

An Empirical Evaluation of an Activity-Based Infrastructure for Supporting Cooperation in Software Engineering

Paolo Tell

Software and Systems Section
IT University of Copenhagen, Denmark
Email: pate@itu.dk

Muhammad Ali Babar

Software and Systems Section
IT University of Copenhagen, Denmark
School of Computer Science
University of Adelaide, Australia
Email: maba@itu.dk

Abstract—[Background] Software Engineering (SE) is predominantly a team effort that needs close cooperation among several people who may be geographically distributed. It has been recognized that appropriate tool support is a prerequisite to improve cooperation within SE teams. In an effort to contribute to this line of research, we have designed and developed an infrastructure, called ABC4GSD, based on the models of Activity Theory (AT) and the principles of the Activity-Based Computing (ABC) paradigm. [Aim] In this paper, we present a study that empirically evaluates the ability of ABC4GSD in supporting teams cooperation. [Method] We designed and executed a study based on a scenario that simulated the Follow-The-Sun (FTS) strategy of Global SE (GSE). Our research design allowed us to ensure cooperation to be both computer-mediated as well as contained within observable short time-windows—the hand-off activities of the FTS strategy. [Results] Overall, the results show that the cooperation support provided by the ABC4GSD system has been positively perceived by the participants. Nonetheless, open issues stimulating further investigations have been raised especially due to a few mixed results. [Conclusions] Aware of the limitations of the simulated scenario, we conclude that the approach followed by the ABC4GSD system based on activities is desirable to improve the cooperation support in SE. Finally, our research approach based on simulating a scenario with geographical and temporal distribution can provide useful ideas for assessing collaborative technologies in SE.

I. INTRODUCTION

Software engineering projects of reasonable size often comprise several people cooperating towards a common purpose of creating a piece of software. To support cooperation within SE teams, it is important to both understand the contextual suitability of SE practices as well as design, develop, and evaluate appropriate tools (e.g., [26][8][17]).

Given the critical role of appropriate tools for supporting cooperation, different solutions have been proposed by both academia and industry [12][18][20]. However, there are still open research issues related to appropriate tool support for cooperative SE. To tackle this research challenge, we have empirically investigated the applicability of Activity Theory (AT) and Activity-Based Computing (ABC) to overcome known intrinsic limitations of currently available systems [10]. Our efforts have resulted in an innovative middleware infrastructure,

called ABC4GSD system¹, designed to be deployed alongside regular operative systems to enhance their support for cooperation. For our research, we define cooperation as a composition of collaboration, coordination, communication, and awareness. In fact, in cooperative SE, two or more practitioners are expected to collaborate to perform different activities and they are likely to have the need of coordinating their actions, which entails engaging in communication in the case where one part is not sufficiently aware of what is required to realize the objective (e.g., due to doubts, misunderstandings, misalignments) [24].

To determine if the ABC4GSD system supports cooperation among software developers, we devised and conducted an evaluation study to assess its ability to support hand-off activities of a team adopting the FTS strategy of GSE. Through a simulation of an FTS arrangement, we were able to assess ABC4GSD supports for cooperation in observable compressed time windows—the hand-offs occurring during the overlapping time of a team engaged in an FTS endeavor.

Three significant contributions of this work are:

- to provide evidence about the feasibility of using a system based on AT and ABC to support SE tasks;
- to empirically show that the ABC4GSD system can support cooperation in SE by improving collaboration, coordination, communication, and awareness within the team; and,
- to present the design and execution details of an evaluation strategy for testing a collaboration infrastructure based on an environment simulating the FTS strategy.

II. RESEARCH BACKGROUND

In this section, we describe the GSE paradigm and the FTS strategy as well as some of the approaches and tools based on the activity metaphor.

¹ABC4GSD: activity-based computing for global software development. The system is available for download as open source software at <https://github.com/crest-centre/ABC4GSD>

A. GSE and FTS

GSE [8], the geographical distribution of SE teams across multiple sites, is a widely accepted and practiced SE paradigm. While the GSE paradigm promises several benefits, it usually requires an intensified degree of collaboration, coordination, communication, and awareness among team members. According to Herbsleb [8], “the fundamental problem of GSD is that many of the mechanisms that function to coordinate the work in a co-located setting are absent or disrupted in a distributed project.” Geographical, temporal, cultural, and linguistic distances contribute, in a complex interdependency, which makes cooperation in GSE settings more troublesome [14][5].

A special case or strategy of GSE is known as Follow-The-Sun (FTS) or round-the-clock development whose goal is to reduce time to market by leveraging the world rotation. In an FTS arrangement, software development tasks can be theoretically performed 24/7 by teams that are located at various geographical locations, which can have little to no overlapping time zones. To achieve this, work is handed over by one team to the time-adjacent one through a hand-off process. Carmel et al. [3] refer to this process as *passing the baton*. The uniqueness of an FTS setting compared to other GSE arrangements lies in this delicate phase; in fact, failing a hand-off can escalate to the loss of an entire working day (vulnerability cost [3]). In an FTS arrangement where time zone overlap is present, practitioners are usually required to dedicate time at the beginning and/or end of their work shift to smoothly and successfully complete hand-offs. In these short time windows, heavy cooperation is necessary and appropriate tool support needs to be in place to facilitate it. During the hand-offs, practitioners are likely to require appropriate tool support in terms of: (i) being able to switch work context from one task to another to accommodate requests expressed by the other team; (ii) enhanced overall awareness; (iii) shared workspace; and, (iv) communication.

B. Approaches Based on Activities

Due to the limitations of the current desktop metaphor that was developed in the ‘70s, industry and academia are exploring different solutions [10]. Solutions based on the *activity* metaphor have recently been attracting significant interest. Examples include Gnome Shell in which the concept of workspace is replaced by the one of activity bundling up digital artifacts and applications connected to them. Yarosh et al. from IBM [27] defined their work on Lotus Activities as Activity-Centric Computing (ACC): an AT loosely inspired approach designed to “[...] address work fragmentation by allowing users to structure their work around the computational construct of an Activity”. Recently another activity centric system called co-Activity Manager was evaluated with knowledge workers [9]; such solution is designed to enhance the Windows operative system by supporting the aggregation of human and digital resources around the concept of activity. Other notable contributions for tool support based on the AT include the desktop manager called Giornata [25], and the Context-Aware Activity Display (CAAD) [19]. In these

examples, the approach followed is the one initially introduced by Norman in the ‘90s called Activity-Based Computing (ABC). In [15], Norman states that “[...] the basic idea is simple; make it possible to have all the material needed for an activity ready at hand, available with little or no mental overhead”. Thus, the core concept of ABC is to provide an automatic, seamless, and non-intrusive support for activities. Bardram successfully demonstrated the applicability of the ABC paradigm for supporting different collaborative activities in hospitals [2]. He developed a framework that provides a replacement for the application-oriented computing paradigm. As shown in [22], the way in which AT is applied in SE differs from the one used for supporting physicians in hospitals; but the models of AT and the core principles of ABC [2] appear to be able to provide a solid foundation for building an infrastructure that can help address many of the existing GSE related challenges by: aggregating human and digital resources around the concept of activity and providing a new interaction mechanism able to facilitate the handling of interruptions.

