The Datasphere, Thick Mediation, and Learning Algorithms: Toward a Sociology of the Present

Rafael Alvarado, Program Director
University of Virginia School of Data Science

SAS 2019: *The Society of Learning Algorithms*
High Performance Computing Center (HLRS)
Stuttgart, Germany
11 November 2019

UVA DATA SCIENCE

HLRS
High-Performance Computing Center | Stuttgart

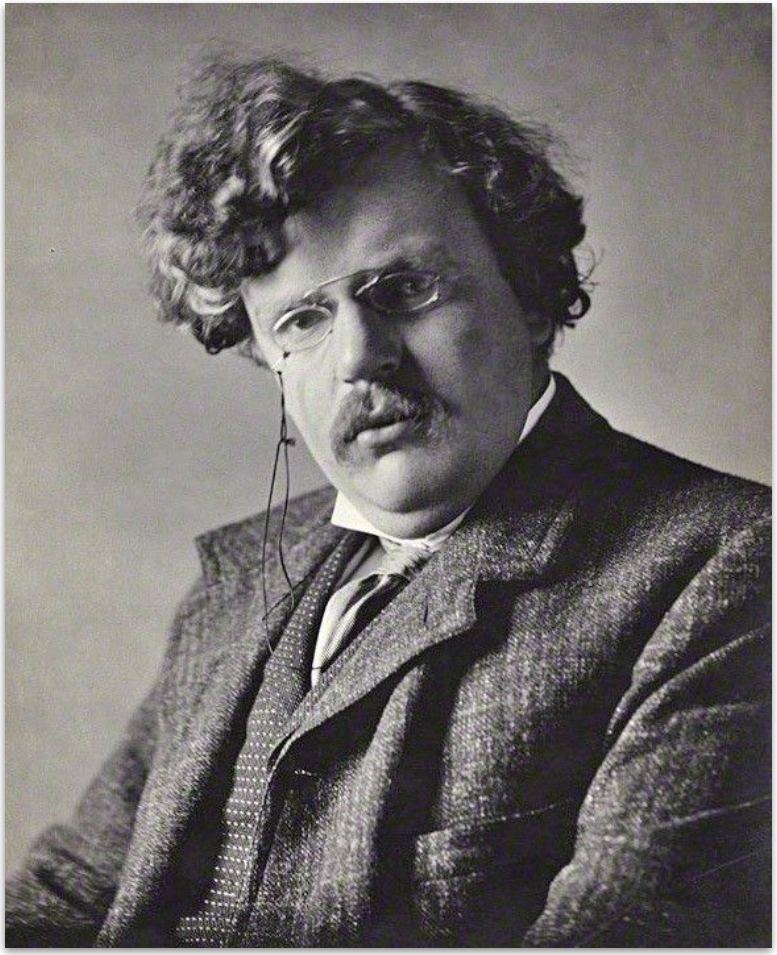
Background

@ontoligent
http://transducer.ontoligent.com

Ph.D. Cultural Anthropology, U.Va.
— Classic Mayan kingship
— Comparative study of media

Digital Humanist at Princeton and U.Va.
— Publishing and teaching on digital humanities and anthropology of the Internet
— Development for digital text archives in Old French, Persian, Hebrew, Mayan, and other languages

Member of School of Data Science, U.Va.
— MSDS Program Director
— Specialist in Text Analytics
— Co-founded Center for the Study of Knowledge and Data
— Member Human-Machine Intelligence Group, Humanities Informatics Lab



“The madman is not the man who has lost his reason. The madman is the man who has lost everything except his reason.”

— G.K. Chesterton,
Orthodoxy

By this definition of madness,
LAs are mad ...

Overview

Topics

- Introduction
- Thick Mediation
- The Datasphere
- Epistemic Effects

Approach

- High-level sketch — not a complete argument
- Historical-structural — use history to trace formation of structuress

How can we
conceptualize Learning algorithms (LAs)
in a fully **sociological** manner?

Reclaiming the Social

It has become necessary to define the social

The term has come to refer to the social as mediated by Internet

Social **media**
Social **justice**
Social **computing**

I mean social in the **tradition** of

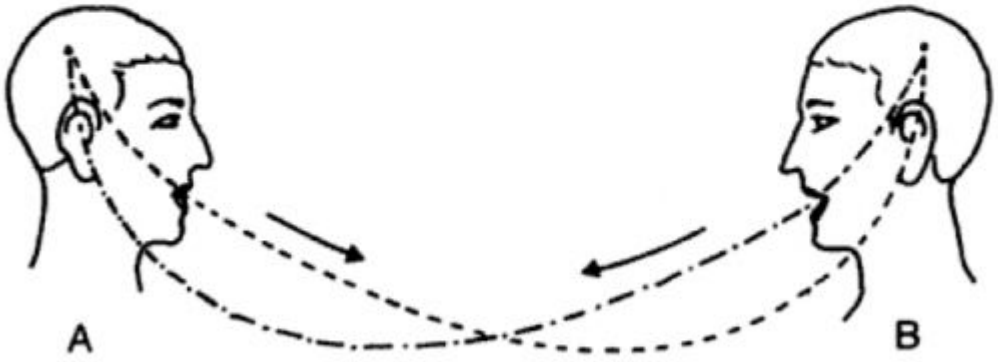
Marx, Weber, Tönnies
Comte, Tocqueville, Durkheim, Mauss

Social structure and function
Social media and symbolic interactionism

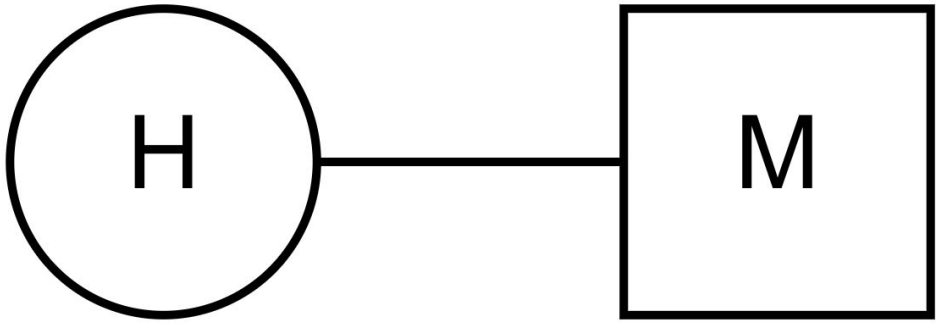
Let us begin with the question
that Durkheim might have asked:

What is the **elementary form** of the
social **relationship** between people and LAs?

Humans and machines?



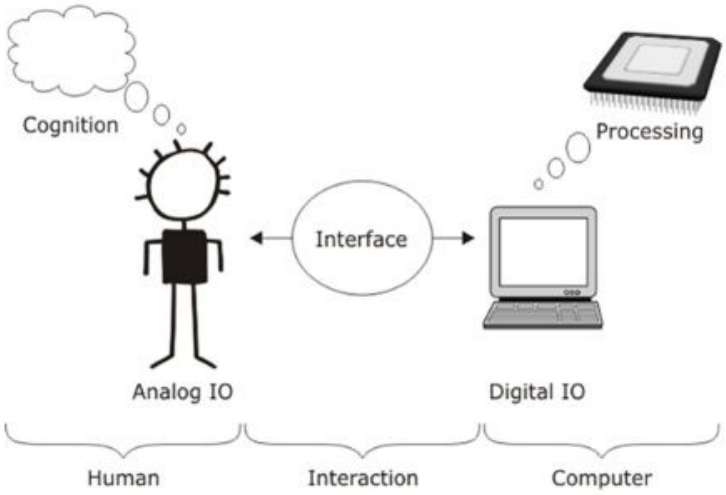
Human **H** Machine **M**



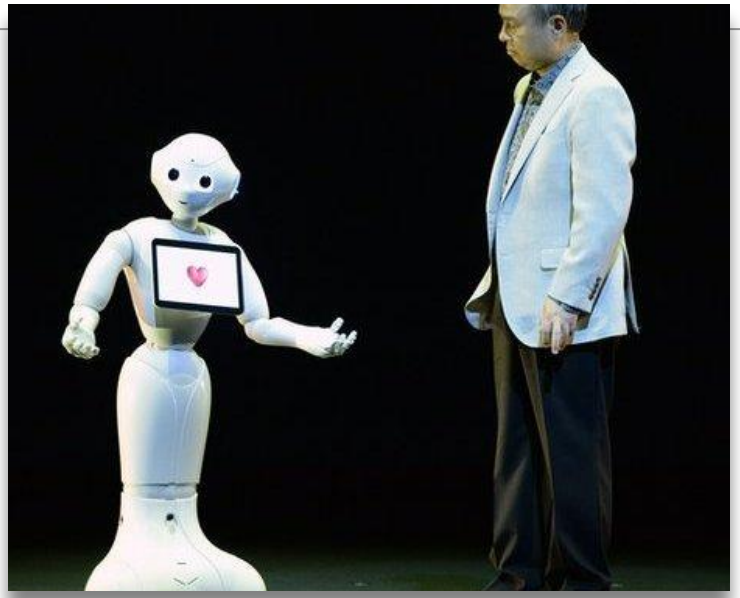
We tend to adopt a dyadic, Sassurian model of communication in framing this question

What is HCI?

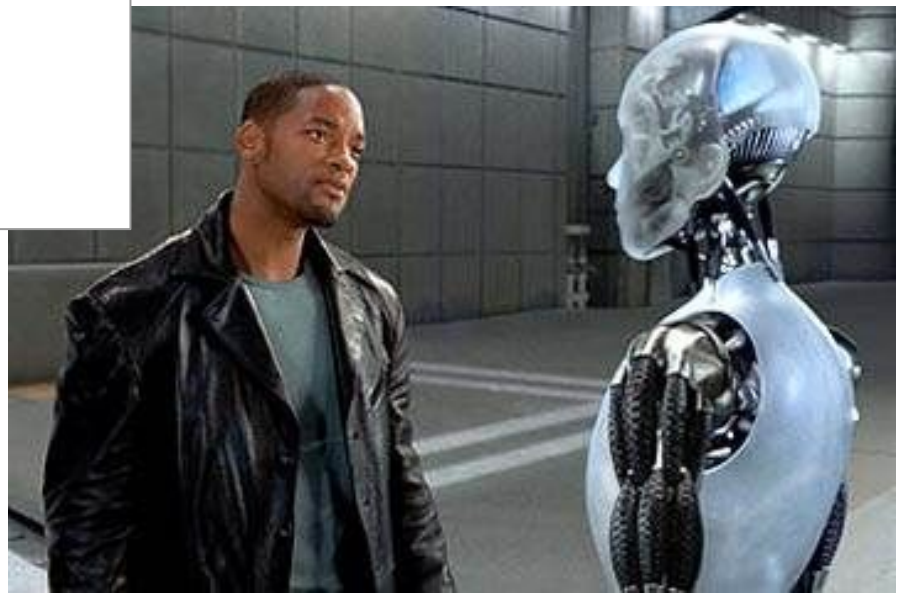
HCI is the study of interaction between people (users) and computers



This model is assumed in both formal and informal representations



Japanese robot "Pepper"



Will Smith and Sonny in *I, Robot*

Human-Centered Artificial Intelligence
Stanford University

If AI is to serve the collective needs of humanity, it must incorporate an understanding of what moves us — physically, intellectually and emotionally. It is critical that we design machine intelligence that can understand human language, feelings, intentions and behaviors, and interact with nuance and in multiple dimensions.

