Carnegie Mellon University

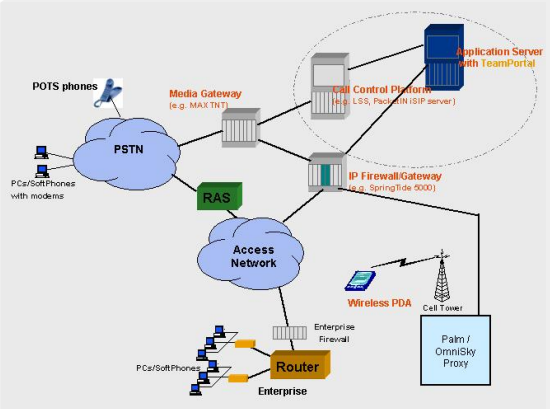
Building a Theory of Coordination: Why and How

Jim Herbsleb

Agenda

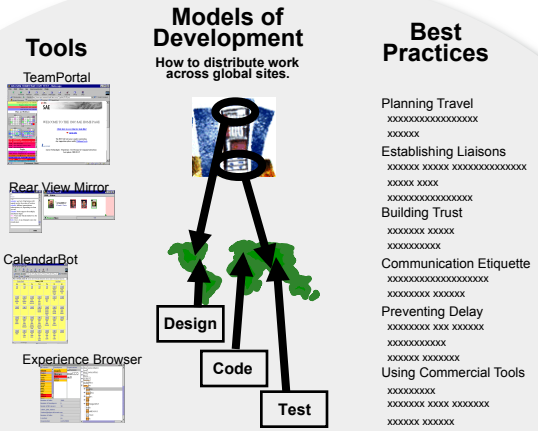
- Bell Labs Collaboratory
  - Management reactions
- The need for science
- What kind of science do we need?
  - Human science of software engineering
- The way forward
  - Barriers
  - Next steps

Bell Labs Collaboratory

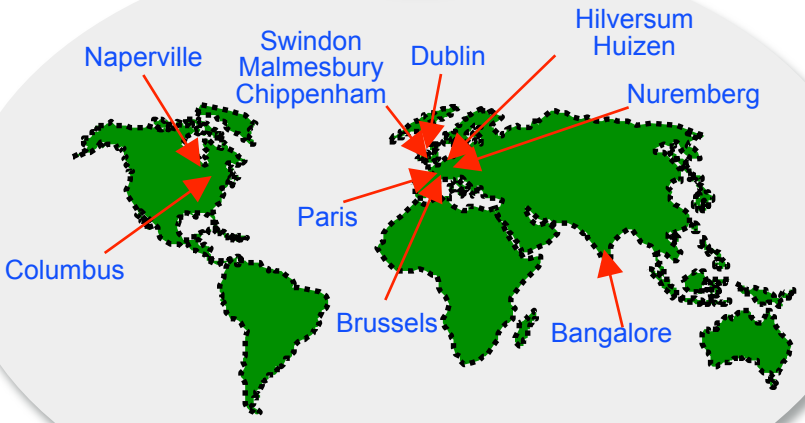


New Products

Research Team



Global Development Solutions



Empirical Studies

Executive VP:

Don’t study the problem! Just solve it!

Problems Uncovered

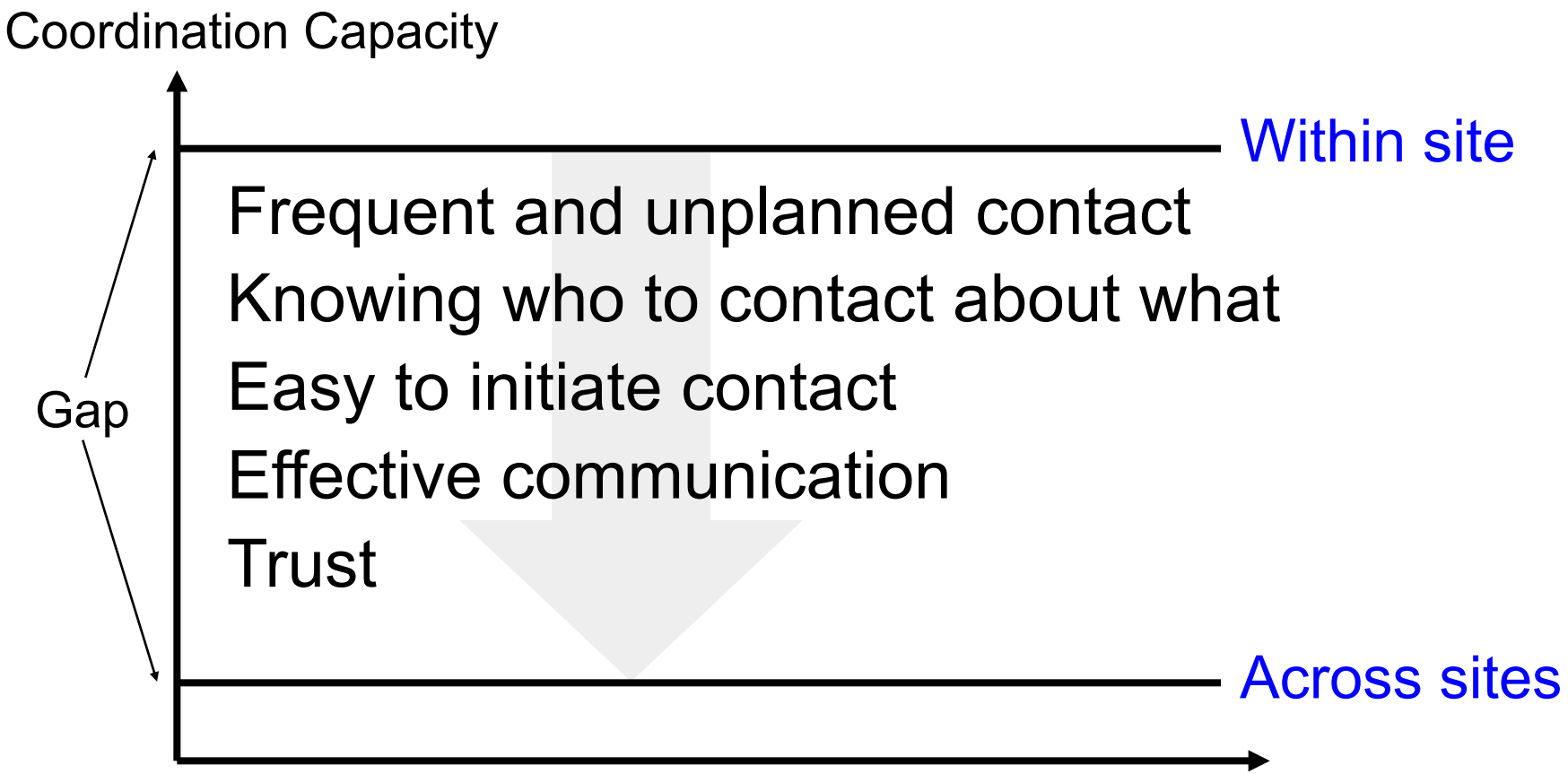
- Issue resolution paralysis, delay
- Very difficult to stay "in the loop"
  - constantly surprised
- Misalignment
  - undiscovered, conflicting assumptions
- Nonexistent or impaired social networks
  - loss of critical problem-solving mechanism
- Ineffective collaborative sessions
  - "What was decided?"

Problems Uncovered

- Isso[illegible] solutio[illegible]ral[illegible]dela[illegible]
- Very d[illegible]
- Mis[illegible]
  - un[illegible]
- [illegible]nexiste[illegible]tworks
  - loss of [illegible]n-[illegible]mech[illegible]m
- Ineffective coll[illegible]rative [illegible]essions
  - "What was deci[illegible]d?"