III. OUR SOLUTION

We have developed a middleware infrastructure based on the ABC paradigm grounded in AT that is able to sit along side regular operative systems to support cooperation by enhancing their capabilities with regards to: collaboration, coordination, communication, and awareness. The system has been designed based on high level requirements identified in [21]; and theoretical foundations detailed in [22]. A preliminary evaluation of the ABC4GSD system was reported in [23] through which we gathered useful feedback regarding the user interface and the overall usability of the system. The findings from our preliminary evaluation study directed our efforts to improve the user interface mechanisms for supporting the four aspects of the hand-offs previously highlighted as critical (i.e., support for switches in the work context, enhanced overall awareness, provision of a shared workspace, and improved communication). An overview of the user interface can be seen in Fig.1. In the following, we briefly describe the improved version of the ABC4GSD user interface.

Work Context Switching. By supporting the organization of work around the concept of activity, the ABC4GSD system is aware of which resources are required for performing a task; and, once an activity is resumed, the system is able to present the user with the status in which the activity was last suspended (i.e., running applications, opened artifacts, position of windows). After one of the activities is selected by a double-click from either the hierarchical or graph view (Fig.1.(1,4)), the selected activity is resumed after suspending the active one. The suspension procedure is performed by sending an event to all running applications, which are integrated in the system (i.e., implement a specific interface), to let them store the information needed for future resumptions before closing; whereas, in the case of applications shallowly integrated, a simple termination signal is sent. The resumption procedure is orchestrated by the ABC4GSD system, and it entails the execution and initialization of all the applications associated with

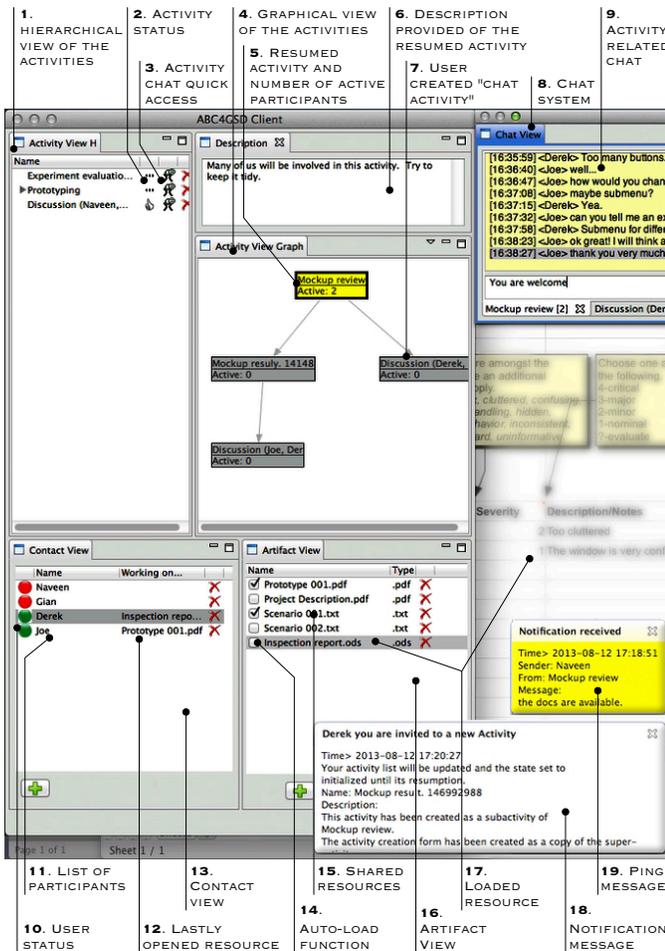


Figure 1. Overview of the ABC4GSD system user interface. The screenshot has been taken from one of the screencast of the evaluation; the notification and ping messages have been taken from other frames and added to show all elements of the ABC4GSD system UI.

the activity. The initialization of the applications implementing the interfaces is obtained by calling a specialized method, which fetches all data needed for the application to be resumed (e.g., position and size) and subscribes to all desired events. Thus, during the resumption all components contributing to the user interface are initialized and populated with consistent data: the Activity Graph View recreates the activity tree using the selected activity as root (Fig.1.(5)); the Contact List component updates the list of team members participating in the activity (Fig.1.(11)); the Artifact List component updates the list of digital resources (Fig.1.(15)); and, the chat system loads the corresponding chat room and retrieves all previous conversations (Fig.1.(8)).

Awareness. The client side of the system is entirely based on an event system orchestrated by the underlying infrastructure, which potentially allows designers to customize the way information is presented to users by developing both extensions to existing components of the user interface (e.g., the integration of an IM like Skype) as well as brand new components (e.g., a new chat system). This version of the ABC4GSD system includes diverse customizations aimed at improving the over-

all awareness: status information of activities (Fig.1.(2)) and activity members (Fig.1.(10)); information about lastly opened artifact of activity members (Fig.1.(12)); automatic notification messages related to major events, e.g., the inclusion of the user to an activity (Fig.1.(18)); permanent notifications, which are sent by users to capture others attention (Fig.1.(19)); on screen information about the number of people active on an activity (Fig.1.(5)).

Shared Workspace. The underpinning theoretical models of AT are used to aggregate together both human and digital resources. Currently the type of artifacts supported by the ABC4GSD system are files and URLs pointing to version control repositories. In the case of regular files, they are automatically uploaded to a git repository and shared among all the members of an activity once linked to it by, for instance, dragging them in the Artifact View (Fig.1.(15)). Therefore, the ABC4GSD system provides a shared workspace for each activity—the Artifact View—through which members of an activity can share digital content in an easy and intuitive way. Drag&Drop capabilities are provided as well as more advanced ones for further customizations.

Communication. Similarly to the shared workspace, each activity is equipped with a dedicated chat room; members of an activity can exchange text messages, which are stored to allow both asynchronous communication and future revisitations. Once an activity is created, a chat room accessible to all members of the activity is automatically linked to it. To quickly communicate with specific colleagues, users have the possibility to create ad-hoc chat rooms by using the chat function accessible from the contextual menu in the Contact View (Fig.1.(13)). This process results in the creation of a sub-activity of the current one to support dedicated conversations, which eventually means that each chat is an activity that after initiation can evolve from simple “chat activity” to an activity comprising artifacts and so on. It is worth noting that chats are kept open on different tabs to ease participation even when engaged in other activities, and their removal from the UI must be explicitly performed by the users.