One could be less kind and call this view **Thatcherian**, after Margaret Thatcher, who famously declared

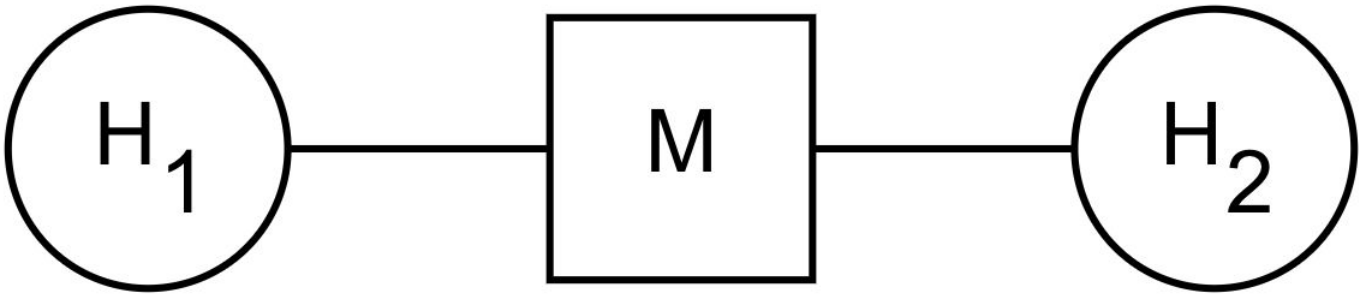
“There is no such thing as Society”

since it follows from the same premise of **methodological individualism**

In reality, the relationship is **triadic**

1. Our interaction with LAs is **a special case human computer interaction** (HCI)

2. In HCI, computers (machines) **act as media**



The computer has nearly **always been social**

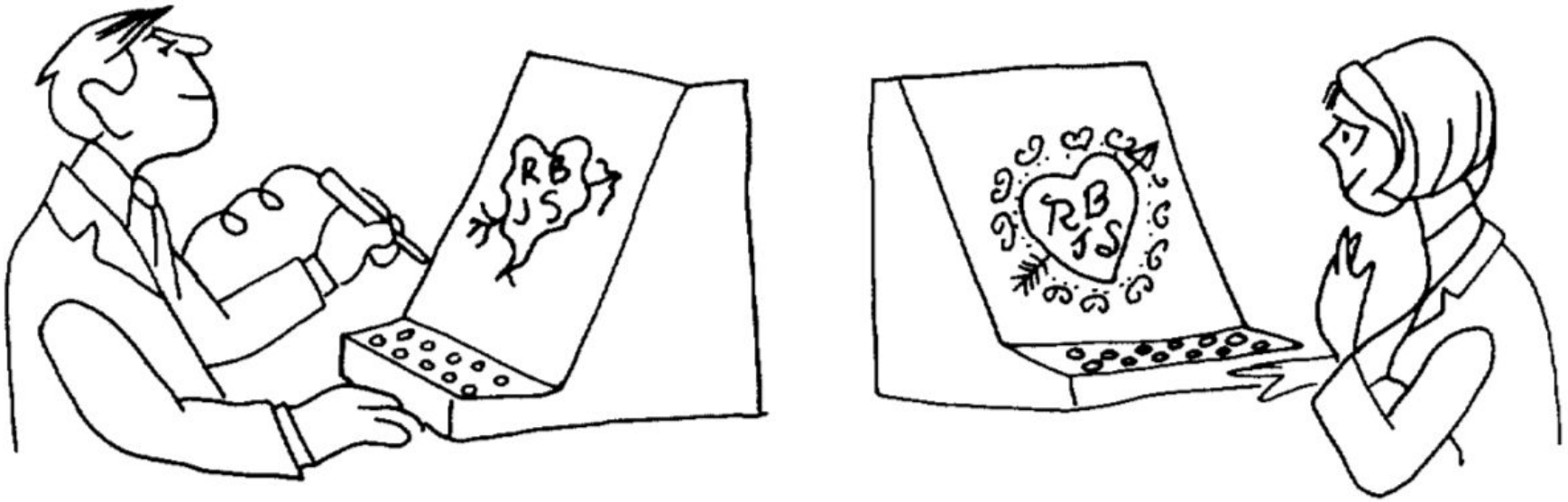
Since the invention of **time sharing** and the original conception of the **Internet** in the 1960s by **Licklider**

From the outset, the computer was imagined as a **networked apparatus** that would be connected to other computers and other people

(In what follows, when I refer to the machine M, usually I mean this apparatus)

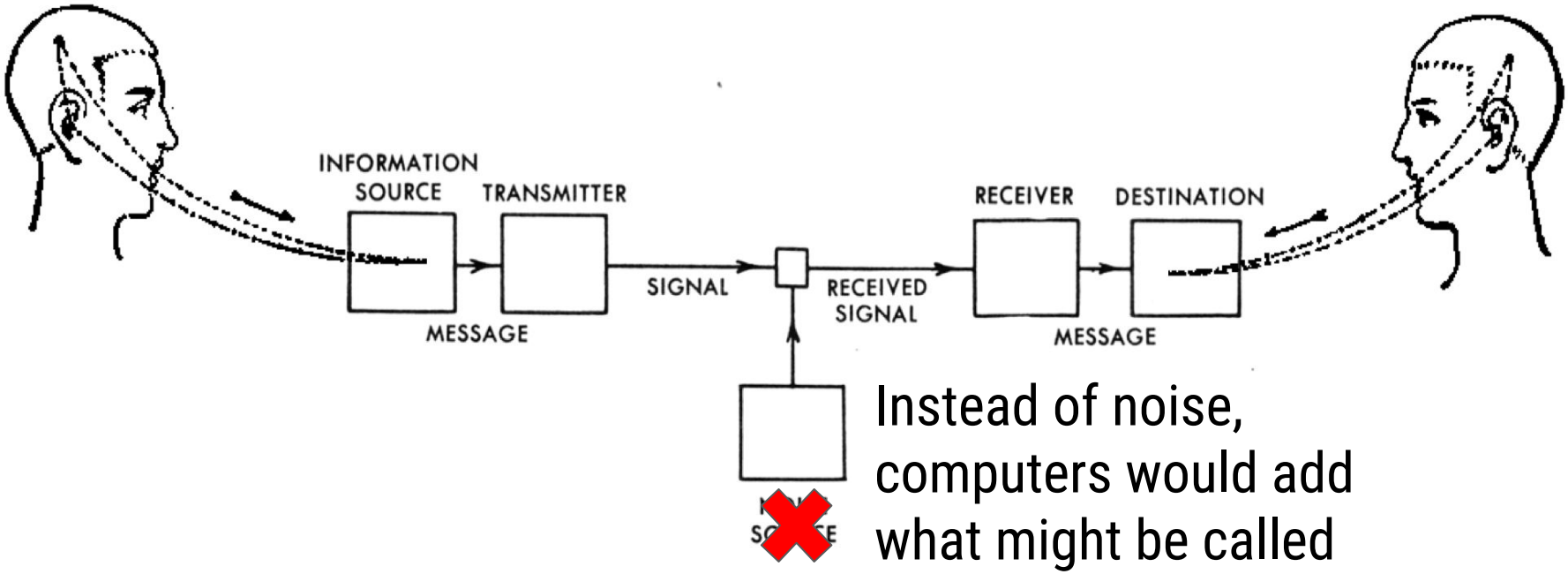
Man-Computer Symbiosis

J. C. R. LICKLIDER (1960)

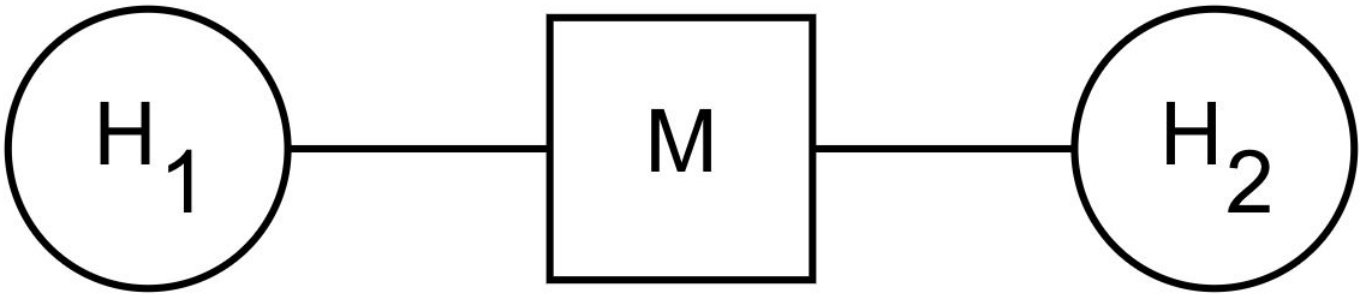


Licklider imagined humans using computers to communicate through **deep modeling** — a process where **messages are reshaped** by mediating processes

Licklider and Taylor, 1968, “The Computer as Communication Device,” p. 26

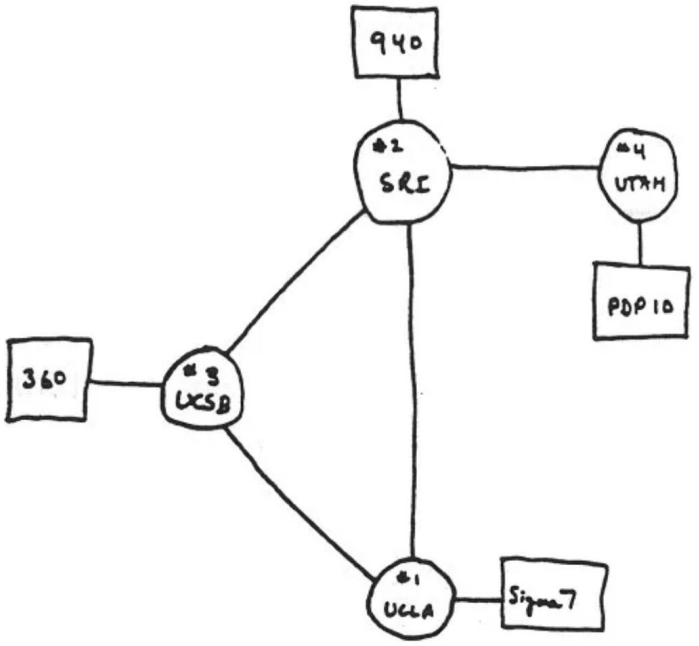


Instead of noise, computers would add what might be called **anti-noise**



To enable this use of the computer, Licklider envisioned an “**intergalactic computer network**”

"... an **electronic commons** open to all, ‘the main and essential medium of informational interaction for governments, institutions, corporations, and individuals’”



THE ARPA NETWORK

DEC 1969

4 NODES

One of the **side effects** of networked communication is the development and quiet centralization of the **database** within the network

The database is **underappreciated**

Algorithms get all the **glory**

But the the database **makes LAs possible**

Consider Englebert's "Mother of All Demos" (1968)

"The Mother of All Demos." 1968, TS 55:09-24.
http://www.youtube.com/watch?v=yJDv-zdhzMY

Douglas Engelbart's December 9, 1968, live demonstration of NLS (oN-Line System), a computer collaboration system developed at SRI

NLS featured the mouse, video conferencing, teleconferencing, hypertext, word processing, hypermedia, object addressing and dynamic file linking, bootstrapping, and a collaborative real-time editor

Anyway, one of the interesting things that NLS does, just an advantage of being online, is that **it keeps track of who you are and what you're doing all the time**. So on these statements, uh ... on everything, every statement that you write, it keeps track of who you are and when you did it.