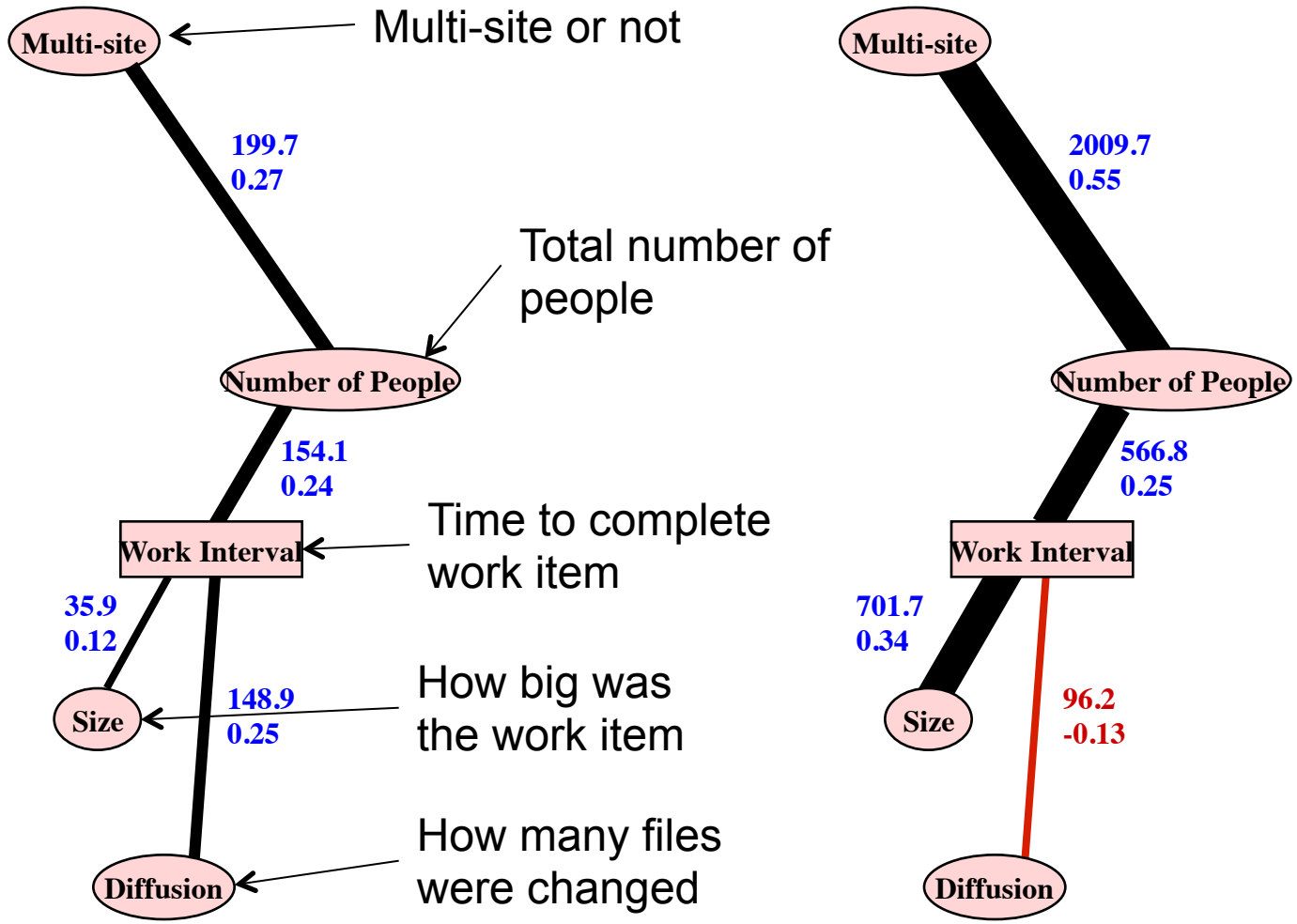
Result: Delay
Work split across sites
took *much* longer

Communication and Coordination



Herbsleb, J. D., & Grinter, R. E. (1999). Splitting the organization and integrating the code: Conway's Law revisited. In *Proceedings, International Conference on Software Engineering*, Los Angeles, CA, May 16-22, pp. 85-95.

Probing Extent and Causes of Delay



Graphical model of work interval for Product A

Replication: Product B

Herbsleb, J.D. & Mockus, A. (2003). An empirical study of speed and communication in globally-distributed software development. IEEE Transactions on Software Engineering, 29, 3, 1-14.

Research VP:

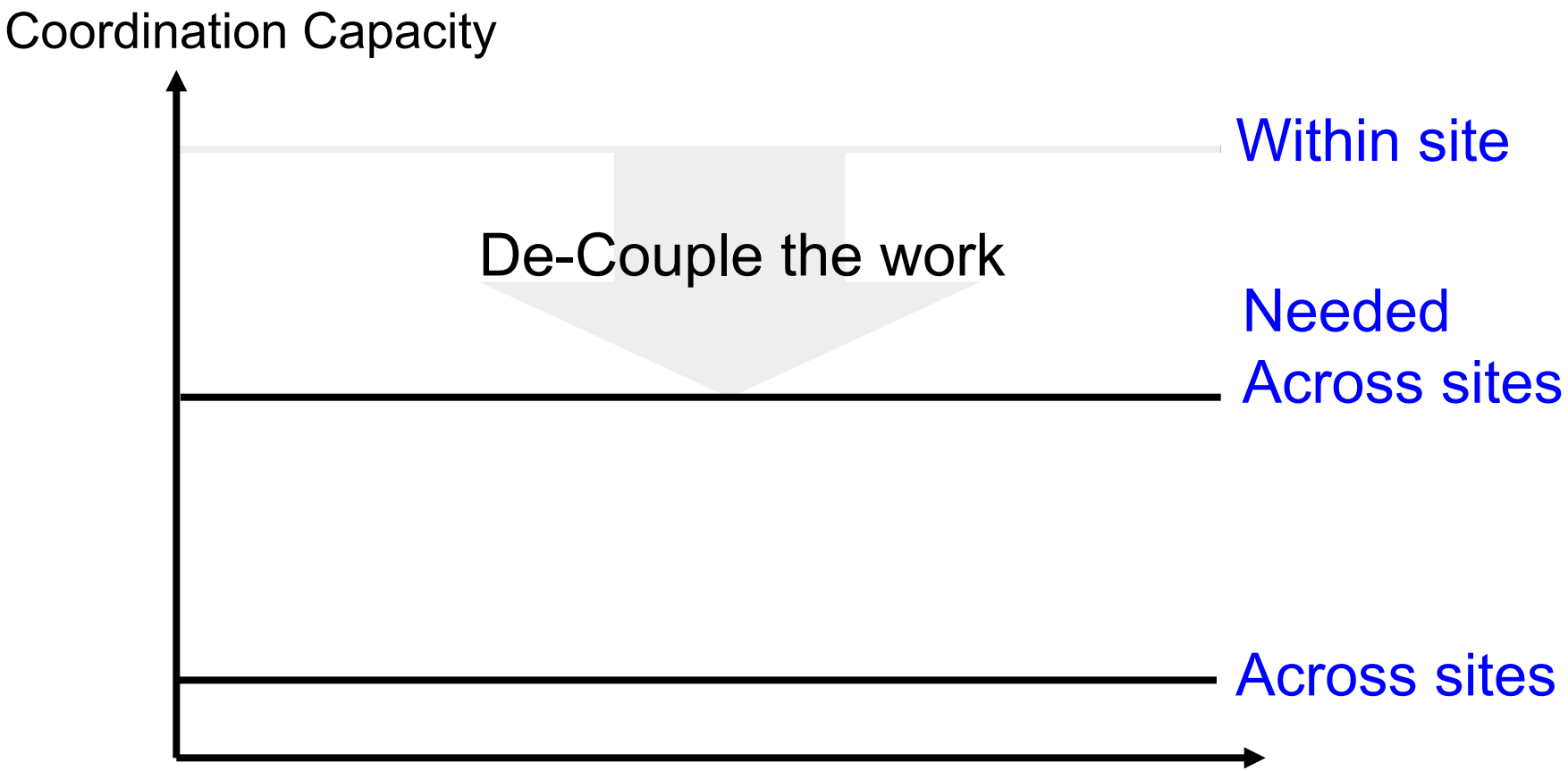
This is a waste of time! Don't work on this – this project counts for nothing on your performance review.

MR Interval
Distance Requires More People?

