



UNIVERSITÀ
DEGLI STUDI DI BARI
ALDO MORO



Facing Communication Challenges in Distributed Software Development

Keynote for VirtuES 2013 workshop, Bari, Italy

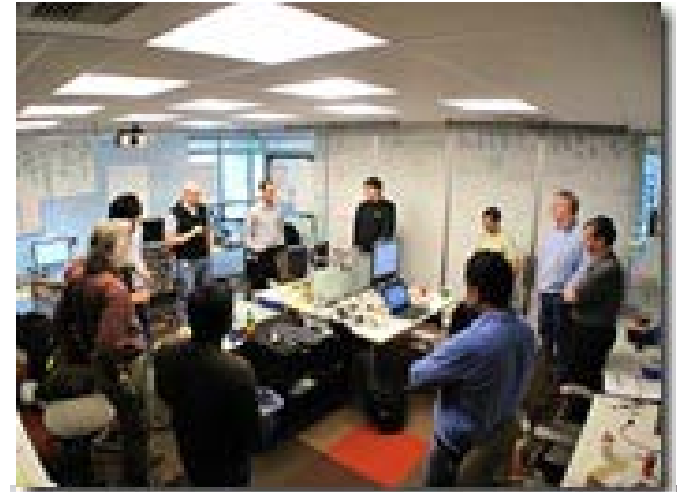
Fabio Calefato
University of Bari, Italy

Collaborative Development Research Group
<http://cdg.di.uniba.it>



Software development as an intense collaborative process

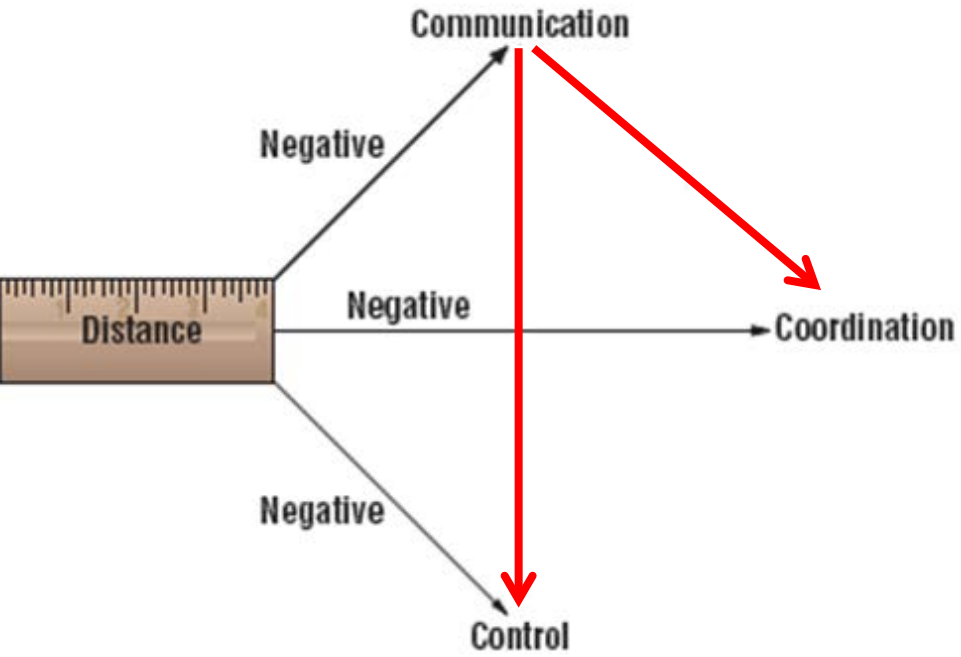
- Software development as “*an exercise in complex interrelationships*” *
- Types of collaboration within a team
 - Communication
 - Coordination
 - Control





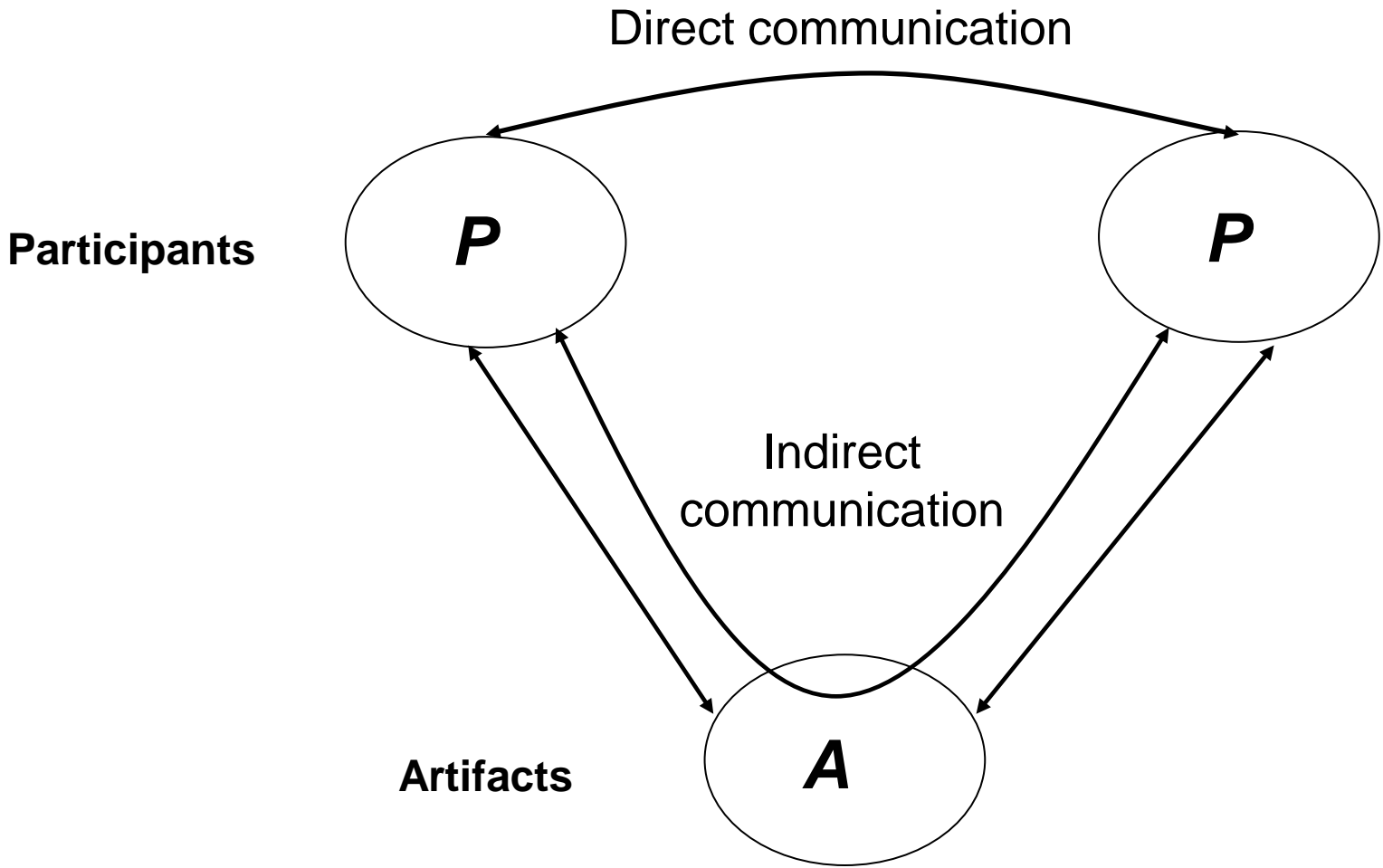
Communication's key role in managing distributed sw projects

