is adequately informed about the capabilities of this year's cameras, software and data sharing can be improved almost invisibly next year.

* It is very hard to provide effective notice of the presence and capabilities of cameras in most public places, much less obtain meaningful consent. Travel through many public places, for example government offices and centralized transportation facilities, is hardly a matter of choice for any individual wishing to live in the modern world. Even in the private sector, many retail industries (groceries, for example) are highly concentrated, so that consumers have little choice but to submit to the dominant company's surveillance practices.

* If face recognition technologies are pioneered in countries where civil liberties are relatively strong, it becomes more likely that they will also be deployed in countries where civil liberties hardly exist. In twenty years, at current rates of progress, it will be feasible for

the Chinese government to use face recognition to track the public movements of everyone in the country.

Responses to arguments in favor of automatic face recognition in public places

"The civilized world has been attacked by terrorists. We have to defend ourselves. It's wartime, and we have to give up some civil liberties in order to secure ourselves against the danger."

We must certainly improve our security in many areas. I have said that myself for years. The fallacy here is in the automatic association between security and restrictions on civil liberties. Security can be improved in many ways that have no effect on civil liberties, for example by rationalizing identification systems for airport employees or training flight attendants in martial arts. Security can be improved in other ways that greatly improve privacy, for example by preventing identity theft or replacing Microsoft products with well-engineered software. And many proposals for improved security have a minimal effect on privacy relative to existing practices, for example searching passengers' luggage properly. The "trade-off" between security and civil liberties, therefore, is over-rated, and I am surprised by the speed with which many defenders of freedom have given up any effort to defend the core value of our society as a result of the terrorist attack.

Once we transcend automatic associations, we can think clearly about the choices that face us. We should redesign our security arrangements to protect both security and civil liberties. Among the many security measures we might choose, it seems doubtful that we would choose the ones that, like automatic face recognition in public places, carry astronomical dangers for privacy. At least any argument for such technologies requires a high standard of proof.

"But the case for face recognition is straightforward. They were looking for two of the terrorists and had photographs of them. Face recognition systems in airports would have caught them."

I'm not sure we really know that the authorities had photographs that were good enough for face recognition, even for those small number of suspects that they claim to have placed on a terrorist watch list. But even if we grant the premise, not much follows from it. First, the fact that the authorities suspected only two of the nineteen hijackers reminds us that automatic face recognition cannot recognize a face until it is in the database. Most hijackers are not on lists of suspected terrorists, and even if those particular hijackers had been prevented from boarding their planes, seventeen others would have boarded.

More importantly, security procedures at the Boston airport and elsewhere were so shoddy, on so many fronts, that a wide variety of improvements would have prevented the hijackings. If you read the white paper about the hijackings from the leading face-recognition company, Visionics, it becomes clear that face recognition is really being suggested to plug holes in identification systems. Terrorist watch lists include the terrorists' names, and so automatic face recognition is only necessary in those cases where the government possesses high-quality facial photographs of terrorists but does not know their names (not very common) or

where the terrorists carry falsified identification cards in names that the government does not know. In fact, some of the terrorists in the recent attacks appear to have stolen identities from innocent people. The best solution to this problem is to repair the immensely destructive weaknesses in identification procedures, for example at state DMV's, that have been widely publicized for at least fifteen years. If these recent attacks do not motivate us to fix our identity systems, then we are truly lost. But if we do fix them, then the role that automatic face recognition actually plays in the context of other security measures becomes quite marginal.

That said, from a civil liberties perspective we ought to distinguish among different applications of face recognition. Those applications can be arranged along a spectrum. At one end of the spectrum are applications in public places, for example scanning crowds in shops or on city streets. Those are the applications that I propose banning. At the other end of the spectrum are applications that are strongly bounded by legal due process, for example matching a mug shot of an arrested person to a database of mug shots of people who have been arrested in the past. When we consider any applications of automatic face recognition, we ought to weigh the dangers to civil liberties against the benefits. In the case of airport security, the proposed applications fall at various points along the spectrum. Applications that scan crowds in an airport terminal lie toward the "public" end of the spectrum; applications that check the validity of a boarding passenger's photo-ID card by comparing it with the photo that is associated with that card in a database lies toward the "due process" end of the spectrum. The dangers of face scanning in public places (e.g., the tracking of potentially unbounded categories of individuals) may not apply to applications at the "due process" end of the scale. It is important, therefore, to evaluate proposed systems in their specifics, and not in terms of abstract slogans about the need for security.