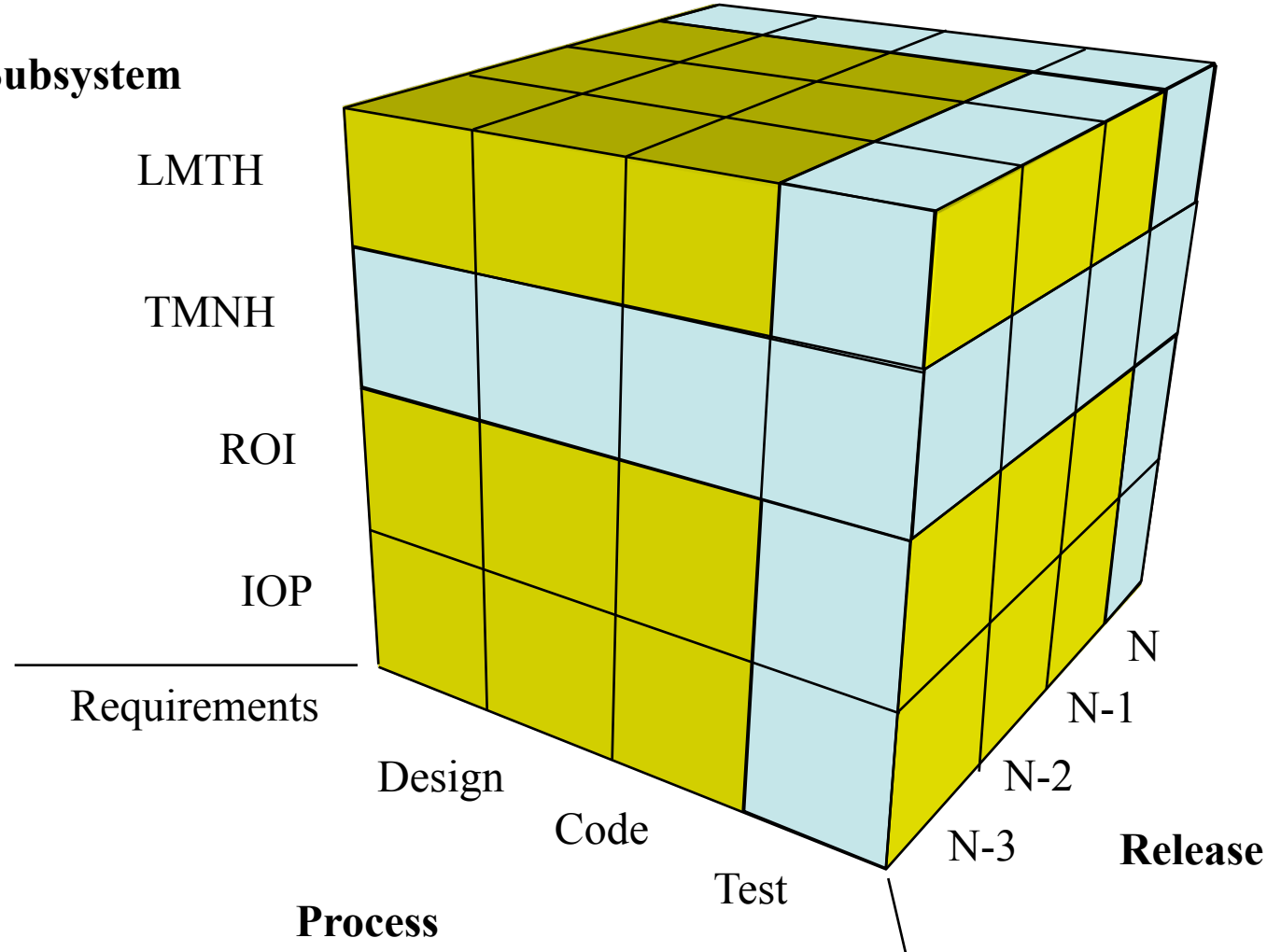
- MR is assigned to "owner" who recruits others
- Finding the right expert
  - Search time
  - If mistaken, reassignment and delay
- Trust and familiarity: Can MR owner get "right person" to do the work?
  - Slow to respond
  - Refuses or gives very low priority