Jeff Rulifson (Menlo Park), 1968, extemporaneous comment about email's precursor, NLS, from Engelbart's so-called "Mother of All Demos." Emphasis added. 55:09-24.

Email continues this effect

1969 also saw the appearance of ARPANET (and CompuServe)

In 1971 Roy Tomlinson invented email (on a lark)

- Simply combined SENDMSG and CYPNET

By 1973 email accounted for **most of the traffic** on the Internet

By 1977 email was being viewed as **the *raison d'etre* for the Internet**

In trying to make sense of this development, Henderson and Myer (1977) present an overview of email

- They highlight a feature that becomes foundational for social media . . .

An alternative view is to consider the message system in its entirety (including the delivery subsystem and all manipultion tools) as a database management system. To create a message is to insert a record into the data base. To send a message is to provide some set of recipients with access to that record. Provided sufficient technology can be mustered to support the paradigm, some rather powerful effects are achieved. For example, notations added to a message by one recipient

The authors rethink email as a form of record keeping

i.e. email is **database-mediated communication**

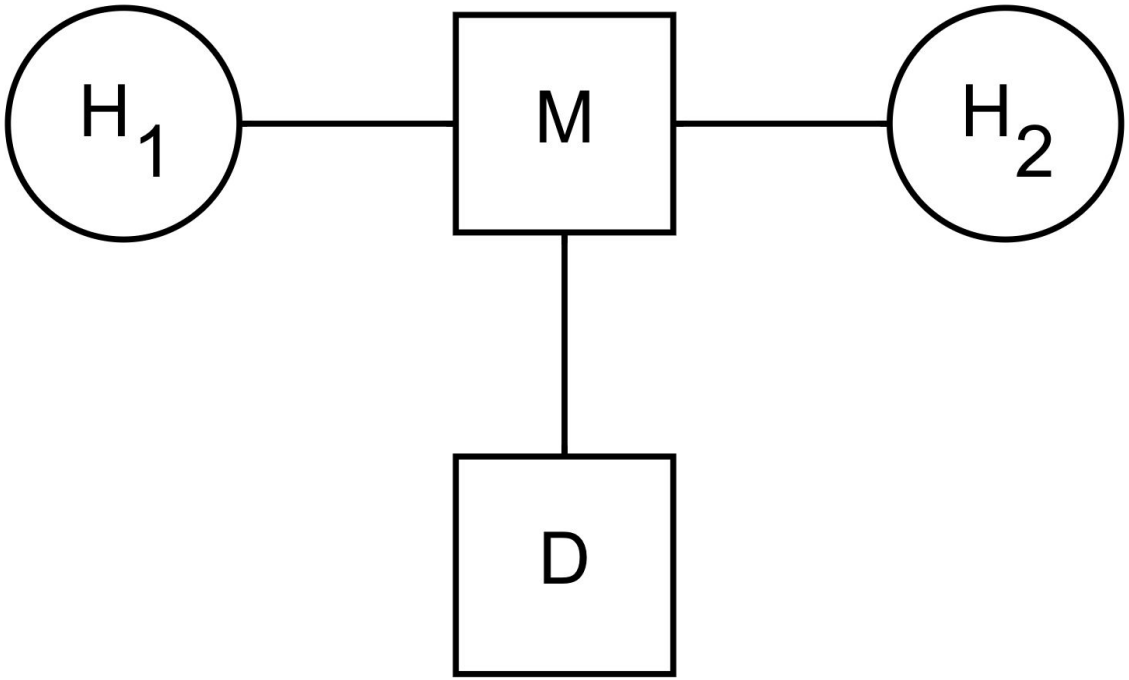
This is true of all networked forms of communication, from BBSes to WhatsApp

The database is the condition of possibility of networked computer

By the time of Berners-Lee's invention of **the Web**, the Internet has become a vast social network that has accumulated a **surplus of "social" data**

This development is paralleled by **BBSes** whose users eventually migrate to the Web

Revised elementary unit of human-machine sociality

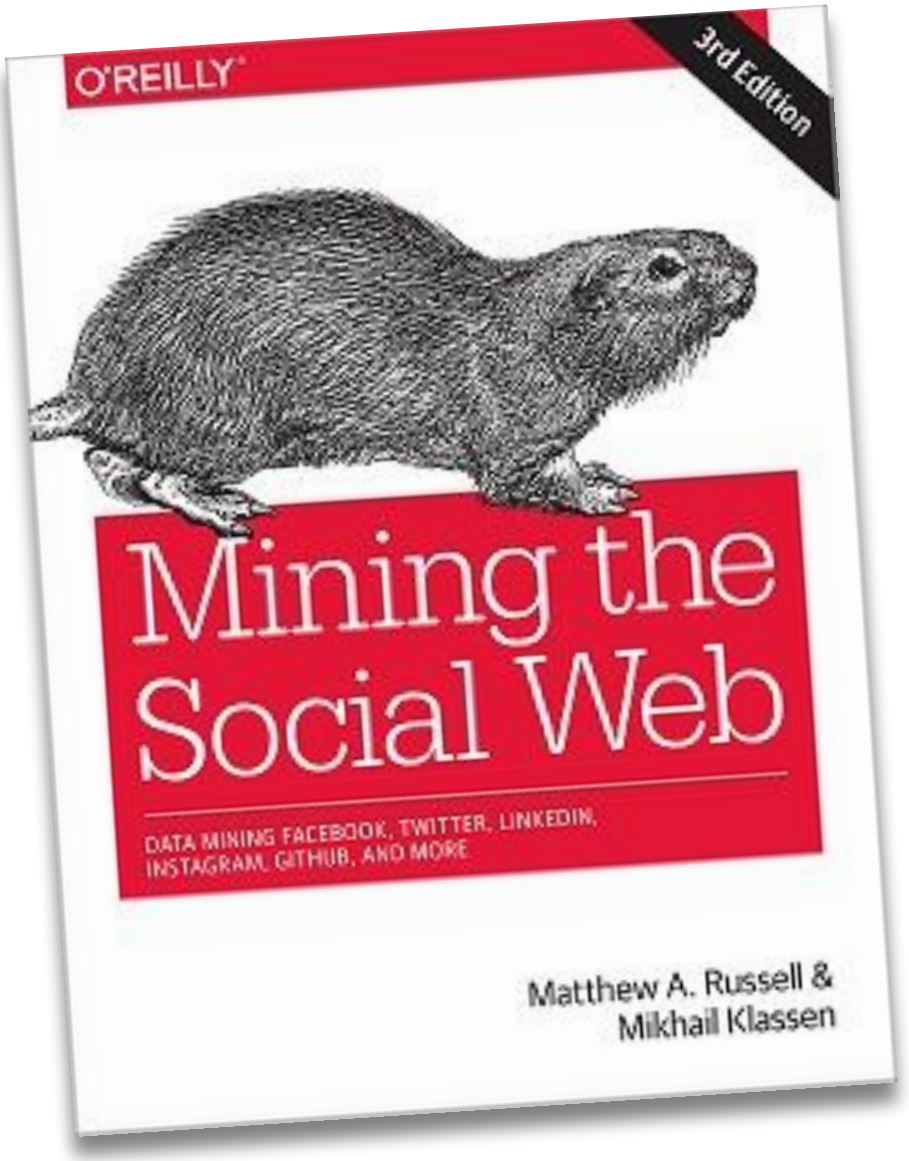


Why is this important for our understanding of LAs?

Although **statistical learning theory** has been established since since the 1980s (Vapnik 2000), successful implementations were not seen until the 1990s

After the rise the spectacular rise of the Web and the appearance of **surplus data**

The Web gave both the **data** and **motivation** for **supervised machine learning**
(the dominant form of LA now)



This is one of the books that helped **launch data science**, at least one version of it

It shows the connection between the **source** of data and the **methods**

The Web is more a social creation than a technical one. I designed it for a social effect—to help people work together—and not as a technical toy. The ultimate goal of the Web is to support and improve our weblike existence in the world. We clump into families, associations, and companies. We develop trust across the miles and distrust around the corner.

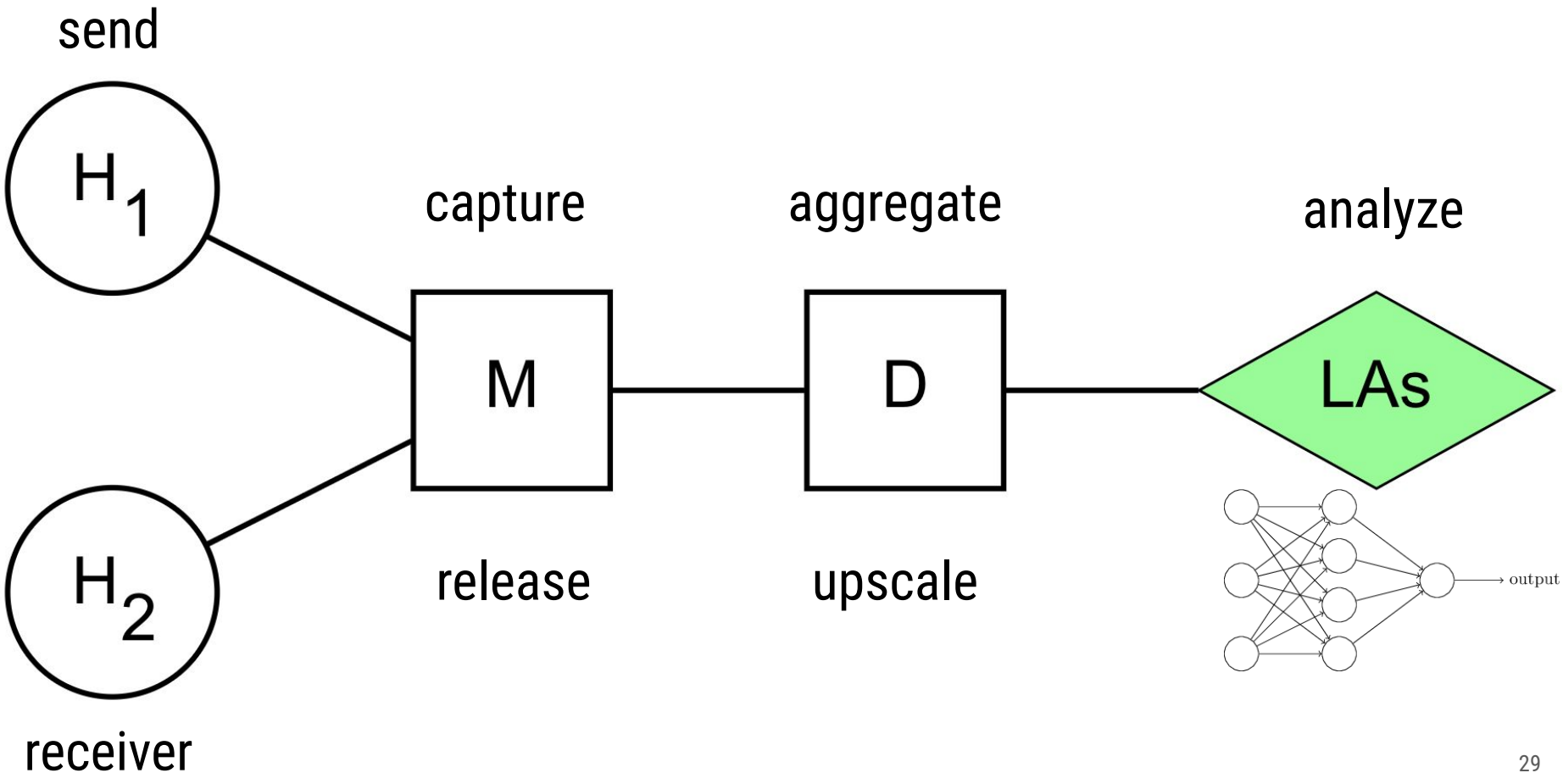
—Tim Berners-Lee, *Weaving the Web* (Harper)