IV. RESEARCH OBJECTIVE

After performing a preliminary user evaluation [23] in which the ABC4GSD system was assessed by one participant at a time, we designed a second evaluation to investigate the ability of ABC4GSD to support cooperation within a team of software engineers who are geographically distributed. We were fully aware that an empirical evaluation of benefits and drawbacks of a collaborative technology can be a quite challenging undertaking [26]. However, gaining inspiration from our efforts and lessons from successfully completing the preliminary evaluation study [23], we decided to design and execute this evaluation by leveraging some parts of the research approach and logistical apparatus used in our preliminary evaluation study. We designed a user evaluation driven by scenarios as described in [4] investigating the perceived usefulness, perceived ease of use, and self reported future use in relation to an environment augmented by ABC4GSD. The

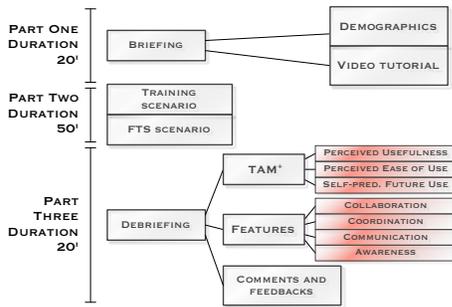


Figure 2. Evaluation breakdown.

research framework chosen to address these aspects has been the technology acceptance model (TAM) [11]. Additionally, four sets of questions addressing collaboration, coordination, communication, and awareness were designed. Finally, the opportunity was also leveraged to gather feedback and opinions about ABC4GSD. The key underlying research questions (RQ) that were addressed by our study are:

- **RQ1:** *How is the ABC4GSD system perceived by the users?* Do users perceive ABC4GSD as useful for supporting the simulation of the FTS strategy? Do users perceive ABC4GSD as easy to use, difficult to use, or are they indifferent? Do users report willingness to use our ABC4GSD system and/or an activity-based approach for supporting their future work?
- **RQ2:** *Is the ABC4GSD system able to support cooperation among SE team members, hence, collaboration, coordination, communication, and awareness?* Do users perceive the ABC4GSD system as able to improve the support in terms of collaboration, coordination, communication, and awareness among team members?

V. EVALUATION DESIGN AND EXECUTION

In this section, we provide details about different elements of the empirical evaluation study reported in this paper. We start by describing the process followed and the research instrument used for data collection. Afterwards, we present the details about the evaluation setup, we detail and motivate the simulated scenario, and conclude presenting the participants of the study.

A. Procedure

Given the rather novel interaction based on the suspension and resumption of activities introduced by the ABC4GSD system, we opted to expose study participants to the system through a scenario-based approach. This method entails the design of scenarios based on realistic settings exposing participants to complex situations, which would otherwise be hard to observe [4]. Fig.2, provides an overview of the procedure followed during this study. The study was organized into three sessions as follows.

In the first session, after signing an informed consent form and filling out a basic demographic sheet (Table I), participants were introduced to the ABC4GSD system through a 10 minutes video tutorial. In particular, they were introduced to the key concepts and functionalities needed to perform the main FTS scenario, e.g., how to create, edit, and delete an

Demographics				
D.1	Age			
D.2	Job title			
D.3	Gender	M	F	
D.4	Most recent education achievem.	M.Sc.	Ph.D.	...
D.5	Computer literacy	Novice	Avg	Exp
D.6	OsX familiarity	None	Avg	Adv
D.7	Exp. in working as part of a team	<1y	1-3y	>3y
D.8	Exp. in working with collab tools	<1y	1-3y	>3y
D.9	Exp. in software engineering	<1y	1-3y	>3y

Table I
DEMOGRAPHIC INFORMATION.

Comments and Feedbacks	
CaF.1	What did you like about ABC4GSD?
CaF.2	What did you dislike about ABC4GSD?
CaF.3	What would you like to see improved in a future version of ABC4GSD?
CaF.4	What would you like to see included in a future version of ABC4GSD?
CaF.5	What was intuitive about the system?
CaF.6	What was confusing about the system?
CaF.7	Please make any additional comments below.

Table II
OPEN-ENDED QUESTIONNAIRE.

activity, how to link human and digital assets to an activity, how to resume and suspend an activity, how members of each activity are organized, and how they are put in communication through the chat system and the ping functionality.

In the second part, participants were asked to perform two scenarios, and they were informed about the beginning of the evaluation by making them aware that the screencast recording would have been started. The first scenario, *Training scenario*, comprised a list of fine-grained tasks to execute sequentially, and was meant to provide participants with some hands-on training on ABC4GSD to improve their confidence. During this task, participants were allowed to ask any question to the observer that was present in the room with them. The second scenario, *FTS scenario*, represents the main task and will be extensively detailed below; during the execution of this scenario participants were asked not to query the observer. Each scenario was executed by administering to the participants a description of their task that would have introduced them to the simulated context and provided all the information required to execute the task independently.

In the third part, participants were administered the questionnaire detailed below (Table III), as well as a final questionnaire including open-ended questions aimed at gathering feedback and comments on the system (Table II).

B. Research Instrument

The Technology Acceptance Model (TAM) [6] has been selected as primary instrument for the data collection. The TAM aims at assessing user beliefs about the usefulness and ease of use of a technology by means of a questionnaire focused on two variables: perceived usefulness and perceived ease of use. According to Davis [6], perceived usefulness is defined as “the degree to which a person believes that using a particular system would enhance his or her job performance”; whereas, perceived ease of use as “the degree to which a person believes that using a particular system would be free of effort”. The model that was used in this study is the extended version of the TAM that was proposed in [11] and has been adopted by previous studies such as [1]. According to [11], “since both usefulness and ease of use are correlated to self-

predicted future usage, they can be considered determinants of tool acceptance behaviors”. Due to the novelty of the AT concepts and the ABC interactions (i.e., suspension/resumption), we opted for testing through the TAM instrument the perception participants have of them. For each of the three variables the related predefined questionnaires were adapted to focus on ABC4GSD. Moreover, in line with the overall instrument, four sets of questions have been added to assess the perceived support to cooperation, hence, collaboration, coordination, communication, and awareness features.

The questions related to each of the investigated variables are reported in Table III, and below we describe the rationale behind the design of the questions related to the last four variables.

Collaboration. As previously argued, to collaborate on a common objective, two (or more) practitioners need to coordinate their actions. This would possibly entail engaging in communication in the case where one part is not sufficiently aware of what is required to realize the objective. This interdependence of the cooperation dimensions makes the isolation of the collaboration one challenging. For this reason, we opted for focusing on the shared workspace used to realize the common motive (in line with the technological lens used by Steinmacher et al. [20] to identify the collaboration feature of GSE tools²), hence, the *process* facilitated by the system to work on the shared workspace (Table III.(Coll.1)), the ability of the system in handling the *digital resources* populating the shared workspace (Table III.(Coll.2)), and the effectiveness of the system in facilitating the understanding of the *common goal* (Table III.(Coll.3)).

