CrossMark

Gathering big data for teamwork evaluation with microworlds

Claudio Miguel Sapateiro[1] · Pedro Antunes[2] · David Johnstone[2] · José A. Pino[3]

Abstract We identify some of the challenges related with conducting research into teamwork, addressing in particular the data gathering problem, where researchers face multiple tensions derived from different viewpoints regarding what data to gather and how to do it. To address this problem, we propose a microworld approach for conducting research into teamwork. We present the main requirements guiding the microworld development, and discuss a set of components that realise the requirements. Then, we discuss a study that used the developed microworld to evaluate a groupware tool, which was designed to support team activities related to infrastructure maintenance. The paper emphasises the range of data gathered with the microworld, and how it contributed to simultaneously evaluate team behaviour and tool design. The paper reflects on the major contributions brought by the microworld approach, emphasising in particular the capacity to gather diverse data, and to combine behaviour and design evaluations. This research contributes to consolidate the microworld approach in teamwork research. It also contributes to reduce the gap between behavioural-oriented and design-oriented research. The combination of the behaviour-oriented and design-oriented views is of particular importance to design science, since it is founded on iterative cycles of development and evaluation.

Keywords Teamwork data · Teamwork evaluation · Microworlds · Design science

1 Introduction and motivation

This paper discusses the challenges of conducting research into teamwork and the huge potential brought by microworlds in helping us understand first, how teams perform when executing a common task, and second, in particular, what impact groupware support has on teamwork. To clarify the research context, we define *team* as a set of two or more people, who are interdependent, have a common set of goals and perform activities in the context of a common task [1]. We define *teamwork* as a construct that involves a set of behaviours and attitudes that occur as team members perform a collective task, which may be supported by groupware tools [2]. Finally, we regard *groupware* as a class of software tools supporting small teams through information-based and communication technologies [3].

Ideally, research into teamwork would involve theory building and testing using widely accepted methods, variables and data gathering instruments. Though, historically, this has rarely been the case. Research in this field reflects a long-lasting division between technologists and social scientists [4,5]. In particular, Jacovi et al. [4] note the lack of citations across the two different groups, while Grudin and Poltrock [5] observe the lack of appreciation for theory and methods developed from each side of the fence.

✉ Pedro Antunes
pedro.antunes@vuw.ac.nz

Claudio Miguel Sapateiro
claudio.sapateiro@estsetubal.ips.pt

David Johnstone
david.johnstone@vuw.ac.nz

José A. Pino
jpino@dcc.uchile.cl

[1] Department of Systems and Informatics Polytechnic of Setubal, Setubal, Portugal

[2] School of Information Management Victoria University of Wellington, Wellington, New Zealand

[3] Department of Computer Science, University of Chile, Santiago, Chile

These differences naturally extend beyond theory and methods towards more specific issues such as research design and evaluation. For instance, even though a wide range of evaluation methods have been developed for teamwork, they usually fall into one of two categories: design-oriented and behaviour-oriented [4,6].

Design-oriented methods typically have a strong focus on the suitability of a specific technology to support teamwork. Reflecting their origins in computer science, the scope of the variables in play when evaluating teamwork is usually narrow. For instance, many design-oriented studies are done in controlled settings, using highly constrained apparatus and specific scenarios and tasks [7]. Some design-oriented methods are completely centred on formative evaluation, often done in very early stages of development, and therefore do not address aspects of collaboration other than technology use [8]. Popular methods such as iterative prototyping and heuristic evaluation fall into this category [9].

By contrast, behaviour-oriented methods adopt the view that social, cultural, workplace, and organizational factors are determinant variables in play when evaluating teamwork. According to this perspective, evaluations should be conducted under conditions that are representative of actual, naturalistic work settings [10]. They should also address a rich set of independent, intervening, extraneous, and dependent variables. Typical examples include field studies and evaluative ethnography, which consider aspects such as social and physical settings, emergent behaviour, and work contingencies [11].

In practice, closing the divide between design-oriented and behaviour-oriented evaluation is a very difficult endeavour. First, a fundamental reason is that the field lacks unifying theory. Second, an even more complicated reason is the different epistemological assumptions about what theory is, in spite of important advancements brought by design science [12] and discussions on the nature of theory from a design perspective [13,14].

Third, unifying the field requires bringing together the wide range of factors involved in the phenomenon of interest into a conceptual framework. Such a framework would have to address various aspects such as the teams' characteristics and dynamics, the social and organisational context in which teams are inserted, the diversity of tasks assigned to teams, the positive and negative effects of technology on the team's tasks and processes, and of course the impact of technology design. Under this umbrella, we would have to specifically account for the multiplicity of factors constraining the teams' performance, such as human–human communication, human–technology interaction, coordination, collaboration, information sharing, visualisation, task and team awareness, and decision-making, just to mention a few.

Unfortunately, in this paper we do not propose such a unifying framework. Our discussion is less ambitious and really targeted at just one small part of the problem: data gathering. In particular, we discuss data gathering at the different granularity levels necessary to investigate teamwork.

1.1 Data gathering problem

Data gathering is a fundamental component of any study. Though, deciding what data to gather in a study addressing teamwork seems to be particularly challenging for researchers. Firstly, data gathering must focus on a set of variables that are explicitly conceptualised, defined, measured and inter-related. Prior research shows that using an incomplete set of variables often leads to conflicting results. For instance, a meta-review of experimental studies in the area of group decision-making attributed conflicting results to a generalised lack of consideration for contextual (e.g. team composition and tools) and intervening variables (e.g. task, process structure, and communication mode) [15]. Prior research also highlights that teams may activate different cognitive processing levels depending on task context, expertise level, social and organizational norms, and established practices [16–18]. Thus the consideration of context-related variables and their impact on data gathering is part of the problem.

Secondly, many of the variables considered of interest for teamwork research are difficult to examine directly. These include, for instance, situation awareness, mental workload, sensemaking and decision-making [19–21]. In these cases, one may have to add different data gathering methods to the mix, e.g. combining usage logs with queries to the participants about the values of various state parameters before, during and after performing a task [22,23].

Thirdly, as previously noted, a significant portion of research into teamwork is centred on the design and development of software tools and applications. According to the design science paradigm [12], researchers are involved in iterative cycles of artefact development and validation, which emphasise formative approaches to evaluation [24]. However, it seems difficult to reconcile the understanding of teamwork as a complex phenomenon with the adoption of iterative, necessarily low-cost, approaches to data gathering [6].

Finally, if we combine the issues related to context, method diversification and iteration, with the consideration that teams are often entangled in multiple, concurrent interaction patterns, moving swiftly between the individual, group and sub-group spheres, we can easily assert there are significant difficulties aggregating data in meaningful ways for research. The acquired data may pertain to different domains (e.g. social, cognitive and technological), encompass different viewpoints (e.g. qualitative and quantitative), targets (e.g. team and individuals), scales (e.g. events occurring in seconds, minutes and hours), timeframes (e.g. before, during

and after accomplishing a task), sizes (e.g., a few pieces of data vs. other very large log files or other data) and research goals (e.g. formative and summative).

All in all, developing data gathering instruments capable of spanning such a wide range of requirements and constraints seems to be a meritorious research endeavour in itself. Our research addresses these challenges by adopting a microworld approach to teamwork data gathering.

1.2 Microworlds as an approach to the problem

Microworlds are task-oriented, synthetic environments used to study human behaviour in quasi-naturalistic scenarios [25]. Two relevant examples consider experimental research on financial decision-making and train driving. In the first case, the microworld consisted of an auction game where traders would be able to buy and sell goods in multiple rounds with increasing difficulty [26]. This microworld supported research on the participants' decision-making capabilities under increasing stress based on questionnaires delivered after the game, eyetracker logging data and video recording. In the second case, a microworld was designed to mimic the dynamic work environment of a single train driver, including control interfaces, warning messages, alarms, etc. [27]. This microworld was then used to analyse the drivers' performance when facing interface designs representing the task in different ways. Measures of speed, time keeping, energy efficiency, attentional capacity and workload were used in the study.

Microworlds are often scrutinised regarding the validity of the variables they operationalize. In general, microworlds are recognised as having high face validity [28], i.e. their measures are assumed to be valid as they exhibit a pragmatic, common-sense relationship with the external variables being studied [29]. They are also seen as halfway between internal and external validity: external validity (measures valid outside the experiment) is high because of the realism of the task, while internal validity (measures caused by independent variables) is also high because of the degree of control over certain variables [30].

Since external validity is a fundamental requirement of naturalistic research [31] and internal validity is a tenet of experimental research [30], microworlds end up balancing the characteristics of both types of studies [32]. In fact, prior studies support the hypothesis that microworlds provide reliability and internal validity of performance measures [33,34], even though later studies suggest the difficulty level of microworlds may decrease external validity [28].

Based on this apparent capacity to balance competing views over research, we suggest that microworlds may contribute to a consensus around what data to gather in teamwork studies. We also suggest that addressing the data gathering problem requires an infrastructure capable of gathering a large amount of data from multiple sources, as well as coping with different levels of detail and formalisation, different time/space frames, and related to contextual variables, all of which seem feasible in microworlds [35].

Acknowledging that teamwork research endeavours require that data gathering leads to representative information unveiling the underlying phenomenon. That brings the need for the ability to systematically collect multidimensional data, and moreover, in some dimensions, the power to collect data volume. With the lack of a unified theory, data gathering posit several challenges that we identify as dichotomous tensions in Sect. 2. As introduced, we root our proposal for dealing with data gathering requirements, on the microworld paradigm given is inherent quasi-naturalistic orientation for teamwork enactment, while maintaining nevertheless, some degree of control over the data gathering process towards internal and external validity.

We consequently frame our overarching research goal as: *to develop a foundational set of building blocks guiding the development of microworlds as a data-gathering instrument for teamwork studies*. We document the development of these building blocks using the case study approach. The overall organization and rationale of the paper is provided below.

The next section discusses a set of fundamental tensions related to data gathering in the teamwork domain. Sect. 3 overviews the role of microworlds in supporting that endeavour. A set of requirements for data gathering with microworlds is developed in Sect. 4. Framed by this set of requirements, some fundamental building blocks for data gathering with microworlds are proposed in Sect. 5. Section 6 describes a study investigating teamwork using the proposed microworld. The study investigated the impact of a groupware tool designed to support teams performing a task related to infrastructure maintenance: handling disruptive events in network infrastructures. The study highlights how teamwork studies often involve the collection and analysis of big data. Details from the experimental results are also provided in Sect. 6. Section 7 discusses the role of microworlds in teamwork evaluation. Finally, in Sect. 8 we summarise the research, provide some concluding remarks and point future work directions.

2 Tensions in data gathering

A fundamental principle of empirical science is to determine the fate of theoretical propositions based on test results [36]. However, many different methods can be adopted to obtain test results, which often reflect different epistemologies and methodologies [37]. In particular, the investigation of teamwork is characterised by a notable diversity of views about what data should be collected and how it should be collected, which reflect different foundations in psychology, cogni-

tive science, computer science, and social science [4,6,38]. Several considerations have to be made to frame the data gathering process in a way that is at the same time valid (according to the requirements of each research community), efficient (from the researcher's point of view), effective (considering data quality) and reliable (considering the research design).

We expand the framework developed by Pinelle and Gutwin [39] by identifying several tensions that should be considered when gathering teamwork data. Table 1 shows these tensions as semantic differentials highlighting established viewpoints. Tensions No. 1 and 2 address the classic qualitative–quantitative divide [37], highlighting the differences between rigorous and relaxed manipulations, and controlled versus naturalistic settings, which are respectively typical of laboratory and field studies.

Tension No. 3 emphasises the researchers' goals either gathering empirical data about team behaviour or gathering design-oriented data, which supports iterative development. Tension No. 4 acknowledges the individual and distributed dimensions of teamwork. Decisions about gathering data at individual or team levels usually imply using very different theoretical scaffolds. Finally, in tension No. 5 we acknowledge that gathering teamwork data involves logging both macro and micro activities.