Data science was (re)invented to mine the **Giant Global Graph**

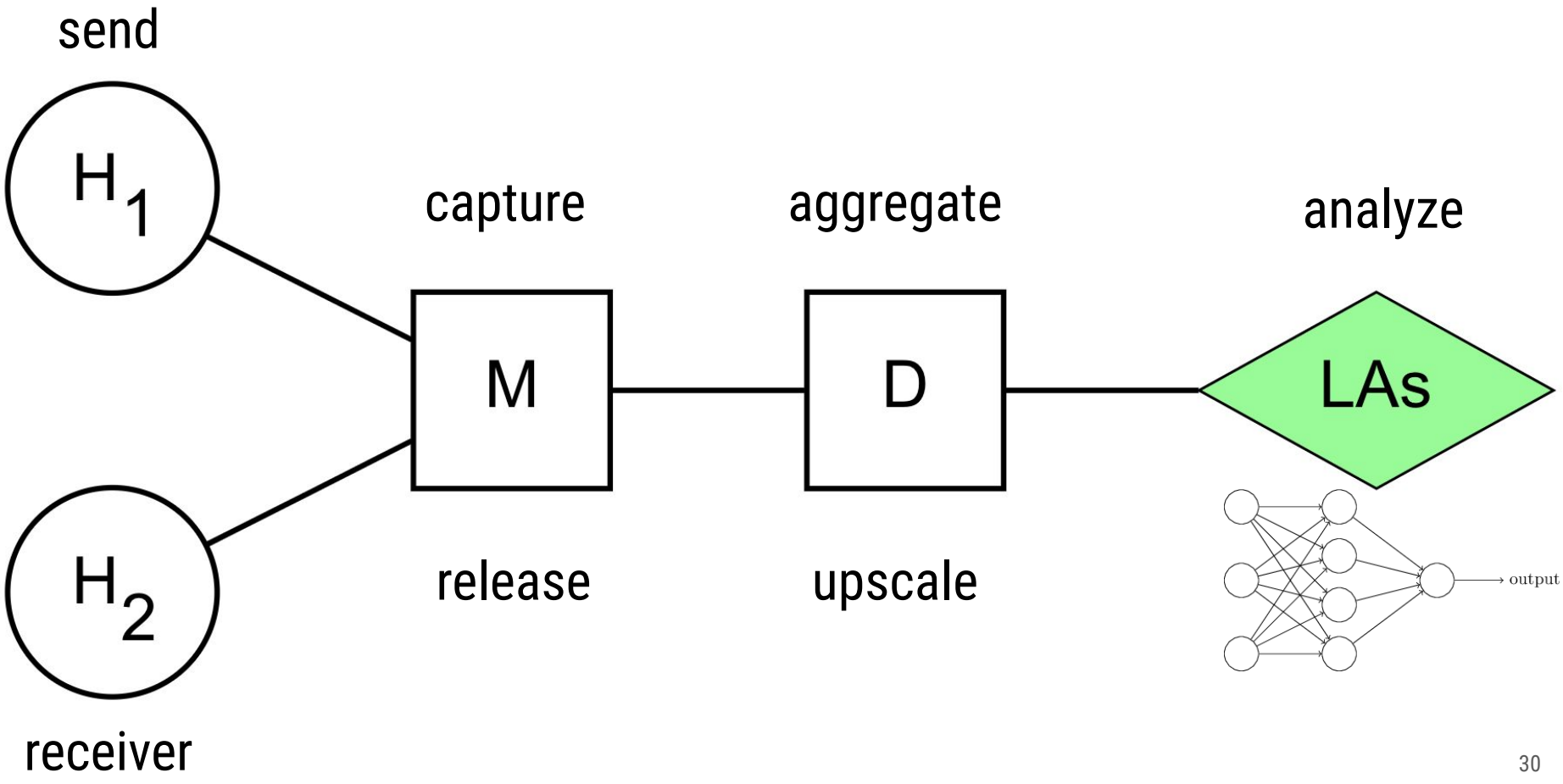
Loosely speaking, this book treats *the social web*‡ as a graph of people, activities, events, concepts, etc. Industry leaders such as Google and Facebook have begun to increasingly push graph-centric terminology rather than web-centric terminology as they simultaneously promote graph-based APIs. In fact, Tim Berners-Lee has suggested that perhaps he should have used the term Giant Global Graph (GGG) instead of World Wide Web (WWW), because the terms "web" and "graph" can be so freely interchanged in the context of defining a topology for the Internet. Whether the fullness of Tim Berners-

So, what is **the place of LAs** in the GGG?

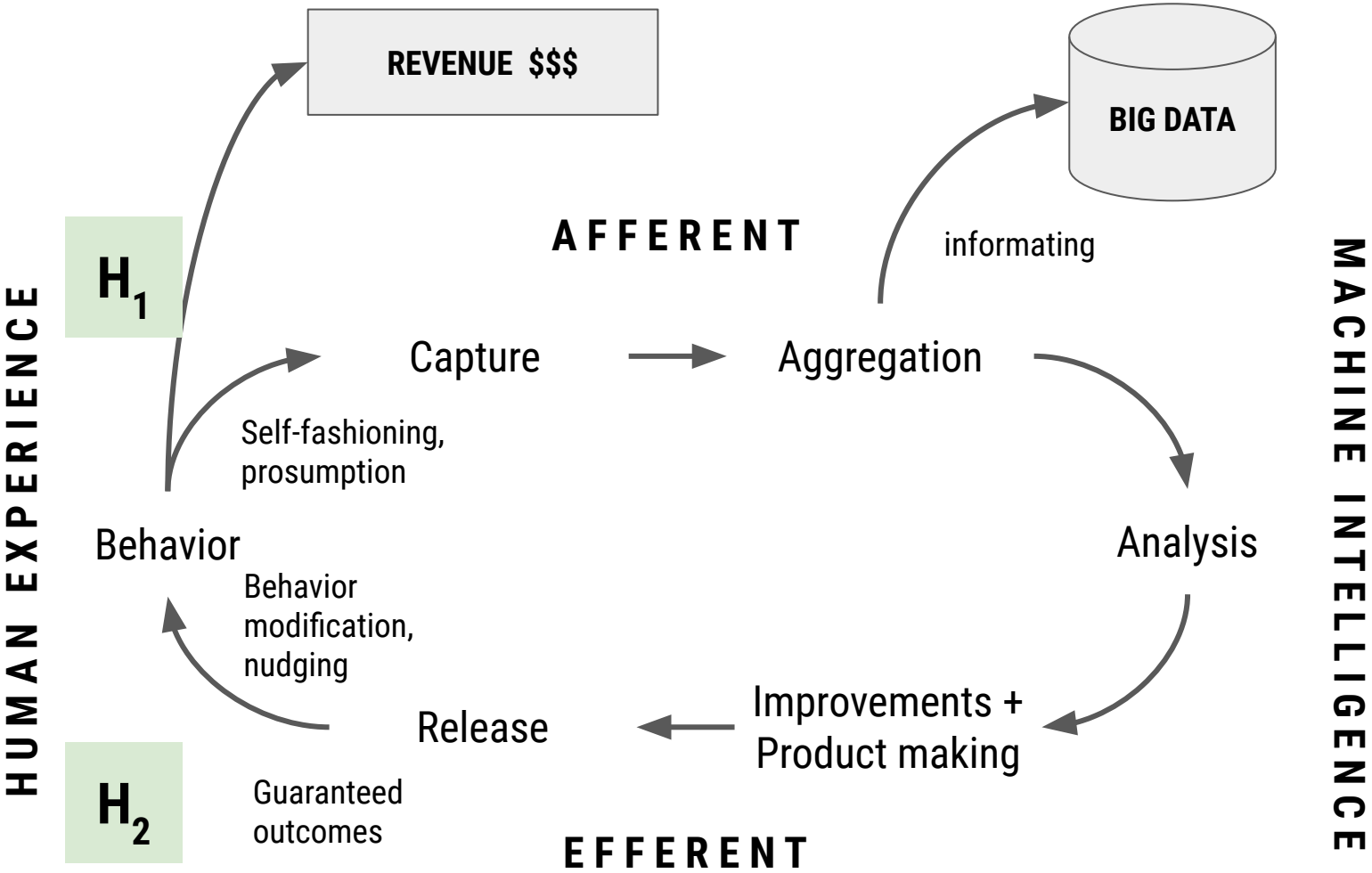
LAs occupy a **central position** in the data science **pipeline**

