



# The Baby Singularity

*Or: Why right now may be the best possible time to be a hacker*

**Gary William Flake, Ph.D.**

Principal Scientist, Yahoo!  
Head, Yahoo! Research Labs



## Back Story

---

- ***Singularity***: the idea that the exponential speed and growth of technology will make the future completely unknowable
  - Introduced by Vernor Vinge, in 1993
  - Primary causes cited as AI and nanotechnology resulting in super-human intelligence
  - Corollary: human society will end as we know it
  - Talked about in both optimistic and pessimistic terms



## Today's Topic

---

- ***Baby Singularity***: the idea that a deeper and tighter coupling between the online and offline worlds will profoundly change how we do science, mathematics, and engineering
  - Introduced by me, today
  - Primary cause is ubiquitous computing
  - Corollary: the increased pace of discovery will fuel amazing innovations in the near future
  - Very optimistic statement



# Outline for Today

---

- Philosophy of science and the role of computer science
- Technology trends, near and long term
- The Internet and its role in the “baby singularity”
- R&D at Yahoo! Research Labs



# Philosophy of Science

... and the role of computer science



## An Old Joke

---

- Engineers study interesting real-world problems, but fudge their results.
- Mathematicians get exact results, but study only toy problems.
- But computer scientists, being neither engineers nor mathematicians, study toy problems and fudge their results.



## How to ruin a joke: Analyze it!

---

- The joke works with substitutions:
  - Engineer: Empiricist / Experimentalist
  - Mathematician: Theorist
  - Computer Science: *Simulationist*



# Flavors of Science

---

- **Experimentation**

- Generate and look at messy data, then try to match it to an existing theory
- Deductive: uses hypothesis to make predictions and compare to observations

- **Theorization**

- Crunch observations into an analytic form so as to better explain the past and predict the future
- Inductive: uses observations to make hypothesis

- **Simulation**

- Experimentation in a universe of theories
- Transductive: combines hypothesis, prediction, and observation into one framework



# Computer Science is Special

---

- Computer science is concerned with the theory and application of modeling real and artificial phenomena — **including models themselves.**
- No other discipline can make this claim.
- Most everything that we know about the universe suggests that Turing-completeness = reality.
- The future, depends on the future of science, which depends on the future of computer science.



# Technology Trends

... near and long term



# Trends (with ~5 year outlook)

---

1. Hardware
2. Networking
3. Input/Output
4. Software
5. Social
6. Economic
7. Data



# Hardware – Trends 1/7

---

- Moore's Law: number of transistors per square inch doubles every 18 months
  - 5 years yields a factor of 10 in growth
  - Directly impacts RAM sizes
  - Impacts sophistication of CPU
  - Indirectly impacts clock speeds
- Super Moore's Law: hard disk capacities doubles every 12 months
  - 5 years yields a factor of 32 in capacity
  - 10 years yields a factor of 1024 in capacity



## Networking – Trends 2/7

---

- Good news: total bandwidth follows super-Moore's law – doubling every 12 months
- Bad news: networks have bottlenecks – bandwidth from *A* to *B* is only as fast as the slowest link
- Many more options for the last hop (in ~5 years):
  - Consumer broadband: cable, DSL, etc. (10 Mb/s soon)
  - WiFi: expect more local regions to adopt (100 Mb/s shared)
  - Other wireless possibilities proposed but have yet to be successful (5-10 Mb/s for handhelds)
- Internal home networks wired at 1 Gb/s as a matter of course



## Input / Output – Trends 3/7

---

- Voice recognition is pretty good today, but it eats up a lot of CPU cycle.
  - 5 years: good enough for everyday use
- LCD
  - Costs drop to be cheaper than CRT
  - Density increases making 800x600 common in handhelds
  - Size increases, yielding both profound and silly applications
- Smart paper
  - As crisp, clear, thin, and with equal resolution



# Software Industry – Trends 4/7

---

- Infrastructure software continues to be commoditized
  - OS for servers (both large and tiny)
  - Databases
  - Web servers
  - Many development platforms
- Desktop will probably **not** be commoditized
  - But the traditional Desktop may not exist anymore
- Expect more new companies to develop “Open Source”-like practices internally
  - Open API
  - Well-defined services and infrastructure
  - Open contribution internally but with some access controls
  - Impacts business models



## Social – Trends 5/7

---

- “Grass roots” is perhaps the single most important social theme:
  - Ebay / Paypal
  - Open source
  - Digital photography
  - Napster and other P2P applications
  - Meetup / flash mobs
  - Friendster
  - Blogs
- Expect this trend to continue
- Each area represent a Pandora’s Box:
  - At best, each empowers individuals in unprecedented ways
  - But they come with a price: fraud, IP issues, theft, spam, poor quality, lack of authenticity or authority.