Along with each tension we provide a list of advantages and drawbacks that can be found in related literature. Overall, what we observe is that every choice pushes data gathering towards a different direction. There are 32 combinations of choices, and considering those combined choices and their implications is certainly one of the major reasons making the study of teamwork a complex endeavour.

3 The microworld approach in more detail

In an attempt to bridge the gap between the overly controlled and the overly naturalistic approaches, a number of researchers in human factors and complex work environments have been adopting the microworld approach. The term and early vision of the paradigm may be traced back to the work, in the 80s, of Seymour Papert and colleagues, MIT Media Labs cofounders [69]. Grounded on Piaget's work on constructivist learning theories, Papert and colleagues created a Lisp-based environment: the Turtle Geometry. Children could use this environment to learn by exploring geometrical and mathematical problems through an elementary language named Logo. Many more of these learning environments have emerged thereafter, sharing the paradigm of a delimited and constrained slice of reality designed as a safe place for exploring and learning.

Today the paradigm has been extended to other applications allowing the study of human behaviour on safe, synthetic, task-oriented environments operating in simulated conditions. When properly designed, microworlds hold some basic representative real-world characteristics while omitting other aspects deemed secondary for the purposes of the study, in order to keep some degree of control over some variables [70].

Moreover, microworld-based experiments are not usually long, expensive or resource demanding. Additionally, since they allow collecting large data sets with different granularity, it is possible to support summative evaluations as well as formative ones. They have been considered particularly pertinent in domains where field studies or full scope simulations present major challenges [71,72]. Indeed microworlds have been adopted in application domains like naval warfare [73], industrial processes control [74], air traffic control [75], and fire-fighting [76]. Additionally, they have supported research on complex phenomena such as dynamic decision-making [77], naturalistic decision-making [78], and learning [79], thus covering a wide range of variables related to teams.

Furthermore, some studies in the training arena, using sophisticated environments, have yielded successful outcomes, as is the case of the Aspire + Packet Tracer [80]. It supports a microworld based training and evaluation program in the computer networks domain for professional compliance with the international Cisco Network Academy global program. In fact, the Aspire module overlaps the Packet Tracer existing network simulator, to furnish the overall accomplished microworld with some enhanced face validity and gamification flavour, intended to favour learning and assessment realism.

This is an emergent trend, which posits that microworlds are complemented with, and may constitute a type of, serious game, where "serious" has been coined given the designated primary purpose other than pure entertainment. Generally, serious games refer to products used by industries like defense, education, scientific exploration, health care, emergency management, city planning, engineering, and politics [81].

Another pragmatic example of gamification in microworlds is the Foldit game, requiring players to learn to fold proteins using direct manipulation tools and user-friendly versions of algorithms. Players are engaging with real science problems achieving real results working collaboratively and competitively to build on each other's solutions and strategies toward uncovering previously unknown structures that have been subsequently published in first tier science journals [82]; and there are many other examples attesting to the external validity of this approach, e.g. [83].

Unfortunately, a common characteristic found in most microworlds developed to this day is they are tailor-made for each specific study. Even though each individual study contributes significant knowledge to their specific domain, the knowledge about the microworld paradigm does not

Table 1 Tensions in data gathering

		Perspective and goals	Main advantages	Main problems
1	Rigorous manipulation	Gather data in the laboratory, through rigorous control over treatment conditions and variables [40]; promote phenomena manifestations	Maximises, internal validity [40]; systematic validation of hypotheses and theory development	Relevance is highly constrained by the artificiality imposed by laboratorial settings [41]; often lack external validity [40]
	Minimal manipulation	Capture unstructured data from the field [42]	Openness, exploration, naturalness [43]	Interpretive nature of data and inter-subjectivity of data capture [42]; may be difficult to rightfully contrast findings [44]
2	Controlled setting	Understand the relationships between dependent and independent variables [43]	Eliminate confounding phenomena through isolation [45]	Cannot be applied to certain work contexts, e.g. high-risk situations [46,47]
	Naturalistic setting	Understand how teams make decisions in real-world settings [48]	Eliminate the constraints imposed by laboratorial settings [49]	Dependence on practical problems related with task and context [50]
3	Behaviour	Validate theories and models explaining how humans behave [12,51]	Complex processes can be analysed with methods such as process tracing and communication analysis [52]	Outcomes may not directly translate to technology development [12,53]
	Design	Gather pragmatic lessons from iterative design [12,54]	Validation through utility assessment [54]	Outcomes rely more on common sense than fully articulated research hypothesis [55]
4	Individuals	Humans as information processing machines [56]	Studies of individual cognitive functions have been enriching the way we understand human behaviour [57,58]	The focus on individuals misses much of the actual team cognition [59]
	Teams	Expand our view from individuals to the relations between individuals and the environment where they operate [60,61]	Better/broadly correlate individual decisions to their actions [62]	Many complicating factors introduced by the aetiology of team's dynamics [63] and context [39]
5	Macro	Evaluate complex tasks at a macro scale of performance [64,65]	Research on micro phenomena lacks correspondence with the scale where teams perform complex tasks [66,67]	Situated nature, dependent on concrete situations [45]; primary emphasis bears on experts [68]
	Micro	Understand how complex cognitive phenomena are entangled regarding the task execution [45]	Allow precise control and measurement [45]; the more we reduce phenomena into elementary components the more general will be the principles [45]	Some cognitive phenomena are difficult to examine directly [19]; primary emphasis on routine tasks [68]

evolve correspondingly because studies usually do not regard microworlds as a phenomenon of interest. In particular, a tailor-made approach makes it more difficult to standardise the instrument, considering for instance the balance between internal and external validity, the data and contextual information that should be gathered, what constraints to data gathering have to be considered, and their implications to theory building and theory testing. Next, we elaborate upon a list of requirements that would contribute to standardising the use of microworlds in teamwork studies.

4 Requirements for teamwork evaluation using microworlds

We can now consider in more detail the possibilities brought by microworlds put forward in Sect. 3 in relation with the tensions discussed in Sect. 2. Microworlds, because of their semi open/closed nature, can combine laboratory experiments (rigorous, controlled) with field observations (naturalistic, minimal manipulation), thus resolving tensions No. 1 and 2. This semi open/closed nature comes from the support to unpredictable behaviour while still controlling and monitoring the participants' interactions [84].

Microworlds may also support the simultaneous evaluation of technology designs and team behaviour, thus resolving tension No. 3. The key issues of tensions No. 4 and 5 concern the capacity to analyse teamwork at individual and team levels by gathering data with different granularity. Since microworlds usually mediate all team/user interactions, they represent an ideal vehicle for overcoming the main problems raised by tensions No. 4 and 5. From this match between tensions in data gathering and specific affordances brought by microworlds, we have derived a set of functional requirements to inform a generic teamwork evaluation platform using microworlds.

4.1 R1—Control external events

This is related to experimental rigor and control noted in tensions No. 1 and 2. The goal is balancing the teams' capacity to make decisions as if in a naturalistic setting with the capacity to capture behavioural data in a rigorous and controlled way. This involves controlling the injection of external events in the experimental scenarios, promoting context changes and unexpected reactions.

4.2 R2—Mediate human–human, human–technology, and human–environment interactions

This requirement concerns tensions No. 2, 3 and 4. A key characteristic of teamwork is interaction, and a key goal of behavioural studies is examining interaction patterns. Three types of interaction are considered: human–human (H–H), human–technology (H–T) and human–environment (H–E). H–H interaction involves information sharing, coordination and decision-making support, and other communication-based phenomena among humans. H–T interaction concerns the use of support tools. Teams often use generic tools like social media software and shared editors. They also use specialised tools designed to support work on the domain. Human interaction with these tools should be considered/captured within the context of the external events by the platform.

H–E interaction considers the physical reality over which teams operate. Teams interact with the elements of their operational physical reality in various dimensions, either mediated through technological systems or directly toward physical elements/resources. An example of the latter is the interaction with elements in the physical world such as mechanical levers. The work setting may also impose constraints that bound physical activities. These interactions can be simulated by the microworld platform in various ways. For example, adopting sophisticated immersion mechanisms to mimic the affordances of the real world, as seen in flight simulators. Other approaches with relaxed face validity may be considered depending on the evaluation purposes and the extent of its coupling with phenomena of interest [85].

4.3 R2.1—Human–human interactions

Interactions in the real world occur through different modalities, most often face-to-face, but video, radio, phone, chatting, e-mailing, and messaging are also common. The platform should reproduce the main characteristics of these modalities and in particular should preserve their one-to-one, one-to-many or many-to-many capabilities. Gathering data according to these modalities is paramount because it affords data richness so necessary to analyse teamwork. Most often an interesting facet of research derives from the analysis of teams' communicational modalities preferences and patterns according to context and exchanged items.

4.4 R2.2—Human–technology interactions

As noted above, the platform should also support the evaluation of the envisioned technology design options, considering in particular the functional features. To accomplish this goal, the platform requires a model and interface of the technology being evaluated. Then, the platform should gather data about every human–technology interaction down to the keystroke level.

4.5 R2.3—Human–environment interactions

The platform must consider two constructs related to the physical environment: locations and work elements. Locations are necessary to model teamwork done in multiple distributed physical places, while work elements provide simulators for the relevant physical interactions with physical world elements, e.g. operations on a physical machine. Then, the platform should gather data about state changes related to locations and all interactions with work elements.

4.6 R3—Data gathering must be contextualised at both macro and micro levels

Considering that teamwork is open and dynamic, with multiple external events injected over time and multiple interactions occurring in parallel, data logs can be quite difficult to analyse. The problem is even more relevant when extensive data is gathered at both macro and micro levels. So an important requirement is keeping a coherent view of the relationships between the captured data and the environmental and task conditions triggered during the evaluation.

4.7 R3.1—Behavioural data must be complemented with direct user inquiries

Even though capturing all user interactions in the microworld already supplies a large amount of data necessary to research teamwork, in many studies that is not enough. Phenomena such as situation awareness, attention, stress, decision-making, and information overload can hardly be inferred from interaction data alone and thus require other complementary data gathering sources. Several complementary approaches can be used, such as debriefings and talk-aloud protocols. Nevertheless, the approach that seems more in line with our perspective is inquiring the users through freeze-probe questionnaires. The main idea is prompting users with some questions before, during or after the task.

5 Proposed microworld architecture

We now put forward a set of building blocks for developing a microworld for teamwork evaluation. In Fig. 1 we overview the proposed logical architecture, which is directly derived from the requirements presented in Sect. 4. The proposed architecture has four main components: operational work environment simulator (OWES), team communication (TC), questionnaires manager (QM), and experimental control manager (ECM).