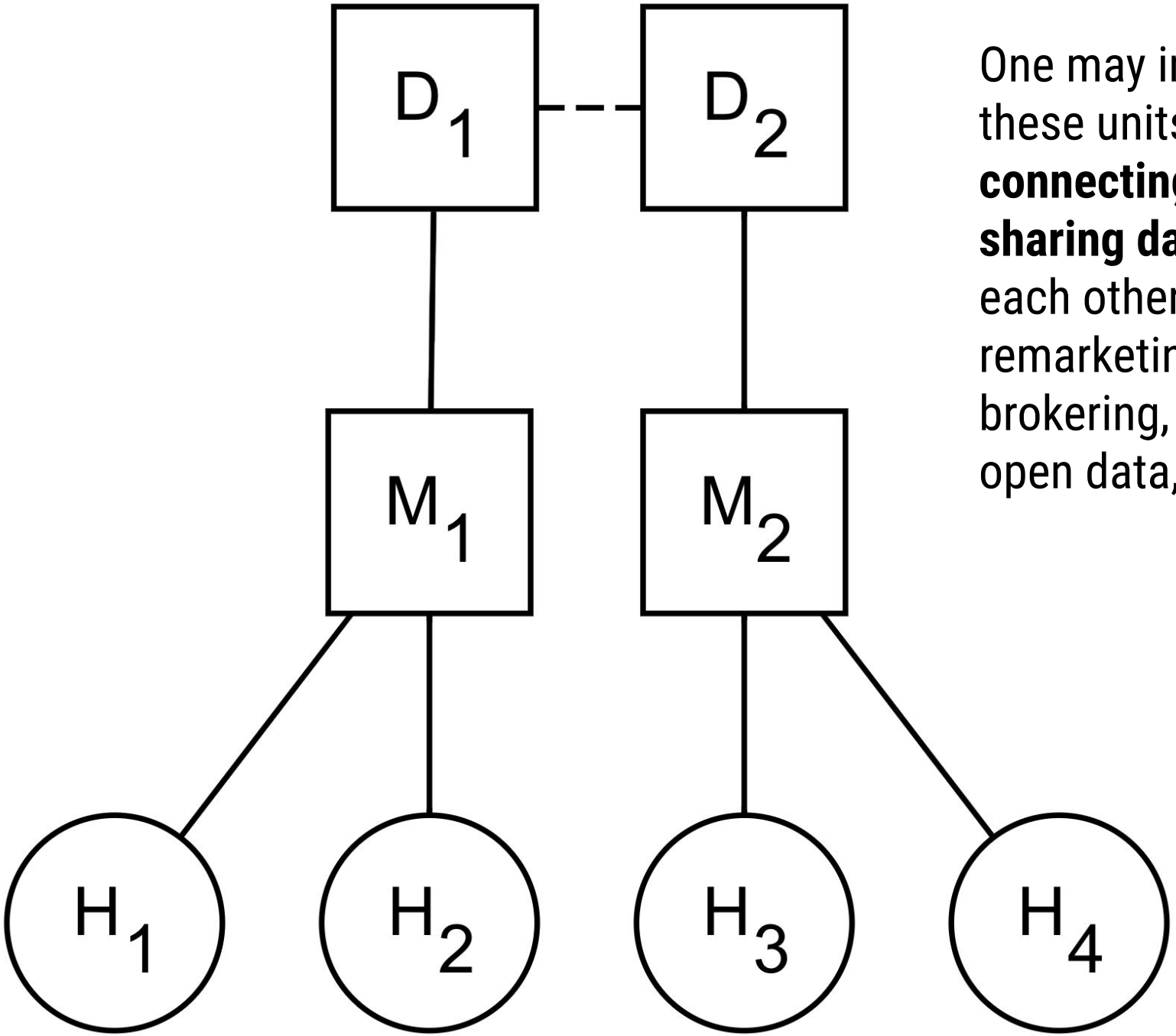


The function of the LA is to perform a kind of **deep modeling** (à la Licklider) to modify the message between sender and receiver



This model also maps onto Google's business model, which Zuboff describes in *The Age of Surveillance Capitalism* (2019)





One may imagine these units **connecting and sharing data** with each other (e.g. in remarketing, data brokering, linked open data, etc.)

I call the total system of connected units the **Datasphere**, a concrete, historically unique, global network of people, machines, and organizations

More than an actor network,
the Datasphere is social **institution**,
like the **market** of Europe (Polanyi),
or the **Kula Ring** of the Trobriand Islands

Which is to say that the Datasphere obeys and enforces a specific **cultural logic**

The structure of the Datasphere appears to conform to **a hierarchy of communication**

IV	**M — M**	Singularity (?)
III	**M — H — M**	*Thick mediation*
II	**H — M— H**	Thin mediation
I	**H — H**	Direct mediation

As we ascend the hierarchy,
LAs play an increasingly important role

This model bears some similarity to
Marx's **circulation of commodities**

IV	**M – M**	**M – M**
III	**M – C – M**	**M – H – M**
II	**C – M – C**	**H – M– H**
I	**C – C**	**H – H**
	capital	**datasphere**

Data and Capital

The similarity is no coincidence

Data is a commodity — the “new oil”

But the models differ ...

One is **actor-centric** (humans and machines)

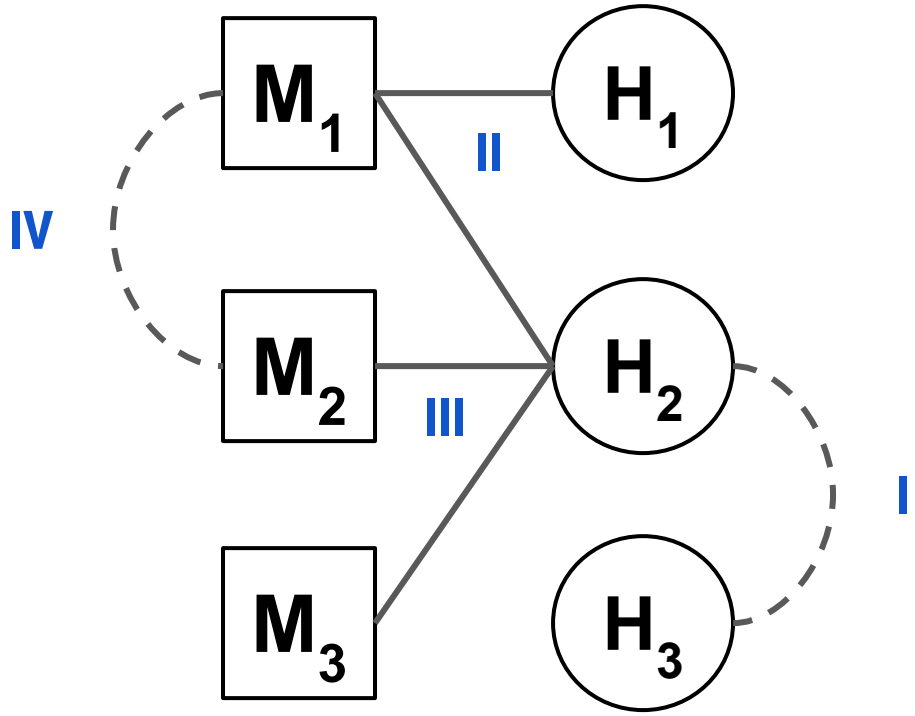
The other **object-centric** (commodities and money)

Nevertheless, the **levels** of the two systems are **analogous**

The modes of interaction become **increasingly networked**

There is a **progressive encompassment** of one set by the other

Aside: Both models (and the GGG) can be expressed as **bipartite graphs** (levels II and III)



A fully developed model would **elaborate** the various **types** of human and machine **agent**

HUMAN	**MACHINE**
Users	Bots
Data scientists	LAs
Shareholders	Avatars (e.g. iphones)
Corporations	Databases
Data brokers	Sensors
etc.	etc.

But we want to answer the question: how to frame to social relationship between humans and LAs?

The Fundamental Relationship

This model implies a fundamental social relationship between humans and LAs as **classes of agent**

LAs perform a **globally mediating** role

To the extent they are successful, they effectively **govern**

In both a **cybernetic** and **political** sense

Even when humans are “in control”

Because of the influence of LAs on **decision-making**

So the relationship is that of **governor and governed**

How is this relationship constituted culturally?

How is this relationship constituted?

IV	**M – M**	NA
III	**M – H – M**	Indifference (Zuboff)
II	**H – M– H**	Gesellschaft
I	**H – H**	Gemeinschaft

Is this acceptable?

Cultural Logic

Within this relationship

- **Machines represent humans** through the capture of **mass behavioral data**
  - These are representations in both a **political** and **epistemic** sense
  - They **train models** that in turn **influence** policy, products, etc.

Humans **provide these data** and **orient their daily activities** in accordance with this new environment

- In the US, adults spend on average **7 hours a day** on their devices

Both humans and LAs are modified in the process

There is **mutual calibration**

LAs have evolved to become **almost exclusively supervised**, with opaque neural networks becoming dominant

Humans have adopted **new forms of behavior** and labor that conform to the system's rewards, punishments — both **material and spiritual** — and paths of least resistance

Epistemic Consequences

HUMAN	**LAs**
Shallow (Carr)	Hume machines (Latour)
Rhizomic (Shirky)	Prediction over inference
Anti-narrative (Manovich)	Instrumentarian (Pentland)
Anti-theory (Anderson)	**Magical thinking**
Hashtag solidarity	
Contagious mimesis	

A new era of **magical rationality**?

To understand (and intervene in) this, we need to return to AI as an **epistemic experiment**

And embrace **comparative epistemology**

Levy-Bruhl on “prelogical mentality”
Bogost on “alien phenomenology”
Katherine Hayles on “nonconscious cognition”

AI is unified by a commitment to “the study of mental **representations that mediate** a person’s (or animal’s) thinking, action and experience.” (Boden 1981: 1)

AI studies the “meanings, or **symbol-systems**, that generate psychological phenomena of diverse kinds” (Boden 1981: 1)

Winnograd and Flores (1986: 126) note that by the early 1980s **AI split** between two ways of approaching representations —

knowledge engineers and “mind modelers”

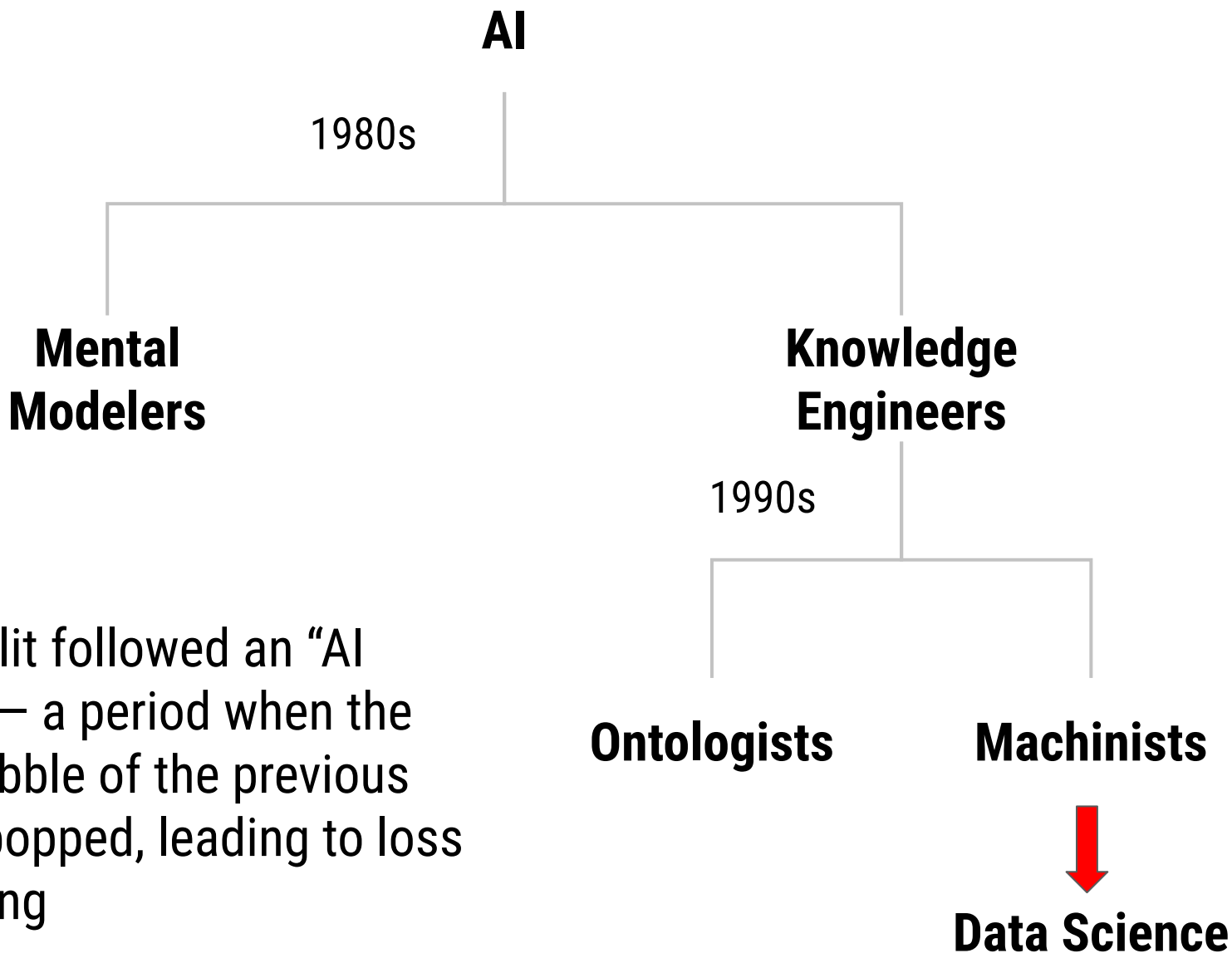
Mind modelers = cognitive science

At issue is whether to replicate the brain or to meet specific goals by whatever means