10

Bridging the Gap

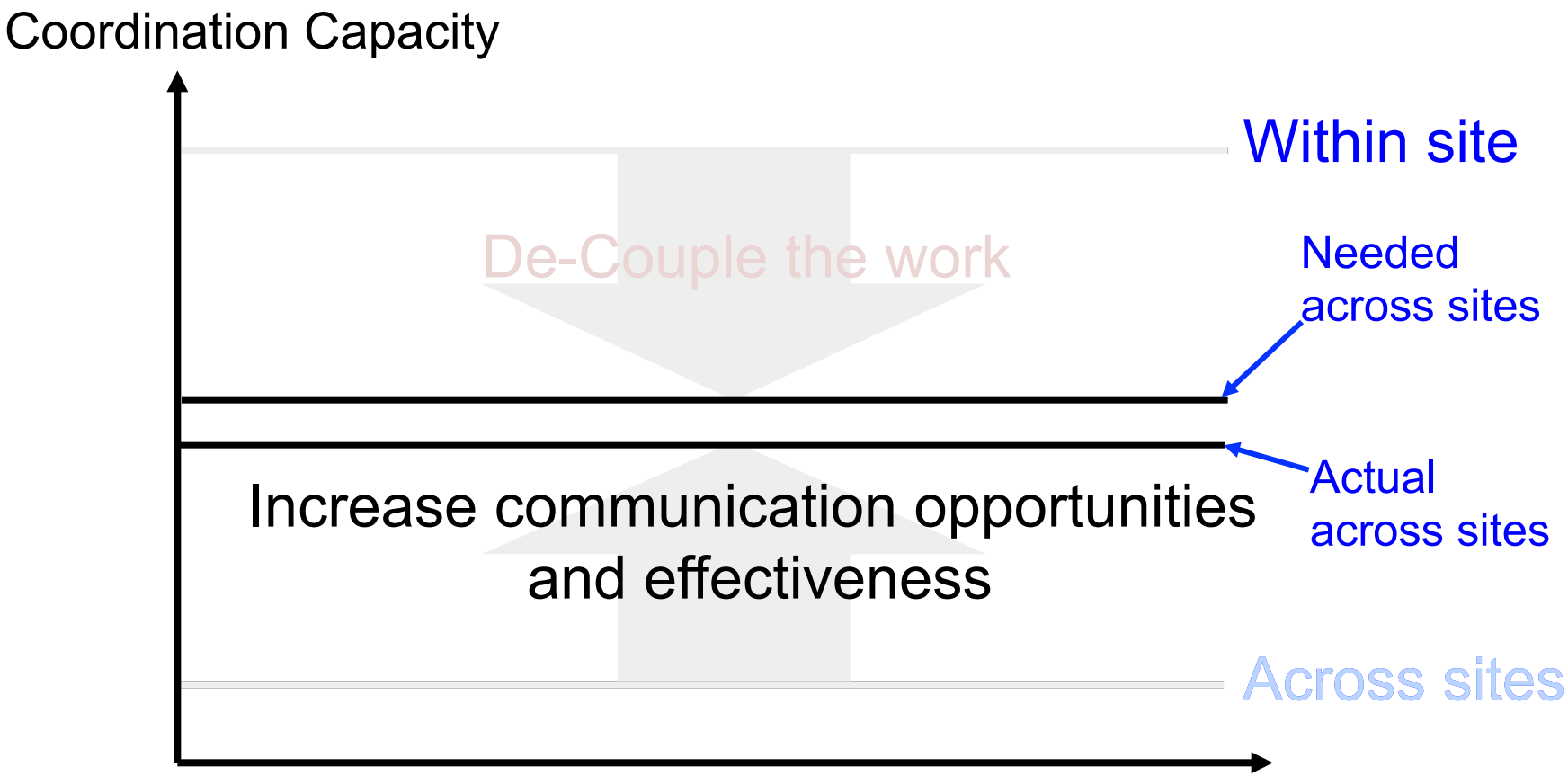


Organizational Models

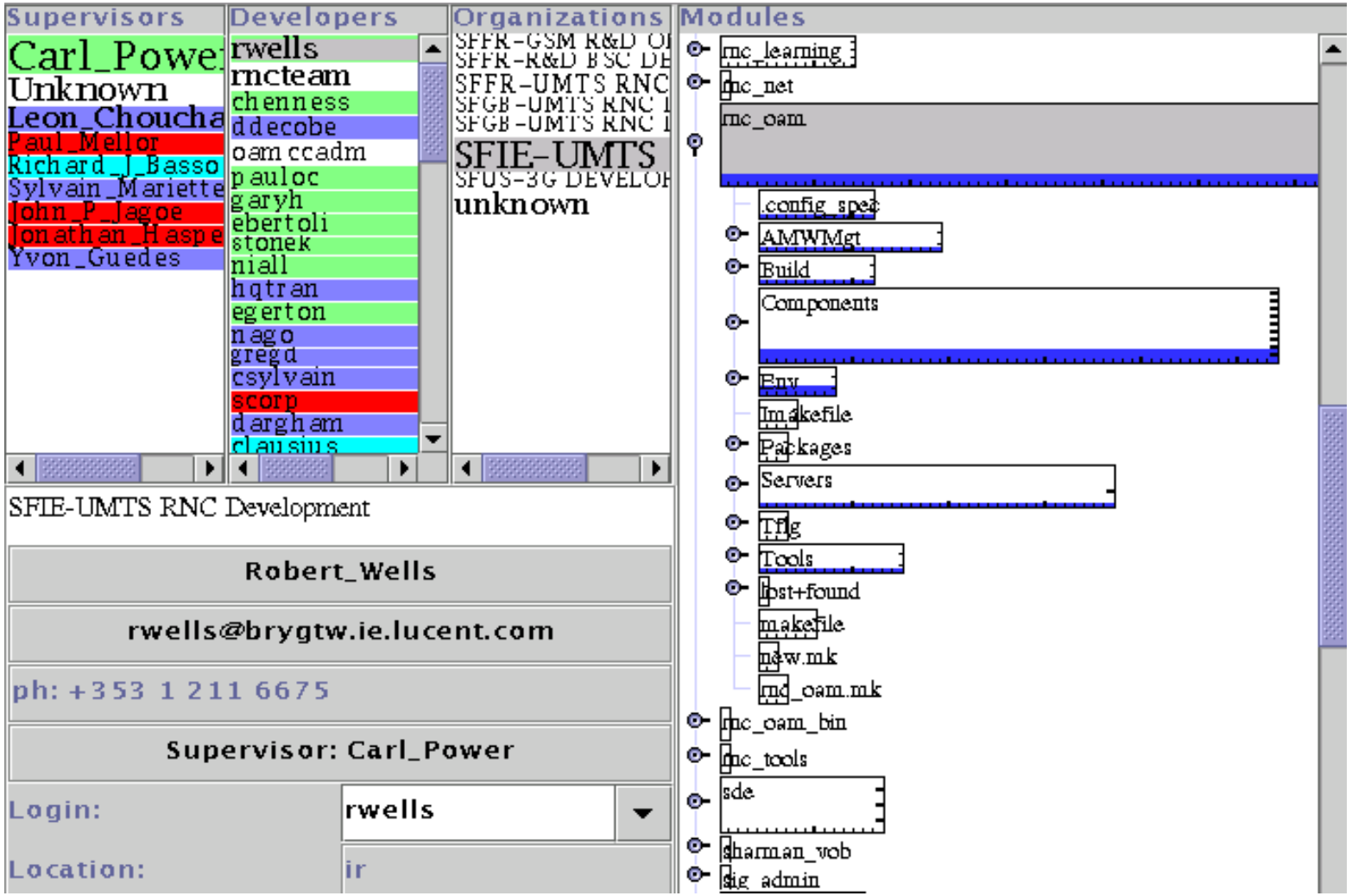


Grinter, R. E., Herbsleb, J. D. and Perry, D. E. The Geography of Coordination: Dealing with Distance in R&D Work. In *Proceedings of GROUP '99* (Phoenix, AZ, November 14-17, 1999).

Bridging the Gap



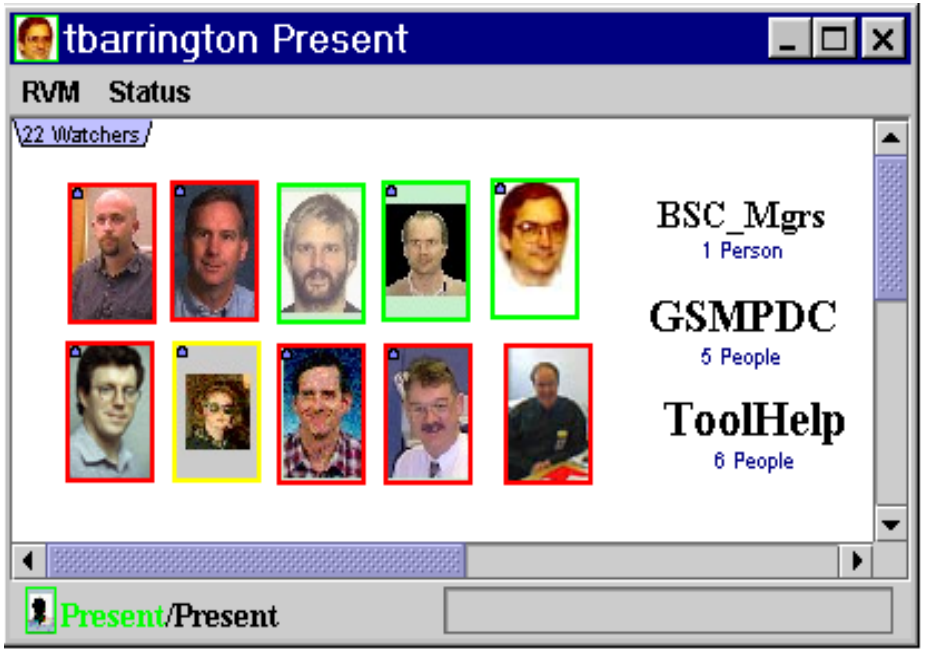
Expertise Browser



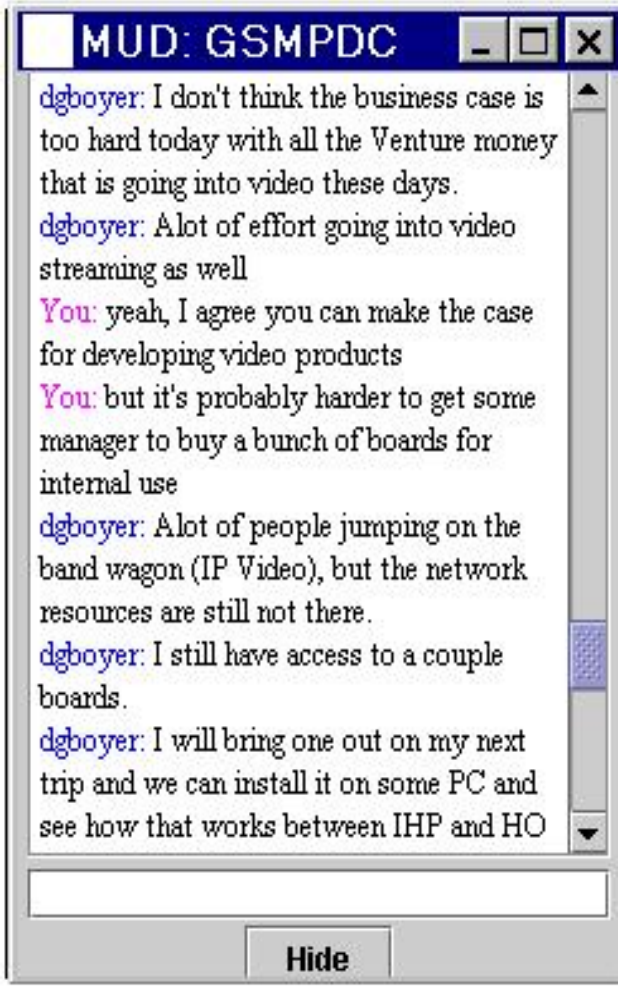
Mockus, A., & Herbsleb, J.D. (2002). Expertise Browser: A quantitative approach to identifying expertise. In Proceedings of *International Conference on Software Engineering*, Orlando, FL, May 19-25, pp. 503-512.