OWES supports virtual representations of the work environment over which teams perform their tasks. Its main purpose is to provide the means necessary to simulate real-world activities in the virtualised environment. Since communication is a fundamental asset of teamwork, TC is dedicated to support several communication modalities used by the team. The integration of QM in the microworld is based on the fact that questionnaires have been widely adopted to
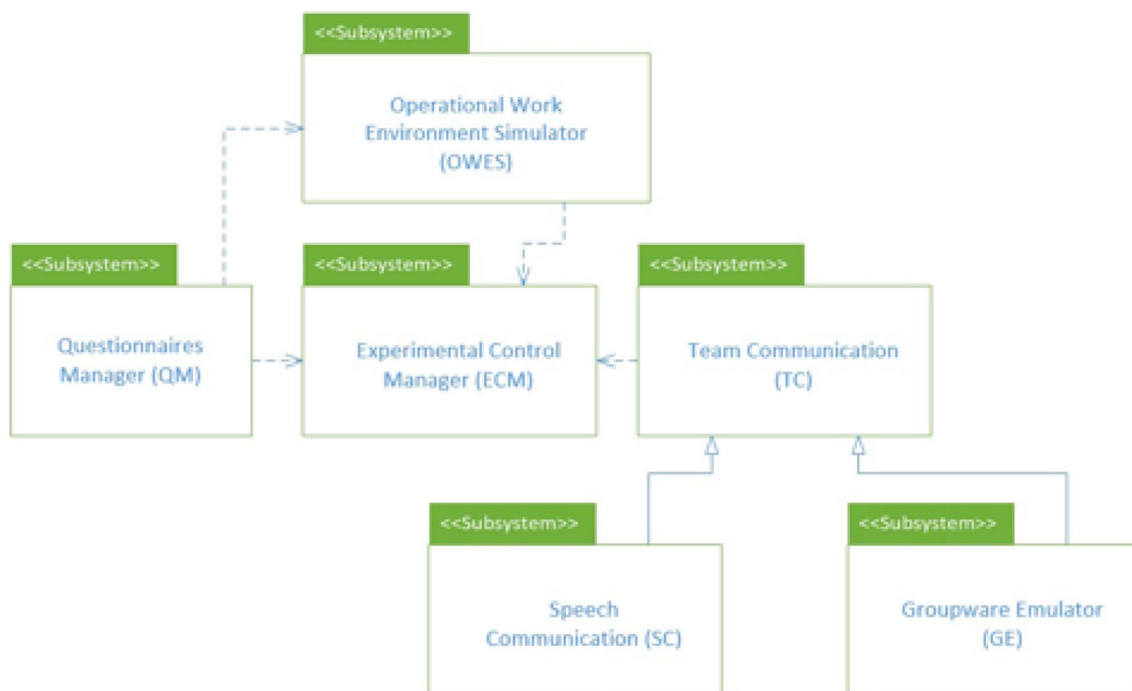
Fig. 1 Proposed microworld architecture
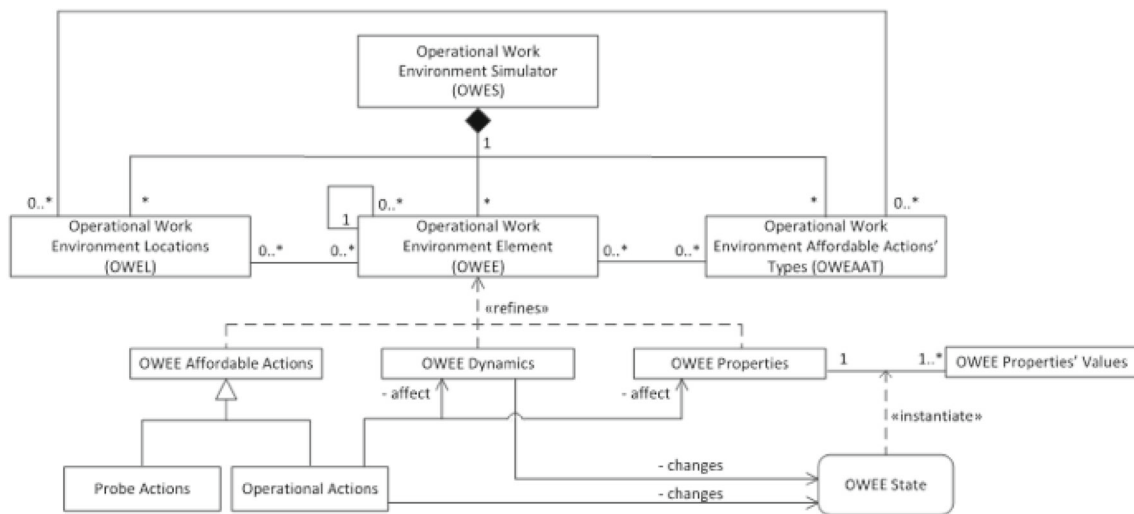
Fig. 2 OWES—operational work environment simulator model

implement freeze probes. Finally, ECM oversees the experimental process, focusing on task assignment and control.

Each of these components is further described in the following subsections.

5.1 Operational work environment simulator (OWES)

The main goal of OWES is to simulate real-world physical elements, their affordances and the actions that users can do on them. The definition of both elements and actions to be accommodated in a synthetic representation of the work environment is inevitably dependent on the specific application domain and simulation purposes [86]. For instance, simulating an online market will necessarily be very different from simulating a firefight.

Though in Fig. 2 we elaborate a domain independent model of OWES (using the UML notation). The model defines three abstract classes: operational work environment locations (OWEL), operational work environment affordable action types (OWEAAT), and operational work environment elements (OWEE). Being abstract classes (in the UML modelling sense), they are not intended to be directly realized by an implementation, but to provide guidance for development on specific application domains. Most importantly, they specify what data will be captured by the microworld in relation with the work setting, i.e. elements, locations and actions.

OWEL represents physical locations, e.g. fields, rooms and buildings. While working in the field, team members can perform a myriad of location-based actions like moving around and arriving to a certain place, which should be logged. OWEE models physical elements of the work environment, like equipment and tools. OWEAAT represents the set of actions that can be associated to each element defined in the work setting.

Relying on the UML abstraction relationship, which defines dependencies between model elements at different levels of abstractions, the proposed model uses the ⟨⟨refine⟩⟩ dependency to show that an OWEE definition is accomplished by three constituents: properties, actions and dynamics. The scope and level of detail of OWEE properties are bounded by domain requirements regarding face validity: if high face validity is required, then the OWEE properties must account for fine-grained attributes such as size, shape and colour; conversely, if face validity can be relaxed, then those properties may not be specified and the OWEE representation may be simpler and more abstract [85].

The specification of actions is of paramount importance in a synthetic environment, since they define what the team members can do with the elements of the work setting. In the proposed architecture, we define two types of action: probe operations (e.g. checking the status of a device, where blue means it is working and read means it failed), and operative actions (e.g. restarting and unplugging a device). We additionally note that some OWEE may require the consideration of their intrinsic dynamics in order to reproduce real world behaviour (e.g. how some properties change through time and/or in respect to interactions between OWEEs or OWEAA).

The OWES specification uses a state space approach. State spaces have been used in software engineering for specifying high-level conceptual requirements and rigorously inform design and implementation [87,88]. Buchner and Funke [89] highlight four strengths of state spaces. First, they afford a rigorous, systematic and reusable formulation of synthetic environments. Second, by representing an environment with a set of internal states, they smoothly accommodate the manipulation of states to promote the manifestations of phenomena of interest, which is paramount given our purposes.

Third, since the elements specified in OWES may themselves be described with state spaces, the representation of the overall environment contributes to internal validity. Fourth, considering that in complex dynamic scenarios users continuously interact with OWEE, framing these interactions within the state space provides a white-box view over the users' interactions, which in turn allows capturing data about every action performed by users within the environmental context.

The definition of an OWEE state is based on a set of properties-domain values. State transitions are accomplished by three means: (1) derived from methods implemented by OWEE objects (internal dynamics); (2) stemmed from user interactions with OWEE objects (team dynamics, which is not modelled); and (3) derived by ripple effects affecting several OWEE objects, where a state change in an object causes a state change in other objects. Furthermore, state changes in OWEE objects may also be caused by the injection of external events.

Given the above considerations one may note that, although an OWES state space collection depends on the application domain, if care is taken on the implementation then a state space engine will be largely reusable across multiple application domains.

5.2 Team communication (TC)

Team communication may take several forms considering established team practices and available mediation tools. Taylor et al. [90] studied how teams adapt to dynamic tasks, and found that most effective teams exhibit specific communication patterns aimed at supporting the information needs of the team and its individual members. Citera et al. [91] and Klein [92] also point out that efficient teams have high levels of communication to improve situation awareness and track other team members' activities.

A number of application domains such as air traffic control [93], control rooms [94], emergency dispatch [95], fire-fighting [96], and network troubleshooting [97], where teamwork is time-critical, recognize the important role of speech communication. Field studies involving mobile professionals, like service technicians [98], police patrols [99], and fire-fighters [100] also suggest that collaboration in these settings is primarily supported by speech. As a consequence, gathering details about the speech communications done by teams is of paramount importance to study teamwork.

For that reason, we propose that all team communications have to be mediated by the microworld, which will log every event. This constraint applies in particular to face-to-face-communication. Therefore the proposed TC component includes a sub-component named Speech communications (SC). SC supports three communication modes: one-to-one, one-to-many and many-to-many. These modes address typ-

ical types of speech communication such as phone calls, conference and radio calls.

Besides speech communication, teams may also communicate using various types of tools sharing text, pictures, etc. Once again, in order to study teamwork, one has to gather data about these communication events, which in turn requires all groupware-based communications done by the team to be mediated by the microworld. Therefore, the proposed TC also includes a sub-component named groupware emulator (GE). The specific characteristics of the GE component cannot be detailed, since they are specific to the tools adopted by each team. So the GE is just an architectural placeholder, which is responsible for logging tool-based events, which in turn have to be specified on a case-by-case basis.

Moreover, one further aspect that has to be considered about the GE is that the groupware tools may not actually exist. That is, the GE may be used as a proxy for evaluating design features of early conceptual prototypes.

The functional implementation of the GE is naturally dependent on the particular groupware or conceptual feature that is considered by a research study. Thus we can say that the GE component has a lesser overall reusability than the SC, which may be used across multiple studies.

5.3 Questionnaires manager (QM)

Studies in human cognition often rely on the administration of questionnaires to understand the phenomenon of interest. In the specification of QM we consider three main aspects constraining the administratio n of questionnaires to the team: (1) the moment at which the questionnaires are administered to the team; (2) the extent that questionnaires are aligned with the users' operational context; and (3) the extent that answers to questionnaires contribute to analyse behaviour at both the individual and the team levels.

Regarding the first issue, we note freeze probes can be problematic because they intrude the users' primary task [101]. Post-trial questionnaires are less intrusive, but on the other hand may be less sensitive to the phenomenon of interest, since users may rationalize their performance.

QM may be configured to prescribe pre-trial, post-trial or freeze probe questionnaires. The latter considers freezing the primary task at a certain point in time to administer a questionnaire to users. Furthermore, considering the amount and types of data collected by the microworld, the QM component can be configured to trigger individual questionnaires, which are triggered after the occurrence of a specific event or a certain number of actions performed by the user on a OWEE object.

Regarding the alignment between the questionnaires and the users' operational context, we observe that, since the microworld traces all users' actions within the environment's state space, it is possible to dynamically generate a ques-
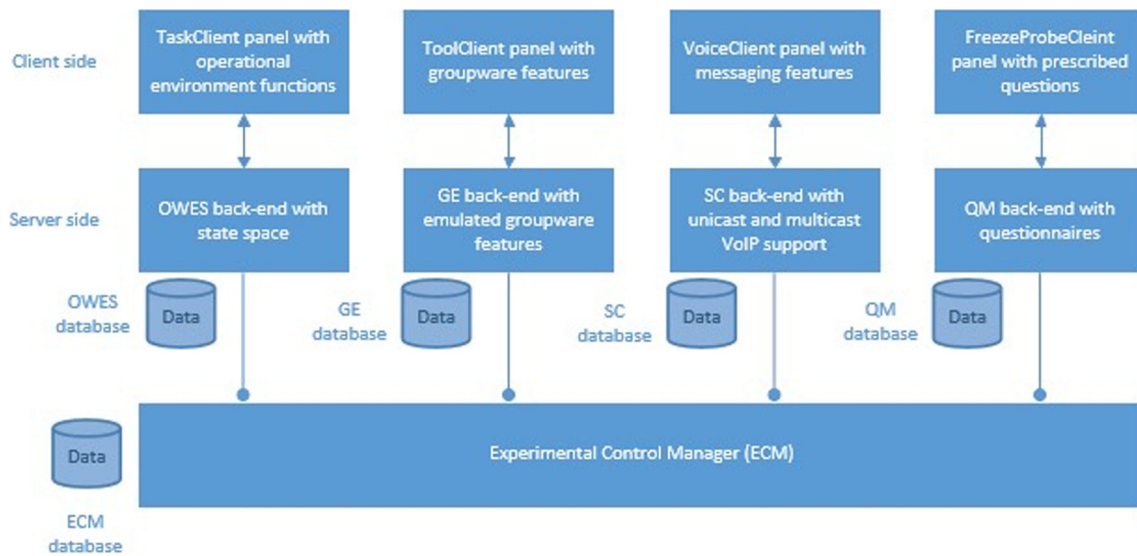
Fig. 3 Instantiation of the microworld architecture

tionnaire that accounts for the users' specific operational contexts. This important feature contributes to capturing teamwork data in context. Finally, we note that the QM can be reused across different studies, since only the questions delivered to users must be tailored to each specific study.