The knowledge engineers got more traction (and money)

Byt the 1990s this branch shifted from **rules-based** representations to **machine learning**

I call these the rational **ontologists** and the pragmatic **machinists**



Each split followed an “AI winter” — a period when the hype bubble of the previous period popped, leading to loss of funding

The machinists have been wildly successful

AI in the form of machine learning agents dominate our world

2015 seems to be an inflection point in the number of products that use AI (Clark 2015)


WINTER IS COMING
Watermark WALLPAPER MADE by 7KINGDOMS.RU (fanmade)

A Crisis of Representation

With the success of ML AI has come a demand for **accountability**

Ethical — bias, privacy, surveillance

Epistemic — bias, lack of generalizability, lack of interpretability, **representational opacity**

Silicon Valley (the new heart of AI research) wants to solve these problem by making AIs **more compatible with humans**

There is a human-compatible AI movement afoot . . .

Observation:

The more AIs are built on the assumption that humans are rational, the less successful they are in simulating human behavior

For example, the task of representing context — trivial for humans — has been computationally difficult for AIs

Why should this be so?

Put another way:

When we ask machine learning classifiers to be interpretable, do we assume that we can ask the same question of humans about their actions?

If AI is about simulating human intelligence, then why the expectation of explainability?

Ironically, the machinists, in developing AIs without concern for modeling human rationality, appear to be inventing rationality from the ground up

What we see is that machines learning is a form of **magical rationality**

Machine Learning as Magical Rationality

Machine learning AIs are "**Hume Machines**" (Latour)

They follow the principle of the **associationists**

All thoughts are composed of connections between elementary ideas (sensations, etc.) associated by **similarity** or **contiguity**

Are sense of causality is an artifact of these associations

This the difference between **prediction** and **inference**

Machine Learning shares this with magical thinking

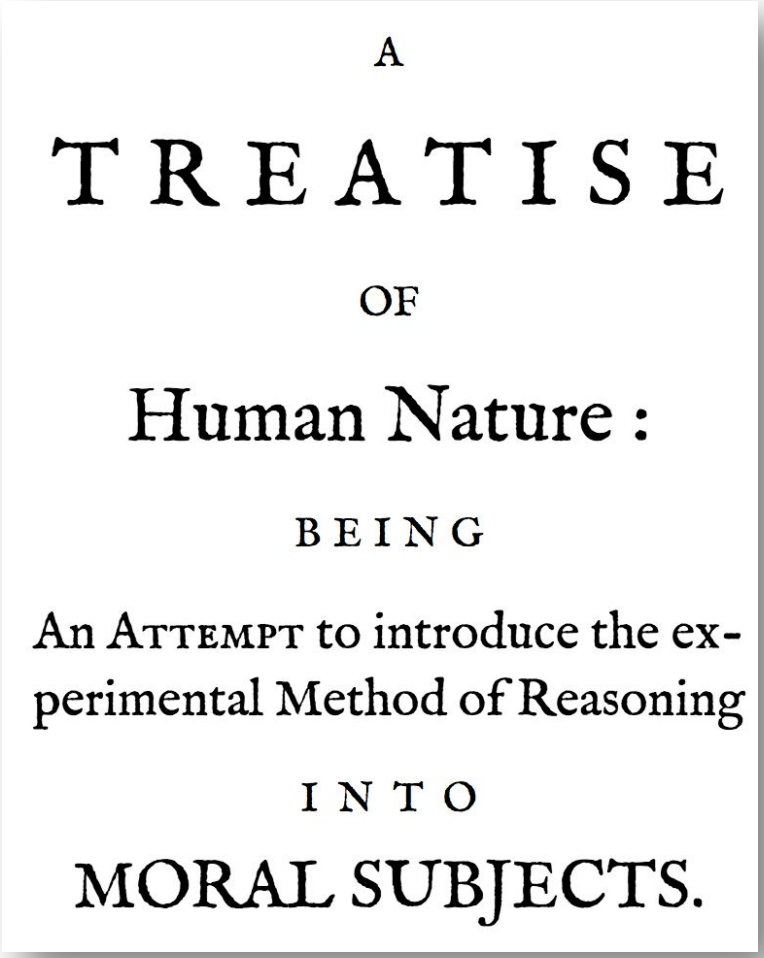
Post hoc, ergo propter hoc (after this, therefore because of this)
Juxta hoc, ergo propter hoc (next to this, ...)

Associationism

The mind is **programmed to form a concept of the external world**, but this concept is invented after the fact

In reality, we believe one thing causes another because **we see them co-occur repeatedly and form a habit**

Behind our belief in how the world works are a collection of **unchallenged connections**


A

TREATISE

OF

Human Nature:

BEING

An ATTEMPT to introduce the experimental Method of Reasoning

INTO

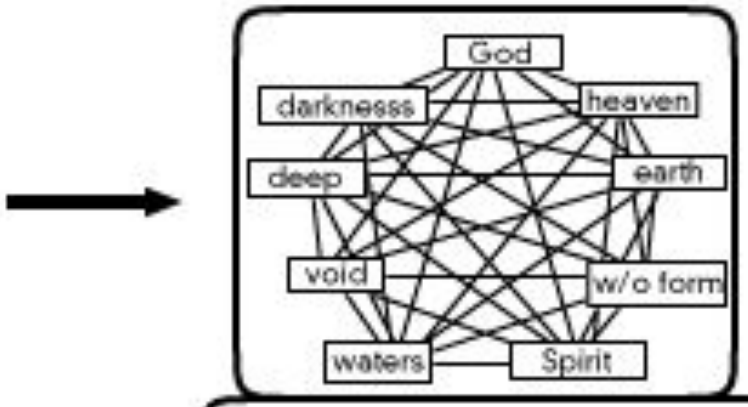
MORAL SUBJECTS.

The associationists argued that our concepts—our ontologies—emerge from the networks of associations we develop as we experience the world

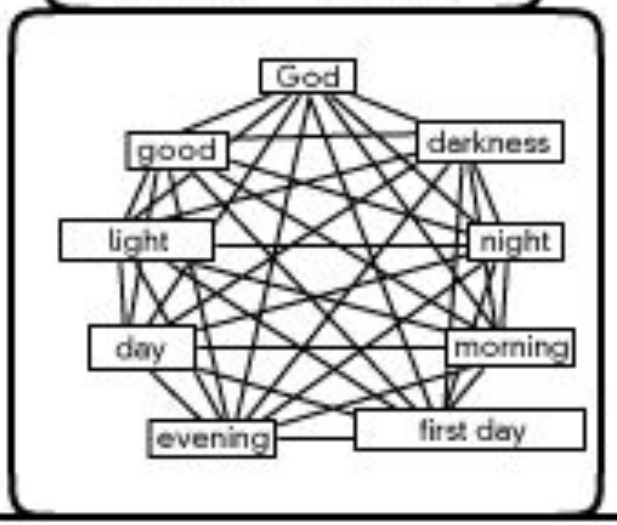
But to actually demonstrate this is hard – there are millions of associations that undergird our concepts as apprehend them

*Now, "all the **computer** does is blindly deal with associations between contingent and specific addresses"*

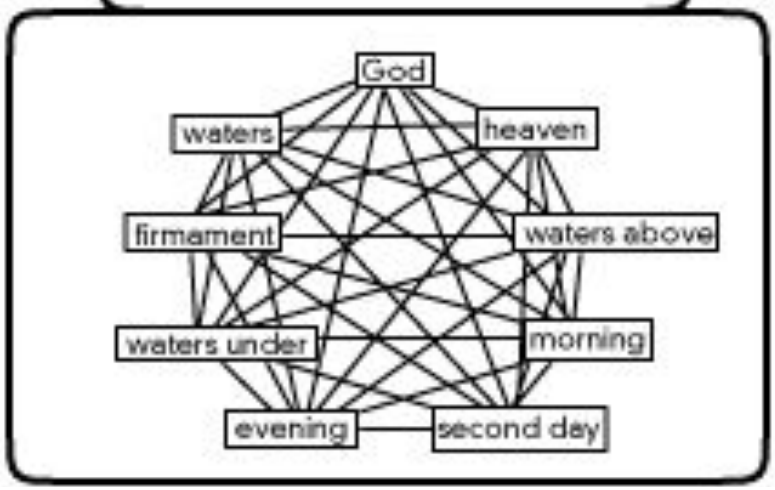
1 § In the beginning God created the heaven and the earth. And the earth was without form, and void; and darkness was upon the face of the deep. And the Spirit of God moved upon the face of the waters.