Instant Messaging

Rear View Mirror



Presence Viewer



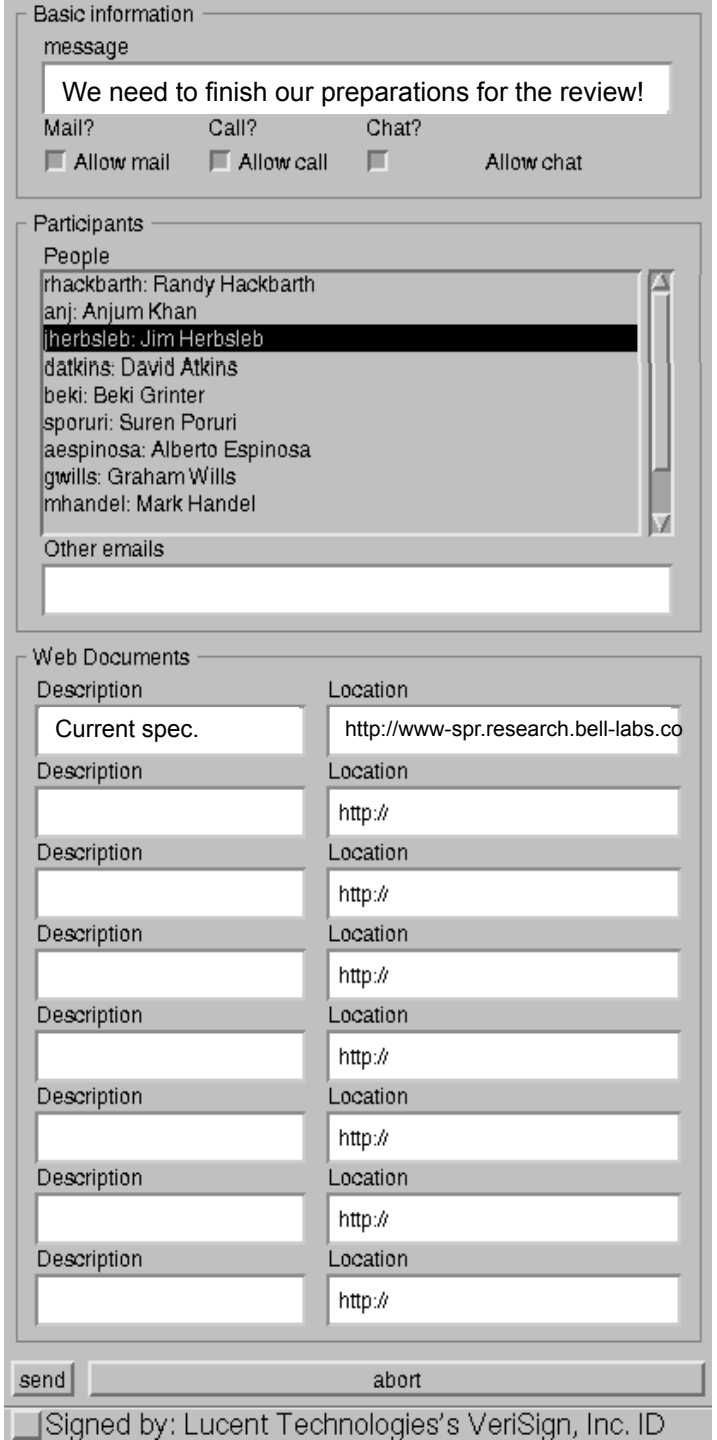
Group Chat

Handel, M. & Herbsleb, J.D. (2002). What is Chat doing in the workplace? Proceedings of ACM *Conference on Computer-Supported Cooperative Work* (CSCW), New Orleans, LA, pp. 1-10.

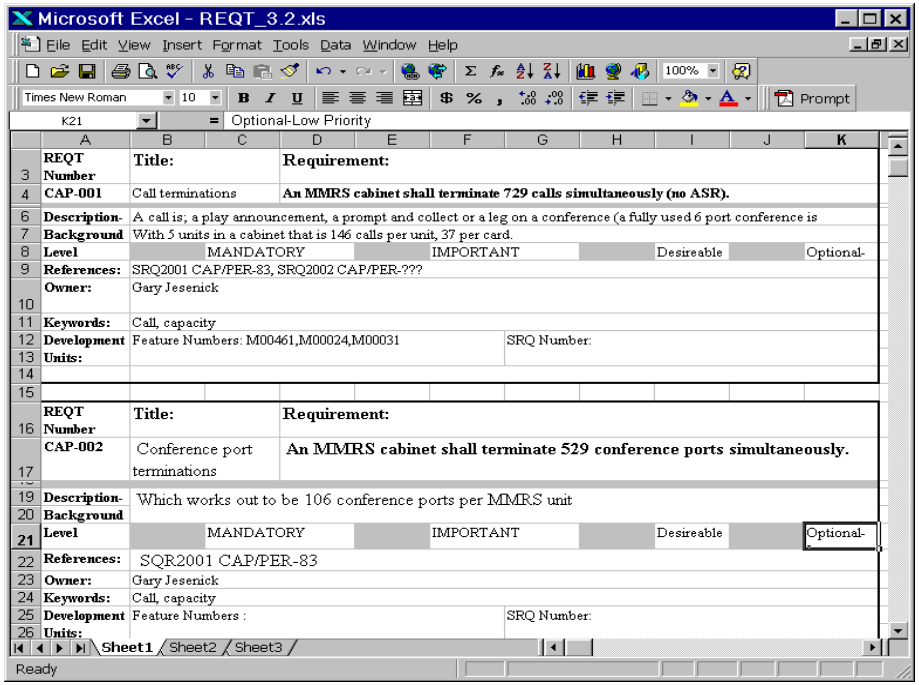
Herbsleb, J.D., Atkins, D.L., Boyer, D.G., Handel, M., & Finholt, T.A. (2002). Introducing Instant Messaging and Chat into the workplace. In Proceedings of ACM *Conference on Computer-Human Interaction*, Minneapolis, MN, April 20-25, pp. 171-178.

ConnectIcon

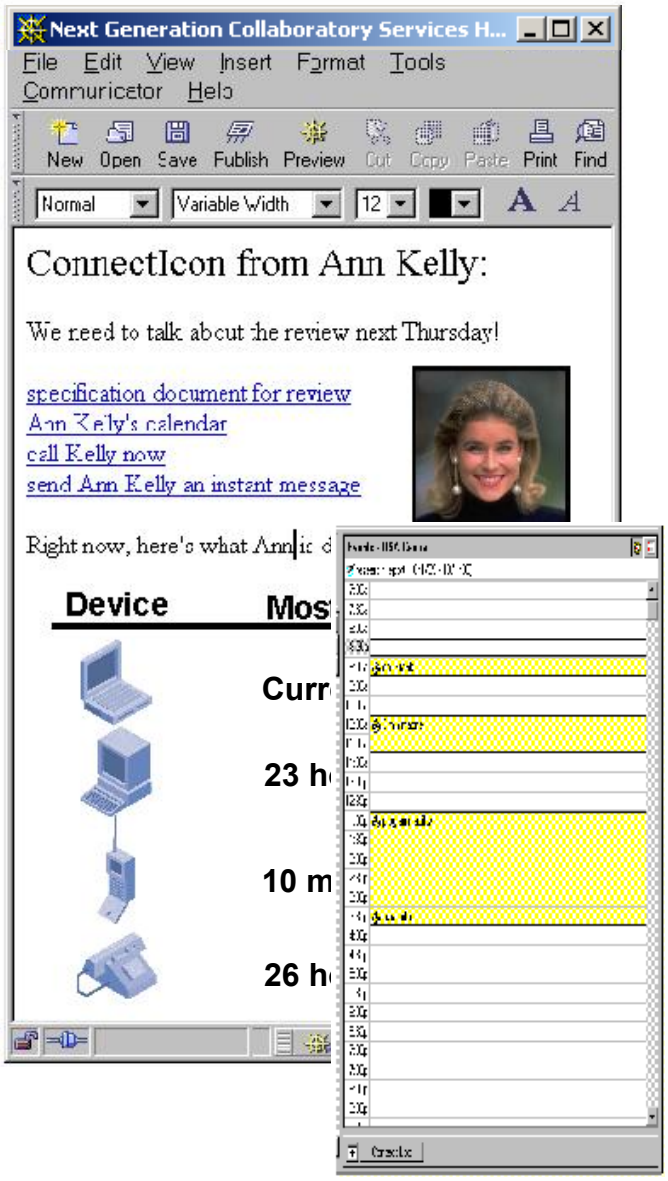
- Antidote for phone tag
- Send presence and contact ability to anyone



ConnectIcon

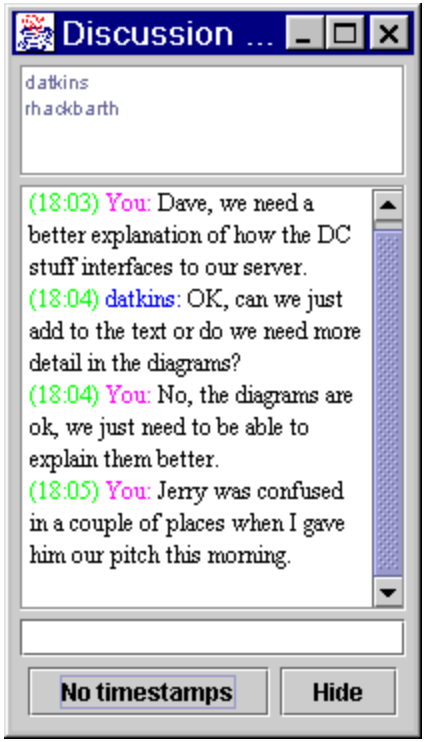


ConnectIcon



ConnectIcon

ConnectIcon

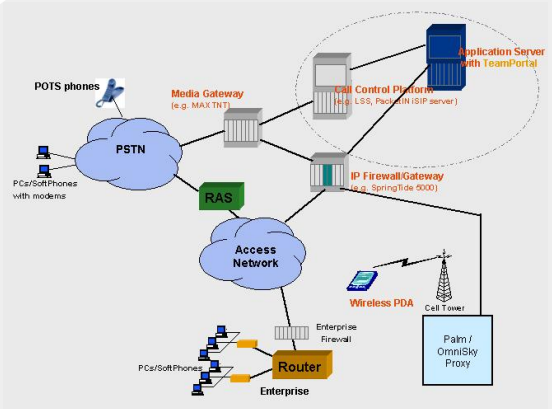


ConnectIcon

Bell Labs Executive Team:

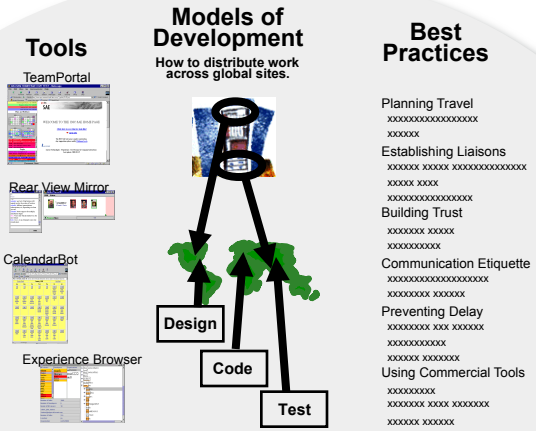
- This is a breakthrough project! You have the attention of the President of Bell Labs.
- We are assigning development teams to productize these technologies.
- Tell us what resources you need!