5.4 Experimental control manager (ECM)

Microworlds, as laboratory instruments, must control the experimental conditions. This involves an integrated control of its constituents, OWES, TC, and QM components, configuring and, enabling or constraining their functional features. As an example, we bring up the set of experiments described in Sect. 6, which compare team performance under two conditions, one relying on phone calls to accomplish a task, and another one using a collaborative tool in addition to phone calls. In this particular case, SC was configured to operate in the one-to-one mode, and GE was enabled for one group of users and disabled for the control group.

Another paramount requirement underlying the ECM component is to accommodate multiple task scenarios, i.e. multiple manipulations of OWES that are intended to promote the manifestation of a phenomenon of interest. The representation of task scenarios relies on an event-driven logic, which combines a set of external events with the state space defined in OWEE. By adding external events, which disrupt the state space, we may intentionally lead the operational behaviour to depart from its expected dynamics towards unexpected behaviour [102].

The present characterization of task scenarios builds upon three main constituents. The first one holds a description, in natural language of the task context that is conveyed to the team through the OWES graphical user interface. The sec-

ond constituent is the set of initial states that the OWEE holds in order to reflect a purported task scenario. The third constituent holds the collection of events that will affect the OWES state space. Such events will override the OWEE state transitions defined in the OWES engine state matrices at certain points in time.

5.5 Architecture realisation

Since the previous architecture description is abstract, we now provide additional details about its realisation.

Figure 3 illustrates how we have instantiated the microworld architecture. The instantiation adopted a client–server approach, where the server is responsible for all the microworld logic and data storage and the client-side provides a set of four independent graphical user interfaces that give users access to a set of functions. Below we describe these instantiations in more detail emphasising the client–server relationships.

5.5.1 Speech communication (SC)

Speech communication is realised by the SC component. From the server perspective, this component uses voice over the Internet protocol (VoIP) to mediate communication between team members. For users, SC provides a VoiceClient control panel that allows them to initiate a communication in either unicast or multicast modes. When the unicast mode is selected, the user can select the receivers from a menu showing the list of team members. Then the voice communication is controlled by pressing "Start" and "Finish" buttons. This realisation is generic and can be used in any study.

5.5.2 Operational work environment simulator (OWES)

This component holds the state space that describes the operational behaviour of the microworld. The client side offers a TaskClient control panel that provides a set of operational possibilities mimicking some real-world characteristics of the operational environment. These include: (1) moving to a certain location; and (2) checking and/or changing the status of an element defined in the environment. When a user invokes one of these operations on the TaskClient panel, the server uses the state space to propagate state changes to every affected element.

Even though the OWES supports three generic operations on the environment (move, check and change status), the full realisation of this component may require specifying other operations relevant for the application domain and phenomena of interested being studied. Furthermore, the microworld has to

be configured with specific sets of locations, elements of the environment, actions, dependencies between them, etc. Nonetheless, the abstract orientation provided by the propose architecture, combined with state-space functionality, provide partial reusability of the OWES.

5.5.3 Groupware emulator (GE)

The client-side of this component is realised by the Tool-Client control panel. This panel provides a set of user-interface controls that invoke groupware functions such as instant messaging and collaborative editing. The back-end component either simulates these functions or implements an application interface to the actual groupware tools. As such, its realisation is particular to the specific application domain and phenomena of interested being studied.

5.5.4 Questionnaires manager (QM)

This component is realised through a FreezeProbeClient pop-up panel, which can be configured to prescribe a set of questions to the users. The FreezeProbeClient interacts with the server to collect information about the OWES state space, which may be necessary to contextualise the questions (e.g. "you have already done operation X twice, please explain why"), and to display the questions and collect answers.

This component can be configured regarding the type and timing or the triggering event. It can also be configured to collect various types of open and closed questions (yes/no and multiple choice). Therefore, this realisation is generic and can be used in any study.

In Table 2 we summarise the characteristics of the developed platform components according to the controlled and naturalistic categories. We also highlight the reusability potential of each component.

Table 2 Characteristics of the proposed microworld architecture

Component	Naturalistic characteristics	Controlled characteristics	Reuse potential
Speech communication (SC)	Can reproduce typical functionality of mobilephones and walkie-talkies	User has to press buttons to select a receiver and to start/stop communication	Yes
Operational work environment simulator (OWES)	Can reproduce specific actions in the physical space such as moving around and operating physical devices	Checking and changing the status of physical devices are substituted by pressing buttons on the user interface. Users also press buttons to simulate when they move around	Only the state space engine can be reused. The physical effort associated to moving around space has to be simulated
Groupware emulator (GE)	Can emulate functionality provided by groupware tools	The interaction is detached from the actual device, e.g. the microworld panel instead of a mobile device	No
Questionnaires manager (QM)	Gathers user data in context. Responses are not affected by hindsight and delays	If not conceived with caution, it may cause disruptions and biases	Yes

6 Case study

We conducted a case study in the field of infrastructure maintenance using the proposed microworld architecture. The following section briefly describes the study's domain of application. Next, we describe how the microworld was applied in the case, and then discuss the obtained results.

6.1 Preliminary considerations

Large and medium organisations usually need specific teams that take the responsibility to maintain information and communication technology (ICT), handling failures, responding to clients' requests, installing and configuring technology, helping users, etc. These teams often have to collaborate to solve disruptive events, sometimes in multiple physical spaces. The primary goal of the case study was to design a mobile tool that would increase teams' situation awareness.

Various types of disruptive events may occur in ICT systems, involving both hardware and software failures, which in turn may originate service-level failures. Some service failures may be regarded by organizations as very serious or even critical, since they may not only compromise the internal operations but also service-level agreements with other organisations. During these events, the responding teams may have to adapt routines to better respond to emergent problems, time constrains, high stress levels, and improvisation of containment and mitigation actions [103].

To better understand teamwork in these scenarios, we conducted several interviews and workshops with practitioners [104,105]. From that study, we identified a research gap: even though responding teams rely on well-known trouble ticket software to support more routine operations, such software was perceived as irrelevant in the support of non-routine scenarios. In most cases, trouble tickets were just used to open an incident; and occasionally they were used for post mortem annotations. However, trouble tickets were never used to support teamwork during the events. Thus a design opportunity has been identified.

Our main goal was then to design a groupware tool that would increase situation awareness by sharing up-to-date information about an on-going disruptive event, sharing information about the individual activities done to identify failures, and information necessary to coordinate the team and solve the problems. Consequently, we decided to design and develop a mobile tool supporting situation awareness [104].

The critical challenge though was evaluating the tool. The adoption of a laboratory approach to validate the tool seemed inadequate because the whole purpose of the tool was supporting teams in realistic, non-routine scenarios. Furthermore, data acquisition would have to range from micro to macro details, regarding both individual and team situation awareness.

Validating such tool in the field also seemed to have some drawbacks: (1) it required various people to gather data; (2) the distributed work setting made it difficult to co-relate the collected data; (3) collecting such data during disruptive events also created logistic and organisational problems; and (4) most importantly, since the tool design could evolve at any time, the whole data gathering effort seemed excessive when pondered against the exploratory nature of design. It was this context that justified using a microworld to evaluate the tool.

6.2 Realisation of the microworld architecture

In Fig. 4 we illustrate the realisation of the OWES for this particular case study. The top of Fig. 4 presents the TaskClient panel that was shown to each team member during the experiments. It shows that users can move between different rooms (from A to E), can select devices available in each room (e.g. room D has 5 computers and 2 routers), and can also operate these devices. For this particular application, the considered
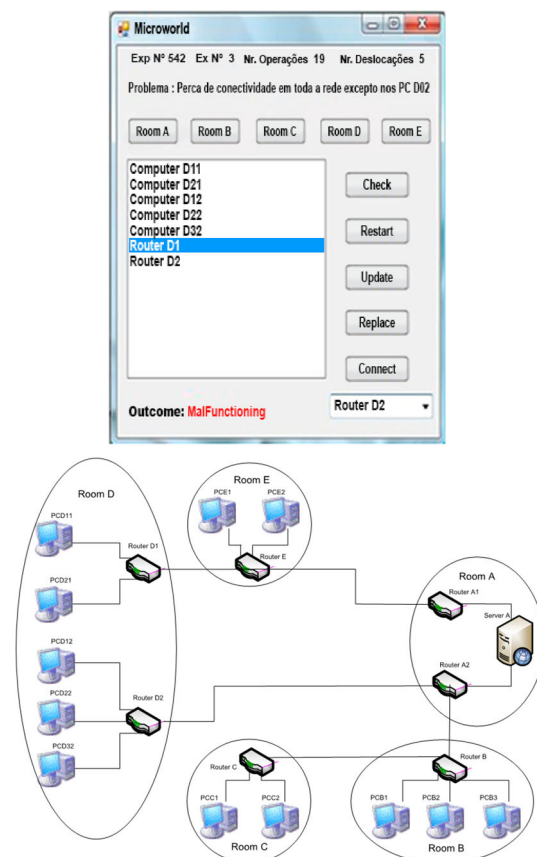
Fig. 4 *Top* TaskClient with a sample network loaded; *Bottom* simulated respective network's infrastructure
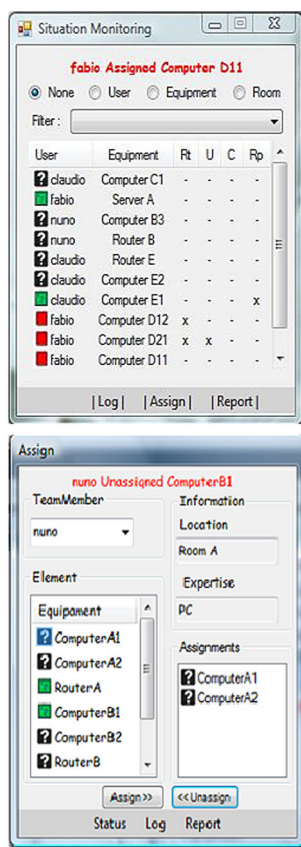
Fig. 5 ToolClient—*Top* Situation-monitoring panel; *Bottom* assignment panel

operations were: check status, restart, update, replace and connect.

The bottom of Fig. 4 illustrates the simulated network infrastructure. The state space defined for this scenario has three types of environmental elements: computers, routers and servers. It also defines relationships between these elements and several locations (rooms A to E). Finally, each element may hold an overall state of "working" or "malfunctioning", which can propagate to other dependent elements according to the network structure being simulated.

The considered operational actions include: check status, restart, update, replace and connect to the network. It is also possible to move to a room where a set of elements was located. The simulation adds a certain amount of time to complete this action to simulate the cost of travelling around.

In Fig. 5 we present the realisation of the GE for this case study. The ToolClient has two panels: assignment and situation monitoring. The assignment panel shows the tasks assigned to team members, while the situation-monitoring panel summarises the results from the checks done by team members (after moving to a room and checking a device). In the assignment panel, a team member can also ask a colleague to check a device.

We note again that the reusability of the abovementioned two components is low. They have to be developed for each specific case study. Though the microworld standardises data collection. In this particular case, data is gathered about every action done by the users on TaskClient and ToolClient.