- Distance has both direct and indirect negative effect on coordination and control
- Communication disruption aggravates coordination and control breakdowns





Direct vs. Indirect communication





Formal vs. Informal communication

	Formal	Informal
Message	<i>Planned</i>	<i>Spontaneous</i>
Content	<i>Work-related</i>	<i>Relational</i>
Purpose	<i>Organizational</i>	<i>Personal</i>
Location / channels	<i>Official</i>	<i>Random</i>

Communication challenges “faced” in this talk



What

- Direct formal communication
 - Finding best fitting media for communication-intensive tasks
 - Overcoming language barriers with machine translation
- Direct informal communication
 - Establishing personal, trust-based connections in distributed teams

How

- Studies to inform software practices or tool designs
- Mixed research methods:
 - Analysis of software artifacts
 - Questionnaires
 - Interviews
 - Content analysis



Research partners:

Daniela Damian (UVic), Filippo Lanubile (Univ. of Bari)

FINDING BEST FITTING MEDIA FOR COMMUNICATION-INTENSIVE TASKS



Research goal

- To assess the support offered by synchronous text-based communication media (CMC) to distributed groups involved in requirements elicitations and negotiations as compared to traditional face-to-face communication (F2F)



Computer-Mediated Communication (CMC)

- Media can be characterized along three dimensions of information exchange:
 - Time (when)
 - Space (where)
 - Richness (how much)

Media-Richness continuum in the Time/Space Matrix

Synchronous (same time)	F2F	Rich
	Videoconference Telephone Chat	
Asynchronous (different time)	Billboard	Lean
	Email Letter	
	Collocated (same space)	Distributed (different space)



Main theories on CMC

- **Social Presence Theory (Short et al., 1976)**
 - Lean single-channel media have low sense of presence (inability to convey non-verbal cues)
 - Lean media better for task-focused communication, rich media for relational communication
- **Media Richness Theory (Daft & Lengel, 1984)**
 - The more complex the task, the richer the media to use
 - Lean media better for uncertain tasks, rich media better for equivocal tasks
- **Common Ground Theory (Clark & Brennan, 1991)**
 - Argues that communications is not simply the sending of messages
 - There is no best medium in absolute



Main theories on CMC

- Channel Expansion Theory (Goodhue et al., 1995)
 - Factors other than channel characteristics affect CMC effectiveness
 - Group's shared experience and media use experience
- Media Synchronicity Theory (Dennis & Valacich, 1998)
 - Highly synch media when convergence is the key process to task accomplishment
 - Lowly synch media when conveyance is the key process to task accomplishment
- Cognitive-Based View (Robert & Dennis, 2005)
 - Sense of presence not as vital as the ability to process information
 - Media Richness Paradox: The richer the media, the harder to properly process information