DotCom Bubble . . .

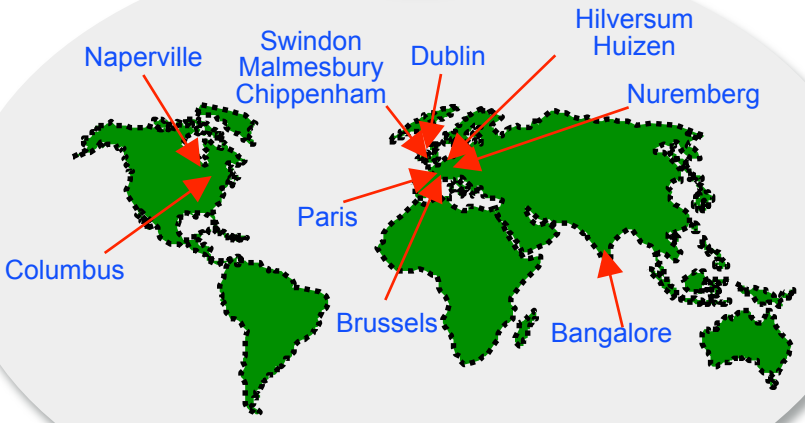


New Products

Research Team



Global Development Solutions



Empirical Studies

DotCom Bubble . . .

- Stock price: $80/share → $0.50/share
- Employees: 150,000 → 35,000
- Oops!

What does this have to do with engineering?

- Identified specific problems
- Crafted tools and practices to address them
- All very ad hoc, not like other, more mature forms of engineering

What Is Engineering?

- Creating cost-effective solutions
  - Engineering is not just about solving problems; it is about solving problems with economical use of all resources.
- to practical problems
  - Engineering deals with practical problems whose solutions matter to people outside the engineering domain-the customers.
- by applying scientific knowledge
  - Engineering solves problems in a particular way: by applying science, mathematics, and design analysis.
- to building things
  - Engineering emphasizes the solutions, which are usually tangible artifacts.
- in the service of mankind.
  - Engineering not only serves the immediate customer, but it also develops technology and expertise that will support the society.

From Shaw, M. Prospects for an engineering discipline of software. *IEEE Software*, 7, 6 (1990), 15-24.

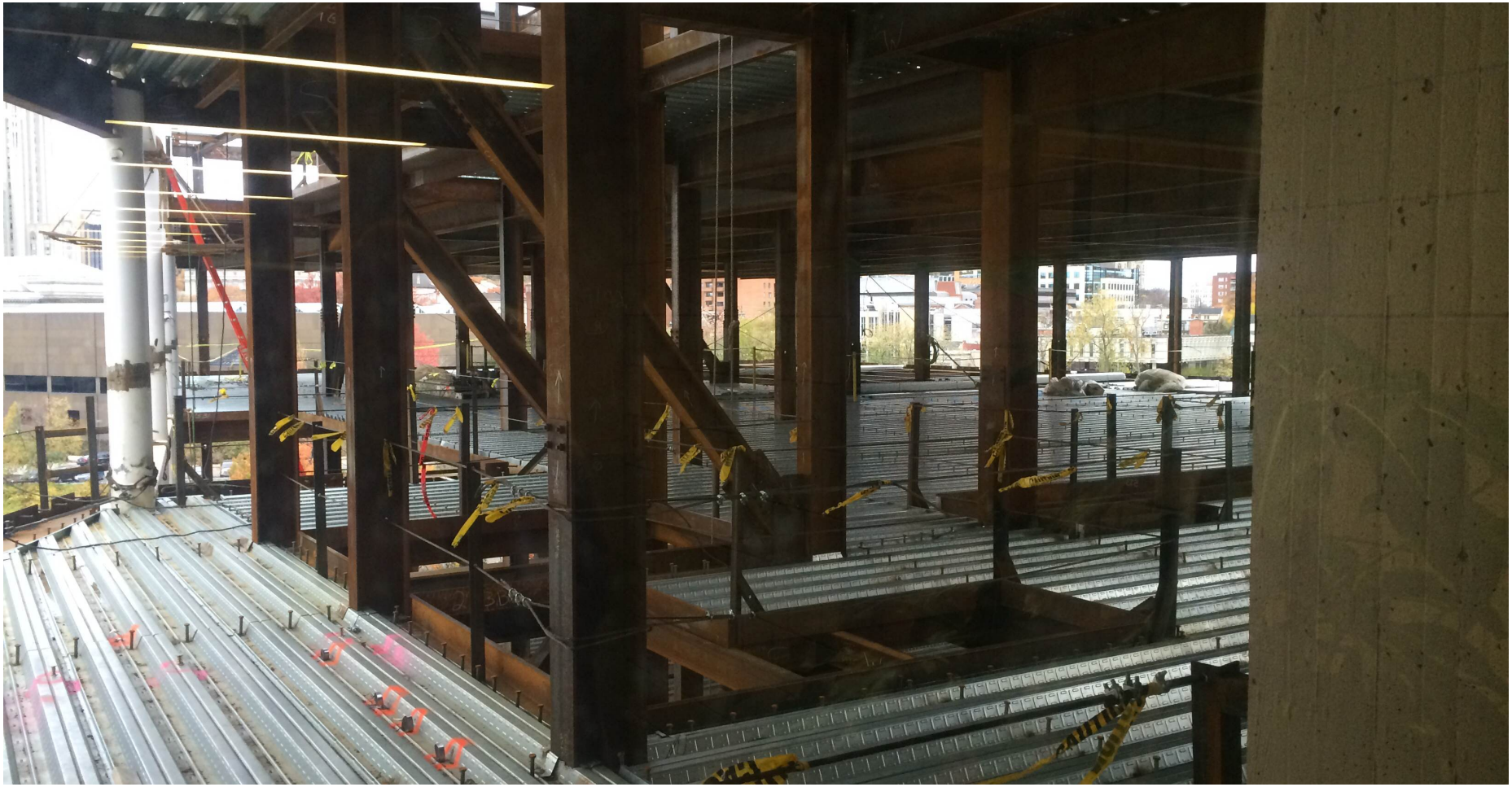
What Is Engineering?

- Creating cost-effective solutions
  - Engineering is not just about solving problems; it is about solving problems with economical use of all resources.
- to practical problems
  - Engineering deals with practical problems whose solutions matter
- by applying scientific knowledge
  - Engineering solves problems in a particular way: by applying science, mathematics, and design analysis.
- to building things
  - Engineering emphasizes the solutions, which are usually tangible artifacts.
- in the service of mankind.
  - Engineering not only serves the immediate customer, but it also develops technology and expertise that will support the society.

From Shaw, M. Prospects for an engineering discipline of software. *IEEE Software*, 7, 6 (1990), 15-24.

We Need a Science

- But what science do we need?



Many Engineering Challenges Are about Physical Components

- Strength of structural members
- Power consumption and output of a motor
- Power storage
- Sensitivity of sensors
- Etc., etc.
- **The science they need:** properties of physical and electronic components and compositions

We Need a Science

- But what science do we need?
- ***A science of humans as designers and builders***

Mental/Social Equipment: Evolution

- Adapted to hunter/gatherer way of life
  - Mental and physical capabilities
  - Evolution is a slow process
  - We did not change much in last 12,000 years (since the agricultural revolution*)
- We need to use mental equipment suited to simple hunter/gatherer life to design and build software

*Harari, Y. N. *Sapiens: A brief history of humankind* Random House, 2014.

Example Cognitive Modules

- Acquiring natural language
  - Chomsky: our brain has a built-in language acquisition device (LAD)
- Visually interpreting 3D space
  - The best terrain modeling and autopilot programs are not there yet
- "Theory of mind"
  - Cognitive module that interprets and predicts behavior of others based on inferred beliefs and desires*

*Herbsleb, J. D. (1999). Metaphorical representation in collaborative software engineering. In *Proceedings, International Joint Conference on Work Activities, Coordination, and Collaboration*, San Francisco, CA, February 22-25, pp. 117-125.