"All of the people in our database are wanted criminals. We don't store any of the images that our cameras capture, except when they match an image in the database. So the only people who have any cause for complaint are criminals."

The problems with this argument are numerous:

(1) We have to trust your word that the only people whose images are stored in the database are wanted criminals, and we have to trust your word that you throw away all of the images that fail to match the database.

(2) You don't really know yourself whether all of the people in the database are criminals. Quality control on those databases is far from perfect, as the database of "felons" that was used to purge some Florida counties' electoral rolls in 2000 demonstrated.

(3) Even if the only people in the database today are criminals, the forces pushing us down a slippery slope of ever-expanding surveillance are nearly overwhelming. Once the system is established and working, why don't we add alleged troublemakers who have been ejected from businesses in the past but have never been convicted of crimes? Then we could add people with criminal records who have served their time, people who have been convicted of minor offenses such as shoplifting, people with court orders to stay away from certain places, prisoners, parolees, gang members, soldiers, people with court summonses for minor offenses such as unpaid parking tickets, foreigners who have outstayed their visas, all foreigners in general, people with a history of mental illness, people who are wanted as material witnesses,

missing persons, children whose parents are worried about them, elders whose children are worried about them, parents who are behind on their child support, employees of the businesses where the system is operating, rich people who are afraid of being kidnapped, alcoholics who want to be kept out of bars, and other individuals who have signed contracts agreeing to be tracked. And once those people are added, it is then a short step to add many other categories of people as well.

"In effect you're saying that face recognition won't work, and that we should ban it because it will work so well. You are contradicting yourself."

Oh come on. Face recognition will work well enough to be dangerous, and poorly enough to be dangerous as well.

"Public is public. If someone happens to notice you walking in the park, you have no grounds for complaint if they decide to tell someone else where you were. That's all we're doing. You don't have any reasonable expectation of privacy in a public place, and I have a free-speech right to communicate factual information about where you were."

A human being who spots me in the park has the accountability that someone can spot them as well. Cameras are much more anonymous and easy to hide. More important is the question of scale. Most people understand the moral difference between a single chance observation in a park and an investigator who follows you everywhere you go. The information collected in the second case is obviously more dangerous. What is more, custom and law have always recognized many kinds of privacy in public. For example, the press cannot publish pictures of most people in personally sensitive situations that have no legitimate news value. It is considered impolite to listen in on conversations in public. Pervasive face recognition clearly lies at the morally most problematic end of this spectrum. The chance of being spotted is different from the certainty of being tracked.

The phrase "reasonable expectation of privacy" comes from a US Supreme Court decision. The phrase has been widely criticized as useless, simply because reasonable expectations of privacy in a situation can disappear as soon as someone starts routinely invading privacy in that situation. The problem is an often-exploited ambiguity in the word "expectation", which can mean either a prediction (with no logical implication that the world morally *ought* to conform to it) or a norm (with no logical implication that the world actually *will* conform to it). In arguing in favor of a ban on automatic face recognition in public places, one is not arguing for a blanket "right of privacy in public", which would be unreasonable and impractical. Rather, one is arguing for a right against technologically mediated privacy invasions of certain types. Technological mediation is key because of its continuous operation, standardized results, lack of other legitimate purposes, and rapidly dropping costs.

The argument about free speech rights is spurious because the proposed ban is not on the transfer of information, but on the creation of certain kinds of electronic records. You still have the right to communicate the same information if you acquire it in other ways.

"Providing proper notice of cameras in public places is easy. In Europe, many public places are plastered with signs that read 'This area monitored by CCTV'. What is the problem?"