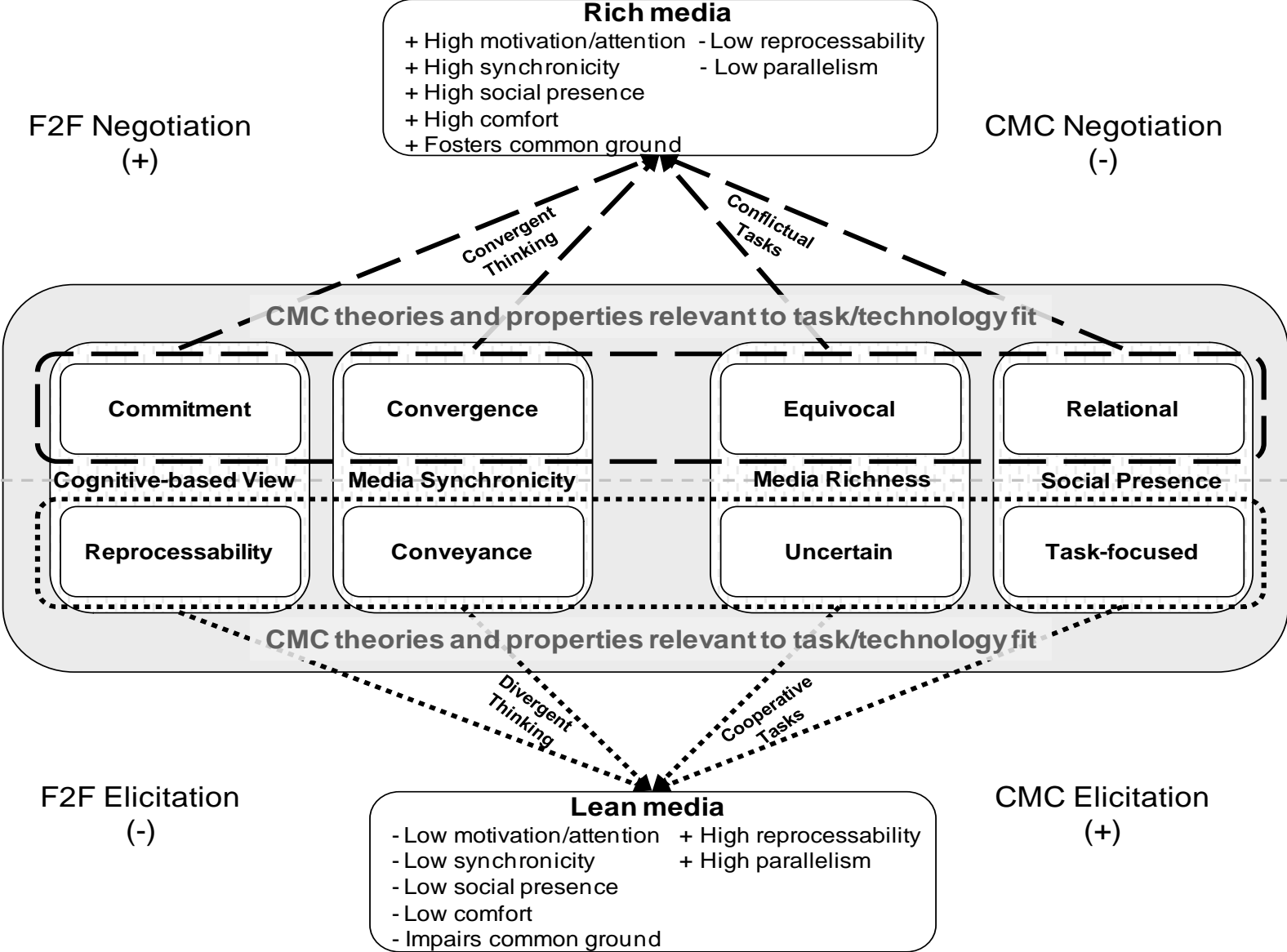
Task/Technology Fit (TTF)



- Effectiveness of CMC varies on the type of task
 - Differences in tasks and media generate differences in group performance
- Rich media do not always provide the best solution for any given task
 - Too much or too few media richness for a given task represents a poor TTF
- Good TTF only when information richness required by task is proportional to that conveyed by media
 - TTF theories suggest how to appropriately match task characteristics to medium properties



Putting all the pieces together





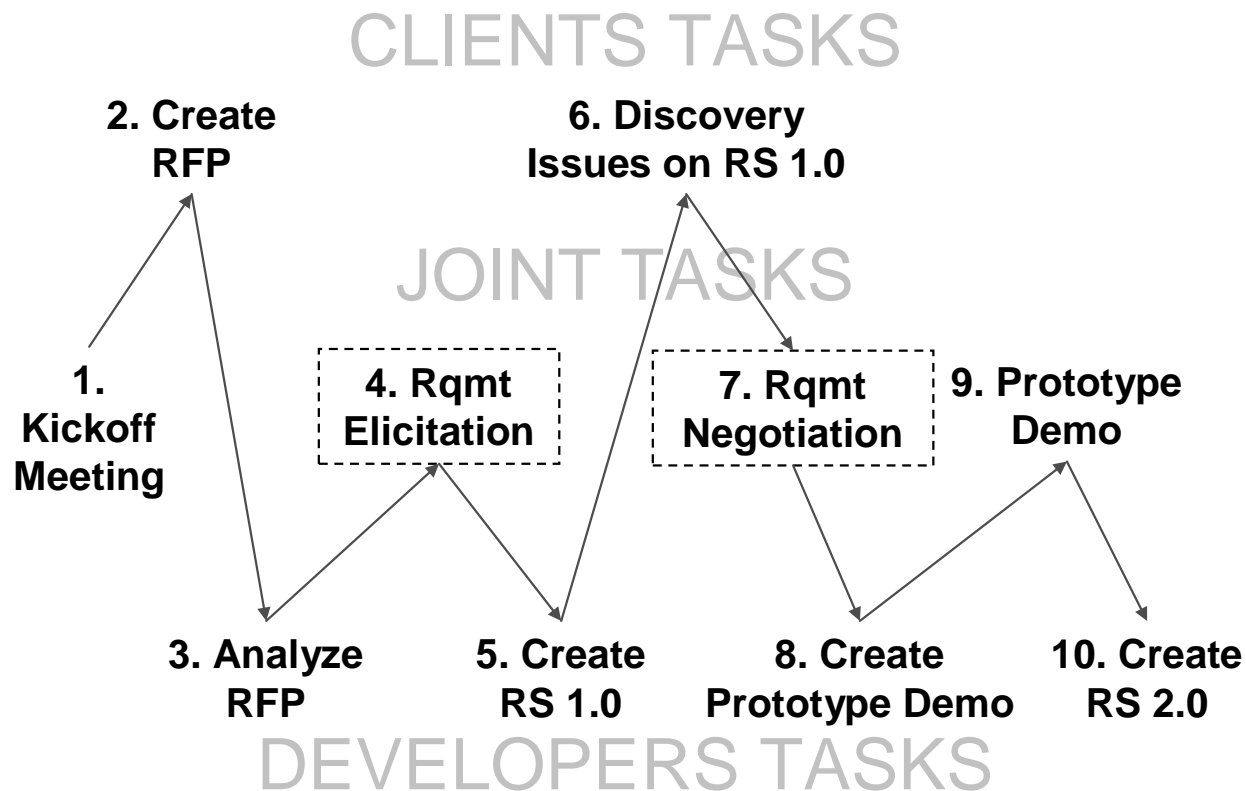
Empirical study

- Goal
 - Compare F2F and synchronous text-based communication (CMC) modes in distributed requirements elicitations and negotiations
- Setting
 - RE undergraduate course at University of Victoria, Canada
 - 40 students in 6 groups of stakeholders involved in the definition of sw requirements for 6 projects
- Research hypotheses
 - **H1** – *F2F requirements workshops are better appreciated (i.e., perceived as more comfortable and satisfying with outcome) than CMC requirements workshops*
 - **H2** – *CMC Elicitation represents a better task/technology fit (i.e., produce better results) than F2F Elicitation. F2F Negotiation represents a better task/technology fit (i.e., produce better results) than CMC Negotiation*