On the other hand, vanilla VoiceClient and FreezeProbeClient components were used in the case study without any modifications. The VoiceClient was configured to simulate the use of mobile phone calls (unicast). And the FreezeProbeClient was configured to suspend the task at three points in time to prompt the team members about a set of task-related factors that would contribute to analyse situation awareness. More details about these questions are provided in the next section.

6.3 Experimental design

The experimental design was focused on understanding the impact of the groupware tool on teamwork. More precisely, we were seeking to understand how the provision of up-to-date information to the team about assigned tasks and device checks affected the team.

The fundamental phenomenon of interest was situation awareness, since it has been found to constitute a fundamental team asset under demanding work settings [106]. Situation awareness is a complex construct with many different analytic lenses. Situation awareness is constructed from the continuous extraction of environmental information and the integration of such information with previous knowledge to form a coherent representation of a situation. Situation awareness guides action and helps make projections about how a situation may evolve [107].

These definitions bring out two dimensions traditionally considered by research in situation awareness: (1) the product dimension, i.e. the information held in the individuals' minds [108]; and (2) the process dimension, which concerns the activities enacted by individuals to build situation awareness [109].

When analysing teams we should bring forward two additional dimensions: (1) a shared dimension, which addresses the overlap of situation awareness among the team members [110]; and (2) a distributed dimension, which regards situation awareness as spread among the team members in a complementary way [111].

Measuring the phenomenon according to these different facets precludes different measurement techniques. While the product and shared dimensions have been mainly assessed through the use of questionnaires [19,110], the process and distributed dimensions have been studied using operational work tracing techniques [111]. A clear advantage of using a microworld in this case is that it supports gathering data related to these two different dimensions. Therefore in our

experiments with the microworld we combined freeze probe questionnaires with activity logs.

A repeated measurements design was adopted and each participating team was submitted to two experimental treatments. Given that the introduction of a groupware tool constituted the independent variable, the teams were subject to two sequential treatments assigned in random order: one having access to the groupware tool (w/ condition) and the other without (w/o condition), which served as control treatment. In each condition, teams performed first a practice test with a different task scenario.

The teams were assembled from final year students of undergraduate courses in informatics. The selected team size considered three elements. Extra course credits and prize money were offered to the best performing teams to encourage deep engagement with the tasks. The participants were informed that their performance was evaluated according to three main factors: time to accomplish the task, number of operations necessary to complete the task, and number of displacements over the (virtual) places necessary to complete the task. Only students that had successfully completed a course on computer networks were selected for the experiment, to ensure they were knowledgeable about the task. All participants signed consent forms and received briefing materials and participated in briefing sessions describing and clarifying the goals of the experiments and the type of task they had to accomplish.

The experiments were done in two rounds, the first one with 12 teams (36 participants) and the second one with 11 teams (33 participants). There were several differences between the two rounds. Besides some minor changes related with the specific variables that were measured, the major change introduced in the second round was increasing the complexity of task, using a more complex network topology and more intricate failure modes. These changes required modifications to the OWES component of the microworld but did not affect the other components. In the following, we will only discuss results from the second round of experiments. Details about other rounds of experiments can be found in [104].

For each experimental treatment in the second round of experiments, the microworld was configured to freeze the task three times to administer a questionnaires with three questions gathering different data about situation awareness: (1) what is the current status of network device X? (awareness about the problem) (2) where the other team members are located? (team awareness) and (3) what are the underlying causes of the problem? (awareness about the problem causes).

Besides gathering data through the questionnaires, we also gathered data on the task completion times, speech communications, actions done in the environment (moving around and checking/operating devices), and operations done in the

groupware tool (assign teams to tasks and report device status). All that data was then analysed according to a set of measures described in more detail in the next section.

6.4 Experimental results and analysis

Our discussion of the experimental results is mainly focussed on illustrating the capacity to analyse the range of data captured by the microworld. Given that the gathered data does not follow a normal distribution, we rely on the non-parametric, distribution free Wilcoxon matched-pairs signed rank test [112] for assessing the statistical significance of the results, considering the minimal threshold as 0.05.

We start by reporting on task completion times and number of operations enacted by teams (Table 2). Regarding task completion times, teams in the w/ condition took more time to complete the task, with an average of 2 more minutes when compared with the control group. These results were statistically significant. Though curiously the differences in number of operations performed by teams were not statistically significant.

We defined individual awareness (IA) as the quotient between the number of correct answers to freeze probe questions and the total number of answers, $\text{IA} = \frac{\#\text{correct answers}}{\#\text{answers}}$. The results are presented in Table 3. They show no statistically significant differences, although results for question 3, which asked if the participants had a perception of the underlying causes of the network problems, were on the threshold to significance.

We defined team awareness (TA) as the quotient between the number of correct answers to freeze probe questions shared by pairs and triplets of participants, and the total number of answers (considering again pairs and triplets),

$$\text{TA} = \frac{\#\text{correct answers shareled by pairs}}{\#\text{questions shared by pairs}}$$
$$+ \frac{2}{3}\frac{\#\text{correct answers shared by pairs}}{\#\text{questions shared by triplets}}$$
$$+ \frac{\#\text{correct answers shared by triplets}}{\#\text{questions shared by triplets}}.$$

The results obtained for team awareness, which are presented in Table 4, also showed no significant differences.

Table 3 Averages (and standard deviations) of task completion times and number of operations

Condition	Completion times (min)	Number of operations
w/o	8.23 (1.85)	38.18 (12.05)
w/	10.55 (1.90)	41.27 (9.52)
p-value	**0.016**	0.262

Bold value indicates statistical significance of the results considering the minimal threshold as 0.05

Table 4 Averages (and standard deviations) of individual awareness (IA) for each question (Q1–Q3) and freeze probe (#1–#3)

Condition	IA—individual awareness								
	Q1			Q2			Q3		
	#1 (N = 33)	#2 (N = 33)	#3 (N = 24)	#1 (N = 33)	#2 (N = 33)	#3 (N = 24)	#1 (N = 33)	#2 (N = 33)	#3 (N = 24)
w/o	0.60 (0.28)	0.77 (0.24)	0.78 (0.25)	0.70 (0.28)	0.58 (0.27)	0.35 (0.23)	0.33 (0.19)	0.31 (0.19)	0.48 (0.29)
w/	0.59 (0.30)	0.74 (0.23)	0.77 (0.27)	0.78 (0.31)	0.57 (0.32)	0.47 (0.31)	0.24 (0.16)	0.36 (0.20)	0.49 (0.27)
p-value	–	–	–	–	–	0.15	0.05	–	–

Given that we delivered three freeze probes to groups at different points in time during the task, we could also analyse how individual and team awareness evolved from one probe to the other. That measure corresponds to what we designate situation awareness improvement ratio (SAIR), which is reported in Table 5. This measure was defined as the ratio between the IA measured in a freeze probe and the IA measured in the previous probe, for each question. A positive value indicates the individual situation awareness has improved, while a negative value indicates it has decreased.

As shown in Table 5, these measures provided some statistically significant differences. Considering Q1, we note that individual team members increased situation awareness from the first to the second probe, but from then on kept a similar level of awareness.

Regarding Q2, this measure suggests that awareness about the location of team members decreased as the task unfolded, even though the ratio was less pronounced in the w/ condition. Considering Q3, we note that gains in situation awareness were higher for the w/ condition in the first probe but then again reached a plateau, suggested by no statistical significances between the two conditions in the transition from probe 2 to probe 3.

We also measured individual and team situation awareness based on the activities done by the participants to diagnose the problem. The individual diagnosis efficiency (IDE) is a quotient between redundant equipment checks and the total number of checks, $IDE = 1 - \frac{\#redundant\ equipment\ checks}{\#equipment\ checks}$. The individual operational efficiency (IOE) is a quotient between redundant operations and the total number of operations,

$$IOE = 1 - \frac{\#redundant\ operations}{\#total\ operations}.$$ Team diagnosis efficiency (TDE) is the quotient between the number of redundant checks done by each member and the total number of checks done by the team, $TDE = 1 - \frac{\sum_{i=1}^{3} \#redundant\ equipment\ checks\ of\ member(i)}{\#equipment\ checks}$. Finally, team operational efficiency (TOE) is the quotient between the number of redundant actions done by each member and the total number of actions done by the team, $TOE = 1 - \frac{\sum_{i=1}^{3} \#redundant\ operations\ of\ member(i)}{\#operational\ actions}$. The obtained results, which are summarised in Table 6, did not exhibit any significant differences between conditions.

Besides analysing activities, we also analysed the speech communication between team members (Table 7). Here, the results were more interesting. We identified significant differences in two variables. One was the number of messages, where the teams not using the groupware tool exchanged more messages than the control teams. Since we could also classify the messages as either related to information sharing or team management, we could also analyse the differences between the two categories. The results show significant dif-

Table 5 Averages (and standard deviations) of team awareness (TA) for each question (Q1–Q3) and freeze probe (#1–#3)

Condition	TA—team awareness								
	Q1			Q2			Q3		
	#1	#2	#3	#1	#2	#3	#1	#2	#3
w/o	0.27 (0.36)	0.66 (0.29)	0.67 (0.21)	0.65 (0.21)	0.45 (0.19)	0.29 (0.17)	0.20 (0.11)	0.17 (0.14)	0.31 (0.23)
w/	0.25 (0.22)	0.63 (0.33)	0.60 (0.26)	0.76 (0.24)	0.46 (0.21)	0.32 (0.21)	0.13 (0.13)	0.26 (0.13)	0.32 (0.13)
p-value	– (N = 11)	– (N = 11)	– (N = 8)	0.109 (N = 11)	– (N = 11)	– (N = 8)	0.203 (N = 11)	0.139 (N = 11)	– (N = 8)

ferences in the number of messages exchanged for team management, where the teams using the groupware tool exchanged fewer messages than the control groups (Table 8).

Overall, the obtained results suggest that the groupware tool slightly modified the team's behaviour regarding information exchange: teams using the tool exchanged fewer messages, especially messages related to team management. Generally speaking, these results suggest the groupware tool did not increase either individual or team situation awareness. However, a more fine-grained analysis of the three freeze probes showed that, for one specific type of awareness—awareness of the problem causes—the groupware tool provided more awareness in the first probe, but teams reached basically the same levels of awareness in the last probe.

7 Discussion

The results from the case study indicate that the groupware tool changed the teams' behaviour by decreasing the number of exchanged messages related to team management. However, it did not increase situation awareness as primarily hypothesized. Reflecting on these results, it seems that adopting situation awareness as a development goal and measure of success, was a poor choice. We optimistically assumed that *increasing* situation awareness was *desirable*, but instead the experimental results showed that the participants could actually fulfil the task with the existing level of situation awareness. Perhaps more dramatic was the realisation that our attempt at increasing situation awareness through the groupware tool had an actual cost, which resulted in teams spending more time to complete the task. This trade-off could then provide a rational explanation for the lack of significant differences in the situation awareness measures.

However, the team members were keen to change communication patterns when new groupware features were available. Using the groupware tool, the team members significantly reduced voice communication. With hindsight, we realise the teams had a clear preference for *reducing* the amount of effort required to accomplish the task over *increasing* awareness about what was going on. The results from the case study suggest that future developments of the groupware tool should focus on optimising the overall usage effort instead of just focussing on situation awareness.