## Economic – Trends 6/7

---

- Monetization of small things (but on massive scales) is the single most important economic theme:
  - Ebay / Paypal
  - Overture
  - Apple iTunes Store
- Ubiquity and Accessibility
  - Lowered access costs
  - Increased penetration globally
  - Crossing of the “digital divide”



## Data – Trends 7/7

---

- More digital data will be generated in the next five years, than the total data ever generated to date
- Massive quantities of data will be generated on small scales
  - RFIDs
  - Smart appliances
  - Home healthcare
- Massive quantities will also be generated online as:
  - More people engage with the Internet
  - More of an individual's life will be coupled to the online world



# The Internet

... and its role in the baby singularity



# Internet in the Near Future

---

- We need to first consider how people will connect (what computers will look like)
- Afterwards, we can consider what the Internet will look like as a result
- Later still, we can guess at how Internet as a computing and social platform will change



# Cell Phone in 5 Years

---

- As powerful as today's desktops on all dimensions (CPU speed, RAM, disk storage, bandwidth, etc.)
- Will combine all current handhelds
  - Phone, PDA, games, camera, iPod, video player, thin web and communication client
  - Modest version available today
- Will be able to trivially “morph” into a desktop
  - Dumb terminal with keyboard, mouse, and monitor
  - Bluetooth or other short distance network protocol
  - Drop your phone next to the dumb terminal
  - Instant desktop with all of your stuff
- May kill traditional desktop because of sales and power
- **Most consumers will literally have an internet connection at their fingertips at all times**



# Home Servers in 5 years

---

- Living room convergence: DVD, television tuning, DVR, gaming, music and video serving
  - Über appliance will integrate the entire home
- At least 10 times faster, at least 10 times as much RAM, with at least 1 TB of disk space
  - > 200 DVDs, >200 thousand MP3s, >200 thousand photos, or 10 million x 100 KB objects
  - As much computing infrastructure as what was behind most search engines 5 years ago
- Deeper integration with the home
  - RFID + refrigerator: smart shopping list
  - Smart toilet + other medical integration: home healthcare
  - Automation and security
- **Data mining and information retrieval for individuals becomes a necessity**



# Other Access Considerations

---

- Distinction between physical network types dissolves
  - Data, television, voice all collapse into one
- Expect more communication convergence
  - Email, IM, voice, video, and other modalities will be unified on one communication platform
- Storage location will slowly become irrelevant
  - You will not really care if your data is on your handheld, your home network, or a common network
- Despite past hype new interfaces may still happen
  - Smart book made out of smart paper
  - Video wallpaper for virtual conferencing



# Changes in Internet Commerce

---

- **A company's data will become one of its most important assets**
- Companies will be architected to optimize the properties of their data
  - In the physical world, physical constraints often dramatically limit how commerce can be improved (e.g. airline industry)
  - In the online world, logical constraints exist, but physical constraints are far less relevant (e.g. capturing intent in Web search)
- **How data is generated (and the network effects behind it) will be the core of Internet commerce**



# What is Important

---

- It's the data. It's all about the data.
- Paradoxically, some problems become more tractable as data volume increases:
  - To understand language, we search for properties in online corpora
  - To understand preference, we instrument and model individual choice
  - To understand society, we look for patterns in its digital mirror, the Web



# The Internet Deconstructed

---

- Entities (people, businesses, and soon agents)
- Content (text, multimedia, and even programs)
- Relationships between things
  - Entities connecting to content (search)
  - Entities connecting to entities (communication)
- Metadata in the aggregate
  - *How* entities connect to content (personalization)
  - *How* entities connect to one another (social patterns)
  - *How* content relates to other content (ontology)



# The Internet as a Mirror

---

- The Internet's content, composition, and participants will eventually reflect a nearly complete and accurate version of the physical world.
- The future of the search engine is to model the human mind in the aggregate.
- The future of the personal agent is to model us as individuals, acting on our behalf and proactively harvesting information.