Data

- Gathered through two satisfaction questionnaires and chat logs
- Questionnaires administered at the end of the elicitation and negotiation sessions





eConference

<http://code.google.com/p/econference4>

Collaborative workbench

File Workbench Chat Options Help

Perspectives

Agenda

- Stop conference
- Epic 1 - See buddies status and send IM
- Epic 2 - Create and join a chat room
- Epic 3 - Organize and join an eConfe

Agenda

Who's on

- Fabio
- Filippo
- Mario (SCRIBE)

Presence

Epic 1 - See buddies status and send IM

Fabio > Epic 1 can be split into two stories of smaller size:
* A user can check the online status of contacts in the roster
* A user can start an IM session with online contacts
Fabio > I think we all agree on this

Filippo > Just a Note: An IM can be sent also to offline contacts, as XMPP server usually store messages sent to offline recipients

Message board

Decisions place

We are splitting epics into smaller, testable

Epic 1 can be split into two stories of smaller
* A user can check the online status of cont
* A user can start an IM session with online

Epic 2 can be split into 3 stories:
* A user can create a chat room
* A user can join a chat room
* A user can send invitations to people for j

Hand raising

Id	Requestor
1	mario

Hand raise

- Reject
- Approve

Dependant variables and measures – H1

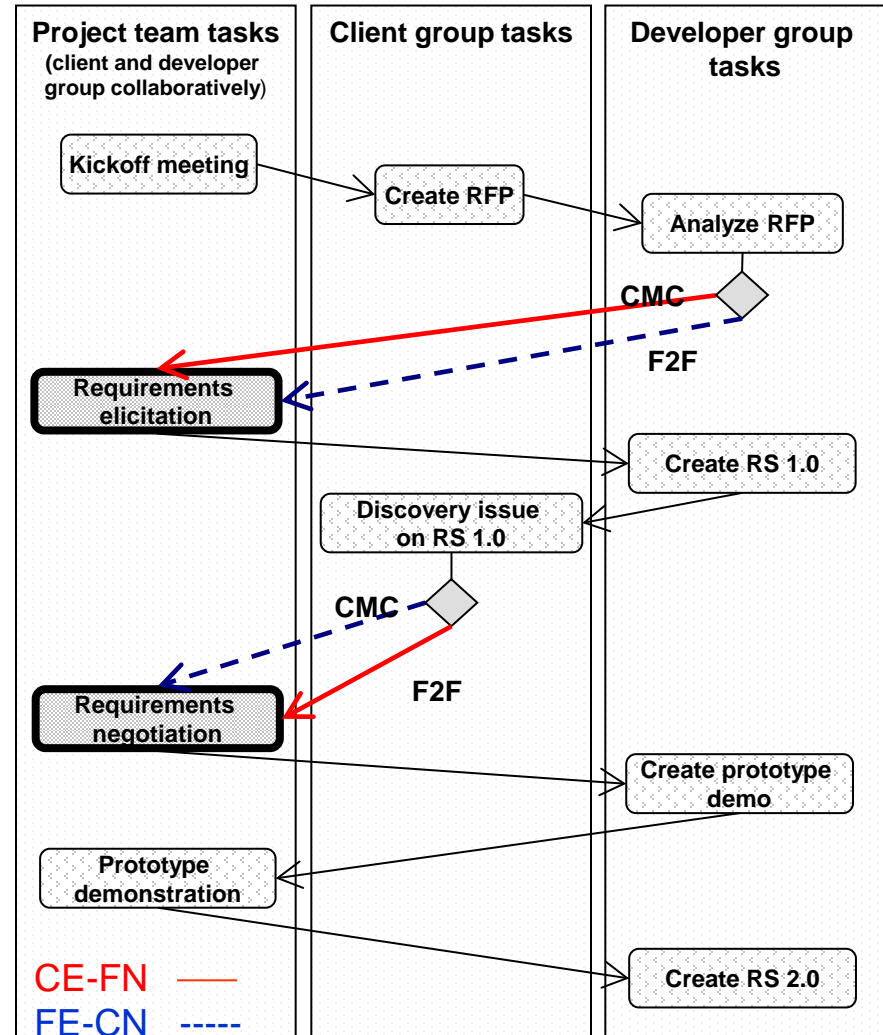


- Subjects' responses to satisfaction questionnaires coded to perform quantitative analysis
 - Used 4-point Likert scales and closed questions
- Differences between stakeholders' perception of requirements workshops conceptualized as:
 - Satisfaction with performance
 - Comfort with communication mode



Dependant variables and measures – H2

- *Requirements workshop* and *Communication mode* factors created two variants in the rqmt definition process
 - **CE - FN**
 - **FE - CN**
- Differences in rqmt workshops effectiveness conceptualized as:
 - Group performance
 - Shared understanding => lack of common ground





Coding schema

- Performed content analysis on the logs of the CMC workshops to collect data on negative evidence and grounding chains constructs
- 9 major categories (thematic unit)

	Category	Description
QUESTION	YES-NO	A question that takes a yes/no answer or just a few words (e.g., Q: "How many beds are available in the hospital overall?", A: "There are 400"). It may initiate a grounding chain.
	COMPLEX	Any other question not covered by the YES-NO QUESTION category, such as the <u>wh</u> -questions or those that aggregate multiple questions in a single utterance (e.g., "What information will each employee use to login to the system? Is a key card a consideration?"). It may initiate a grounding chain.
ANSWER	YES-NO	An answer to a question that takes a few words or a yes-no utterance (e.g., yes, no, yep, nope, y, n, yeah, "yeah, 400"). Normally appears after a YES-NO QUESTION or CHECK. It may close a grounding chain.
	COMPLEX	Any other answer to a question not covered by the YES-NO ANSWER category (e.g., Q: "Are you going to do the display public events portion for the final product?", A: "Yes"). It may close a grounding chain.
CHECK	PROVISIONAL	Any utterance that explicitly looks for confirmation of acceptance through provisional, try-marked statements (e.g., "So we decided for 400 beds, right?"). It is normally followed by an AGREEMENT or an ANSWER.
	VERBATIM COPY	Any utterance that explicitly gives confirmation of acceptance through by verbatim copying a previous utterances (e.g., "Backup monthly on a tape", "Ok, once a month on tape"). It is normally followed by an AGREEMENT.
	MISUNDERSTANDING	Any utterance that provides evidence that a previously entered utterance was not accepted (e.g., "I'm not sure I get the question", "What?"). It may initiate a grounding chain and is normally followed by a TASK or an ANSWER.