The phrase "This area monitored by CCTV" does not properly convey what the cameras can do, much less what will be done with the images that they capture. As cameras and their capabilities become more diverse, notifications will have to become either more detailed or more vague. Likewise with the expanding range of potential secondary uses.

"Automatic face recognition is not all bad. It has positive uses. For example, as the technology gets miniaturized you could put a device in your glasses to remind you of people's names when you meet them. No doubt our inventive society will come up with other positive uses as well. Don't stigmatize the technology as simply a tool of oppression."

The technology does have positive uses. At the outset I acknowledged some of the positive uses that don't involve involuntary scanning of people in public places. The argument is: (1) the positive uses in public places are outweighed by the dangers, (2) even the positive uses in public places generally involve scanning people without their consent, and (3) the positive uses that do involve people's consent can almost always be done just as well with alternative technologies that do not lend themselves so easily to abuse.

"You can't outlaw technology. The technology will get out there anyway."

This same argument, if it made sense, would work against any law. Outlawing murder doesn't mean that no murders get committed. By passing laws against murder, society expresses its views of right and wrong, creates a deterrent, gets people who commit murder off the street so they'll be less likely to murder again, and makes murder much more difficult and expensive. Automatic face recognition in public places is not as bad as murder, but the analogy is clear: outlawing it would express public disapproval of it and make it harder than it would be otherwise.

"The real solution is to make sure that everyone is subject to surveillance. Once society is completely transparent, the powerful won't be able to use technology for repression, because their repressive scheming will be under surveillance too."

This scenario is unrealistic and immoral. The powerful by definition are the ones with the greatest capacity to escape surveillance, and so even in the greatest possible epidemic of surveillance the powerful would be the last to succumb. A regime of total surveillance would itself be extreme repression, and because a large proportion of the population would resist it, it could only be enforced through extreme repression. Promising that the repression will be turned against the powerful as well resembles nothing so much as the repression of the ruling classes in the French, Russian, and Chinese Revolutions. I am a democrat and an egalitarian myself, but I recognize that a generalized repression is the worst way to promote those values.

"Automatic face recognition stops crime. Police say they want it. By automating some of their more tedious jobs, it will free them to allocate their limited resources more effectively. And if it prevents one child from being killed then I support it."

A free society is a society in which there are limits on what the police can do. If we want to remain a free society then we need to make a decision. Once a new surveillance technology is installed, it is nearly impossible to stop the slippery slope toward ever broader law enforcement use of it. The case of automatic toll collection makes this clear. Absent clear legal protections, then, we should assume from the beginning that any technology that captures personal information will be used for law enforcement purposes, and not only in cases where lives are immediately at stake. The potential for abuse should then be figured into our decision about whether the technology should be deployed at all. That said, it is hardly proven that face recognition stops crime, when face recognition is being added to a world that already contains many other crime-fighting technologies. The range of crime detection technologies available to the police has grown immensely in recent years, and even if one encountered a case where a crime was solved using a given technology it by no means follows that the crime would not have been solved equally well using some other technology. And even if face recognition causes additional crimes to be prevented or solved, that effect should be weighed against the number of additional crimes that abuse of face recognition makes possible.

"I've been in the military and the police, and if you had seen some of the things that I've seen then you would change your mind."

You don't know what I've seen. Besides, everyone knows, having been reminded daily by the news, that evil crimes are committed every day. The real problem with your argument is that, like the argument I just addressed, it could be applied to support giving absolute power to the military and police. But then, by definition, we would no longer be a free society. We need principled arguments about the place of government force in a free society, and my purpose here is to suggest what some of those arguments might be.

"Why are you anti-law enforcement? The only thing that's keeping you, your families and your property safe is a robust law enforcement system. Without law enforcement your belongings would be stolen in no time."