Coordination. To support the dimension of coordination, a system needs to support a computational construct that allows the aggregation of resources in a rational, common, and intuitive manner (Table III.(Coor.1)). Moreover, we deemed appropriate to apply Mintzberg’s work on coordination mechanisms within organizations [13]. Even though not designed to address SE environments, we believe that the framework provides a fine-grained explanation of what coordination is, hence, a better understanding of the diverse mechanisms that occur between people cooperating and that needs to be supported. Mintzberg’s coordination model comprises six basic mechanisms: *mutual adjustment* (Table III.(Coor.2)), *direct supervision* (Table III.(Coor.3)), and four degrees of standardization (i.e., *standardization of work processes* (Table III.(Coor.4)), of outputs, of skills, and of norms)³.

Communication. To understand how the communication features were perceived, we opted for questioning the main

²Refer to [24] for a discussion about the different terminology used.

³Each one of these mechanisms contribute in different ways to the way practitioners can coordinate in an organization; however, we did not consider the last three. In fact: (i) given that the ABC4GSD system supports coordination from a higher abstraction the standardization of outputs can be imposed independently from the infrastructure (e.g., by linking an entire code repository instead of single files); (ii) the standardization of skills should be handled by the management and focus of awareness and knowledge management in relation to expertise and experience is not the focus of this work yet; and, (iii) the standardization of norms is environmentally supported rather than technologically [22].

Perceived Usefulness	
PUsf.1	Using ABC4GSD in my job would enable me to accomplish tasks more quickly.
PUsf.2	Using ABC4GSD would improve my job performance.
PUsf.3	Using ABC4GSD in my job would increase my productivity.
PUsf.4	Using ABC4GSD would enhance my effectiveness on the job.
PUsf.5	Using ABC4GSD would make it easier to do my job.
PUsf.6	I would find ABC4GSD useful in my job.
Perceived Ease of Use	
PEoU.1	Learning to operate ABC4GSD would be easy for me.
PEoU.2	I would find it easy to get ABC4GSD to do what I want it to do.
PEoU.3	My interaction with ABC4GSD would be clear and understandable.
PEoU.4	I would find ABC4GSD to be flexible to interact with.
PEoU.5	It would be easy for me to become skillful at using ABC4GSD.
PEoU.6	I would find ABC4GSD easy to use.
Self-Predicted Future Use	
SPFU.1	Assuming ABC4GSD would be available on my job, I predict that I will use it on a regular basis in the future.
SPFU.2	I would prefer using ABC4GSD to one dependent on multiple applications for organizing my tasks within a collaborative environment.
Collaboration	
Coll.1	While completing the tasks, I did not feel enforced by ABC4GSD in any pre-specified manner.
Coll.2	Sharing a document with a colleague through ABC4GSD is easy.
Coll.3	Using ABC4GSD makes the understanding of common outcomes within a team easy.
Coordination	
Coor.1	ABC4GSD allows me to organize work and consolidate information around the concept of activity.
Coor.2	Being aware of events happening that pertain my work through ABC4GSD is simple.
Coor.3	Delegating a task to a team through ABC4GSD is easy.
Coor.4	Having a common way to organize work through ABC4GSD is useful.
Communication	
Com.1	ABC4GSD allows me to communicate with my colleagues both on work related issues and more informal topics.
Com.2	Being able to leave messages via the ABC4GSD system to colleagues even when offline is valuable.
Com.3	Having a text based communication facility is valuable.
Com.4	I would prefer having also an audio or video based communication facility.
Awareness	
Aw.1	When using ABC4GSD, I am aware of who is doing what.
Aw.2	When using ABC4GSD, I am aware of who is available.
Aw.3	When using ABC4GSD, I am able to understand who is responsible for what.
Aw.4	When using ABC4GSD, I am able to be aware of the people I am working with in each task and their status.

Table III
QUESTIONNAIRES.

communication channel capabilities. Hence, we analyzed the chat system in terms of *formality* (Table III.(Com.1)), degree of *synchronicity* (Table III.(Com.2)), and communication *mechanism* (Table III.(Com.3,4)). In particular, with the last question, we wanted to investigate whether text based communication is sufficient for the tasks simulated or if practitioners perceive the need for richer communication media like voice or video based facilities.

Awareness. An extensive literature is available with regards to awareness and its importance. To structure the evaluation of the awareness features we have used the classification by Gutwin et al. [7], which was further extended by Omoronyia et al. [16], comprising: *workspace awareness* (“the up-to-the-minute knowledge of other participants interactions with the shared workspace” [7]) (Table III.(Aw.1)); *informal awareness* (“the general sense of who is around and what they are up to” [7]) (Table III.(Aw.2)); *group-structural awareness* (“knowledge about such things as peoples roles and responsibilities, their positions on an issue, their status, and group processes” [7]) (Table III.(Aw.3)); *social awareness* (“information that a person maintains about others in a social or conversational

context” [7]) (not supported); and, *context awareness* (“the evolving internal and external state information that fully characterizes the situation of each entity in a shared environment” [16]) (Table III.(Aw.4)).

C. Data Collection

Similarly to what has been used in [6] and [11], each question was measured with a seven-point Likert scale⁴, allowing to capture positive, negative, and neutral evaluations. Self-reported quantitative data obtained through questionnaires was the main data collected during the evaluation. Qualitative data was also collected to better interpret the results. In particular, through: (i) the open-ended questions administered during the debriefing; (ii) notes taken by the observers; (iii) screencasts (including audio, but no video) of the sessions; (iv) logs automatically stored by the system on users’ interactions; and, (v) a follow up questionnaire sent via email to the participants that gave their consent on being further contacted.

D. Setup

During the entire evaluation two observers were situated in two distant rooms, each with one of the two participants as per `In-situ` arrangement detailed in Fig.3. Participants were informed only about the `Virtual` arrangement (Fig.3) and no information was shared about the real arrangement. Two teams, Bangalore and London (in red in Fig.3), were impersonated by participants of the study, while the Melbourne and New York teams (in blue in Fig.3) were enacted by confederates (i.e., the two observers) to initiate and finish the entire simulation. According to Convertino et al. [4], a confederate is a member of the evaluation team trained to encourage participants to perform certain activities from within the simulation; participants are unaware of the fact that such participants are part of the evaluation and perceive them as regular participants. From hereafter, we will refer to confederate only to emphasize their role from the simulated environment perspective; therefore, even if the same person, the *observer* would take notes, while the *confederate* would send a message to a participant.