Summary of study findings (1/3)



- *The role (i.e., being a customer or a developer) has no effect on media selection*
 - No significant differences in satisfaction and comfort between CMC and F2F requirements meetings

Summary of study findings (2/3)



- *No Conclusive Evidence that F2F is The Most Preferred Medium for Communication*
 - F2F provided more opportunity to familiarize with other participants, better ability to express complex ideas and to understand others' opinions in both elicitations and negotiations
 - CMC was more comfortable during both elicitations and negotiations to better participate and openly discuss conflicting issues more openly with group members
 - Stakeholders more satisfied with the performance in the F2F negotiations than in the CMC negotiations
 - No differences in the perceived satisfaction with performance between F2F and CMC elicitations



Summary of study findings (3/3)

- *Group performance not affected by communication medium*
 - No significant differences in the number of defects found in the final RS docs produced at the end of the process
- *Computer-Mediated Elicitations Offer Support to Achieving Shared Understanding*
 - Data on grounding chains did not allow to draw conclusions about support given by CMC elicitations as compared to the CMC negotiations
 - CMC elicitations had less negative evidence than the CMC negotiations



Research Partners:

*Tayana Conte (UFAM), Filippo Lanubile (Univ. of Bari),
Rafael Prikladnicki (PUCRS)*

OVERCOMING LANGUAGE BARRIERS WITH MACHINE TRANSLATION



Motivation

- Global software projects challenged by language differences
 - Especially requirements meetings
- Machine translation technology for remote meetings in countries with
 - Opportunities for global projects
 - Lack of English speaking professionals
- Example: Brazil
 - Large pool of IT professionals
 - Only 10M speaking English (< 6% of the population)



Research questions

- RQ1: *Can MT services be used in distributed multilingual requirements meetings? (instead of English)*
- RQ2: *How does the adoption of MT affect group interaction? (in distributed multilingual requirements meetings)*
- RQ3: *Do individuals with a low English proficiency level benefit more than individuals with a high level from MT?*



1st study: Simulation

F. Calefato, F. Lanubile, and P. Minervini. “*Can Real-Time Machine Translation Overcome Language Barriers in Distributed Requirements Engineering?*”, ICGSE 2010.

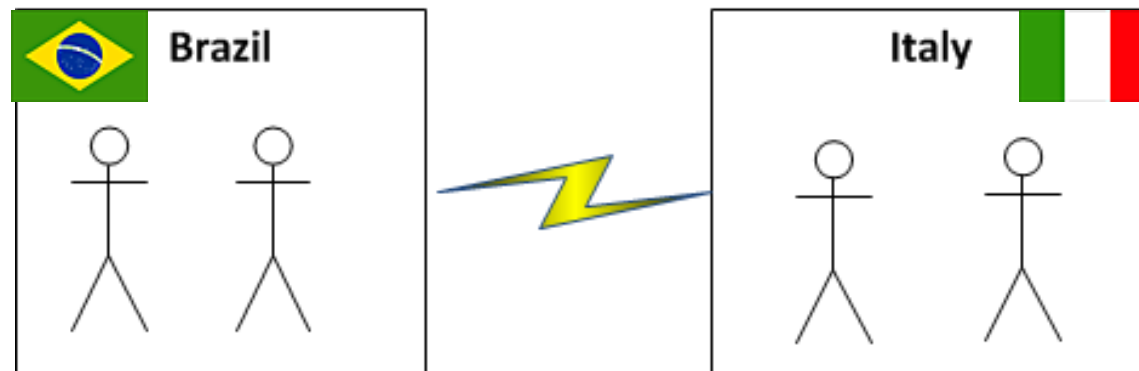
- Simulation
 - Requirements meetings logs as data source
- Assessment of time performance:
 - Delay is negligible
- Assessment of translation quality:
 - Google Translate (corpus-based) more accurate than Apertium (rule-based)

2nd study: Controlled experiment



F. Calefato, F. Lanubile, T. Conte and R. Prikladnicki, "*Assessing the Impact of Real-Time Machine Translation on Requirements Meetings: A Replicated Experiment*", ESEM'12

- 32 participants: 16 students from Bari (IT)
8 from PUCRS, Porto Alegre (BR)
8 from Fed. Univ. of Amazonas, Manaus (BR)
- Multilingual groups arranged by proficiency level of English (high vs. low)



Experimental tasks



T1 – requirements prioritization (30 min.)

- Customer’s perspective
 1. Assign 16 mobile phone features to 3 piles:
very important, important, less important
 2. Rank the features within piles

T2 – release planning (60 min.)