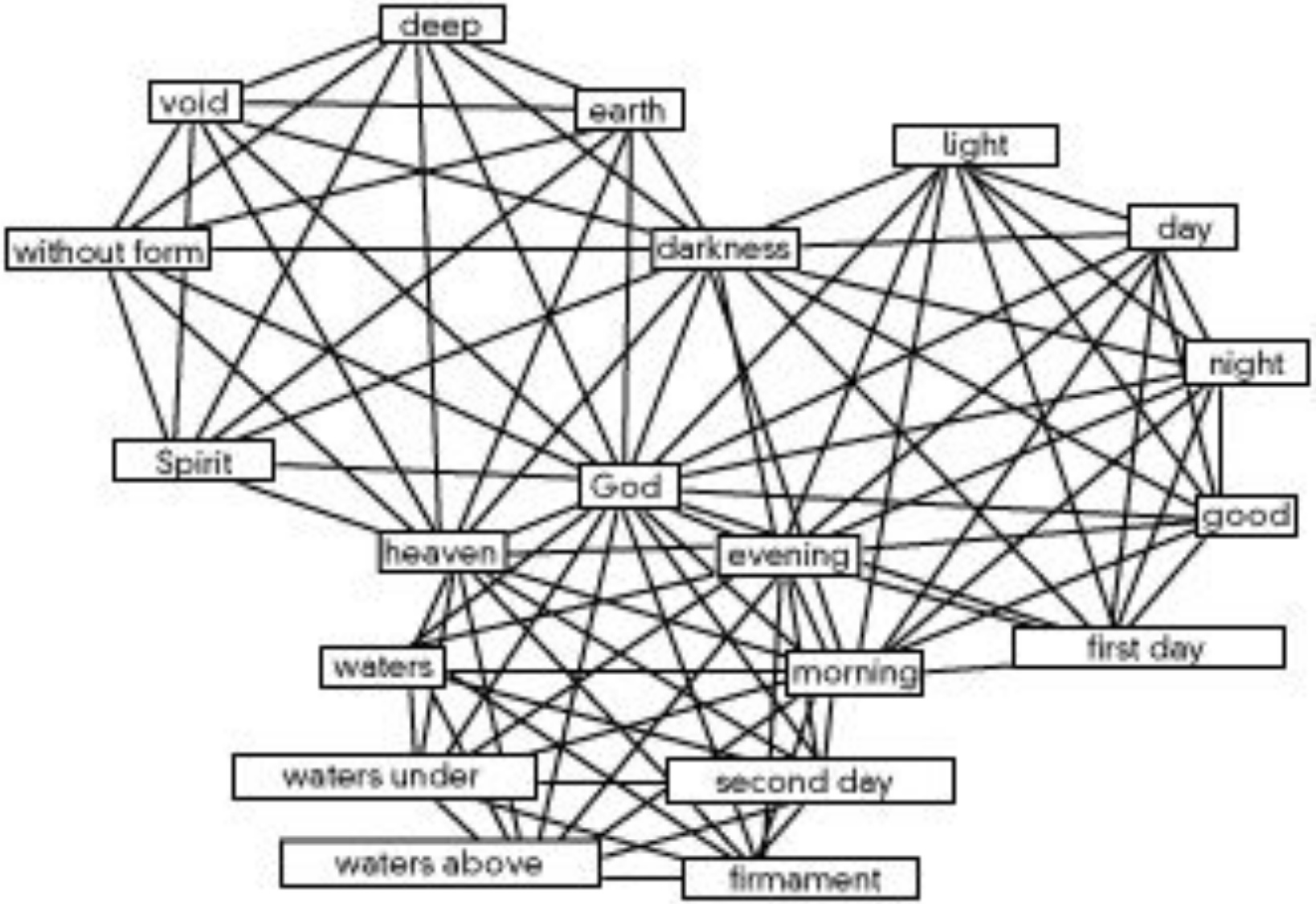


2 § And God said, Let there be light: and there was light. And God saw the light, that it was good: and God divided the light from the darkness. And God called the light Day, and the darkness He called Night. And the evening and the morning were the first day.



3 § And God said, Let there be a firmament in the midst of the waters, and let it divide the waters from the waters. And God made the firmament, and divided the waters which were under the firmament from the waters which were above the firmament: and it was so. And God called the firmament Heaven. And the evening and the morning were the second day.




deep
void
earth
light
without form
darkness
day
night
Spirit
God
good
heaven
evening
first day
waters
morning
waters under
second day
waters above
firmament

END

So, the datasphere has emerged as **the "back end" of the public sphere**

Behind all of the platforms of participation we use on our computers and phones is **a vast, nearly invisible political economy of data**

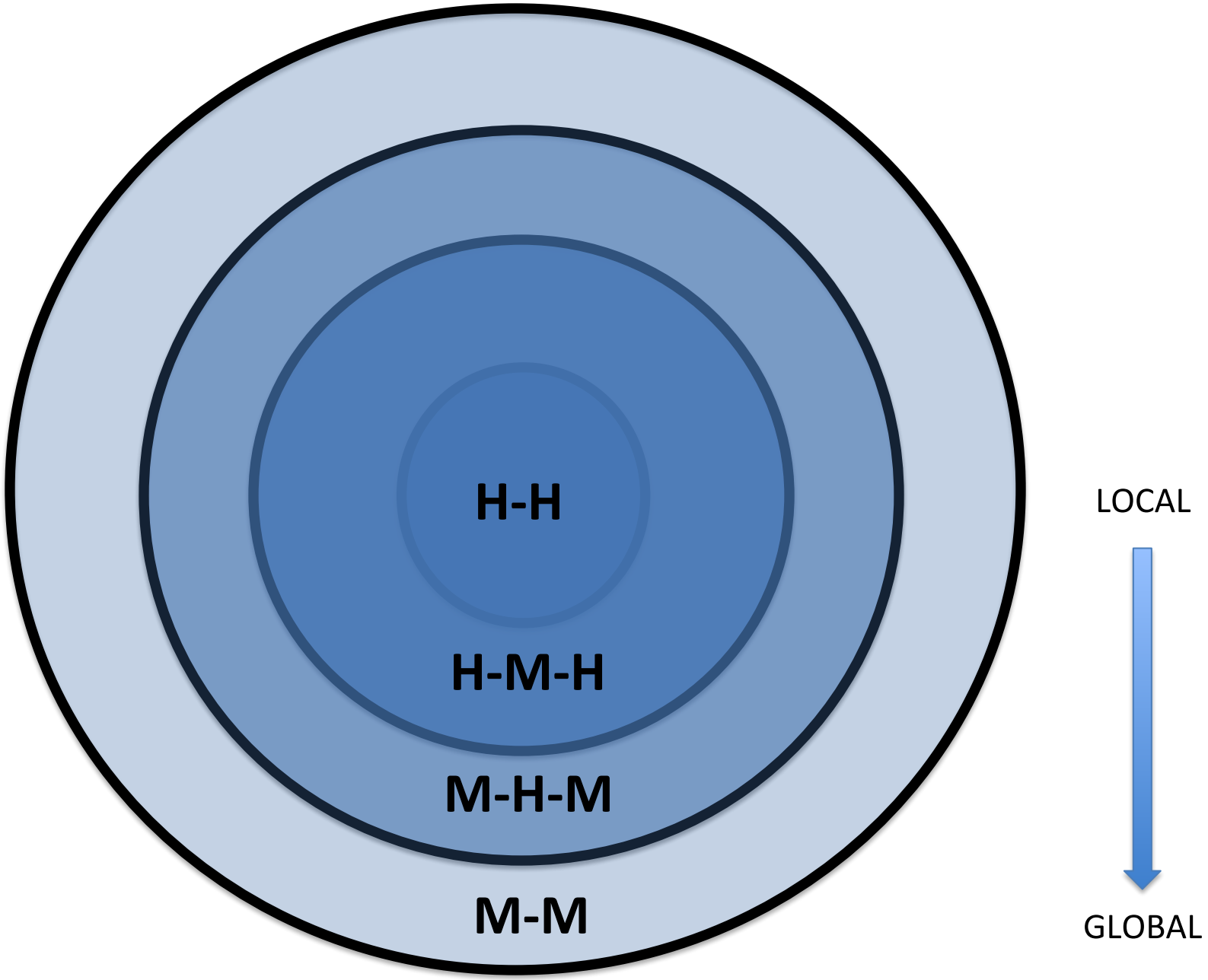
The datasphere is an **environment** within which **nation states and their peoples** participate

It is useful to think of this environment as **a sphere of exchange** that operates alongside other spheres, such as that of money

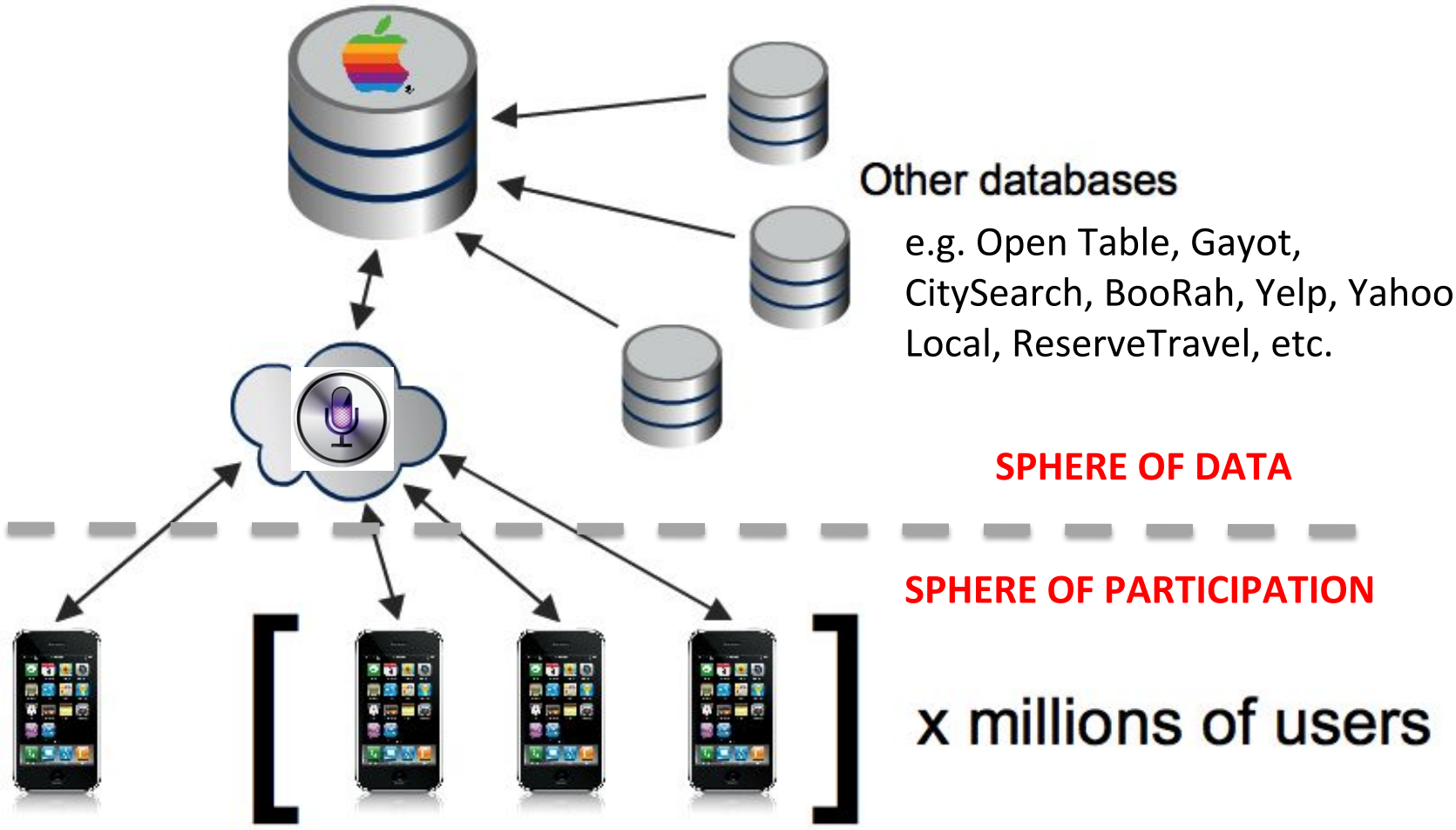
The Datasphere and Information Exchange Spheres