A large, faded purple watermark of the Yahoo! logo is centered in the background. It consists of a purple circle containing a white 'Y' and a purple exclamation point to its right.

# R&D at Yahoo! Research Labs

... what we think is important, and why



# R&D Thematic Priorities

---

- **Machine Learning**
  - How to do things smartly
  - Classification, regression, modeling, and forecasting
- **Efficient Algorithms**
  - How to do thing faster
  - Clustering, dimensionality reduction, sparsification, approximation
- **Information Retrieval / Data Mining**
  - How to find and use systematic patterns
  - Extraction, matching, relations, sensitivity analysis
- **Collective / Distributed Intelligence**
  - How to use many simple units for complex tasks
  - Auction and market theory, collaborative filtering, ensemble behavior



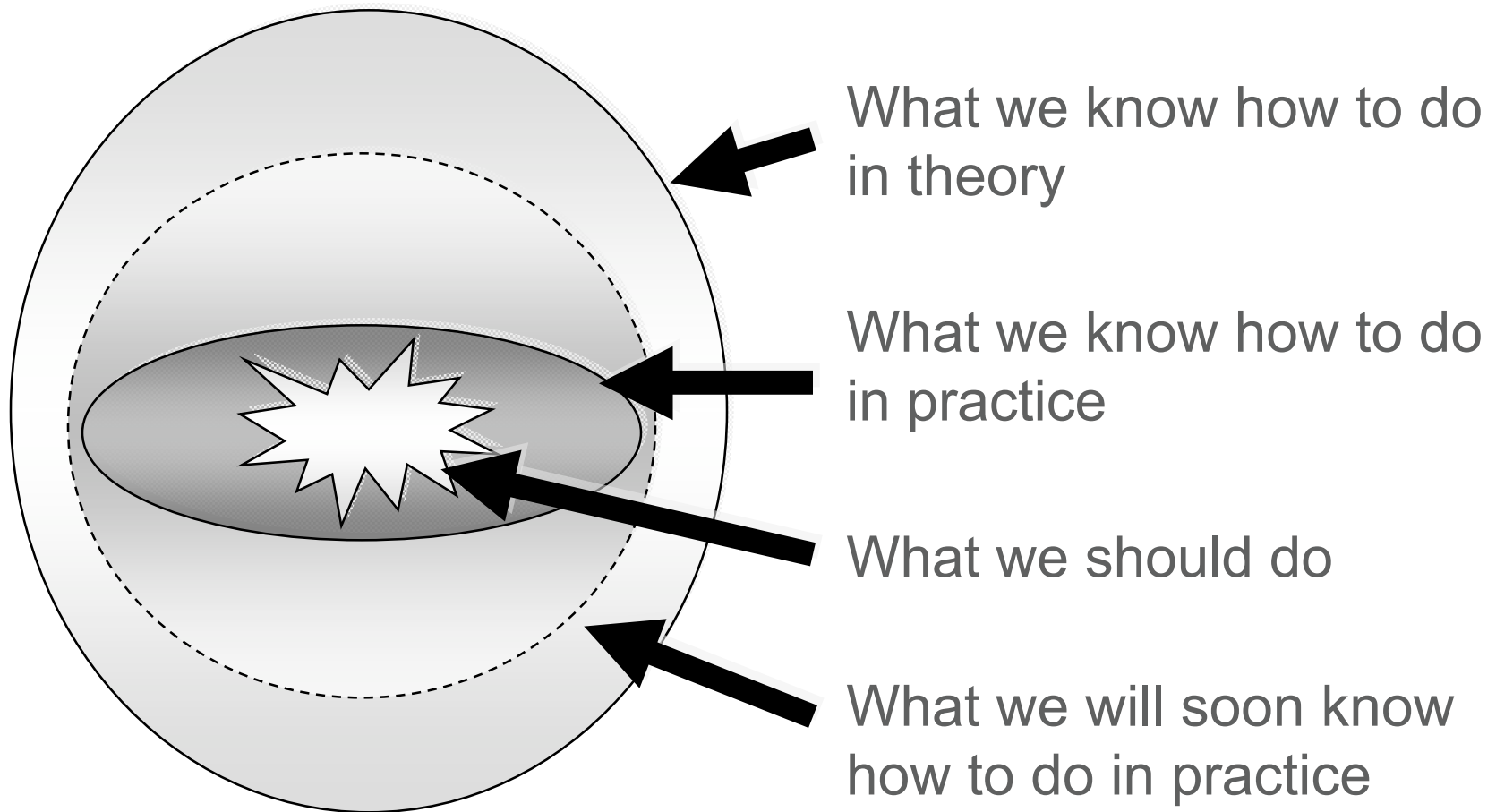
## State of the Art

---