To speak in terms of pro- versus anti-law enforcement is a simplistic dichotomy. Society should relate to the police the same way that we relate to the military: of course we need it, but if it becomes the central organizing principle of our culture then we are in trouble. It is dangerous to create a government bureaucracy, in this case the police, and tell it, "your one and only job is to suppress crime, and to that end we will give you absolutely whatever you ask for". That is a recipe for authoritarianism, which in the long run is no better for the police than it is for anyone else. Democracy means reckoning balances, not choosing between extremes, and the argument here is simply that one particular technology, automatic face recognition used in public places, creates such powerful imbalances that a democracy cannot tolerate it.

"Your arguments are scare tactics. Rather than trying to scare people with scenarios about slippery slopes, why don't you join in the constructive work of figuring out how the systems can be used responsibly?"

The arguments in favor of automatic face recognition in public places are "scare tactics" too, in that they appeal to our fear of terrorism. But some fears are justified, and it is reasonable to talk about them. Terrorism is a justifiable fear, and so is repression by a government that is given too much power. History is replete with examples of both. Plenty of precedents exist to suppose that automatic face recognition, once implemented and institutionalized, will be applied to ever-broader purposes. The concern about slippery slopes is not mere speculation, but is based on the very real politics of all of the many issues to which automatic face recognition could be applied. My argument here *is* intended to contribute to the constructive work of deciding how automatic face recognition can be responsibly used. It can be responsibly used in contexts where the individuals involved have been provided with due process protections, and it cannot be responsibly used in public places. I fully recognize that literally banning automatic face recognition in public places is a major step. The reason to ban it, though, is simple: the civil liberties dangers associated with automatic face recognition are virtually in a class by themselves.

"Liberty is not absolute. It is reasonable for the government to curtail liberty to a reasonable degree for the sake of the collective good."

Certainly so. The question is which curtailments of liberty provide benefits that are worth the danger. The argument here is simply that automatic face recognition in public places does not meet that test.

"The technology doesn't create anything new. If the government wants to follow you around now, they get plain-clothes cops to do it. The technology may make following you cheaper, but it doesn't make anything possible that wasn't possible before."

That's true with most information and communication technologies, which people use to amplify forces that already existed in society. The argument against automatic face recognition is not that it creates something qualitatively new, but that it amplifies existing dangers, such as political repression, beyond a level that a democracy can tolerate.

"What are you talking about? Your face already is a bar code. Everyone's face is unique, and people can use your face to recognize you. That's all the technology does."

Well, obviously, to say that your face is not a bar code is first and foremost a moral statement. Your face should not be *treated* as a bar code. But in fact, you face really is not a bar code. When a person sees your face, that is different from a machine reading a bar code because the person who sees your face cannot easily communicate to a third party what the face they saw looks like. That is why the police need skilled interviewers with specialized artistic techniques to recover facial images from eyewitnesses. An automatic face recognition machine, on the other hand, computes a digital representation of your face that is easily communicated, compared, stored, and associated with other information. So the technology does something more than what people do. If several different people spot you in several

different locations, then they cannot connect the different sightings unless they all know your name, or they are all shown a photograph of you, or your appearance is very distinctive in some way. Even then, the effort required to put the different sightings together is considerable. Machines can remember identities in industrial quantities, which people cannot do without special training, and they can assemble data across great distances much more quickly and efficiently than people can. The differences between human and machine face recognition, then, are so extensive that they cannot be treated as interchangeable.

"The evils that you envision are all speculative. This technology has not hurt anybody, and you can't go imposing a death sentence on it without evidence that it's dangerous."

The dramatic improvements in the underlying technology are hardly speculative. We know what technologies are in the lab, and we know roughly how long it will take before those technologies reach the market. We are therefore justified in extrapolating historical cost trends into the foreseeable future. The capabilities of the technology in the next couple of decades are hardly in doubt.

Nor can there be much doubt about the potential for abuse. We have abundant precedents from other technologies, and the burden is really on the person who would argue that automatic face recognition in public places will be an exception to these precedents. Databases will leak, technologies will exhibit function creep, information will be diverted to secondary uses, law enforcement will make use of technologies originally designed for other purposes, repressive governments will make use of technological advances pioneered in relatively free societies, and people's lives will be disrupted by quality control problems in the data. The argument here is not that automatic face recognition in public places will turn society into Orwell's *1984* overnight, or at all. The harms from automatic face recognition will develop slowly because the technology will not be deployed instantaneously, and because institutions change slowly. But the danger is great enough, and backed up by enough history and logic, and will be hard enough to reverse if it does materialize, that we are justified in acting now.