Even though these results are interesting by themselves, what also emerged from the study was the capacity of the microworld to gather rich data about various aspects of team behaviour. In particular, it was the cross-analysis of communication-related and awareness-related data framed in the activity log with regard for the task context that allowed us to reach these results. Furthermore, it was the ability of the

Table 6 Measures of individual situation awareness improvement ratio (SAIR) between freeze probes, for each question (Q1–Q3)

Probes	SAIR—Situation awareness improvement ratio					
	Q1		Q2		Q3	
	w/o	w/	w/o	w/	w/o	w/
#1	0.60 (0.28)	0.59 (0.30)	0.70 (0.28)	0.78 (0.31)	0.33 (0.19)	0.24 (0.16)
#2	0.77 (0.24)	0.74 (0.23)	0.58 (0.27)	0.57 (0.32)	0.31 (0.19)	0.36 (0.20)
SAIR (#2 / #1)	1.28	1.25	0.83	0.73	0.94	1.5
p-value	**0.0043** (N = 33)	**0.01** (N = 33)	**0.017** (N = 33)	**0.0029** (N = 33)	– (N = 33)	**0.002** (N = 33)
#2	0.76 (0.25)	0.74 (0.23)	0.61 (0.25)	0.57 (0.32)	0.32 (0.21)	0.36 (0.20)
#3	0.78 (0.25)	0.75 (0.27)	0.35 (0.23)	0.45 (0.31)	0.48 (0.29)	0.52 (0.27)
SAIR (#3 / #2)	1.02	1.01	0.57	0.79	1.5	1.4
p-value	– (N = 24)	– (N = 33)	**0.0046** (N = 24)	0.09 (N = 33)	**0.026** (N = 24)	**0.003** (N = 33)

Bold values indicate statistical significance of the results considering the minimal threshold as 0.05

Table 7 Averages (and standard deviations) of individual (IA) and team (TA) situation awareness measures derived from an analysis of team members' activities (IDE, IOE, TDE and TOE)

Dependent variables		Conditions		p-value
		w/o	w/	
IA (N = 33)	IDE	0.71 (0.24)	0.73 (0.25)	–
	IOE	0.68 (0.41)	0.80 (0.25)	0.937
TA (N = 11)	TDE	0.63 (0.10)	0.66 (0.16)	–
	TOE	0.78 (0.19)	0.82 (0.12)	0.575

microworld to collect fine-grained data from the users during the experiments that allowed us to fully understand what was really happening with teamwork and how the teams directed their usage of the groupware tool. Specifically, (1) different types of situation awareness evolved differently, and (2) for some types, the groupware tool could increase situation awareness, but the teams could find other ways to fulfil their awareness needs and therefore the value brought by the tool on this matter decreased as the task evolved.

We also realised that these conclusions could have only been achieved, given the holistic perspective put forward on the data gathering for the study. We gathered more data than strictly necessary to fulfil the original experimental aims. Had we been restricted to the set of data concerning situation awareness, e.g. the questionnaires, the conclusions would instead lean towards merely abandoning the tool. We therefore argue that data variety has been important to support the cycles of iterative development as posited by the tenets of design science. According to the foundations of design science, knowledge and understanding of problem domain is interwoven with the solution domain and may only be achieved through building, deploying and evaluating technology use [113]. Microworlds provide a platform that seems quite adequate to simultaneously support the cycles of building knowledge and designing technology under the design science paradigm. Accordingly, microworlds focus on

problem-solving interventions that require formative feedback on how the development is progressing and how it should proceed, and such feedback inherently requires the ability to gather rich multidimensional data.

Indeed we suggest that microworld platforms contribute to teamwork evaluation in two different ways. One is to provide a clear baseline on how to evaluate progress. By controlling a set of variables that express different aspects of the team's performance, e.g. individual and team awareness, time to complete the task, number and nature of communications, actions done by the team members on the environment, and interactions with the groupware prototype, researchers can objectively assess how the prototype development is progressing throughout the iterations. The second contribution is to provide a platform that eases repeating the experiments with different prototypes. In particular, the current microworld referent proposed (and the respective developed and deployed instance) allows changing the groupware prototype design while preserving many other aspects related to teamwork, such as the features of the working environment, communication channels used by teams, and instruments necessary to run the questionnaires.

Based on lessons taken from extensive usage of the microworld, described in this paper, we may now discuss more generic issues related to using microworlds in the research of teamwork. In this discussion, we take especially into consideration the tensions identified in Sect. 2. We highlight three major advantages we find in the microworld approach.

7.1 Capacity to gather wide-band experimental data about teamwork

Through various evaluation actions with the microworld, we collected very detailed indirect data about (1) movements of team members in simulated physical places; (2) actions done by the team members in the work environment; (3)

exchange of voice messages between the team members, as the task enfolds; (4) operations done by team members on groupware tools; and (5) time necessary to complete the task. Furthermore, we also collected data directly from the users, through questionnaires, which would have to be answered at different points during the task and with consideration of its context.

This wide-band approach to data collection addresses tensions No. 4 and 5. Because of inherent complexities associated to experimental designs, often researchers have to be very economic when deciding what data to collect, for instance restricting data collection to either macro or micro activities, and to individual or team measurements. Though the case discussed in this paper suggests that experimental designs can actually cover a wider set of variables, mainly because the microworld platform facilitates data collection and relating the information in meaningful ways (e.g. by relating voice communication with activities done by the team members, relating direct and indirect measures, and segmenting tasks in multiple phases). Then, the whole data set can be used to analyse various aspects of team behaviour.

7.2 Capacity to combine behaviour and design research

This combination of research goals addresses tension No. 3. Such a combination seems to be one of the most promising aspects resulting from our case study: exploring phenomena important to understanding team behaviour, like situation awareness, while at the same time exploring more design-oriented issues, such as groupware support. We suggest this combination of behaviour and design oriented research is unusual and can only be supported by adopting evaluation strategies that combine the requirements of both research methods, e.g. blending rigorous manipulation of variables with reuse of experimental design.

Of course microworlds bring some degree of artificiality in teamwork that challenges behaviour research. For instance, in our case we disallowed face-to-face communication and required teams to instead use an unfamiliar voice channel. Perhaps, more importantly, the microworld also substituted actual physical movement with an artificial feature where the team members had to press buttons to move from one place to the other. And our prior experiments with the microworld simulating physical movement showed that if the feature is not properly developed, it changes teamwork behaviour (e.g., if there is no time associated to simulated movements, the team members will continually jump from one place to the other [104]). However, the platform did not impose significant constraints on the overall practice of defining a plan, assigning activities to team members, diagnosing network failures, and reporting to the group. We suggest that as more research into teamwork adopts the microworld approach, some of the main constraints to behavioural research may

Table 8 Averages (and standard deviations) of the number of speech communications by nature: total, information sharing related and team management related

Dependent variables	Conditions		p-value
	w/o	w/	
# messages	10.09 (2.43)	7.73 (3.07)	**0.007**
# messages related to information sharing	2.36 (0.48)	1.46 (0.83)	**0.005**
# messages related to team management	0.72 (0.30)	0.80 (0.45)	0,789

Bold values indicate statistical significance of the results considering the minimal threshold as 0.05

be better understood and perhaps better controlled. This, of course, would imply increasing the researchers' attention on microworlds as a phenomenon of interest.

7.3 Capacity to reuse experimental components

We note in particular the reuse of the communication and freeze-probe components and their pertinence across studies. Regarding the latter, our study underlined the advantages of suspending a collaborative task so as to ask users about the task and the team. Suspending teamwork can be difficult to achieve in truly naturalistic settings but it is easy to orchestrate in a more controlled environment provided by a microworld. Furthermore, considering the ability to deploy dynamic and contextually bounded questions, we find there are plenty of opportunities to further research into teamwork behaviour. This is especially important when data collection concerns fine-grained cognitive phenomena, such as group attention, task awareness, mental load, memory, impact of interruptions, etc.

Concerning the reusability of the communication component, we recognise the current limitations of it, which only supports two types of speech communication. Video and text messaging are obvious gaps. However, we do not foresee technical constraints that would disallow further enriching the set of communication channels, which could then be reused across experiments. Overall, we observe that the proposed microworld components addresses tension No. 1 by combining repeatability with openness and exploration (Table 1).

8 Conclusions

In this paper we present a contribution towards the data-gathering problem in teamwork research. Especially in the teamwork arena, many phenomena of interest from different fields converge, like decision making, cognition, information sharing, and communication. The evaluation of new theoretical propositions and technologies is not straightforward,

since the very promising theories and technological solutions may change the current status of the problem domain, as well as its understanding. As such, theory, design and evaluation should be combined and such combination requires evaluating readily and regularly. Because of the divided nature of the teamwork research, gathering evaluation data about teamwork faces many tensions, which make both behaviour and design oriented assessments difficult to integrate and operationalize. In this paper we suggest that microworlds, which inherently provide quasi-naturalistic task scenarios, can afford to bring together the assessment of teamwork behaviour and the assessment of technology designs supporting teamwork.

We therefore proposed a set of microworld components aimed at teamwork studies. We discussed, in particular, four components: one dedicated to simulate operations in the physical environment, one supporting voice communication between the team members, one dedicated to simulate the functionality of groupware tools, and finally a component that questions participants at certain points during the experiments.

We described in detail the research study that was used to investigate the microworld based approach for data gathering in a teamwork related research endeavour. The study emphasises the capacity of the microworld to generate big data sets covering diversified aspects about team behaviour. In particular, we discussed how the microworld allowed the simultaneous gathering of both behaviour data and design-oriented data. Regarding the former, the study was mainly focussed on gathering data about situation awareness using a mix of questionnaires and logs documenting the team members' activities within task scenarios. Addressing the design perspective, the microworld also logged the team members' interactions with the simulated groupware tool intended to support teamwork.

The diversity of data that was gathered with the microworld, and its affordance for cross analyses, provide a rich insight over the enacted teamwork. It allowed us to conduct not only positivist oriented hypothesis testing, but further supported an interpretivist perspective, supporting an over-reaching insight on overall team behaviour and the role of the introduced groupware tool prototype. In particular, we analysed the evolution of different types of situation awareness as the team task evolved. We also observed significant changes in voice communication patterns caused by the introduction of the groupware tool.

The lack of statistical significance of multiple measures collected through the microworld highlights the complexity of doing studies in the teamwork area, and confirms that multiple design changes and evaluations are often necessary. This suggests researchers need methods to ease the research design effort, readily support consequent iterations, and to gather more data about the phenomena of interest.

Reflecting on our experiments, both developing the microworld and using the microworld for research, we brought forward positive and negative aspects of the microworld approach. We suggest the microworld approach allows for the capture of a wide range of data about teamwork, combining the quantitative and qualitative views, addressing micro and macro phenomena, focussing on individual and team activities, setting some naturalistic and some controlled actions, and support behavioural and design considerations. According to the present proposal, we further posit that the microworld building blocks may significantly contribute to the development of repeatable and more systematic studies in the teamwork area. Although, as discussed, not at all a new concept (microworlds are in fact used in several research studies on multiple application domains), there is still a lack of focus on microworlds, especially in the teamwork field.

We take here a first step toward a referent on the building blocks required to develop microworlds for teamwork research in order to promote research findings that enable both comparison and generalization. Nevertheless, one must point out that more research is necessary to clearly define the specific bounds where both rigorous manipulation of variables and naturalness of experiments do not compromise face validity and become accepted by the research community.

This work further suggests some additional directions for future research. In a practical perspective, we envisage having a microworld platform with more features, especially regarding the support to communication modes, more realistic support to mobility and more operations in the environment, and more control over a wider range of variables. The possibility of combining humans with simulated actors in teams seems also very interesting [114]. In a more theoretical perspective, we envisage that a more widespread adoption of microworlds may help to fill the gaps between behavioural and design research. We believe such endeavour would certainly contribute to consolidating teamwork research into a more distinctive integrated field, with reliable and widely accepted innovative research methods.