- Almost everything we know about scale has been learnt in the past decade
- The Internet is the mother of all scaling problems
- Hardware is now catching up to software
- We are on the cusp of changing the game completely



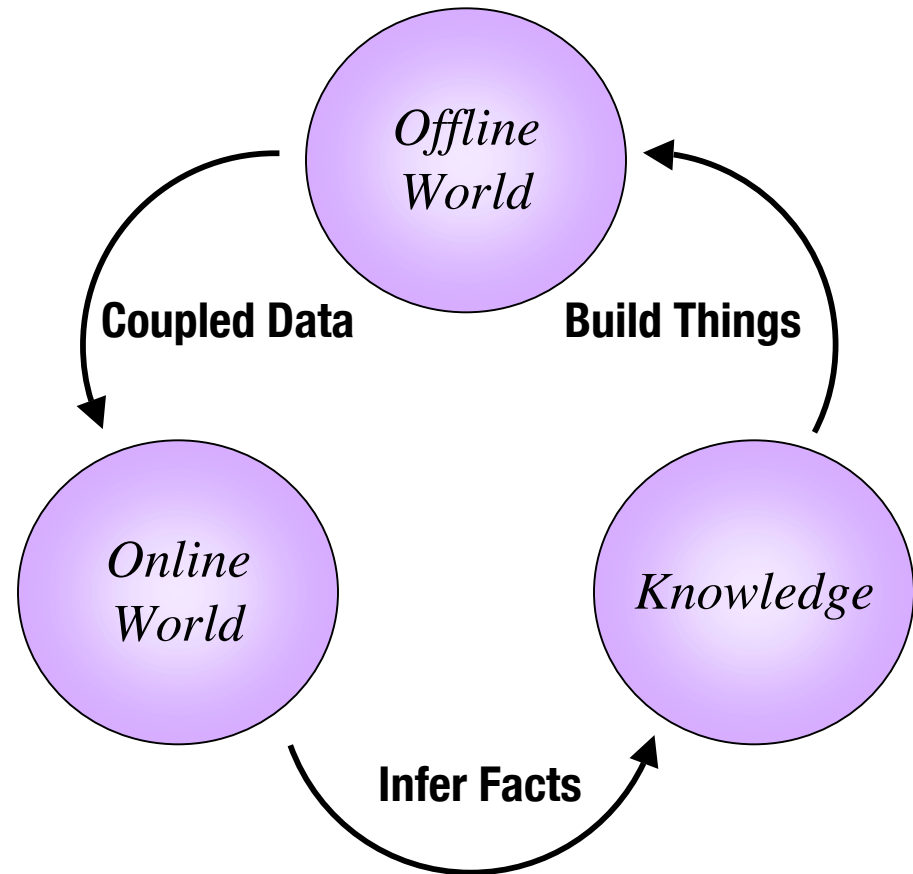
# How to View the Current World





# Speed up of Inflection Points

- The Internet Instruments the world
- The world can be analyzed in completely new ways
- New things can be built from what we learn
- The world is changed by what we add
- Science, math, and engineering will be further intertwined





# Final Motivation

---

*Prediction is difficult, especially of the future.*

—**Mark Twain** (also attributed to **Niels Bohr**)

*All stable processes we shall predict. All unstable processes we shall control.*

—**John von Neumann**



# Thank You!

---

- Questions?
- See <http://research.yahoo.com/>



**LIFE ENGINE™**