"When an automatic face recognition system produces a match, it is not the judge, jury, and executioner. If your name comes up wrongly, you'll be cleared in the same way that you'd be cleared after any other sort of mistaken identification. Automatic face recognition may not be perfect, but it's a lot more accurate than identification by human beings, and I don't see you trying to outlaw that."

As the cost of the underlying technologies drops exponentially, face recognition systems can easily become pervasive. As that happens, the number of opportunities for false identification will become pervasive as well. Identification by people and identification by machine cannot really be compared anyway because the conditions under which the police receive identifications from people and from machines are quite different. People can't easily be programmed to recognize large numbers of faces of people they have never met. And when a machine does produce a false match, the reputation of technology for accuracy will create a greater stigma than would a false identification by a person. In any event, the potential for false positives would not be a sufficient argument by itself against automatic face recognition

in public places. Combined with the other strong arguments, it is one part of a decisive case against them.

"Privacy prevents the marketplace from functioning efficiently. When a company knows more about you, it can tailor its offerings more specifically to your needs. Of course if you ask people whether scary face recognition systems should be banned then they'll say yes. But you're asking the wrong question. The right question is whether people are willing to give up information in exchange for something of value, and most people are."

This is a non sequitur. Few proposals for privacy protection prevent people from voluntarily handing information about themselves to companies with which they wish to do business. The problem arises when information is transferred without the individual's knowledge, and in ways that might well cause upset or harm if they became known. What distinguishes automatic face recognition from many other equally good identification technologies is that it can be used without the individual's permission (and therefore without the individual having agreed to any exchange). That is why it should be banned.

"A preoccupation with privacy is corrosive. Democracy requires people to have public personae, and excessive secrecy is unhealthy."

Privacy does not equal secrecy. Privacy means that an individual has reasonable control over what information is made public, and what is not. Any decent social order requires that individuals be entrusted with this judgement. Even if particular individuals choose to become secretive in a pathological way, forcing them to change will not help the situation and is intrinsincally wrong anyway. As to the value of public personae, we should encourage the development of technologies that give people the option to appear publicly where and how they want.

"What do you have to hide?"

This line is used against nearly every attempt to protect personal privacy, and the response in each case is the same. People have lots of valid reasons, personal safety for example, to prevent particular others from knowing particular information about them. Democracy only works if groups can organize and develop their political strategies in seclusion from the government and from any established interests they might be opposing. This includes, for example, the identities of people who might travel through public places to gather for a private political meeting. In its normal use, the question "What do you have to hide?" stigmatizes all personal autonomy as anti-social. As such it is an authoritarian demand, and has no place in a free society.

For more responses to bad arguments against privacy, see:
http://dlis.gseis.ucla.edu/people/pagre/arguments.html

Discussion of face recognition in the wake of the terrorist attacks.

comments advocating the use of face recognition in public places
http://www.nytimes.com/2001/12/06/national/06SURV.html?pagewanted=print
http://www.washingtontechnology.com/news/16_15/state/17338-1.html
http://www.washingtonpost.com/wp-dyn/articles/A53844-2001Oct25.html
http://www.washingtonpost.com/wp-dyn/articles/A14273-2001Sep23.html
http://www.nytimes.com/2001/09/19/nyregion/19TECH.html
http://www.nytimes.com/2001/09/16/nyregion/16SECU.html?pagewanted=all
http://www.nytimes.com/2001/09/15/national/15CIVI.html
http://news.cnet.com/news/0-1003-200-7141717.html

Biometrics in Airports: How To, and How Not To, Stop Mahommed Atta and Friends
http://www.anu.edu.au/people/Roger.Clarke/DV/BioAirports.html

Biometrics: Facing Up to Terrorism
http://www.rand.org/publications/IP/IP218/

Biometrics: A Look at Facial Recognition
http://www.rand.org/publications/DB/DB396/