- Developer’s perspective
 1. Distribute 1000 story points to each feature as an estimate of implementation costs
 2. Plan 3 releases based on priorities (T1) and cost estimates



Experimental design*

	Original experiment (high proficiency)		Replicated experiment (low proficiency)	
	MT	EN	MT	EN
Run 1	Gr1, Gr3 execute T1	Gr2, Gr4 execute T1	Gr6, Gr8 execute T1	Gr5, Gr7 execute T1
Run 2	Gr2, Gr4 execute T2	Gr1, Gr3 execute T2	Gr5, Gr7 execute T2	Gr6, Gr8 execute T2

Data sources:

- post-task questionnaires
- meeting logs

* We have doubled the # of groups to 16, new data still to be analyzed



eConferenceMT

<http://code.google.com/p/econference-mt-plugin>

eConference

File Workbench Chat Options Help

MTConference IM

Agenda

Stop conference

Task 1 (prioritization)
Task 2 (planning)

Current Item: Task 2 (pl)

Who's on

- Fabio
- Filippo
- Rafael

*Task 2 (planning)

Rafael

não é tão complicado

Click on Top-Right Button to close

Filippo> Quindi per PHONEBOOK
Rafael> non è così complicato
Fabio> ok, quindi mettiamo quanto? 50?
Rafael> va bene

*Decisions place

Task1

VERY IMPORTANT

- SMS
- Phonebook
- Calendar
- Wi-Fi

RELEASE 2

- Alarm (50 pts)
- Notes (34 pts)
- Expandable internal memory (44 pts)
- Calculator (17 pts)
- Ring tone customization (34 pts)
- Bluetooth (40 pts)

RELEASE 3

- Timer (34 pts)
- MMS (67 pts)
- Games (117 pts)

Send

Hand raising

Connected as econferencetest1@gmail.com



Findings

- *RQ1 - Can MT services be used in distributed multilingual requirements meetings?*
 - Machine translation is not disruptive of the conversation flow and is accepted with favor
- *RQ2 - How does the adoption of MT affect group interaction?*
 - More balanced discussions when using native language with MT
- *RQ3 - Do individuals with a low English proficiency level benefit more than individuals with a high level from MT?*
 - So far **NO**, although people with low English skills are more prone to use MT again



However...

- Messaging is easier than talking for a non-native English speaker
- Therefore, as future work we should:
 - Replicate with voice conferences
 - Compare with groups including native English speakers
 - Replicate with more distant languages couples (e.g., Chinese, Russian)



Research Partner:

Filippo Lanubile (Univ. of Bari)

ESTABLISHING PERSONAL TRUST-BASED CONNECTIONS IN DISTRIBUTED TEAMS

Challenges of distributed software development



	Geographical Distance	Temporal Distance	Sociocultural Distance
Communication	Decreased frequency of communication Lack of informal exchange Providing right technical infrastructure	Delay in responses	Language differences and misunderstandings
Coordination	Lack of awareness Reduced trust	Reduced hours for same time collaboration	Doubtful of others' capabilities
Control	Push for heavy-weight processes	Management of project artifacts subject to delays	Jobs perceived as under threat Different perceptions of authority



Group awareness

*“An understanding of the activities of others which provides a context for your own activity” **

- Informal awareness
- Group-structural awareness
- Workspace awareness
- Social awareness

How to increase awareness & build trust in distributed teams?



Meeting Room



Remote Conferencing



Informal Communication



Social Networks



How can social software mitigate the negative effects of distance in globally distributed development?



- Current approach: Have all team members use a single, project-oriented, social networking site
- Problems
 - Project-oriented networking sites do not capture the full social identity of a software engineer
 - On the other hand, nobody would like to get project news feeds into the personal timelines of Facebook or Twitter
 - In large global teams, you might do not know people you should be aware of (awareness network)
 - The awareness network can be very dynamic*

* de Souza, C.R.B., Redmiles, D.F. *The Awareness Network, To Whom Should I Display My Actions? And, Whose Actions Should I Monitor?*, IEEE Trans. on Sw Eng, 37(3), 2011.

Is social awareness needed?



- Previous research

- J. M. Costa, M. Cataldo, C. R.B. de Souza. The scale and evolution of coordination needs in large-scale distributed projects: implications for the future generation of collaborative tools. CHI '11, 2011.
- Bradner, E. , Mark, G. Why distance matters: effects on cooperation, persuasion and deception. CSCW '02, New Orleans, USA, 2002
- Shami, N.S. et al. Making sense of strangers' expertise from signals in digital artifacts. CHI '09, Boston, USA
- DiMicco, J. et al. Motivations for Social Networking at Work. CSCW'08, San Diego, CA, 2008



Theoretical Model

Cognitive Trust

PREDICTABILITY

The degree to which a person is liable and accountable and meets the expectation of another person

ABILITY

Capability of a trustee (based on knowledge, competence, skills) to perform tasks within a specific domain

Trustee's antecedents to trust

Affective Trust

INTEGRITY

The adherence to intrinsic moral norms which makes a trustee reliable

BENEVOLENCE

The perceived level of courtesy and positive attitude

Trustor's antecedent to trust

PROPENSITY TO TRUST

A general, not experience-based inclination to display faith and adopt a trusting attitude toward others

So, is social awareness needed???



YES!

*We are in the **social era** of software engineering where **social media ecosystems** are an integral component of software repositories, and vice versa*