What Is the Problem?

- Our most severe problems and limitations do not arise from physical components
- We need computer science, obviously, but:
- Most limitations come from our own limited capacities
  - What can we understand?
  - What languages, abstractions, algorithms, and data structures can we dream up?
  - What are our cognitive and communication limitations and how can we compensate for them?
  - How can we act together in a coordinated way?

Three Examples

- Transactive memory systems
- Gatekeepers and social networks
- Socio-technical theory of coordination

Transactive Memory Systems (TMS)

- Group level phenomenon
- Arises naturally
- Specialization + index
  - People take responsibility for group knowledge and memory in some area
  - Everyone shares an index of “who knows what”
  - Origins in people watching each other work
- Very powerful impacts on how well groups function

TMS: Benefits and Conditions

- Specialization gives better performance
- Better coordination, agree on responsibilities
- Facilitates adaptation to new situations or tasks
- Facilitates creativity
- Develops under right conditions
  - Observe each other working
  - Communication

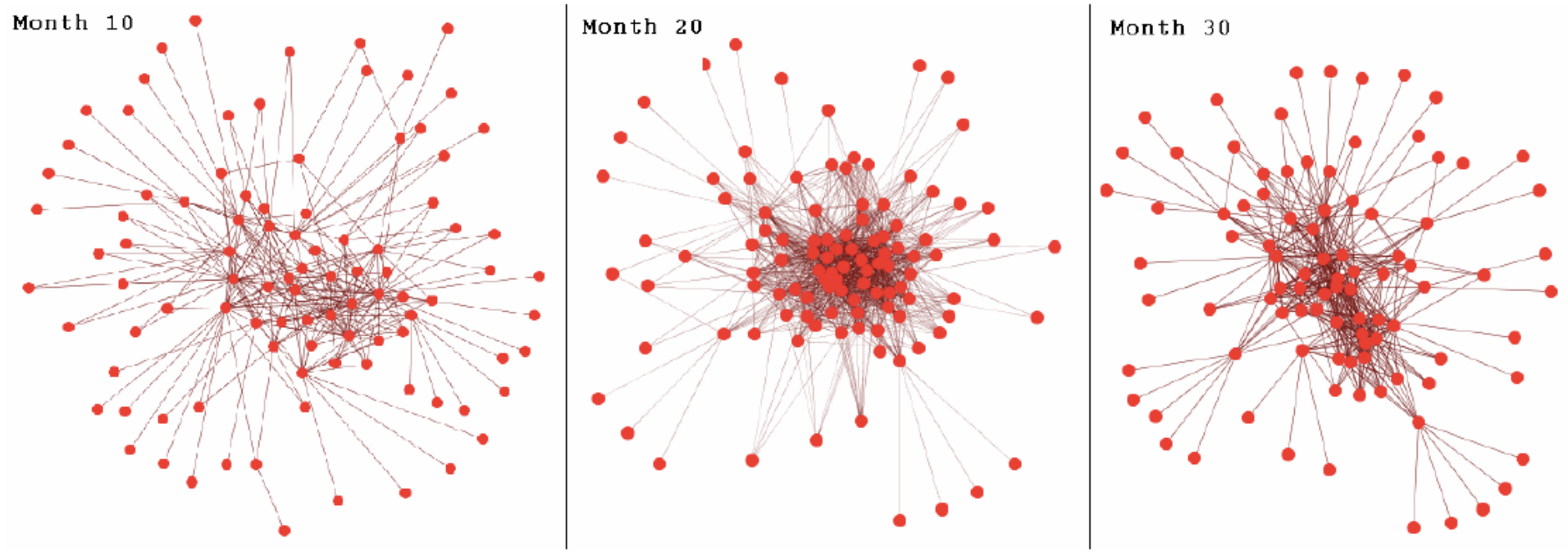
Argote, L. and Ren, Y. Transactive memory systems: A microfoundation of dynamic capabilities. *Journal of Management Studies*, 49, 8 (2012), 1375-1382.

Gatekeepers and Social Networks

- Small number of trusted people become information hubs
- Connected to information sources inside and outside organization
- People go to them with questions
- They form their own network, know each other's expertise

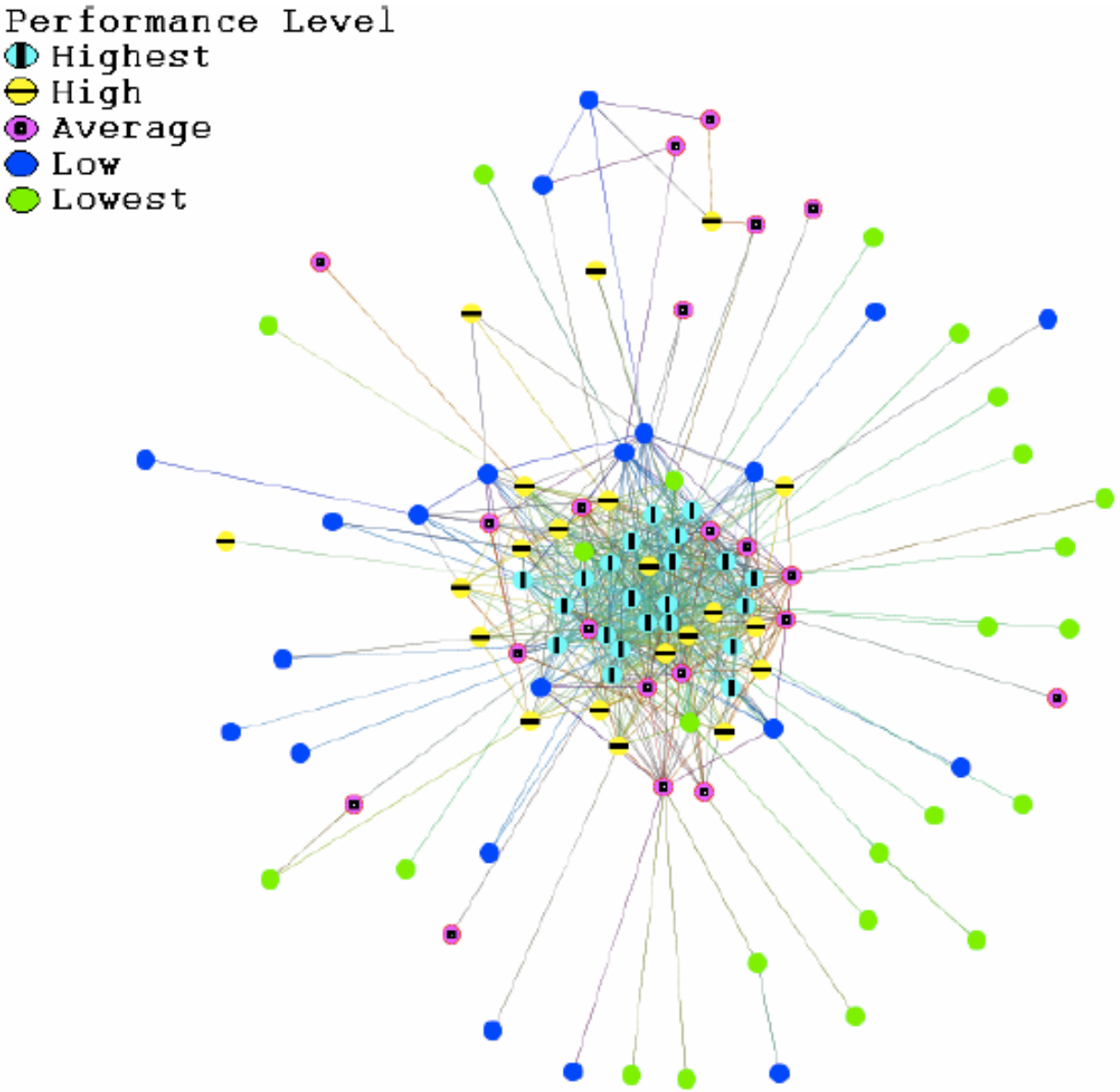
Origin of Gatekeeper findings: Allen, T. J. *Managing the Flow of Technology*. Cambridge, MA: MIT Press, 1977.

Core-Periphery Topology



Cataldo, M. & Herbsleb, J.D. (2008). Communication networks in geographically distributed software development. In Proceedings, ACM Conference on Computer-Supported Cooperative Work, San Diego, CA, Nov. 8-12, pp. 579-588.