References

1. Salas, E., Dickinson, T., Converse, S., Tannenbaum, S.: Toward an understanding of team performance and training. In: Swezey, R.W., Salas, E. (eds.) Teams: Their Training and Performance. Ablex Publishing, Westport (1992)
2. Wilson, K., Salas, E., Priest, H., Andrews, D.: Errors in the heat of battle: taking a closer look at shared cognition breakdowns through teamwork. Hum. FactorsC **49**(2), 243–256 (2007)
3. Ellis, C., Gibbs, S., Rein, G.: Groupware: some issues and experiences. Commun. ACM **34**(1), 39–58 (1991)
4. Jacovi, M., Soroka, V., Gilboa-Freedman, G., Ur, S., Shahar, E., Marmasse, N.: The chasms of CSCW: a citation graph analysis of the CSCW conference. In: Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work 2006, pp. 289–298. ACM (2006)

5. Grudin, J., Poltrock, S.: Taxonomy and theory in computer supported cooperative work. In: Kozlowski, S. (ed.) Handbook of Organizational Psychology, pp. 1323–1348. Oxford University Press, Oxford (2012)

6. Antunes, P., Herskovic, V., Ochoa, S., Pino, J.: Structuring dimensions for collaborative systems evaluation. ACM Comput. Surv. **44**(2), 1–28 (2012)

7. Antunes, P., Xiao, L., Pino, J.: Assessing the impact of educational differences in HCI design practice. Int. J. Technol. Des. Educ. **24**(3), 317–335 (2014). doi:10.1007/s10798-013-9254-8

8. Herskovic, V., Pino, J., Ochoa, S., Antunes, P.: Evaluation methods for groupware systems. In: Haake, J., Ochoa, S., Cechich, A. (eds.) Groupware: design, implementation, and use. 13th International Workshop, CRIWG 2007, Bariloche, Argentina, September 2007 Proceedings, vol. 4715. LNCS, pp. 328–336. Springer, Heidelberg (2007)

9. Hamadache, K., Lancieri, L.: Strategies and taxonomy, tailoring your CSCW evaluation. In: International Conference on Collaboration and Technology, pp. 206–221. Springer, Heidelberg (2009)

10. Grudin, J.: Why CSCW applications fail: problems in the design and evaluation of organizational interfaces. In: Proceedings of the 1988 ACM Conference on Computer-Supported Cooperative Work, Portland, pp. 85–93. ACM (1988)

11. Hughes, J., King, V., Rodden, T., Andersen, H.: Moving out from the control room: ethnography in system design. In: Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work, Chapel Hill, pp. 429–439. ACM Press (1994)

12. Hevner, A., March, S., Park, J., Ram, S.: Design science in information systems research. Manag. Inf. Syst. Q. **28**(1), 75–105 (2004)

13. Gregor, S.: The nature of theory in information systems. MIS Q. **30**(3), 611–642 (2006)

14. Gregor, S., Jones, D.: The anatomy of a design theory. J. Assoc. Inf. Syst. **8**(5), 312–335 (2007)

15. Fjermestad, J., Hiltz, S.: An assessment of group support systems experimental research: methodology and results. J. Manag. Inf. Syst. **15**(3), 7–149 (1999)

16. Klein, G.: Naturalistic decision making. Hum. Factors **50**(3), 456–460 (2008)

17. Hollan, J., Hutchins, E., Kirsh, D.: Distributed cognition: toward a new foundation for human–computer interaction research. ACM Trans. Comput. Hum. Interact. **7**(2), 174–196 (2000)

18. Suchman, L.: Plans and Situated Actions: The Problem of Human–Machine Communication. Cambridge University Press, Cambridge (1987)

19. Endsley, M., Garland, D.: Situation Awareness Analysis and Measurement. CRC Press, Boca Raton (2000)

20. Endsley, M., Jones, W.: Situation awareness. In: Lee, J.D., Kirlik, A. (eds.) The Oxford Handbook of Cognitive Engineering, pp. 88–108. Oxford University Press, New York (2013)

21. Antunes, P., Herskovic, V., Ochoa, S., Pino, J.: Reviewing the quality of awareness support in collaborative applications. J. Syst. Softw. **89**, 146–169 (2014)

22. Wickens, C.: Situation awareness: Review of Mica Endsley's 1995 articles on situation awareness theory and measurement. Hum. Factors **50**(3), 397–403 (2008)

23. Salmon, P., Stanton, N., Walker, G., Jenkins, D., Ladva, D., Rafferty, L., Young, M.: Measuring situation awareness in complex systems: comparison of measures study. Int. J. Ind. Ergon. **39**(3), 490–500 (2009)

24. Collins, A., Joseph, D., Bielaczyc, K.: Design research: theoretical and methodological issues. J. Learn. Sci. **13**(1), 15–42 (2004)

25. Lew, R., Boring, R., Ulrich, T.: A prototyping environment for research on human-machine interfaces in process control use of Microsoft WPF for microworld and distributed control system development. In: 7th International Symposium on Resilient Control Systems. IEEE (2014)

26. Jercic, P., Astor, P., Adam, M., Hilborn, O., Schaaff, K., Lindley, C., Sennersten, C., Eriksson, J.: A serious game using physiological interfaces for emotion regulation training in the context of financial decision-making. In: European Conference on Information Systems, p. 207 (2012)

27. Naweed, A., Hockey, G., Clarke, S.: Designing simulator tools for rail research: the case study of a train driving microworld. Appl. Ergon. **44**(3), 445–454 (2013)

28. Kluge, A.: Performance assessments with microworlds and their difficulty. Appl. Psychol. Meas. **32**(2), 156–180 (2008)

29. Mosier, C.: A critical examination of the concepts of face validity. Educ. Psychol. Meas. **7**, 191–205 (1947)

30. Wastell, D., Peckover, S., White, S., Broadhurst, K., Hall, C., Pithouse, A.: Social work in the laboratory: using microworlds for practice research. Br. J. Soc. Work **41**, 744–760 (2011)

31. Rolo, G., Diaz-Cabrera, D.: Decision-making processes evaluation using two methodologies: field and simulation techniques. Theor. Issues Ergon. Sci. **6**(1), 35–48 (2005)

32. Gray, W.: Simulated task environments: the role of high-fidelity simulations, scaled worlds, synthetic environments, and laboratory tasks in basic and applied cognitive research. Cogn. Sci. Q. **2**(2), 205–207 (2002)

33. Rigas, G., Carling, E., Brehmer, B.: Reliability and validity of performance measures in microworlds. Intelligence **30**(5), 463–480 (2002)

34. DiFonzo, N., Hantula, D., Bordia, P.: Microworlds for experimental research: Having your (control and collection) cake, and realism too. Behav. Res. Methods Instrum. Comput. **30**(2), 278–286 (1998)

35. Chen, C., Zhang, C.: Data-intensive applications, challenges, techniques and technologies: a survey on big data. Inf. Sci. **275**, 314–347 (2014)

36. Popper, K.: Science: conjectures and refutations. In: Introductory Readings in the Philosophy of Science, pp. 38–47. Prometheus Books, Amherst (1998)

37. Venkatesh, V., Brown, S., Bala, H.: Bridging the qualitative–quantitative divide: guidelines for conducting mixed methods research in information systems. MIS Q. **37**(1), 21–54 (2013)

38. Horn, D., Finholt, T., Birnholtz, J., Motwani, D., Jayaraman, S.: Six degrees of Jonathan Grudin: a social network analysis of the evolution and impact of CSCW research. In: Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work, pp. 582–591. ACM (2004)

39. Pinelle, D., Gutwin, C.: A review of groupware evaluations. In: Proceedings of 9th IEEE WETICE Infrastructure for Collaborative Enterprises (2000)

40. Grant, A., Wall, T.: The neglected science and art of quasi-experimentation: why-to, when-to, and how-to advice for organizational researchers. Org. Res. Methods **12**(4), 653–686 (2008)

41. Winter, R.: Design science research in Europe. Eur. J. Inf. Syst. **17**(5), 470–475 (2008)

42. Baskerville, R., Wood-Harper, A.: A critical perspective on action research as a method for information systems research. In: Willcocks, L.P., Sauer, C., Lacity, M.C. (eds.) Enacting Research Methods in Information Systems, pp. 169–170. Springer, Berlin (2016)

43. Jenkins, A.: Research methodologies and MIS research. In: Mumford, E., et al. (eds.) Research Methods in Information Systems, pp. 103–117. North-Holland Publishing Co. Amsterdam, The Netherlands (1985)

44. Sinkovics, R., Ghauri, P.: Enhancing the trustworthiness of qualitative research in international business. Manag Int Rev **48**(6), 689–714 (2008)

45. Flach, J.: Mind the gap: a skeptical view of macrocognition. In: Schraagen, J., Militello, L., Ormerod, T., Lipshitz, R. (eds.) Naturalistic Decision Making and Macrocognition. Ashgate, Hampshire (2008)

46. Wallace, B., Ross, A. (eds.): Beyond Human Error—Taxonomies and Safety Science. CRC Taylor and Francis Group, New York (2006)

47. Dekker, S.: The Field Guide to Understanding Human Error. Ashgate, Hampshire (2006)

48. Klein, G.: A recognition-primed decision (RPD) model of rapid decision making. In: Klein, G., Orasanu, J., Calderwood, R., Zsambok, C. (eds.) Decision Making in Action: Models and Methods. Ablex, Westport, CT (1993)

49. Salas, E., Fiore, S., Warner, N., Letsky, M.: Emerging multidisciplinary theoretical perspectives in team cognition: an overview. Theor. Issues Ergon. Sci. 11(4), 245–249 (2010)

50. Lipshitz, R., Klein, G., Orasanu, J., Salas, E.: Taking stock of naturalistic decision making. J. Behav. Decis. Mak. 14, 331–352 (2001)

51. Cleven, A., Gubler, P., Hüner, K.: Design alternatives for the evaluation of design science research artifacts. In: Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology, Philadelphia, pp. 1–8. ACM (2009)

52. Patrick, J., James, N.: Process tracing of complex cognitive work tasks. J. Occup. Org. Psychol. 77(2), 259–280 (2004)

53. Sá, M., Carriço, L., Antunes, P.: Ubiquitous psychotherapy. IEEE. Pervasive Comput. 6(1), 20–27 (2007)

54. Piirainen, K., Gonzalez, R., Kolfschoten, G.: Quo Vadis, design science?—a survey of literature. In: Global Perspectives on Design Science Research, vol. 6105. Lecture Notes in Computer Science, pp. 93–108. Springer (2010)

55. Briggs, R.: On theory-driven design and deployment of collaboration systems. Int. J. Hum Comput Stud. 64(7), 573–582 (2006)

56. Card, S., Moran, T., Newell, A.: The Psychology of Human–Computer Interaction. Lawrance Elrbaum, Hillsdale (1983)

57. Reason, J.: Human Error. Cambridge University Press, Cambridge (1990)

58. Cacciabue, P.: Guide to Applying Human Factors Methods. Springer, London (2004)

59. Cooke, N., Gorman, J., Myers, C., Duran, J.: Interactive team cognition. Cogn. Sci. 37(2), 255–285 (2013)

60. Gibson, J.: The Senses Considered as Perceptual Systems. Houhton Mifflin, Boston (1966)

61. Antunes, P., Zurita, G., Baloian, N.: Key indicators for assessing the design of geocollaborative applications. Int. J. Inf. Technol. Decis. Mak. 13(2), 361–385 (2014)

62. Turvey, M., Shawn, R.: Toward an ecological physics and physical psychology. In: Solso, R., Massaro, S. (eds.) The Science of the Mind: 2001 and Beyond, pp. 144–169. Oxford University Press, New York (1995)

63. Salas, E., Sims, D., Burke, C.: Is there a "Big Five" in teamwork? Small Group Res. 36(5), 555–599 (2005)

64. Davis, F.: A technology acceptance model for empirically testing new end-user information systems: theory and results. Ph.D. Thesis, Massachusetts Institute of Technology (1986)

65. Read, A., Hullsiek, B., Briggs, R.: The seven layer model of collaboration: an exploratory study of process identification and improvement. In: 45th Hawaii International Conference on System Science, pp. 412–420. IEEE (2012)

66. Barnard, P., May, J., Duke, D., Duce, D.: Systems, interactions, and macrotheory. ACM Trans. Comput. Hum. Interact. 7(2), 222–262 (2000)

67. Klein, G., Ross, K., Moon, B., Klein, D., Hoffman, R., Hollnagel, E.: Macrocognition. IEEE Intell. Syst. 18(3), 81–85 (2003)