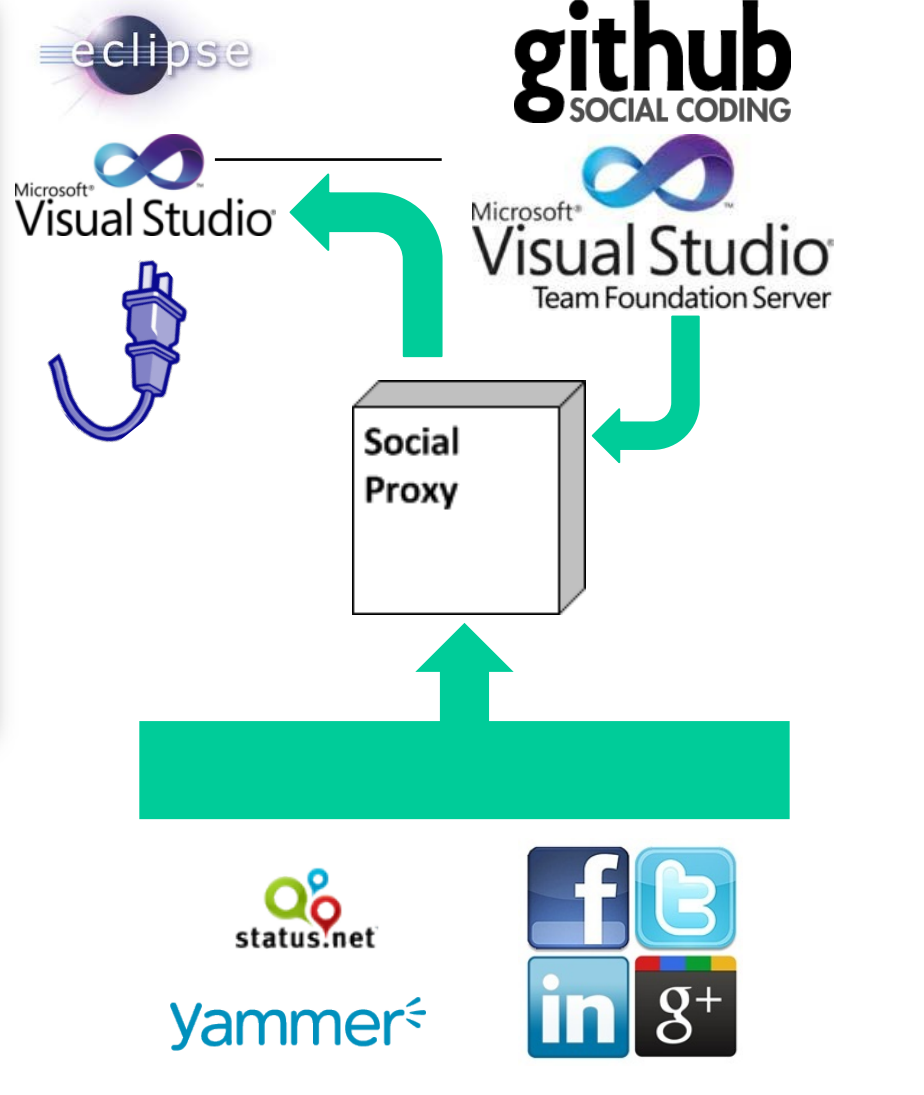
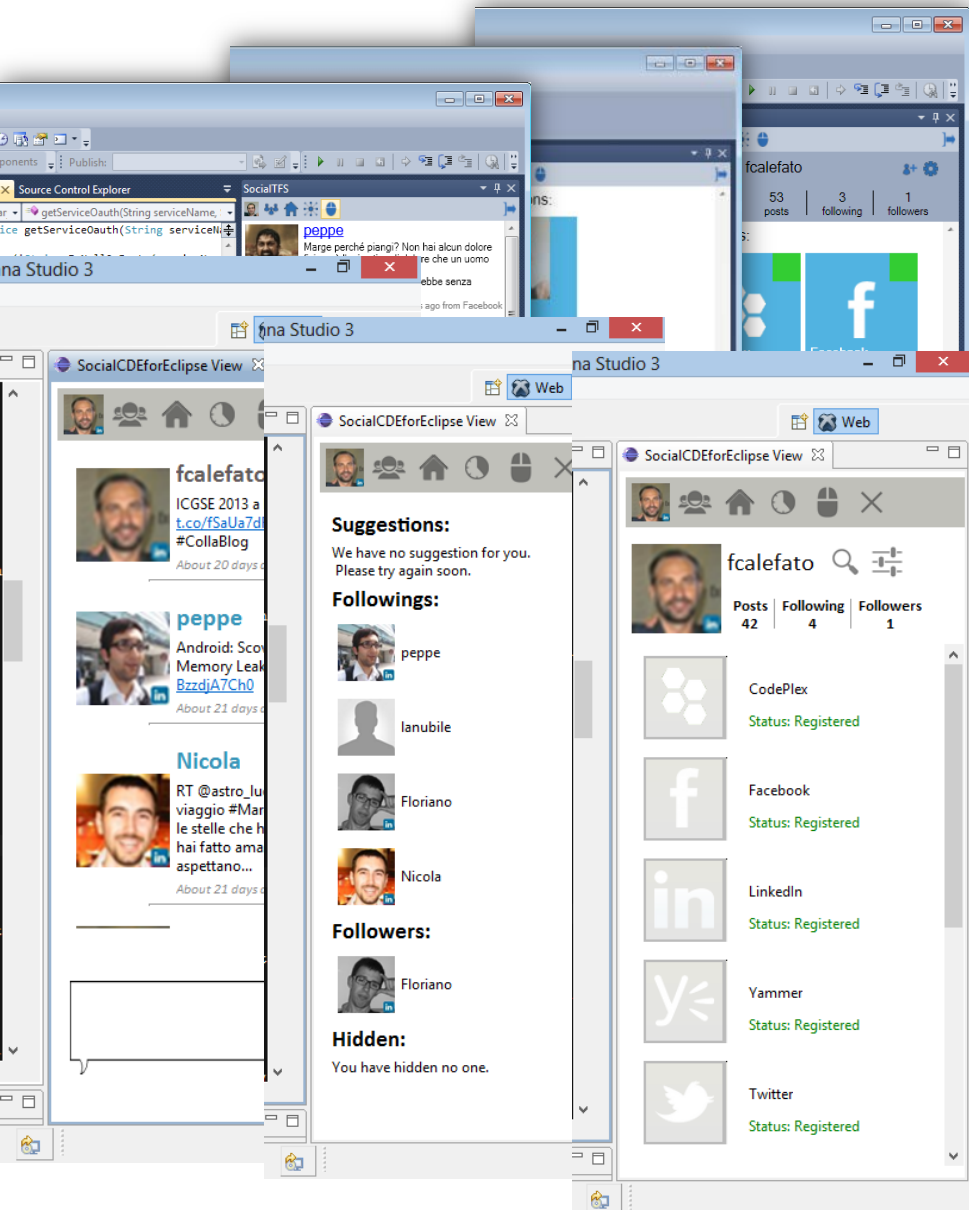
– M.A. Storey, MSR'12 Keynote



Our approach

- Information shared on social media can surrogate the social awareness on which affective trust grows
 - Developers' personal content from social media aggregated and made it contextually available into the shared workspace
- Great impact of social software on collaborative development environments
 - How to provide social awareness?
- Challenge for empirical studies
 - How can we measure the actual benefit of being personally connected to other developers, wherever they are located?