Passports and Visas to Add High-Tech Identity Features
http://www.nytimes.com/2003/08/24/national/24IDEN.html?pagewanted=print

Consistent Security Is Elusive Airport Goal
http://www.washingtonpost.com/wp-dyn/articles/A13786-2002Feb15.html

European Commission's Proposal on Biometric Identifiers
http://europa.eu.int/rapid/start/cgi/guesten.ksh?p_action.gettxt=gt&doc=IP/03/1289|0|RAPID&lg=EN&display=

Viisage Director Takes Homeland Security Post
http://www.corporate-ir.net/ireye/ir_site.zhtml?ticker=VISG&script=410&layout=0&item_id=330130

skepticism about face recognition's use for preventing terrorism
http://www.wired.com/news/culture/0,1284,56878,00.html
http://www.reason.com/0210/fe.dk.face.shtml
http://www.wired.com/news/print/0,1294,54423,00.html
http://www.theregister.co.uk/content/55/25400.html
http://www.nytimes.com/2002/01/20/business/yourmoney/20PROF.html?pagewanted=print
http://www.nytimes.com/2002/01/15/science/physical/15FACE.html
http://www.zdnet.com/zdnn/stories/news/0,4586,5101223,00.html
http://www.pbs.org/cringely/pulpit/pulpit20011220.html
http://ComputerBytesMan.com/facescan/presentation/index.htm
http://www.nytimes.com/2001/10/07/magazine/07SURVEILLANCE.html?pagewanted=all

News articles with background on face recognition.

Face-Recognition Technology Improves (March 2003)
http://www.nytimes.com/2003/03/14/technology/14FACE.html

Facial ID Systems Raising Concerns About Privacy (August 2001)
http://washingtonpost.com/wp-dyn/articles/A12629-2001Jul31.html

Smile, You're on In-Store Camera (August 2002)
http://www.wired.com/news/print/0,1294,54078,00.html

New Side to Face-Recognition Technology: Identifying Victims (June 2002)
http://www.nytimes.com/2002/01/15/science/physical/15FACE.html

Your Face-Scan Dollars at Work (August 2001)
http://www.wired.com/news/technology/0,1282,46018,00.html

Facial-Recognition Tech Has People Pegged (July 2001)
http://www.cnn.com/2001/TECH/ptech/07/17/face.time.idg/

Face Scanners Turn Lens on Selves (July 2001)
http://wired.com/news/privacy/0,1848,45687,00.html

article about a biometric industry public relations initiative (September 2001)
http://www.wired.com/news/print/0,1294,46539,00.html

How Facial Recognition Software Finds Faces (July 2001)
http://abcnews.go.com/sections/scitech/CuttingEdge/cuttingedge010706.html

Law Enforcement Agencies Working on 3D Face Recognition Technology (September 1999)
http://asia.cnn.com/TECH/computing/9909/24/3d.face.recognition.idg/

Face-Recognition Technology Raises Fears of Big Brother (February 2000)
http://www.deseretnews.com/dn/view/0%2C1249%2C150015975%2C00.html

Smile, You're on Scan Camera (March 2001)
http://www.wired.com/news/technology/0,1282,42317,00.html

Face Recognition Via Cell Phones (March 2002)
http://www.internetnews.com/infra/print.php/999361

Other sites with background information on face recognition technology and its potential for privacy invasion.

Electronic Privacy Information Center Face Recognition Page
http://www.epic.org/privacy/facerecognition/

Coalition Declares December 24, 2001 to Be "World Subjectrights Day"
http://wearcam.org/wsd.htm

Facial Recognition Vendor Test 2002
http://www.frvt.org/FRVT2002/default.htm

Facial Recognition Vendor Test 2000
http://www.dodcounterdrug.com/facialrecognition/DLs/FRVT_2000.pdf
http://www.dodcounterdrug.com/facialrecognition/FRVT2000/frvt2000.htm

Selected Facial Scan Projects
http://www.facial-scan.com/selected_facial_scan_projects1.htm