68. Fiore, S., Smith-Jentsch, K., Salas, E., Warner, N., Letsky, M.: Towards an understanding of macrocognition in teams: developing and defining complex collaborative processes and products. Theor. Issues Ergon. Sci. 11(4), 250–271 (2010)

69. Papert, S.: Microworlds: transforming education. Artif. Intell. Educ. 1, 79–94 (1987)

70. Brehmer, B., Dorner, D.: Experiments with computer-simulated microworlds: escaping both the narrow straits of the laboratory and the deep blue sea of the field study. Comput. Hum. Behav. 9, 171–184 (1993)

71. Johansson, B., Trnka, J., Granlund, R.: The Effect of geographical information systems on a collaborative command and control task. In: Proceedings of the 4th International Conference on Information Systems for Crisis Response and Management (ISCRAM). Delft (2007)

72. Schraagen, J., Van den Ven, J.: Improving decision making in crisis response through critical thinking support. J. Cogn. Eng. Decis. Mak. 2, 311–327 (2008)

73. Arthur, W., Day, E., Villado, A., Boatman, P., Kowollik, V., Bennet, W., Bhupatkar, A.: The effect of distributed practice on immediate posttraining, and long-term performance on a complex command-and-control simulation task. Hum. Perform. 23(5), 428–445 (2010)

74. Lew, R., Boring, R., Ulrich, T.: A prototyping environment for research on human-machine interfaces in process control use of Microsoft WPF for microworld and distributed control system development. In: International Symposium on Resilient Control Systems, pp. 1–6. IEEE (2014)

75. O'Brien, K., O'Hare, D.: Situational awareness ability and cognitive skills training in a complex real-world task. Ergonomics 50(7), 1064–1091 (2007)

76. Berggren, P., Johansson, B., Svensson, E., Baroutsi, N., Dahlbäck, N.: Statistical modelling of team training in a microworld study. Proc. Hum. Factors Ergon. Soc. Annu. Meet. 58(1), 894–898 (2014)

77. de Heer, J.: How Do architects think? A game based microworld for elucidating dynamic decision-making. In: Complex Systems Design & Management, pp. 133–142. Springer (2016)

78. Chapman, T., Nettelbecka, T., Welsha, M., Millsab, V.: Investigating the construct validity associated with microworld research: a comparison of performance under different management structures across expert and non-expert naturalistic decision-making groups. Aust. J. Psychol. 58(1), 40–47 (2006)

79. Mavrikis, M., Dragon, T., Abdu, R., Harrer, A., De Groot, R., McLaren, B.: Learning to learn together through planning, discussion and reflection on microworld-based challenges. In: European Conference on Technology Enhanced Learning, pp. 483–488. Springer, Heidelberg (2012)

80. Frezzo, D., DiCerbo, K., Behrens, J., Chen, M.: An extensible micro-world for learning in the data networking professions. Inf. Sci. 264, 91–103 (2014)

81. Djaouti, D., Alvarez, J., Jessel, J.: Classifying serious games: the G/P/S model. In: Felicia, P. (ed.) Handbook of Research on Improving Learning and Motivation Through Educational Games: Multidisciplinary Approaches, pp. 118–136. IGI Global, Hershey (2011)

82. Cooper, S., Khatib, F., Treuille, A., Barbero, J., Lee, J., Beenen, M., Leaver-Fay, A., Baker, D., Popović, Z.: Predicting protein structures with a multiplayer online game. Nature 466(7307), 756–770 (2010)

83. Sonnleitner, P., Brunner, M., Greiff, S., Funke, J., Keller, U., Martin, R., Hazotte, C., Mayer, H., Latour, T.: The Genetics Lab. Acceptance and psychometric characteristics of a computer-based microworld to assess complex problem solving. Psychol. Test Assess. Model. 54, 54–72 (2012)

84. Gonzalez, C., Vanyukov, P., Martin, M.: The use of microworlds to study dynamic decision making. Comput. Hum. Behav. **21**(2), 273–286 (2005)

85. Brehmer, B.: Micro-worlds and the circular relation between people and their environment. Theor. Issues Ergon. Sci. **6**(1), 73–93 (2005)

86. Endsley, M., Bolté, B., Jones, D.: Designing for Situation Awareness. Taylor & Francis, London (2003)

87. Wagner, F., Schmuki, R., Wagner, T., Wolstenholme, P.: Modeling Ssoftware with Finite State Machines: A Practical Approach. CRC Press, New York (2006)

88. Lee, M.: A testing framework based on finite automata for object-oriented software specification. J. Inf. Technol. Theory Soc. **1**, 59–88 (2004)

89. Buchner, A., Funke, J.: Finite state automata: dynamic task environments in problem solving research. Q. J. Exp. Psychol. **46A**(1), 83–118 (1993)

90. Taylor, M., Endsley, M., Henderson, S.: Situational awareness workshop report. In: Hayward, B., Lowe, A. (eds.) Applied Aviation Psychology: Achievement, Change and Challenge, pp. 447–454. Ashgate Publishing, Aldershot (1996)

91. Citera, M., McNeese, M., Brown, C., Selvaraj, J.: Fitting information systems to collaborating design teams. J. Am. Soc. Inf. Sci. Technol. **46**(7), 551 (1995)

92. Klein, G., Zsambock, C., Thordsen, M.: Team decision training: five myths and a model. Mil. Rev. **73**(4), 36–42 (1993)

93. Berndtsson, J., Normark, M.: The coordinative functions of flight strips : air traffic control revisited. In: Proceedings of the International ACM SIGGROUP Conference on Supporting Group Work, pp. 101–110. ACM, New York (1999)

94. Heath, C., Luff, P.: Collaboration and control crisis management and multimedia technology in London Underground Line Control Rooms. Comput. Support. Coop. Work **1**(1–2), 69–94 (1992)

95. Pettersson, M., Randall, D., Helgeson, B.: Ambiguities, awareness and economy: a study of emergency service work. In: Proceedings of the ACM Conference on Computer Supported Cooperative Work, New Orleans, pp. 286–295. ACM (2002)

96. Monares, A., Ochoa, S., Pino, J., Herskovic, V., Rodriguez-Covili, J., Neyem, A.: Mobile computing in urban emergency situations: improving the support to firefighters in the field. Expert Syst. Appl. **38**(2), 1255–1267 (2011)

97. Whittaker, S., Amento, B.: Seeing what you are hearing: coordinating responses to trouble reports in network troubleshooting. In: European Conference on Computer Supported Cooperative Work, Helsinki, pp. 219–238. Springer (2003)

98. Fallman, D.: Enabling physical collaboration in industrial settings by designing for embodied interaction. In: Proceedings of the Latin American Conference on Human–Computer Interaction, Rio de Janeiro, pp. 41–51. ACM (2003)

99. Nulden, U.: Investigating police patrol practice for design of IT. In: CHI '03 Extended Abstracts on Human Factors in Computing Systems, Ft. Lauderdale, pp. 820–821. ACM (2003)

100. Landgren, J.: Supporting fire crew sensemaking enroute to incidents. Int. J. Emerg. Manag. **2**(3), 176–188 (2005)

101. Salmon, P., Stanton, N., Walker, G., Green, D.: Situation awareness measurement: a review of applicability for C4i environments. Appl. Ergon **37**(2), 225–238 (2006)

102. Bai, X., Tsai, W., Paul, R., Feng, K., Yu, L.: Scenario-based modeling and its applications. In: Proceedings of the Seventh International Workshop on Object-Oriented Real-Time Dependable Systems, pp. 253–260. IEEE Computer Society (2002)

103. Barrett, R., Kandogan, E., Maglio, P., Haber, E., Takayama, L., Prabaker, M.: Field studies of computer system administrators: analysis of system management tools and practices. In: Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work, pp. 388–395. ACM (2004)

104. Sapateiro, C.: Evaluating mobile collaborative applications support of teamwork in critical incidents response management. Ph.D. Thesis, University of Lisbon (2013)

105. Sapateiro, C., Antunes, P.: An emergency response model toward situational awareness improvement. In: International Conference on Information Systems for Crisis Response and Management. Göteborg (2009)

106. McManus, S., Seville, E., Brunsdon, D., Vargo, J.: Resilience management: a framework for assessing and improving the resilience of organisations. Research Report 2007/01. Resilient Organizations (2007)

107. Hollnagel, E., Woods, D.: Joint Cognitive Systems: Foundations of Cognitive Systems Engineering. CRC Press, Boca Raton (2005)

108. Endsley, M.: Toward a theory of situation awareness in dynamic systems. Hum. Factors **31**(7), 32–64 (1995)

109. Smith, K., Hancock, P.: Situation awareness is adaptive, externally directed consciousness. Hum. Factors **37**, 137–148 (1995)

110. Bolstad, C., Cuevas, H., Gonzalez, C., Schneider, M.: Modeling shared situation awareness. In: Proceedings of the 14th Conference on Behavior Representation in Modeling and Simulation. Los Angles (2005)

111. Stanton, N., Stewart, R., Harris, D., Houghton, R., Baber, C., McMaster, R., Salmon, P., Hoyle, G., Walker, G., Young, M., Linsell, M., Dymott, R., Green, D.: Distributed situation awareness in dynamic systems: theoretical develoment and application of an ergonomics methodology. Ergonomics **49**(12–13), 1288–1311 (2006)

112. Tamhane, A., Dunlop, D.: Statistics and Data Analysis: From Elementary to Intermediate. Prentice Hall, Upper Saddle River (2000)

113. Hevner, A.: A three cycle view of design science research. Scand. J. Inf. Syst. **19**(2), 87–92 (2007)

114. Kurbalija, V., Ivanović, M., Bernstorff, C., Nachtwei, J., Burkhard, H.: Matching observed with empirical reality-what you see is what you get? Fundam.Inf. **129**(1–2), 133–147 (2014)

Claudio Miguel Sapateiro is currently a lecturer at the Systems and Informatics department of the School of Technology of the Polytechnic Institute of Setubal, Portugal. He holds a Ph.D. in Informatics Engineering and his main research interests are on Business Intelligence and Human Factors. Claudio has published several papers in journals, books and conferences, including the journal Group Decision and Negotiation.

Pedro Antunes is Associate Professor at Victoria University of Wellington, School of Information Management. He has a Ph.D. in Electrical and Computer Engineering from the Technical University of Lisbon and Habilitation in Informatics Engineering from the University of Lisbon. Pedro's research is mainly focused on human/organizational facets of decision support and business process management systems. Pedro has published articles in

ACM Computing Surveys, Behavior and Information Technology, Information Systems Frontiers, IEEE Pervasive Computing, IEEE Transactions on System, Man, and Cybernetics, Journal of Systems and Software, Expert Systems With Applications, Group Decision and Negotiation, and other journals. Pedro is currently Associate Editor at IEEE Transactions on Human–Machine Systems.

David Johnstone is a Senior Lecturer in Information Systems at Victoria University of Wellington, New Zealand. Originally trained in mathematics, statistics and operations research, he went on to hold a variety of IT roles in both the public and private sectors. He completed a Ph.D. in Information Systems at Victoria in 2010. He has published in a variety of IS journals and conferences, including European Journal of Information Systems, Behavior and Information Technology, and Information Systems Frontiers. Current research interests include IT project governance, information behaviour and crowdsourcing.

José A. Pino is a full professor of Computer Science at the University of Chile. His research interests include ubiquitous computing, human–computer interaction, and socio-technical systems. He has served as President of the Chilean Computer Science Society (SCCC) and President of CLEI (the Latin American Association of Universities Concerning Information Technology). He has co-authored six books and published research papers in international conferences and journals, including Journal of the ACM, Communications of the ACM, ACM Computing Surveys, Decision Support Systems, Group Decision and Negotiation, and Future Generation Computer Systems. He is currently a member of the Editorial Boards of the Journal of Network and Computer Applications and the Journal of Educational Technology and Society.