Each confederate was given a precise script to follow to consistently enact the remote colleague role throughout the execution of the evaluation. For example, in the case that no communication would have happened from Naveen (participant), Gian (confederate) would have explicitly invited Naveen through the chat system to engage in a discussion about the requirements. Moreover, the scripts given to the confederates included exact processes describing how to interact with the participants, e.g., Gian possessed a hard copy of the correct priorities and had to respond according to a mechanism designed to stimulate synchronous interaction while at the same time promoting a fast conflict resolution.

To permit the simulation of an entire work shift for all teams, time zones were reproduced through scaling; in particular, 1 virtual hour was compressed in 3 real minutes. Therefore, each evaluation session lasted 42 minutes (9 working hours in

Bangalore, 9 working hours in London, and 4 hours overlap for a total of 14 hours). Time alignment was crucial; for this reason observers were communicating with each other about the main events happening at each side. For instance, an sms on the personal phone would have been sent by the observer in room two to inform that Naveen started the 10 minutes video tutorial after which Derek had to be welcomed by the observer in room one.

The study setup comprised a laptop equipped with mouse and wired to the university local network for each participant. The participants were given machines with MacOS X on which the client side of the ABC4GSD system was deployed. Whereas, the server side of the infrastructure was hosted within the IT University of Copenhagen in which the evaluation study was conducted.

E. FTS Scenario

Given the difficult task of assessing pros and cons of collaboration technologies [26] and the unique nature of ABC4GSD, our understanding is that a system like ABC4GSD can be evaluated based on the following assertions: (i) it is appropriate to simulate GSE scenarios as it allows to guarantee the absence of physical presence, which entails that all interactions need to be computer mediated; (ii) it is appropriate to simulate FTS scenarios as it stimulates participants to interact in a short time window—the hand-offs during the overlapping time of the FTS teams. In these time slots, collaboration technologies can be put to test in a shorter, simpler, more manageable, and more controlled situation for what concerns supporting the dimensions of cooperation, i.e., collaboration, coordination, communication, and awareness.

Fig.3 presents a detailed overview of the evaluation setup, which was designed to simulate a distributed setting in which four teams virtually located in Melbourne, Bangalore, London, and New York had to cooperate on a project using an FTS arrangement. A description of the main activities comprising the simulation follows.

Init. The ABC4GSD system was initialized with four activities aimed at supporting the tasks that the teams had to perform. In particular, the `Requirement prioritization` activity was mainly used by Naveen (Bangalore team) to prioritize requirements; the `Mockup review` activity was chiefly used by Derek (London team) to perform the usability review on a mock-up; and, the `Experiment evaluation Bangalore` and the `Experiment evaluation London` ones contained the questionnaire reported in Table III.(b). Each activity was linked to the required human and digital resources, e.g., the requirement document containing the tentative prioritization performed by Gian (Melbourne team) was linked to the activity `Requirement prioritization`.

Fig.3 A (Naveen). Naveen was instructed to critically assess the priorities chosen by Gian (Melbourne team) and discuss changes with him in order to collaboratively find agreement on the 6 requirements given.

Fig.3 A (Gian). If not contacted by Naveen, Gian was instructed to contact him asking about the status of the prioritization after five minutes from Naveen’s log-in. Such

⁴Likert scale parameters: (1) extremely likely/agree, (2) quite likely/agree, (3) slightly likely/agree, (4) neither, (5) slightly unlikely/disagree, (6) quite unlikely/disagree, (7) extremely unlikely/disagree.

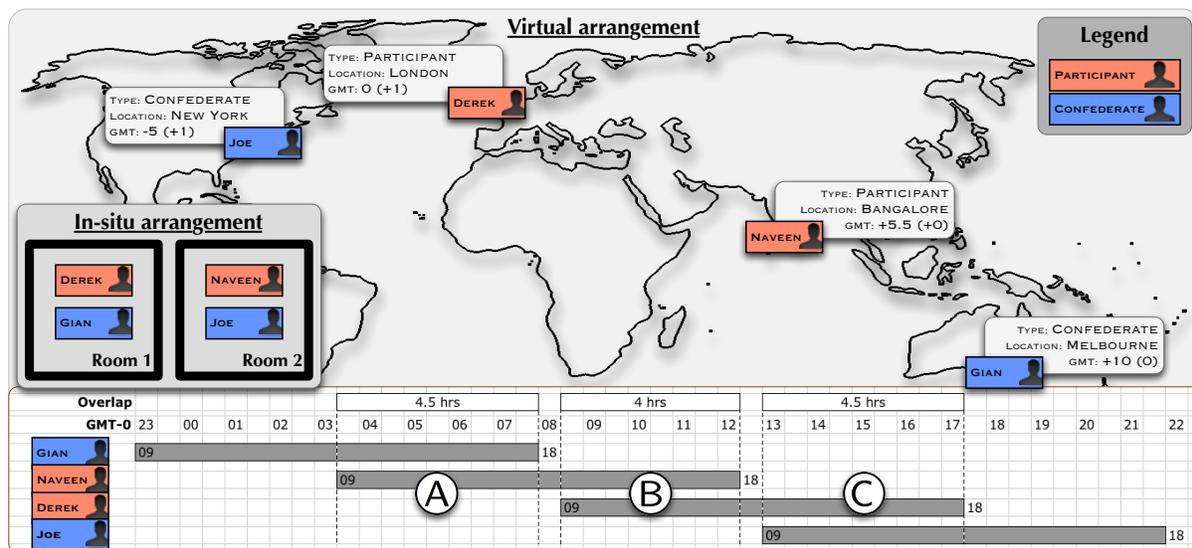


Figure 3. Overview of the FTS scenario. The arrangement included four teams distributed across the world; time distances and overlapping time-frames are highlighted. It can be seen that participants (red cards) and confederates (blue cards) were located in two separate rooms according to the schematic overview in the “In-situ arrangement” box. This setting has been kept throughout the 15 runs of the evaluation.

contact was done to force the initiation of the synchronous discussion over the requirements priorities.

Fig.3 ② (Naveen). Once the prioritization was completed, Naveen had to prepare a text document describing a scenario to be shared in the *Mockup review* activity. One scenario was already linked to the activity to be used both as a reference for Naveen and for the inspection review by Derek (described later). Responsibility of Naveen was also to make sure that Derek had become aware of the material shared for the review process. For the remaining time of the work shift, Naveen was instructed to answer the open ended questionnaire contained in the activity *Experiment evaluation Bangalore*, while remaining available for any further request that might be raised by Derek (simulating the “other admin/work tasks” described in [3]).

Fig.3 ③ (Derek). Derek was instructed to perform the review on the sample scenario present in the *Mockup review* activity against a pdf document containing a screenshot of the user interface which was linked to the activity, and archive the results using the provided templates (i.e., the spreadsheet document visible in Fig.1.(17)).