The SocialCDE Project



<http://socialtfs.codeplex.com>
<https://github.com/collab-uniba/socialcde4eclipse>



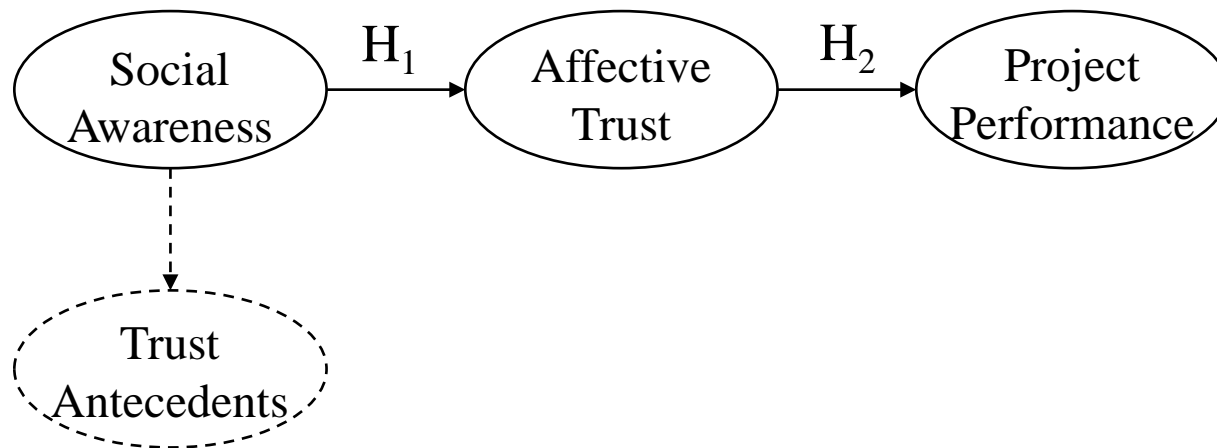
Group awareness in Collaborative Development Environments

	Informal awareness	Group-structural awareness	Workspace awareness	Social awareness
Trac	—	YES	YES	—
Google Code	—	YES	YES	—
Assembla	—	YES	YES	—
Jazz	YES	YES	YES	partial (IBM Connections)
TFS	YES	YES	YES	—
CodePlex	—	YES	YES	—
GitHub	—	YES	YES	partial (view developers' connections and @mentions)



Research Model & Hypotheses

- H_1 - *There is a positive relationship between the amount of social awareness gained through social media and the level of affective trust mutually established among distant teams*
- H_2 - *There is a positive relationship between the level of affective trust mutually established among distant teams and project performance*





CONCLUDING REMARKS



Conclusions

Finding best fitting media for communication-intensive tasks

- Lean media can be effectively used instead of F2F for communication-intensive tasks
- Tasks mostly involving idea generation (divergent thinking)

Overcoming language barriers with machine translation

- State-of-the-art MT solutions still far from perfect
- Can be used to complete communication-intensive tasks with some delay due to mistranslations

Establishing personal, trust-based connections in distributed teams

- The actual benefit of being personally connected to other developers still under investigation

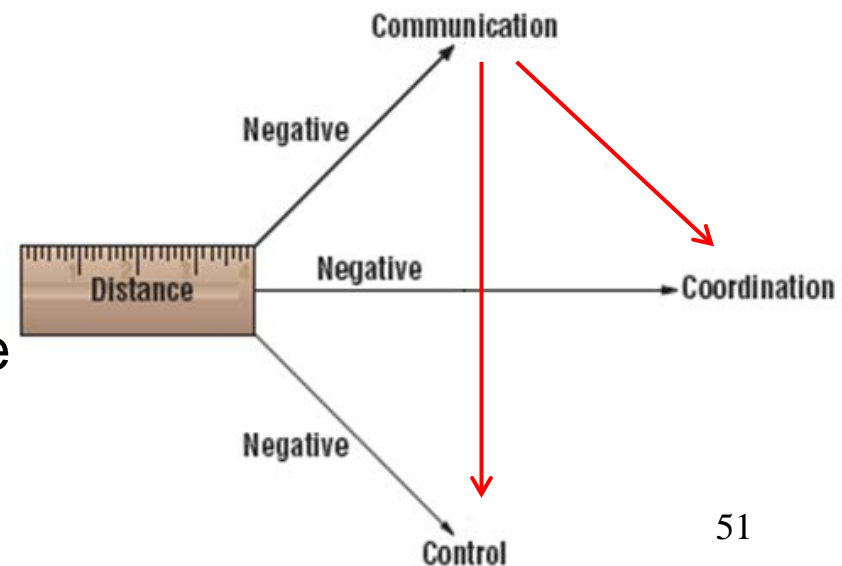


Facing communication issues takeaway

- Communication, coordination and control are all equally important to collaboration within (virtual) teams

YET

- Communication should be first among equals!
 - Avoiding communication disruption and breakdowns is paramount
 - Any effort put to resolve coordination and control issues is ineffective otherwise





Thanks for your attention!

Questions & comments

Contacts:

- fabio.calefato@uniba.it
- <http://cdg.di.uniba.it/calefato>
- https://www.researchgate.net/profile/Fabio_Calefato



References

[ESEJ12] F. Calefato, D. Damian, and F. Lanubile, "*Computer-Mediated Communication to Support Distributed Requirements Elicitations and Negotiations Tasks*", Empirical Software Engineering Journal, Vol. 17, No. 6, pp. 640-674, 2012.

[ESEM12] F. Calefato, F. Lanubile, T. Conte and R. Prikladnicki, "*Assessing the Impact of Real-Time Machine Translation on Requirements Meetings: A Replicated Experiment*", ESEM'12, Lund, Sweden, Sep. 19–20, 2012.

[IEEESW13] F. Lanubile, F. Calefato, and C. Ebert, "*Group Awareness in Global Software Engineering*", IEEE Software 2013.

[ESEC13] F. Calefato and F. Lanubile, "*SocialCDE: A Social Awareness Tool for Global Software Teams*", ESEC/FSE'13, St. Petersburg, Russia, 18-26 Aug. 2013.