Core Membership and Productivity



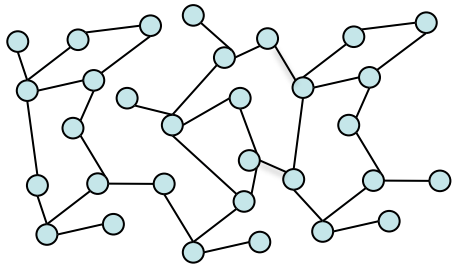
The Point . . .

- When people organize, under the right conditions they spontaneously form
  - Transactive memory systems
  - Gatekeeper networks
- Why this matters
  - Working with them provides powerful capability
  - Working against them will be difficult

Example: GitHub

- Why so successful?
- Provides means for humans to form and use social capabilities
  - TMS: activity traces, profiles, consistent across repositories
  - Gatekeeper networks: Watching, starring, following, curating, “asynchronous mentoring”

Socio-Technical Coordination



Decisions and Constraints

Technical coordination is a Constraint satisfaction problem (CSP) over decisions

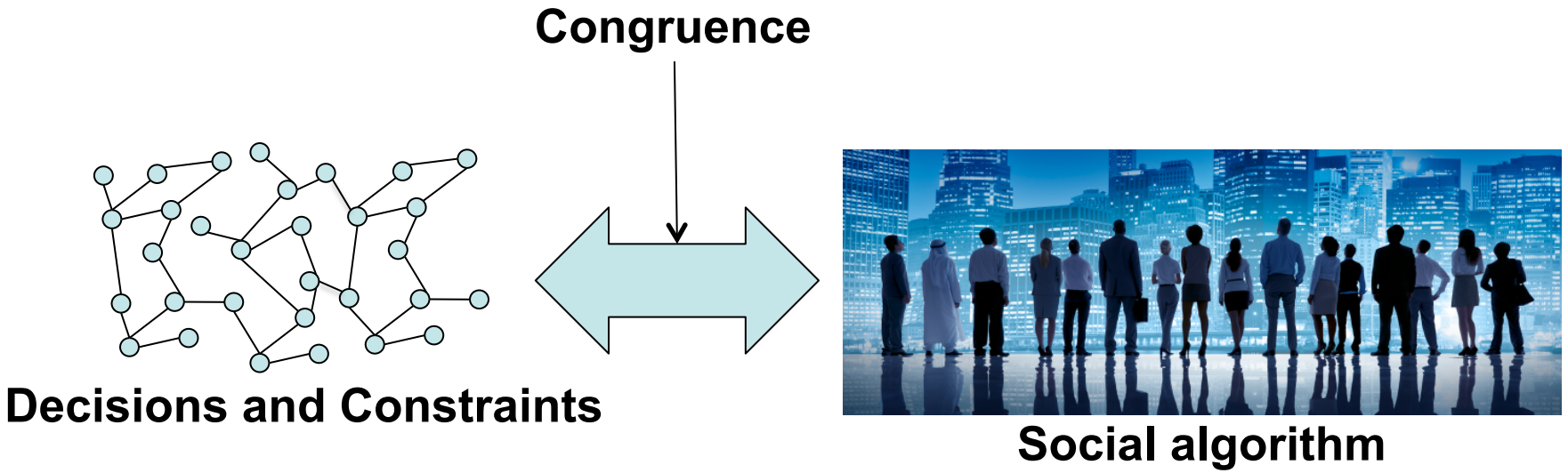
Decisions distributed over people (DCSP)

Social algorithm to solve DCSP

Herbsleb, J.D., & Mockus, A. (2003). Formulation and preliminary test of an empirical theory of coordination in software engineering. In Proceedings, *ACM SIGSOFT Symposium on the Foundations of Software Engineering*, Helsinki, Finland, September 1-5, pp. 112-121

Herbsleb, J.D., Mockus, A., Roberts, J.A. (2006). Collaboration in Software Engineering Projects: A Theory of Coordination. *International Conference on Information Systems*, Milwaukee, WI.

Socio-Technical Coordination



Cataldo, M., Wagstrom, P. A., Herbsleb, J. D. and Carley, K. M. (2006). Identification of coordination requirements: implications for the Design of collaboration and awareness tools. In Proceedings, Computer supported cooperative work, Banff, Alberta, Canada, pp. 353-362.

Cataldo, M., Herbsleb, J. D. and Carley, K. M. (2008). Socio-Technical Congruence: A Framework for Assessing the Impact of Technical and Work Dependencies on Software Development Productivity. In Proceedings, International Symposium on Empirical Software Engineering and Measurement, Kaiserslautern, Germany, pp. 2-11.

Cataldo, M. and Herbsleb, J. D. Coordination Breakdowns and Their Impact on Development Productivity and Software Failures. IEEE Transactions on Software Engineering 39, 3 (2013), 343-360.

Social Algorithms

- Can take advantage or fail to take advantage of powerful capabilities
- Can be derailed by people using capabilities mismatched to task
- We need a much sharper picture of these capabilities and how software tasks map onto them

The Science We Need

- Psychology, sociology, etc. are a starting point
- Only moderately useful by themselves
  - Stretched by complexity of environment and task
  - Stretched by rapid change
  - Stretched by capabilities of digital tools and materials: social reach, free copying, absence of geographic boundaries
- We need a socio-technical perspective to create our own behavioral science!
- ***Theory-driven studies of people using technology to collaborate on technical tasks***

Barriers to Human Science

- The universal principle of interdisciplinary contempt
  - Intellectual worth is evaluated on a single dimension from math to BS
- DPHB* principle: everything I don't understand is simple
  - Behavioral science is fuzzy and just common sense
- Culture does not always appreciate behavioral theory
  - Theory seen as mere decoration and distraction on top of statistical model
  - Statistics used to test relations between theoretical constructs
  - Not just associations among variables
- Border defense, antibodies
  - Is that really computer science?
- Necessity to argue for practical application of each result

*Dilbert's pointy-haired boss

equipment, and to Dr. G. E. R. Deacon and the captain and officers of R.R.S. *Discovery II* for their part in making the observations.

[1] Young, F. B., Gerrard, H., and Jevons, W., *Phil. Mag.*, **40**, 149 (1920).

[2] Longuet-Higgins, M. S., *Mon. Not. Roy. Astro. Soc., Geophys. Supp.*, **5**, 285 (1949).

[3] Von Arx, W. S., Woods Hole Papers in Phys. Oceanog. Meteor., **11** (3) (1950).

[4] Ekman, V. W., *Arkiv. Mat. Astron. Fysik.* (*Stockholm*), **2** (11) (1905).

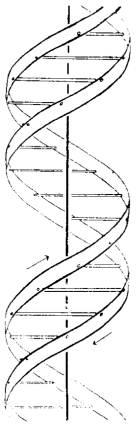
MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

WE wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey[1]. They kindly made their manuscript available to us in advance of publication. Their model consists of three intertwined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.



This figure is purely diagrammatic. The two ribbons symbolize the two phosphate—sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate diester groups joining β-D-deoxyribofuranose residues with 3′,5′ linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furberg's[2] model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it is close to Furberg's 'standard configuration', the sugar being roughly perpendicular to the attached base. There is a residue on each chain every 3·4 A. in the z-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical z-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined.

It has been found experimentally[3,4] that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid.

It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

The previously published X-ray data[5,6] on deoxyribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereochemical arguments.

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on interatomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their co-workers at

The demand for immediate relevance rather than overall contribution . . . a hypothetical rejection letter:

Drs. Watson and Crick:

I regret to inform you that we are unable to accept your paper.

I personally find it very interesting that the DNA molecule has the shape of a double helix held together by paired bases. But the reviewers felt that you have not demonstrated any practical application for this discovery, so it was decided that the contribution was insufficient.

Next Steps Toward a Behavioral Science of Software Engineering

- Work toward a community
- Workshops
- Collect readings
- Develop course/curriculum
- Special issue

A few readings I have found useful

- Argote, L. and Ren, Y. Transactive memory systems: A microfoundation of dynamic capabilities. *Journal of Management Studies*, 49, 8 (2012), 1375-1382.
- Cataldo, M. & Herbsleb, J.D. (2008). Communication networks in geographically distributed software development. In *Proceedings, ACM Conference on Computer- Supported Cooperative Work*, San Diego, CA, Nov. 8-12, pp. 579-588.
- Clark, A. From folk psychology to naive psychology. *Cognitive Science*, 11, 2 (1987), 139-154.
- Dunbar, R. The social brain hypothesis. *Brain*, 9, 10 (1998), 178-190.
- Harari, Y. N. *Sapiens: A brief history of humankind* Random House, 2014.
- Kahneman, D. *Thinking, fast and slow* Macmillan, 2011.
- Pinker, S. *The language instinct: The new science of language and mind* Penguin UK, 1995.