Fig.3 ④ (Derek). Derek had to create an activity called *Review results* (sub-activity of *Mockup review*) for presenting all the results from the usability review session also linking himself and Joe (New York team) to it. Derek had to make sure that Joe had taken over the results of the review for the prototyping stage; and, while remaining available for any further request that might be raised by Joe for the remaining time of the work shift, he was instructed to answer the open ended questionnaire contained in the activity *Experiment evaluation London* again simulating the “other admin/work tasks” described in [3].

Fig.3 ⑤ (Joe). Joe was instructed to question issues discovered in the mockup as soon as the review document was shared in the *Mockup review* activity created by Derek.

		Low	Medium	High
Computer literacy	D.5	0.0 %	13.3 %	86.7 %
OsX familiarity	D.6	43.3 %	30.0 %	26.7 %
Exp. in w. as part of a team	D.7	6.7 %	43.3 %	50.0 %
Exp. in w. with collab tools	D.8	20.0 %	56.7 %	23.3 %
Exp. in software engineering	D.9	13.3 %	30.0 %	53.3 %

Table IV
PARTICIPANT EXPERIENCE.

F. Participants

For the recruitment of study participant, neither gender nor age distinction was applied. In total, 30 participants were recruited for this study (mean age 28.9) to impersonate Naveen and Derek (Fig.3). Participants included PhD students (17), PostDocs (2), professors (2), and practitioners (9) all engaged in computer science or SE related areas. Table IV, shows the reported experience level in relation to the different experience variables.

It can be seen that 80% of the participants had worked with collaborative software tools (D.8) for more than one year. Only 2 participants (out of 30) had less than one year experience in working as part of a team (D.7). The choice of opting for an OsX deployment to mitigate problematics related with the operative system familiarity turned out to be wrong as almost half of the participants had never or very sporadically used such operative system (D.6). However, this had no visible effect on the results of the evaluation, probably because of the very high confidence of the participants with regards to the technical background (D.5). Finally, the expertise in SE was also reasonably good (D.9). In fact, 4 participants out of 29 (1 did not answer this question) reported less than one year experience; however, all the participants held at least an M.Sc. and, in the specific case of those 4 participants, 3 are currently performing their Ph.D. and 1 is a developer, suggesting an underestimation of their skill-set.

VI. RESULTS

Table V provides an overview of the quantitative results from the questionnaires. A discussion of the results including

(PUSf)	μ	\bar{x}	σ
1	3.17	3.00	1.61
2	3.30	3.00	1.46
3	3.63	3.00	1.56
4	3.47	3.00	1.48
5	3.57	3.00	1.67
6	3.07	2.50	1.61
(PEoU)	μ	\bar{x}	σ
1	1.70	2.00	0.69
2	2.30	2.00	1.00
3	2.27	2.00	0.93
4	2.93	2.00	1.36
5	1.70	2.00	0.64
6	1.87	2.00	1.06
(SPFU)	μ	\bar{x}	σ
1	2.93	2.50	1.50
2	3.10	3.00	1.80

(Coll)	μ	\bar{x}	σ
1	3.28	3.00	1.68
2	1.40	1.00	0.49
3	2.67	2.00	0.91
(Coord)	μ	\bar{x}	σ
1	1.73	2.00	0.57
2	2.93	3.00	1.15
3	2.32	2.00	1.00
4	2.03	2.00	0.87
(Com)	μ	\bar{x}	σ
1	2.27	2.00	1.50
2	1.43	1.00	0.72
3	1.63	1.00	1.02
4	2.50	2.50	1.20
(Aw)	μ	\bar{x}	σ
1	3.23	3.00	1.56
2	1.90	2.00	0.91
3	3.43	3.00	1.73
4	1.87	2.00	0.76

Table V

RESULTS FOR: (PUSF) PERCEIVED USEFULNESS; (PEoU) PERCEIVED EASE OF USE; (SPFU) SELF-PREDICTED FUTURE USE; (COLL) COLLABORATION; (COORD) COORDINATION; (COM) COMMUNICATION; (AW) AWARENESS. SCORES ARE ON A 7-POINT LIKERT SCALE FROM 1 (EXTREMELY LIKELY/AGREE) TO 7 (EXTREMELY UNLIKELY/DISAGREE). NOTE: (μ) AVERAGE; (\bar{x}) MEAN; (σ) STANDARD DEVIATION.

insights from the open-ended questions administered in the debriefing session and the responses from the follow up questionnaire is provided in Section VII.

Perceived Usefulness. The ABC4GSD system received mixed results related to its perceived usefulness; on average the results are positive in favour of ABC4GSD. From the standard deviation it can be seen that responses rather than being concentrated around the average value, were either very positive or very negative. Apart from scoring slightly better with regards to perceiving the ABC4GSD system as potentially useful (Table V.(a).PUSf.6), participants agreed on disagreeing in all the questions related to this group. In fact, the perception related to the ability of the system to support the achievement of a task more quickly (Table V.(a).PUSf.1), in an easier way (Table V.(a).PUSf.5), the perception related to improved performance (Table V.(a).PUSf.2), productivity (Table V.(a).PUSf.3), and effectiveness (Table V.(a).PUSf.4) were all perceived positively by slightly more than half of the participants and negatively by the others. A more elaborated discussion will be presented in Section VII.

Perceived Ease of Use. All participants managed to appropriate the system with more or less confidence in a brief period of time. On the one hand, participants found the ABC4GSD easy to learn and use (Table V.(b).PEoU.1,5,6) and rewarded the intuitiveness of the ABC paradigm. On the other hand, they had slightly different responses with regards to the flexibility of the system (Table V.(b).PEoU.2,4) The novel interaction driven by the suspension and resumption of activities was unclear and difficult to understand for some participants (Table V.(b).PEoU.3); but on average all of them scored well.

Self-Predicted Future Use. Similar to the results for the perceived usefulness responses, by aggregating the results there is a slightly positive average response. However, it is clear also in the case of the self-predicted future use variable that the participants disagreed in all the questions related to this group. In fact, slightly more than half of the participants reported a

good likelihood of adopting ABC4GSD for supporting their future work (Table V.(c).SPFU.1), and expressed an interest in using a system based on activity over the currently available ones (Table V.(c).SPFU.2); whereas, the other participants were negative with regards to such questions. We will further discuss these results in Section VII.

Collaboration. Collaboration support was perceived well especially with what concerns the shared workspace and the ability of the ABC4GSD system to share digital resources in an easy way (Table V.(d).Coll.2); most of the participants also rated well the ability of the system to make the understanding of common outcomes clear within a team (Table V.(d).Coll.3). However, the participants felt somewhat forced with regards to the flexibility with which the ABC4GSD system supports users in achieving their goal (Table V.(d).Coll.1). This last point will be further discussed in Section VII.

Coordination. The theoretical models of AT proved to be useful (Table V.(e).Coord.4) and effective (Table V.(e).Coord.1) in aggregating digital and human resources around the concept of activity. The activity metaphor and the ability to hierarchically organize work, assign tasks, and manage assets was also perceived as well supported (Table V.(e).Coord.3), and participants were slightly positive about the ability to be aware of events related to their work happening while connected (Table V.(e).Coord.2).