US government site for biometric technology (including face recognition)
http://www.biometrics.org/

Facing the Truth: A New Tool to Analyze Our Expressions
http://www.hhmi.org/bulletin/may2001/faces/

the two dominant face recognition companies
http://www.visionics.com/faceit/
http://www.viisage.com/

The Many Faces of Viisage
http://www.notbored.org/viisage.html

other companies
http://www.visionspheretech.com/menu.htm
http://www.cognitec-ag.de/
http://www.c-vis.com/htdocs/english/facesnap/
http://www.neurodynamics.com/
http://www.imagistechnologies.com/
http://www.spiritcorp.com/face_rec.html
http://www.bioid.com/
http://www.keyware.com/
http://www.bionetrix.com/

Web pages about technical research projects on face recognition.

Face Recognition and Detection
http://home.t-online.de/home/Robert.Frischholz/face.htm

Fully Automatic Upper Facial Action Recognition
ftp://whitechapel.media.mit.edu/pub/tech-reports/TR-571.pdf

DoD Counterdrug Program Face Recognition Technology Program
http://www.dodcounterdrug.com/facialrecognition/Feret/feret.htm
http://www.itl.nist.gov/iad/humanid/feret/feret_master.html

Handheld Face Identification Technology in a Pervasive Computing Environment
http://www.ai.mit.edu/projects/cbcl/publications/ps/pervasive-2002.pdf

Wearable Face Recognition and Detection
http://www.gvu.gatech.edu/ccg/projects/face/

Identification of Faces From Video
http://staff.psy.gla.ac.uk/~mike/videoproj.html

Evaluation of Face Recognition Algorithms
http://www.cs.colostate.edu/evalfacerec/

slides from an MIT course on human and artificial face recognition
http://web.mit.edu/9.670/www/

Gesture Recognition Home Page (related technology)
http://www.cybernet.com/~ccohen/

Articles about face-recognition controversies in various places, roughly in reverse chronological order.

Borders stores

first Borders says it "suspended any plans to implement" face recognition ...
http://www.computerworld.com/storyba/0,4125,NAV47_STO63359,00.html

... then it denies that it ever had any such intention
http://www.politechbot.com/p-02447.html

Borders is planning to use face recognition to identify shop-lifters
http://www.sundayherald.com/18007/

casinos

OPP uses secret cameras in casinos
("police are secretly scanning the faces of customers at all Ontario casinos")
http://www.efc.ca/pages/media/2001/2001-01-16-a-torontostar.html

Boston

Airport Anti-Terror Systems Flub Tests
http://usatoday.printthis.clickability.com/pt/cpt?action=cpt&expire=&urlID=7387802&fb=Y&partnerID=1664

Face Recognition Fails in Boston Airport
http://www.theregister.co.uk/content/55/26298.html

Logan Will Test Face-Data Security
http://www.boston.com/dailyglobe2/298/metro/Logan_will_test_face_data_security+.shtml

Virginia Beach, Virginia

Virginia Beach Installs Face-Recognition Cameras
http://www.washingtonpost.com/wp-dyn/articles/A19946-2002Jul3.html

Oakland, California

Oakland Airport: "Smile for the Camera"
http://www.usatoday.com/life/cyber/tech/2001/10/18/airport-camera.htm

Huntington Beach, California

Imagis and ORION Chosen to Install Biometrics by Huntington Beach Police
http://cipherwar.com/news/01/imagis_big_brother.htm

Providence, Rhode Island

Airport Chief Reconsiders Face Recognition Technology for Green
http://www.projo.com/cgi-bin/story.pl/news/06877271.htm

Australia

SmartGate: A Face Recognition Trial at Sydney Airport
http://www.anu.edu.au/people/Roger.Clarke/DV/SmartGate.html

Passengers Secretly Filmed in Anti-Terror Trial
http://www.smh.com.au/articles/2003/01/04/1041566268528.html

Colorado

Colorado Governor Doesn't Want Face Recognition Technology Abused
http://www.thedenverchannel.com/den/entertainment/stories/technology-87985620010719-070716.html