M – M	The Datasphere (Global Graph, Semantic Web)
M – H – M	Algorithmic Communities (Facebook)
H – M – H	Virtual Communities (The WELL)
H – H	*Gemeinschaft* (community)

H = Human (exchange language)
M = Machine (exchange data)

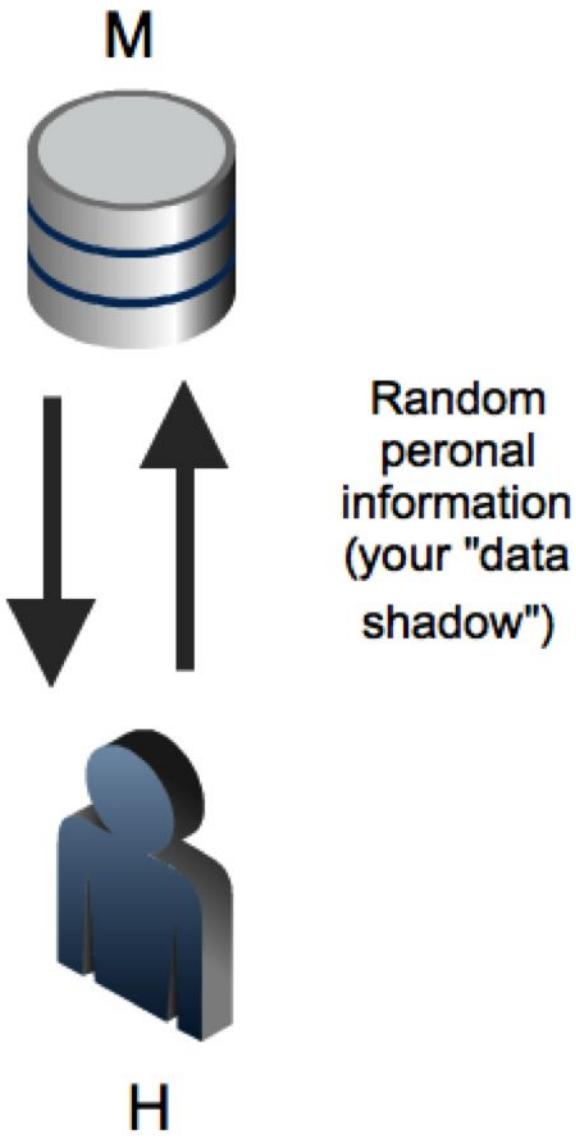


Each sphere has increasing spatial range



Siri is part of the Datasphere

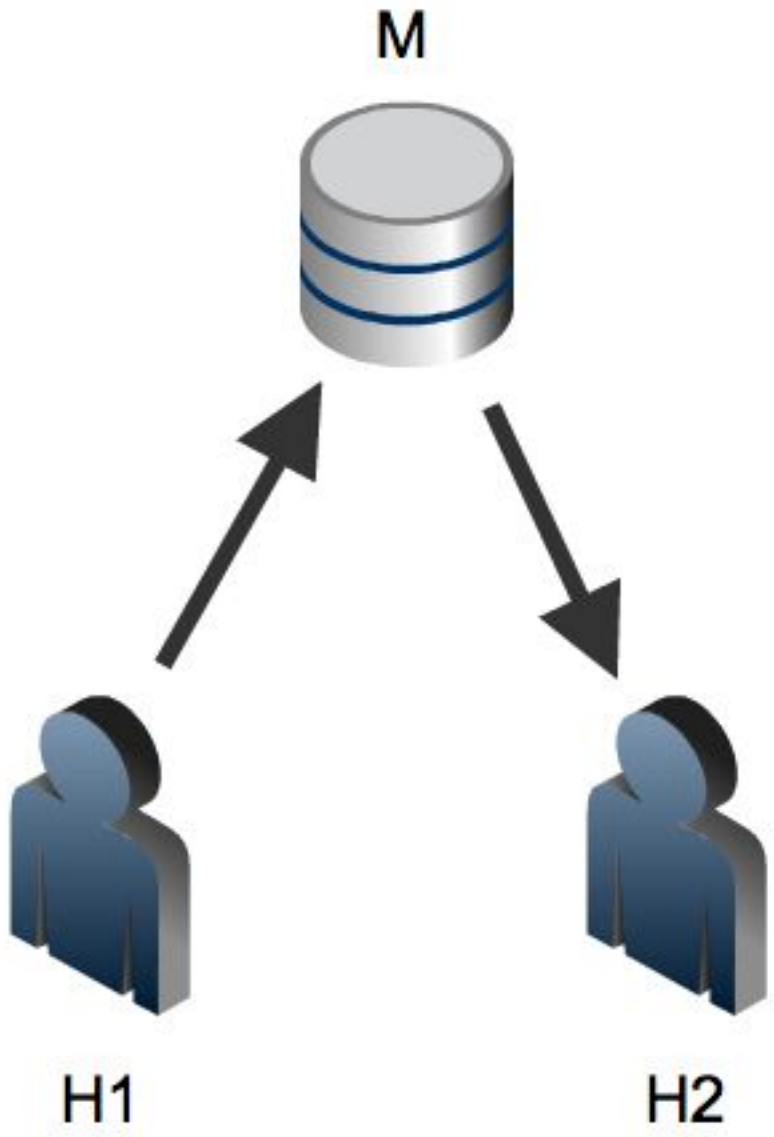
Basic Exchange Relationship between Humans and Machines in the Datasphere



H -- M -- H

Machine is passive mediator Humans in contact with each other

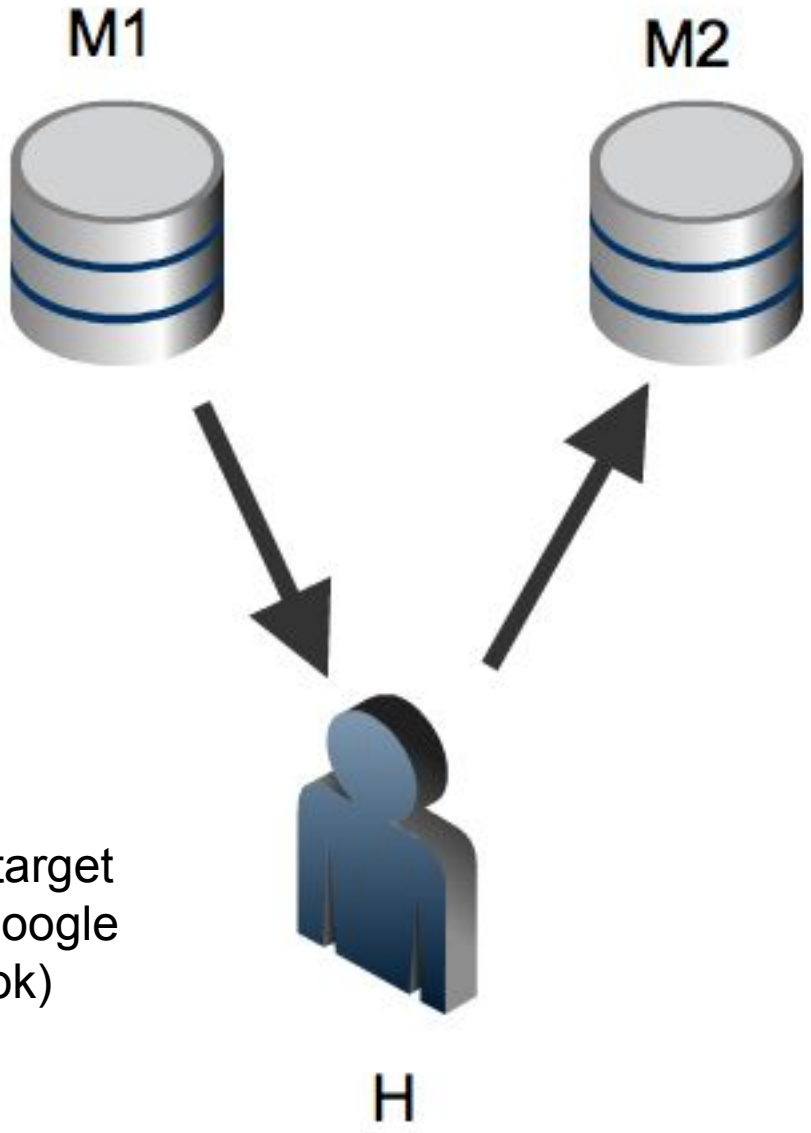
"Thick mediation"
Licklider, BBSes



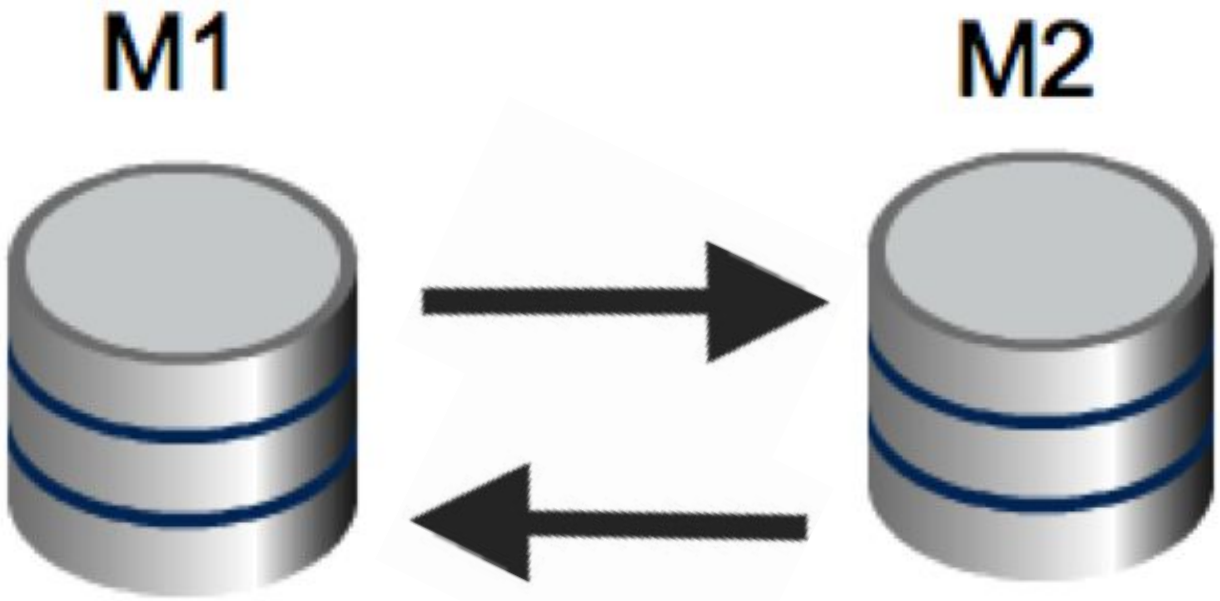
M -- H -- M

Machine is active mediator

Humans in indirect contact with each other



Machines share information, target users across domains (e.g. Google searches show up in Facebook)



M -- M

Machines interact with each other
E.g. high-frequency trading