Communication. Participants agreed about the fact that having the chat system is valuable (Table V.(f).Com.3), and the ability to leave messages to offline users (asynchronous communication) was also valued well (Table V.(f).Com.2). Most of the participants did not feel constrained to work related topics in their chats (Table V.(f).Com.1); interestingly, one participant raised concerns about having an ad-hoc communication system, which was correctly guessed to be persistent hence controllable by the management. Finally, most of the participants expressed the preference of having additional communication means like audio or video facilities (Table V.(f).Com.4).

Awareness. Awareness mechanisms are clearly supported differently. On the one hand, most of the participants felt confident about knowing who is available, also with regards to single activities, and their status (Table V.(g).Aw.2,4). On the other hand, the participants had conflicting judgments with regards to the support for knowing who is doing what (Table V.(g).Aw.1) and who is responsible for what (Table V.(g).Aw.3): workspace and group-structural awareness respectively. Reasons for this result will be further discussed in Section VII.

VII. DISCUSSION

The overall results from the evaluation study are positive about the ability of the ABC4GSD system to support different dimensions of SE as a large majority of the participants scored positively. However, for some of the questions, we received mixed responses. In the following, we will analyze such results by relating them to the responses obtained through the debriefing questionnaire and the follow up explanation sought

from the participants via email⁵. Moreover, we make use of some of the comments collected to describe how the system was perceived with regards to the hand-offs aspects previously identified.

The **Perceived Usefulness and Self-Predicted Future Use** aspects showed similar dichotomous perceptions among the participants. Thus, we discuss them together. After computing the correlation between the results of each couple of questions, there was no statistically significant relation; however, by analyzing the responses to the open ended questions and the discussion that occurred after the debriefing session with some of the participants a possible justification emerged. Most of the participants who perceived the usefulness of the ABC4GSD system negatively explained that their responses were based on the nature of their current work that was focused on long-term milestones being done rather individually and not requiring fine-grained cooperation as was expected during the evaluation study. While performing the main tasks (i.e., requirement prioritization and usability review), these participants reported concerns like “[...][the] chat window popping up abruptly while working on a task was more of a distraction, as I felt it disrupted the work flow [...]”. In [27], Yarosh et al. identify the flexibility of the activity metaphor as one of the main strengths of a system based on activities. Such systems need to be appropriated by users in ways that fit their work needs; such ways cannot be imposed, but only supported. Moreover, participants were asked to strictly follow the directions of the evaluation study and this may have been another reason for some participants to feel forced in the ways they were supported by the system in fulfilling their needs (**Collaboration (Coll.1)**). This could explain the reasons behind users scoring the system in a significantly more positive way during the preliminary evaluation [23]; in fact, in [23] the scenarios had almost no fine-grained directives and participants were allowed to experiment with the system. Finally, in the case of **Awareness (Aw.1,3)**, no significant correlation was found with the perceived usefulness or self-predicted future use. Looking at the open-ended questions, some participants mentioned that they neither noticed information about who was doing what (workspace awareness) nor understood who was responsible for what (group-structural awareness), and they assumed that they should have if the question was posed, hence, the negative response. With regards to (Aw.1), the ABC4GSD system provides information about the status of the members of an activity as well as shows the name of the last digital artifact selected (Fig.1.(10,12)). Whereas, with regards to (Aw.3), the ABC4GSD system currently has no implicit support for this feature; however, the system allows to contextualize tasks in different ways, e.g., the activity description, delegation of tasks, and in general by facilitating text-based communication for alignment.

We cannot generalize the the **hand-offs** support outcomes (i.e., (i) ability to switch work context; (ii) enhanced overall

awareness; (iii) shared workspace; and, (iv) communication). However, comments received from participants are encouraging; some of the more enthusiastic were: “*It was able to supporting me in handing over the work to colleagues as it opened a communication channel to the colleagues and the channel was connected with all the documents and conversations belonging to the the corresponding activity.*”; “*I think it is interesting that a specific document is associated to a particular task. That I value very positively*”; “*I liked integration of tools on ABC4GSD. Having the artifacts, associated with the task and supported through IM was good experience and helps to avoid switching between different tools and handling scattered information*”; and, “*ABC4GSD makes it very easy to collaborate on shared documents*”.

VIII. THREATS TO VALIDITY

Construct Validity. Aware of the complexity of assessing collaboration technologies [26], we carefully designed this study by leveraging established methods and, partially, instruments (i.e., [4] and [6]). We extensively reviewed the related literature and thoroughly discussed which activity to simulate to be as realistic as possible in the simplifications that were applied. This led us to the decision of simulating an arrangement with four teams as well as of selecting the requirement prioritization and the usability review as activities to simulate. The decision of having at the beginning and at the end of the simulated working day the confederates allowed us to both control (Fig.3.(A),(C)) and observe (Fig.3.(B)) the way participants cooperated. Moreover, to mitigate the problem of using the TAM instrument with novel technologies, we provided to the participants both a video tutorial as well as an hands-on experience through the `Training` scenario. The video tutorial was chosen over a presentation to limit bias and ensure repeatability. Finally, to avoid bias observer-participant interaction was reduced at a minimum and confederate-participant interaction was enforced by a predesigned script given to the confederates.

Internal Validity. Firstly, we based our work on the assumption that providing appropriate support to the dimensions of cooperation (i.e., collaboration, coordination, communication, and awareness) would mitigate time and space distance. While we have reasons to believe that these alleviate the problems associated with distributed SE, and there is a vast body of literature supporting such assumption, a causal relationship is hard to verify. Secondly, if the questionnaire used for the TAM has been empirically assessed, the questions used to focus on the dimensions of cooperation have been designed by the authors and were only reviewed and discussed with colleagues knowledgeable and experienced in empirical research. However, rather than claiming that such observed variables are correlated to the investigated unobserved ones, we have explained and justified the rationale behind their selection.

External Validity. No statistical significance, scalable, or generalizable results were sought by this study, as the key objective was to systematically gather and interpret empirical evidence about (i) the feasibility of utilizing a system based on

⁵Among the 30 participants only 27 gave the consent to be contacted, and 17 out of those contacted replied to the questionnaire.

the AT models and the ABC paradigm to support SE tasks, (ii) the ability of the ABC4GSD system in providing support to the dimensions of cooperation in SE; and, (iii) the applicability of the methodology designed for the evaluation of collaboration infrastructure, which we argue being a necessary phase before moving to a study in an industrial setting. Unfortunately, only 30% of the participants were practitioners, hence, it can be argued that the sample was not fully representative of the intended population of SE practitioners. However, all the participants were involved in computer science or SE related areas. Finally, we based our conclusions on the assumption that the simulation of an FTS arrangement would have allowed us to discuss findings in a SE key rather than solely FTS, hence, the results related to cooperation in distributed setting would hold also in collocated ones.