Colorado Won't Use Facial Recognition Technology on Licenses
http://www.thedenverchannel.com/den/entertainment/stories/technology-86955020010712-110740.html

Minnesota

Security Face-Scanning Coming to Airport?
http://www.channel4000.com/msp/news/stories/news-131098520020319-070306.html

New York

Cameras to Seek Faces of Terror In Visitors to the Statue of Liberty
http://www.nytimes.com/2002/05/25/nyregion/25CAME.html

Missouri

Nuke Reactor: Show Me Your Face
http://www.wired.com/news/print/0,1294,54423,00.html

Super Bowl

Face Scans Match Few Suspects
http://www.sptimes.com/News/021601/TampaBay/Face_scans_match_few_.shtml

ACLU Protests High-Tech Super Bowl Surveillance
http://www.usatoday.com/life/cyber/tech/2001-02-02-super-bowl-surveillance.htm

Super Bowl Surveillance: Facing Up to Biometrics
http://www.rand.org/publications/IP/IP209/IP209.pdf

Feds Use Biometrics Against Super Bowl Fans
http://www.theregister.co.uk/content/6/16561.html

Cameras Scanned Fans for Criminals
http://www.sptimes.com/News/013101/TampaBay/Cameras_scanned_fans_.shtml

Jacksonville, Florida

Police Snooper Camera Fight Still Alive
http://www.jacksonville.com/tu-online/stories/083101/met_7161286.html

Tampa, Florida

Tampa Police Eliminate Facial-Recognition System
http://www.palmbeachpost.com/news/content/news/0820camera.html

Face Recognition Technology a Proven Farce
http://www.theregister.co.uk/content/6/23559.html

Facial Frisking in Tampa
http://www.privacyfoundation.org/commentary/tipsheet.asp?id=46&action=0

"Big Brother" Cameras on Watch for Criminals
http://www.usatoday.com/life/cyber/tech/2001-08-02-big-brother-cameras.htm

"They made me feel like a criminal"
http://www.sptimes.com/News/080801/TampaBay/_They_made_me_feel_li.shtml

Civil Rights or Just Sour Grapes?
http://www.sptimes.com/News/080301/TampaBay/Civil_rights_or_just_.shtml

Click. BEEP! Face Captured
http://www.sptimes.com/News/071901/Floridian/Click_BEEP_Face_captu.shtml

Tampa Gets Ready For Its Closeup
http://www.time.com/time/nation/article/0,8599,167846,00.html

Masked Protesters Fight Face Scans
http://www.sptimes.com/News/071501/TampaBay/Masked_protesters_fig.shtml

Tampa Puts Face-Recognition System on Public Street
http://www.usatoday.com/life/cyber/tech/2001-07-13-tampa-surveillance.htm

Tampa Scans the Faces in Its Crowds for Criminals
http://www.nytimes.com/2001/07/04/technology/04VIDE.html

public radio report about the controversy
http://www.npr.org/ramfiles/atc/20010702.atc.14.rmm

Ybor Police Cameras Go Spy-Tech
http://www.sptimes.com/News/063001/TampaBay/Ybor_police_cameras_g.shtml

Palm Beach, Florida

Palm Beach Airport Won't Use Face-Scan Technology
http://www.local6.com/orlpn/news/stories/news-148124920020526-160533.html

Face Recognition Kit Fails in Florida Airport
http://www.theregister.co.uk/content/55/25444.html

Britain

Think Tank Urges Face-Scanning of the Masses
http://www.theregister.co.uk/content/6/20966.html

face recognition technology in the UK
http://www.urban75.com/Action/cctv.html
http://www.sourceuk.net/articles/a00624.html

Iceland

Iceland Places Trust in Face-Scanning
http://news.bbc.co.uk/hi/english/sci/tech/newsid_1780000/1780150.stm

Iceland's Keflavik Airport Upgrades CCTV System with Visionics' FaceIt
http://ir.shareholder.com/vsnx/ReleaseDetail.cfm?ReleaseID=45325