IX. CONCLUSIONS

Building on prior work [21][22][23], this study has provided empirical evidence about the feasibility of using the AT models and the ABC paradigm in SE environments and empirically demonstrated that the ABC4GSD system is able to mitigate time and space distance related issues by providing support to the dimensions of cooperation in SE, i.e., collaboration, coordination, communication, and awareness. Furthermore, in an effort to tackle Whitehead's vision of improving the available assessment instruments for collaboration technologies [26], we showed how to devise and apply an evaluation strategy by leveraging existing methods and, partially, established instruments (i.e., [4] and [6]).

The results from the evaluation study have also highlighted several research questions related to the understanding of what practitioners perceive as an activity; which awareness mechanisms are necessary; and, how to properly perform a comparative study of a collaboration infrastructure. We expect that these research questions will stimulate interesting research efforts in the area of tool support for cooperation in SE in general and GSE in particular. Finally, we hope researchers and practitioners will experiment with ABC4GSD to further validate it. To this end, the system was released as an open source project¹ and we hope others will contribute to determine the pros and cons of ABC4GSD and activity-based approaches to improve cooperation in SE.

ACKNOWLEDGMENTS

This research has been funded by the Danish Agency for Science, Technology and Innovation under the project "Next Generation Technology for Global Software Development", #10-092313.

REFERENCES

- [1] M. A. Babar, D. Winkler, and S. Biffi. Evaluating the Usefulness and Ease of Use of a Groupware Tool for the Software Architecture Evaluation Process. In *Empirical Software Engineering and Measurement, 2007. ESEM 2007. First International Symposium on*, 2007.
- [2] J. E. Bardram. Activity-based computing for medical work in hospitals. *ACM Trans. on Computer-Human Interaction (TOCHI)*, 2009.
- [3] E. Carmel, Y. Dubinsky, and A. Espinosa. Follow The Sun Software Development: New Perspectives, Conceptual Foundation, and Exploratory Field Study. In *Hawaii International Conference on System Sciences, Proceedings of the 41st Annual*, pages 1–9, 2009.
- [4] G. Convertino, D. C. Neale, L. Hobby, J. M. Carroll, and M. B. Rosson. A laboratory method for studying activity awareness. In *Proceedings of the third Nordic conference on Human-computer interaction*, pages 313–322, New York, NY, USA, 2004. ACM.
- [5] D. Damian and D. Moitra. Guest Editors' Introduction: Global Software Development: How Far Have We Come? *IEEE Software*, 2006.
- [6] F. D. Davis. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Q.*, 13(3):319–340, 1989.
- [7] C. Gutwin, S. Greenberg, and M. Roseman. Workspace Framework, Widgets, and Evaluation. In *Proc. of HCI on People and Computers XI*. Springer-Verlag, 1996.
- [8] J. Herbsleb. Global software engineering: The future of socio-technical coordination. *Future of Sw. Eng.*, 2007.
- [9] S. Houben, J. E. Bardram, J. Vermeulen, K. Luyten, and K. Coninx. Activity-centric support for ad hoc knowledge work: a case study of co-activity manager. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*. ACM, 2013.
- [10] V. Kaptelinin and M. Czerwinski. *Beyond the Desktop Metaphor: Designing Integrated Digital Work Environments*, volume 1. MIT Press Books, 2007.
- [11] O. Laitenberger and H. M. Dreyer. Evaluating the usefulness and the ease of use of a Web-based inspection data collection tool. In *Software Metrics Symposium, Metrics 1998 Proceedings Fifth International*, 1998.
- [12] F. Lanubile, C. Ebert, R. Prikladnicki, and A. Vizzaino. Collaboration Tools for Global Software Engineering. *Software, IEEE*, 27(2):52–55, Mar. 2010.
- [13] H. Mintzberg. *Mintzberg on Management: Inside Our Strange World of Organizations*. Free Press, New York, Jan. 1989.
- [14] J. Noll, S. Beecham, and I. Richardson. Global software development and collaboration: barriers and solutions. *ACM Inroads*, 1(3):66–78, 2010.
- [15] D. Norman. *The Invisible Computer: Why Good Products Can Fail, the Personal Computer Is So Complex, and Information Appliances Are the Solution*. Aug. 1999.
- [16] I. Omoronyia, J. Ferguson, M. Roper, and M. Wood. A review of awareness in distributed collaborative software engineering. *Software: Practice and Experience*, 2010.
- [17] J. Paer and B. Fitzgerald. Flexible And Distributed Software Processes: Old Petunias In New Bowls? *Communications of the ACM*, 2006.
- [18] J. Portillo-Rodriguez, A. Vizcaino, M. Piattini, and S. Beecham. Tools used in Global Software Engineering: A systematic mapping review. *IST*, 2012.
- [19] T. Rattenbury and J. Canny. CAAD: an automatic task support system. In *Proc. of the SIGCHI conf. on Human factors in computing systems*, 2007.
- [20] I. Steinmacher, A. Chaves, and M. Gerosa. Awareness Support in Distributed Software Development: A Systematic Review and Mapping of the Literature. *Computer Supported Cooperative Work (CSCW)*, 2013.
- [21] P. Tell and M. A. Babar. Requirements for an Infrastructure to Support Activity-Based Computing in Global Software Development. In *6th IEEE International Conference on Global Software Engineering Workshop (ICGSEW)*, 2011.
- [22] P. Tell and M. A. Babar. Activity Theory Applied to Global Software Engineering: Theoretical Foundations and Implications for Tool Builders. In *IEEE 7th International Conference on Global Software Engineering (ICGSE)*, 2012.
- [23] P. Tell, M. A. Babar, and J. Grundy. A Preliminary User Evaluation of an Infrastructure to Support Activity-Based Computing in Global Software Development (ABC4GSD). In *Global Software Engineering (ICGSE), 2013 IEEE Eight International Conference on*, 2013.
- [24] P. Tell, R. Giuffrida, and H. Shah. Revisiting the Global Software Engineering Terminology. Technical Report TR-2013-168, Dec. 2013.
- [25] S. Voidsa, E. D. Mynatt, and W. K. Edwards. Re-framing the desktop interface around the activities of knowledge work. *Proceedings of the annual ACM symposium on User Interface Software and Technology (UIST '08)*, 2008.
- [26] J. Whitehead. Collaboration in software engineering: A roadmap. *Future of Sw. Eng. (FOSE '07)*, 2007.
- [27] S. Yarosh, T. Matthews, T. P. Moran, and B. Smith. What Is an Activity? Appropriating an Activity-Centric System. In *Proc. of the IFIP TC 13 Int. Conf. on Human-Computer Interaction: Part II (INTERACT)*, 2009.